Blind Deconvolution via Sequential Imputations

Jun S. Liu and Rong Chen*

The sequential imputation procedure is applied to adaptively and sequentially reconstruct discrete input signals that are blurred by an unknown linear moving average channel and contaminated by additive Gaussian noises, a problem known as *blind deconvolution* in digital communication. A rejuvenation procedure for improving the efficiency of sequential imputation is introduced and theoretically justified. The proposed method does not require the channel to be nonminimum phase and can be used in real time signal restoration. Two simulated systems are studied to illustrate the proposed method. Our result shows that the ideas of multiple imputations and flexible simulation techniques are as powerful in engineering as in survey sampling.

KEY WORDS: Bayesian model; Communication; Gibbs sampling; Importance sampling; Predictive distribution; Signal transmission.

1. INTRODUCTION

Signal transmissions and various associated adaptive filtering procedures have been the major focus of communication techniques for more than 50 years. In this article we consider the following linearly degraded moving average system commonly seen in digital communication:

$$y_t = \sum_{i=0}^{q} h_i x_{t-i} + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{1}$$

where $\varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and the y_t are sequentially observed output signals. The unobserved input signal x_t are assumed to be discrete valued with known levels $s_1, \ldots, s_m$. The coefficients of the blurring mechanism, $h_0, \ldots, h_q$, are assumed unknown (i.e., *blind*). The main objective is to recover the input signals, x_t, $t = 1, 2, \ldots$, and the coefficients, h_i, $i = 0, \ldots, q$.

There are two requirements for a desirable solution to this problem. The first is, of course, the *blindness*. The second is the adaptivity; that is, the deconvolution procedure must be sequential and adaptive. With these features, the system can recover itself while receiving and processing the transmission and has the ability to recover after a rapid system coefficient change. This combination of blindness and adaptivity creates the challenge.

In this article we propose an adaptive blind deconvolution algorithm based on a Bayesian formulation of the problem and on the sequential imputation method developed by Kong, Liu, and Wong (1994). The main idea is to treat the input signals ($\{x_t\}$'s) as missing data and to sequentially impute multiples of them based on currently observed output signals ($\{y_t\}$'s). For each imputed signal sequence, its *importance weight* is computed according to its relative "predictability" of the (future) observed blurred signals. The imputed sequences with their importance weights provide Bayesian estimates of the input signal sequence. The coefficients for the linear system can be estimated based on these estimates. The algorithm is easy to implement, and the system is not required to be minimum/nonminimum phase. Because the procedure is developed under a Bayesian framework, prior information can be easily incorporated.

Adaptive blind deconvolution has important applications in many fields such as digital communications, seismology, and underwater acoustics. Linear restoration procedures are widely used for recovering input signals when a training sequence of signals is available for an initial estimation of the system coefficients. But training sequences are not always available. For instance, transmission monitoring and digital microwave radio link modems are usually required to start up or restart their adaptive equalizer without a training sequence. In a multipoint network, if a channel from the master transmitter to one of its tributary stations goes down after the initial training period, then the channel must be recovered without training so as to avoid interrupting the communication between the master transmitter and other stations.

Blind deconvolution problems that do not involve additive noises have been addressed by several researchers. Most of the existing algorithms belong to one of two groups. Sato (1975), Godard (1980), and Donoho (1981) minimized nonconvex cost functions that differ from the classical mean squared error criterion. Another group of algorithms uses higher-order cumulants, spectrum, or correlation structure to identify a nonminimum phase system (Cheng 1990; Giannakis and Mendel 1989; Hatzinakos and Nikias 1991; Lii and Rosenblatt 1982). Some of the above approaches can be extended to cases with additive Gaussian noise. Li (1992, 1993) used inverse filtering procedures; Chen and Li (1993) described a Gibbs sampling method for signal recovery in system (1).

The rest of the article is organized as follows. Section 2 gives a brief overview of sequential imputation. Section 3 provides computational details for implementing the method in blind deconvolution. Sections 4 and 5 contain some modifications of the plain sequential imputation method making it more suitable for the blind deconvolution problem. In particular, Section 4 describes and studies a procedure called *rejuvenation*, and Section 5 contains some other modifications. Finally, Section 6 includes two simulation studies to illustrate the properties of the procedures.

2. THE SEQUENTIAL IMPUTATION

We begin with a brief description of sequential imputation in a general Bayesian framework. Let θ be the parameter of

* Jun S. Liu is Assistant Professor, Department of Statistics, Stanford University, Stanford, CA 94305. Rong Chen is Assistant Professor, Department of Statistics, Texas A&M University, College Station, TX 77843. This research is supported in part by National Science Foundation Grants DMS-9301193 and DMS-9404344. The authors thank Yingnian Wu, Jane Harvill, the editor, the associate editor, and two anonymous referees for very insightful suggestions, one of which leads to the theoretical treatment of rejuvenation in Section 4. Part of the manuscript was prepared when the first author was Assistant Professor, Department of Statistics, Harvard University.

interest and let $\mathbf{z}$ denote the (complete) data set. The posterior distribution $\pi(\theta|\mathbf{z})$ is assumed to be simple. In many practical situations, however, $\mathbf{z}$ is only partially observed. In this case we partition $\mathbf{z}$ as $(\mathbf{x}, \mathbf{y})$, where $\mathbf{y}$ denotes the observed part and $\mathbf{x}$ represents the missing part. Now suppose that $\mathbf{x}$ and $\mathbf{y}$ can each be further decomposed into n corresponding components so that

$$\mathbf{z} = (z_1, z_2, \ldots, z_n) = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n), \quad (2)$$

where $z_i = (x_i, y_i)$ for $i = 1, \ldots, n$. Typically the z_i are iid or follow some linear time series models given θ.

If N copies (ideally, independent ones), $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)$, can be drawn from the conditional distribution $\pi(\mathbf{x}|\mathbf{y})$, then, because $\pi(\theta|\mathbf{y}) = \int \pi(\theta|\mathbf{x}, \mathbf{y})\pi(\mathbf{x}|\mathbf{y})\,dx$, the posterior distribution $\pi(\theta|\mathbf{y})$ can be approximated by $\sum_{j=1}^{N} \pi(\theta|\mathbf{z}(j))/N$, where $\mathbf{z}(j)$ is the augmented complete data set $\{\mathbf{x}(j), \mathbf{y}\}$. But drawing from $\pi(\mathbf{x}|\mathbf{y})$ directly is usually difficult. The Gibbs sampler (see Gelfand and Smith 1990 for details) evolves a Markov chain to generate dependent samples $\mathbf{x}(j), j = 1, \ldots, N$, from the target distribution $\pi(\mathbf{x}|\mathbf{y})$. Sequential imputation, however, achieves a similar goal by imputing the x_i sequentially and then adjusting them by using importance weights. In general, sequential imputation involves two steps:

Step A. Draw x_1^* from $\pi(x_1|y_1)$, and then for $t = 2, \ldots, n$, draw x_t^* recursively from the conditional distribution

$$\pi(x_t^*|x_1^*, y_1, \ldots, x_{t-1}^*, y_{t-1}, y_t).$$

Note that the unobserved signal x_t is imputed by a draw x_t^* from its predictive distribution based on the current observations and imputations.

Step B. For $t = 1, \ldots, n-1$, compute $\pi(y_{t+1}|x_1^*, y_1, \ldots, x_t^*, y_t)$. The *importance weight* of the augmented sequence is computed as

$$w = \pi(y_1)\prod_{t=1}^{n-1} \pi(y_{t+1}|x_1^*, y_1, \ldots, x_t^*, y_t). \quad (3)$$

Here the predictive probability of the next observation y_{t+1} is computed and accumulated to credit the imputed sequence.

Suppose that A and B are done repeatedly and independently N times and let the imputed samples be denoted by $\mathbf{x}^*(1), \ldots, \mathbf{x}^*(N)$, and the respective weights by $w(1), \ldots, w(N)$, where $\mathbf{x}^*(j) = (x_1^*(j), \ldots, x_n^*(j))$ for $j = 1, \ldots, N$. The posterior $\pi(\theta|y)$ is estimated by the weighted mixture

$$\hat{\pi}(\theta|\mathbf{y}) = \sum_{j=1}^{N} w^*(j)\pi(\theta|\mathbf{z}^*(j)), \quad (4)$$

where $w^*(j) = w(j)/\sum_{k=1}^{N} w(k)$ and $\mathbf{z}^*(j) = (\mathbf{x}^*(j), \mathbf{y})$ for $j = 1, \ldots, N$, and the expected values of the "missing part" x_t is estimated by the weighted average $\hat{x}_t = \sum_{j=1}^{N} w^*(j)x_t^*(j)$. The maximum a posteriori estimator can be obtained similarly.

The method can be regarded as a special importance sampling scheme. To see this, note that each imputed sequence $\mathbf{x}^*(j)$ is drawn from the 'trial distribution' $\pi^*(\mathbf{x}|\mathbf{y})$ as described in Step A instead of the ideal distribution $\pi(\mathbf{x}|\mathbf{y})$. The weight computed in Step B is equivalent to the importance sampling weight $\pi^*(\mathbf{x}|\mathbf{y})/\pi(\mathbf{x}|\mathbf{y})$ associated with the scheme (see Kong et al. 1994 for more details). Geweke (1989) also provided some general discussions of the use of importance sampling in Bayesian computation.

3. SEQUENTIAL IMPUTATION ALGORITHM FOR BLIND DECONVOLUTION

3.1 Model Setting and Priors

For the system defined by (1), the likelihood function, up to time n, can be written as

$$L(\mathbf{X}_n, \mathbf{H}, \sigma^2|\mathbf{Y}_n) \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(y_t - \sum_{i=0}^{q} h_i x_{t-i}\right)^2\right\}, \quad (5)$$

where $\mathbf{Y}_n = (y_1, \ldots, y_n)^T$, $\mathbf{X}_n = (x_{1-q}, \ldots, x_n)^T$, and $\mathbf{H} = (h_0, \ldots, h_q)^T$ are column vectors.

Taking a Bayesian viewpoint, we a priori assume that $\mathbf{H} \sim N(H_0, \Sigma_0)$, where H_0 is a $(q+1)$-dimensional column vector and Σ_0 is a $(q+1)\times(q+1)$ positive definite matrix. Furthermore, because the x_t are iid from a discrete distribution with probability vector $(\theta_1, \ldots, \theta_m)$ on the m possible states $(s_1, \ldots, s_m)$, where $\theta_i \geq 0$ and $\theta_1 + \cdots + \theta_m = 1$, we put a Dirichlet prior on Θ, [i.e., $\Theta = (\theta_1, \ldots, \theta_m) \sim$ Dirichlet$(b_1, \ldots, b_m)$] when Θ is unknown.

Ideally, the posterior distribution of the linear coefficients can be obtained by "summing out" those x's in (5). But the impracticability of this can be easily seen. Furthermore, there are no simple conditional independence structures to facilitate a Kalman filtering–type algorithm. The Gibbs sampler, a Monte Carlo method for overcoming this difficulty, has been implemented by Chen and Li (1993) for this problem. But the Gibbs sampler is not an adaptive procedure and has difficulty dealing with sequentially observed data. With new data coming in, the whole computation must be repeated to incorporate new information. Sequential imputation provides a more efficient alternative. To present our result, we introduce the following notation:

- n is the total number of observed signals, m is the number of states of input signals, and $q + 1$ is the number of moving average lags.
- The observed and the input signals, up to time t, are denoted by $\mathbf{Y}_t = (y_1, \ldots, y_t)^T$ and $\mathbf{X}_t = (x_{1-q}, \ldots, x_t)^T$.
- $\mathbf{x}_t = (x_t, x_{t-1}, \ldots, x_{t-q})^T$ is a $(q+1)$-dimensional column vector.
- $\mathcal{X}_t = (\mathbf{x}_1, \ldots, \mathbf{x}_t)^T = \begin{pmatrix} x_1 & x_0 & \cdots & x_{1-q} \\ x_2 & x_1 & \cdots & x_{2-q} \\ \vdots & & \ddots & \vdots \\ x_t & x_{t-1} & \cdots & x_{t-q} \end{pmatrix}$

  is a $t \times (q+1)$ matrix.
- $B = b_1 + \cdots + b_m$.
- $n_{it} = \sum_{j=1-q}^{t} \delta_i(x_j)$, where $\delta_i(x) = 1$ if $x = s_i$ and zero

otherwise. In other words, n_{it} is the number of times of $x_j = s_i$ for $j = 1 - q, \ldots, t$.

- Function $\phi(\cdot)$ is the density function of the standard normal distribution.
- Throughout the article, $\pi(\cdot)$ and $\pi(\cdot \mid \cdot)$ are used to denote the densities or conditional densities.

3.2 Two Steps of the Sequential Imputation

For simplicity, assume that the system variance σ^2 is known. Experience shows that an approximate knowledge of σ^2 suffices. To initiate, presignals $x_{1-q}, \ldots, x_0$ are imputed from the prior distribution. Namely, a Θ is drawn from its prior distribution Dirichlet$(b_1, \ldots, b_m)$, and then $x_{1-q}, \ldots, x_0$ are drawn independently from the multinomial distribution with parameter Θ. The procedure is repeated N times to start the N imputation sequences. As illustrated in Section 2, two steps (A and B) are needed to implement the method. When a new signal y_{t+1} comes in, however, the two steps are usually done simultaneously with a reverse order in programming:

Step B. Compute the predictive probability of the observed value y_{t+1} and update the weight

$$w_{t+1} = w_t \cdot \pi(y_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_t).$$

Step A. Impute the unobserved value x_{t+1} by a draw from the distribution

$$\pi(x_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_{t+1}).$$

The predictive distributions $\pi(y_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_t)$ and $\pi(x_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_{t+1})$ can be written explicitly. Note that the posterior distribution of $\mathbf{H}$, after observing $\mathbf{X}_t$, $\mathbf{Y}_t$, is

$$[\mathbf{H} \mid \mathbf{X}_t, \mathbf{Y}_t] \sim N(H_t, \boldsymbol{\Sigma}_t),$$

where

$$\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathcal{X}_t^T \mathcal{X}_t \quad \text{and}$$

$$H_t = \boldsymbol{\Sigma}_t \left(\boldsymbol{\Sigma}_0^{-1} H_0 + \frac{1}{\sigma^2} \mathcal{X}_t^T \mathbf{Y}_t \right).$$

Thus by integrating out $\mathbf{H}$, we obtain that

$$[y_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_t, x_{t+1} = s_i] \sim N(\mu_{t+1[i]}, \sigma^2_{t+1[i]}),$$

where

$$\mu_{t+1[i]} = [\mathbf{x}_{t+1}^T H_t] \mid_{x_{t+1} = s_i} \tag{6}$$

and

$$(\sigma_{t+1[i]})^2 = [\mathbf{x}_{t+1}^T \boldsymbol{\Sigma}_t \mathbf{x}_{t+1} + \sigma^2] \mid_{x_{t+1} = s_i}. \tag{7}$$

Because x_{t+1} and $\mathbf{Y}_t$ are conditionally independent given $\mathbf{X}_t$, Steps B and A only involve computing

$$\begin{aligned} &\pi(y_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_t) \\ &= \sum_{i=1}^{m} \pi(y_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_t, x_{t+1} = s_i) \pi(x_{t+1} = s_i \mid \mathbf{X}_t) \\ &= \sum_{i=1}^{m} \frac{1}{\sigma_{t+1[i]}} \phi\left(\frac{y_{t+1} - \mu_{t+1[i]}}{\sigma_{t+1[i]}} \right) \frac{n_{it} + b_i}{q + t + B} \end{aligned} \tag{8}$$

and drawing x_{t+1} from the following updated discrete distribution

$$\pi(x_{t+1} = s_i \mid \mathbf{X}_t, \mathbf{Y}_{t+1}) \propto \frac{1}{\sigma_{t+1[i]}} \phi\left(\frac{y_{t+1} - \mu_{t+1[i]}}{\sigma_{t+1[i]}} \right) (n_{it} + b_i), \quad i = 1, \ldots, m. \tag{9}$$

Furthermore, H_t and $\boldsymbol{\Sigma}_t$ enjoy nice recursive updating formulas that greatly reduce the computational burden. Precisely, after x_{t+1} has been imputed, the inverse of the posterior covariance matrix of $\mathbf{H}$ can be updated by

$$\boldsymbol{\Sigma}_{t+1}^{-1} = \boldsymbol{\Sigma}_t^{-1} + \frac{1}{\sigma^2} \mathbf{x}_{t+1} \mathbf{x}_{t+1}^T.$$

Let μ_{t+1} and σ^2_{t+1} be the quantities computed by (6) and (7) for the imputed x_{t+1}. It is easily seen that

$$\begin{aligned} \boldsymbol{\Sigma}_{t+1} &= \boldsymbol{\Sigma}_t - \frac{\boldsymbol{\Sigma}_t \mathbf{x}_{t+1} \mathbf{x}_{t+1}^T \boldsymbol{\Sigma}_t}{\sigma^2 + \mathbf{x}_{t+1}^T \boldsymbol{\Sigma}_t \mathbf{x}_{t+1}} \\ &= \boldsymbol{\Sigma}_t - \frac{1}{\sigma^2_{t+1}} \boldsymbol{\Sigma}_t \mathbf{x}_{t+1} \mathbf{x}_{t+1}^T \boldsymbol{\Sigma}_t. \end{aligned} \tag{10}$$

For the posterior mean of $\mathbf{H}$ at time $t + 1$,

$$\begin{aligned} H_{t+1} &= \boldsymbol{\Sigma}_{t+1} \left(\boldsymbol{\Sigma}_t^{-1} H_t + \frac{1}{\sigma^2} y_{t+1} \mathbf{x}_{t+1} \right) \\ &= H_t + \frac{y_{t+1} - \mathbf{x}_{t+1}^T \mathbf{H}_t}{\sigma^2 + \mathbf{x}_{t+1}^T \boldsymbol{\Sigma}_t \mathbf{x}_{t+1}} \boldsymbol{\Sigma}_t \mathbf{x}_{t+1} \\ &= H_t + \frac{y_{t+1} - \mu_{t+1}}{\sigma^2_{t+1}} \boldsymbol{\Sigma}_t \mathbf{x}_{t+1}. \end{aligned} \tag{11}$$

The sequential imputation algorithm for blind deconvolution can be summarized as follows:

Step 0. To initiate sequential imputations, draw $\Theta(1), \ldots, \Theta(N)$ from the prior distribution Dirichlet$(b_1, \ldots, b_m)$, and for each $\Theta(j), j = 1, \ldots, N$, impute $\mathbf{x}_0(j) = \{x_{1-q}(j), \ldots, x_0(j)\}$ with each $x_t(j)$ iid drawn from the distribution $\Theta(j)$ on $s_1, \ldots, s_m$. Set $w_0(j) = 1$, for $j = 1, \ldots, N$.

Step 1. Repeat the following two procedures for every new observation starting from $t = 0$:

- Compute $\pi(y_{t+1} \mid \mathbf{X}_t(j), \mathbf{Y}_t)$ by (8), let $w_{t+1}(j) = w_t(j) \pi(y_{t+1} \mid \mathbf{X}_t(j), \mathbf{Y}_t)$, and impute $x_{t+1}(j)$ by a draw from $\pi(x_{t+1} \mid \mathbf{X}_t(j), \mathbf{Y}_{t+1})$ computed by (9), for $j = 1, \ldots, N$.
- For $j = 1, \ldots, N$, update $H_{t+1}(j)$ and $\boldsymbol{\Sigma}_{t+1}(j)$, $j = 1, \ldots, N$, by (11) and (10), in which $\mathbf{x}_{t+1}$ is replaced by $\mathbf{x}_{t+1}(j)$.

Step 2. For a fixed number of observed signals $y_1, \ldots, y_n$, the final importance weights $w(j)$ is equal to $w_n(j)$, for $j = 1, \ldots, N$. An estimate of the posterior conditional expectation of the input signal x_t is then

$$\hat{x}_t = \frac{w(1) x_t(1) + \cdots + w(N) x_t(N)}{w(1) + \cdots + w(N)}, \quad t = 1 - q, \ldots, n. \tag{12}$$

Similarly, for $i = 1, \ldots, m$, the marginal posterior probability can be approximated by

$$P(x_t = s_i \mid \mathbf{Y}_n) \approx \frac{w(1)I[x_t(1) = s_i] + \cdots + w(N)I[x_t(N) = s_i]}{w(1) + \cdots + w(N)}, \quad t = 1 - q, \ldots, n, \quad (13)$$

where $I(\cdot)$ is the indicator function. Therefore, an alternative estimate of the input signal x_t is the *maximum a posteriori* (MAP) estimate $\tilde{x}_t = s_{i_0}$; that is, a maximizer of $P(x_t = s_i)$. The posterior mean vector and covariance matrix of $\mathbf{H}$ are provided similarly by the weighted averages of $H_n(j), j = 1, \ldots, N$, and $\mathbf{\Sigma}_n(j), j = 1, \ldots, N$.

Remarks.

1. It is noted that the proposed sequential imputation procedure can be easily modified to accommodate complex variables, which are common in digital communication.

2. In many practical situations, the mechanism that generates the unobserved input signals is known. This information should be used. For example, if it is known that the x are iid *uniformly* distributed on $\{s_1, \ldots, s_m\}$, then formulas (8) and (9) become

$$\pi(y_{t+1} \mid \mathbf{X}_t, \mathbf{Y}_t) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{\sigma_{t+1[i]}} \phi\left(\frac{y_{t+1} - \mu_{t+1[i]}}{\sigma_{t+1[i]}}\right)$$

and

$$\pi(x_{t+1} = s_i \mid \mathbf{X}_t, \mathbf{Y}_{t+1}) \propto \frac{1}{\sigma_{t+1[i]}} \phi\left(\frac{y_{t+1} - \mu_{t+1[i]}}{\sigma_{t+1[i]}}\right).$$

If the signals are transmitted according to a Markov transition function, then the predictive distribution of y_{t+1} and the sampling distribution of x_{t+1} can be revised accordingly.

4. REJUVENATION PROCEDURE

4.1 Rejuvenated Sequential Imputation

The importance sampling weight $w(j)$ evaluates the quality of an imputed signal. A relatively small importance weight for the imputed sequence implies that it is drawn far from the main body of the posterior, and the resulting imputation will have a small effect in the final estimation. Such an imputed sequence is said to be "ineffective." If there are too many ineffective imputations, then the Monte Carlo scheme will be inefficient. This can be detected by observing a large coefficient of variation in the importance weights (defined later). Kong et al. (1994) showed that the importance weight resulting from sequential imputation forms a martingale sequence. As more and more data are processed, the coefficient of variation of the weights *increases;* that is, the number of ineffective imputations increases. To cope with this difficulty, we introduce a restarting strategy, called *rejuvenation.*

Let the current N independent imputations be $\{\mathbf{X}_t(1), \ldots, \mathbf{X}_t(N)\}$, with weights $\{w_t(1), \ldots, w_t(N)\}$. To rejuvenate sequential imputations at time $t + 1$ (after observing y_{t+1}), we first compute the weight $w_{t+1}(j)$ for $j = 1, \ldots, N$ and generate N iid random numbers, $j_1, \ldots, j_N$, from the multinomial distribution $P(j_k = j) \propto w_{t+1}(j)$ for $j = 1, \ldots, N$. Then the N original imputations are replaced by $\mathbf{X}_t^*(k) = \mathbf{X}_t(j_k)$ for $k = 1, \ldots, N$, and the weights are replaced by $w_{t+1}^*(k) = 1/N$; that is, identical weights. After the replacement, the $x_{t+1}(j)$ are imputed according to $\pi(x_{t+1} \mid \mathbf{X}_j^*(j), \mathbf{Y}_{t+1})$. The intuition is that the "good" imputations with large weights have high probability to be resampled, whereas the ineffective imputations have high probability to be discarded. When N is large, it is easy to show that $\mathbf{X}_t^*$ approximately follows $\pi(\mathbf{X}_t \mid \mathbf{Y}_{t+1})$, which is "closer" to the target distribution $\pi(\mathbf{X}_t \mid \mathbf{Y}_n)$ than $\pi(\mathbf{X}_t \mid \mathbf{Y}_t)$ is.

Rejuvenation can be done at any stage of the process. But rejuvenating too often adds computational burden, and information may be lost in the process. On the other hand, rejuvenating too rarely may result in loss of efficiency. We found from simulation results (shown in Sec. 6) that rejuvenation is important, but the threshold chosen is not crucial.

To study the actual benefit of rejuvenation, we need to quantify the efficiency of an importance sampling scheme. Geweke (1989) provided such a discussion when a particular function of interest is considered. Here, because we consider an importance sampling scheme that is used to estimate many quantities of interest, we adopt a "rule of thumb" as discussed below.

Suppose that in an importance sampling scheme, $\pi(X)$ is the target distribution and $p(X)$ is the trial distribution from which we draw independent samples. The *importance weight* is defined as $W(X) = \pi(X)/p(X)$, in which X is drawn from $p(X)$. In the following discussion we use E_p and var_p to indicate the expectation and variance taken under p, and use E_π and var_π to indicate those taken under π. Then $\text{var}_p\{W(X)\}$ can be viewed as a measure of the chi-squared distance between π and p. The larger the value of $\text{var}_p(W)$, the less efficient the corresponding importance sampling scheme. This variance cannot be computed directly in practice, however. Now suppose that $w(1), \ldots, w(N)$ are the normalized importance weights (i.e., $\sum w(j) = 1$) in an importance sampling scheme; then the *coefficient of variation,* $C(w)$, is computed as

$$C^2(w) = \frac{1}{N} \sum_{j=1}^{N} \{Nw(j) - 1\}^2.$$

When N is large, $C^2(w)$ is a reasonable approximation of $\text{var}_p\{W(X)\}$ because of the fact that $E_p\{W(X)\} = 1$. In the discussion of Section 4.2, we use $C^2(w)$ and $\text{var}_p(W)$ interchangeably.

A heuristic measure of efficiency for the sampling scheme is expressed by the "rule of thumb":

$$\text{Effective sample size (ESS)} = \frac{\text{total Monte Carlo sample size } (N)}{1 + C^2(w)}. \quad (14)$$

The intuition of this "rule" is as follows. Suppose that N samples have been drawn, each associated with a weight $w(k)$ that is either zero or $1/M$, where M is the number of nonzero weights. Then $C^2(w) = N/M - 1$, or $M = N/(1 + C^2(w))$. On the other hand, the effective sample size is the number

of samples with nonzero weights; that is, ESS = $M = N/(1 + C^2(w))$. Therefore, a large $C^2(w)$ indicates that the effective sample size is small and memory and computation are being wasted by carrying too many ineffective sequences.

4.2 Losses and Gains From Rejuvenation

In this section the losses and gains from the rejuvenation procedure are studied. To simplify the presentation, we study a case where only two observations are involved. Its analysis will shed some light on the general situation of sequential imputation.

Suppose that V_1 and V_2 are incomplete observations with missing parts U_1 and U_2. We are interested in estimating $\mu_1 = E_\pi\{g_1(U_1)|V_1, V_2\}$ and $\mu_2 = E_\pi\{g_2(U_2)|V_1, V_2\}$ by Monte Carlo integration, where g_1 and g_2 are some general functions. The plain sequential imputation can be applied to this problem and involves two stages. First, N iid samples $U_1(1), \ldots, U_1(N)$ are drawn from $\pi(U_1|V_1)$. Note that the importance weights at time $t = 1$ are $w_1(j) = \pi(V_1) \propto 1$. Second, the new weights $w_2(j) \propto \pi(V_2|U_1(j), V_1)$ are computed and renormalized to 1, and $U_2(j)$ is independently drawn from $\pi(U_2|U_1(j), V_1, V_2)$ for $j = 1, \ldots, N$. Note that $w_2(j)$ can be regarded both as the importance weight for $U_1(j)$ and as the importance weight for $(U_1(j), U_2(j))$ jointly. In the rejuvenation procedure we resample with replacement from the $U_1(j)$ with the probability $w_2(j)$ to obtain $U_1^*(1), \ldots, U_1^*(N)$, and draw $U_2^*(j)$ from $\pi(U_2|U_1^*(j), V_1, V_2)$ for $j = 1, \ldots, N$, each associated with equal weight.

We argue that rejuvenation loses efficiency in estimating μ_1 but gains efficiency in estimating μ_2 in many situations, especially when $C^2(w_2)$ is large, in which case the loss in the first stage is comparatively small.

4.2.1 Losses From Rejuvenation. An estimate of μ_1 from sequential imputation is

$$\hat{\mu}_1 = \sum_{j=1}^{N} w_2(j) g_1(U_1(j)),$$

where $w_2(j)$ are the normalized importance weights. Another estimate can be constructed after rejuvenation; that is,

$$\tilde{\mu}_1 = \frac{1}{N} \sum_{j=1}^{N} g_1(U_1^*(j)).$$

To simplify the notation, we use π to denote the true posterior marginal distribution $\pi(U_1|V_1, V_2)$ and let p be the trial distribution $\pi(U_1|V_1)$, from which we draw samples $U_1(j)$. Let $W(U_1) = \pi(U_1)/p(U_1)$; then when N is large, the normalized weight w_2 is an approximation of W/N.

By the variance decomposition, we have

$$\operatorname{var}_p(\tilde{\mu}_1) = \operatorname{var}_p\{E(\tilde{\mu}_1|U_1(1), \ldots, U_1(N))\} + E_p\{\operatorname{var}(\tilde{\mu}_1|U_1(1), \ldots, U_1(N))\},$$

where the conditional expectation and variance are taken with respect to the random resampling. The first term on the right side is equal to $\operatorname{var}_p(\hat{\mu}_1)$, whereas the second term is equal to

$$\frac{1}{N} E_p\left[\sum_{j=1}^{N} w_2(j)\{g_1(U_1(j)) - \hat{\mu}_1\}^2\right],$$

which is the extra variation introduced by the resampling. But the key quantity here is the ratio of this extra variation to the intrinsic variation $\operatorname{var}_p(\hat{\mu}_1)$. It is seen that

$$\frac{\operatorname{var}_p(\tilde{\mu}_1)}{\operatorname{var}_p(\hat{\mu}_1)} \approx 1 + \frac{\frac{1}{N}[E_p[W\{g_1(U_1) - \mu_1\}^2] - \operatorname{var}_p(\hat{\mu}_1)]}{\operatorname{var}_p\{\hat{\mu}_1\}} = \frac{N-1}{N} + \frac{E_p[W\{g_1(U_1) - \mu_1\}^2]}{\operatorname{var}_p\{Wg_1(U_1)\}}.$$

Now that $E_p[W\{g_1(U_1) - \mu_1\}^2] = \operatorname{var}_\pi\{g_1(U_1)\}$, and by an argument of Liu (1992),

$$\operatorname{var}_p\{Wg_1(U_1)\} = \operatorname{var}_\pi\{g_1(U_1)\}\{1 + \operatorname{var}_p(W) + r\}. \quad (15)$$

The term r tends to be small when the function g is relatively flat but is not guaranteed to be small in general. A nice consequence of the approximation (15) is the following "rule of thumb" for a reference of efficiency:

$$\frac{\operatorname{var}_p\{\tilde{\mu}_1\}}{\operatorname{var}_p\{\hat{\mu}_1\}} \approx \frac{N-1}{N} + \frac{1}{1 + \operatorname{var}_p(W)}. \quad (16)$$

When $W \equiv 1$ (i.e., the trial density is the same as the target density), the foregoing formula coincides with the familiar result for resampling at random with replacement. Because N is usually large, we regard $\operatorname{var}_p(W)$ as equivalent to $C^2(W)$. This formula implies that when the coefficient of variation in the importance weights is large, rejuvenation does not sacrifice much.

4.2.2 Gains From Resampling. A natural estimate of μ_2 is

$$\hat{\mu}_2 = \sum_{j=1}^{N} w_2(j) g_2(U_2(j)).$$

The conditional variance of this estimator, given the first-stage imputation $U_1(j)$ for $j = 1, \ldots, N$, is

$$\operatorname{var}(\hat{\mu}_2) = \sum_{j=1}^{N} w_2^2(j)\sigma^2(j), \quad (17)$$

where $\sigma^2(j) = \operatorname{var}_\pi\{g_2(U_2)|U_1(j), V_1, V_2\}$. We also let $\mu(j) = E_\pi\{g_2(U_2)|U_1(j), V_1, V_2\}$.

Suppose that we resample with replacement to obtain $U_1^*(1), \ldots, U_1^*(N)$, among which there are $M_1 U_1(1)$'s, $\ldots$, and $M_N U_1(N)$'s such that $M_1 + \cdots + M_N = N$. Then another estimator is

$$\tilde{\mu}_2 = \frac{1}{N} \sum_{j=1}^{N} g_2(U_2^*(j)),$$

and its variance, conditional on the first-stage sampling $U_1(1), \ldots, U_1(N)$, is computed by the variance decomposition formula as

$$\operatorname{var}(\tilde{\mu}_2) = \operatorname{var}\left(\frac{1}{N}\sum_{j=1}^{N} M_j\mu(j)\right) + \frac{1}{N^2}E\left(\sum_{j=1}^{N} M_j\sigma^2(j)\right)$$
$$= \frac{1}{N}\left\{\sum_{j=1}^{N} w_2(j)\mu(j)^2 - \left(\sum_{j=1}^{N} w_2(j)\mu(j)\right)^2\right\}$$
$$+ \frac{1}{N}\sum_{j=1}^{N} w_2(j)\sigma^2(j). \tag{18}$$

It is not easy to compare (17) and (18) in general. Here we compare $\operatorname{var}(\hat{\mu}_2)$ and $\operatorname{var}(\tilde{\mu}_2)$ in several extreme cases.

1. When all the weights $w_2(j)$ are equal, $w_2(j) = 1/N$, then V_2 is conditionally independent of U_1 given V_1, and it is easily seen that $\operatorname{var}(\hat{\mu}_2) \le \operatorname{var}(\tilde{\mu}_2)$.
2. When the $\mu(j)$ and the $\sigma(j)$ are all the same (in this case, U_2 is conditionally independent of U_1 given V_1 and V_2), it can be shown that $\operatorname{var}(\hat{\mu}_2) \ge \operatorname{var}(\tilde{\mu}_2)$ by the Cauchy–Schwarz inequality.
3. When the weights $w_2(j)$ are either zero or a constant, and there are $K < N$ nonzero weights, each being $1/K$, assume that the $\mu(j)$ corresponding to the nonzero $w_2(j)$ are close to each other, so that the first term of (18) can be neglected (e.g., when U_2 takes binary value). Then $\operatorname{var}(\hat{\mu}_2) = \sum_j \sigma^2(j)/K^2$ and $\operatorname{var}(\tilde{\mu}_2) \approx \sum_j \sigma^2(j)/NK$, implying that the rejuvenated scheme is about $(1 + C^2(W))$ times more efficient than the standard one.

More generally, because

$$\sum_{j=1}^{N} w_2(j)\mu(j)^2 - \left(\sum_{j=1}^{N} w_2(j)\mu(j)\right)^2 \approx \operatorname{var}_\pi[E_\pi\{g_2(U_2)\,|\,U_1, V_1, V_2\}\,|\,V_1, V_2]$$

and

$$\sum_{j=1}^{N} w_2(j)\sigma^2(j) \approx E_\pi[\operatorname{var}_\pi\{g_2(U_2)\,|\,U_1, V_1, V_2\}\,|\,V_1, V_2],$$

we have that $\operatorname{var}(\tilde{\mu}_2) \approx \operatorname{var}_\pi\{g_2(U_2)\,|\,V_1, V_2\}/N$ for the rejuvenated second-stage sampling. Similarly, the second-stage sampling variance for $\hat{\mu}_2$ is

$$\operatorname{var}(\hat{\mu}_2) \approx \frac{1}{N}E_\pi[W \operatorname{var}_\pi\{g_2(U_2)\,|\,U_1, V_1, V_2\}\,|\,V_1, V_2],$$

where $W = \pi(U_1\,|\,V_1, V_2)/\pi(U_1\,|\,V_1)$. Therefore, comparing the two schemes amounts to comparing

$$E_\pi[W \operatorname{var}_\pi\{g_2(U_2)\,|\,U_1, V_1, V_2\}\,|\,V_1, V_2]$$

and

$$\operatorname{var}_\pi\{g_2(U_2)\,|\,V_1, V_2\}.$$

Which one is larger depends on the problem at hand. A rule of thumb, however, is that the latter quantity is usually smaller when $C^2(W)$ is large and W is positively correlated with the conditional variance $\operatorname{var}_\pi\{g_2(U_2)\,|\,U_1, V_1, V_2\}$.

We also note that the potential gains are not limited to U_2. If there are more observations coming in, then all the future imputations and estimations will gain efficiency from rejuvenation. More precisely, after rejuvenation, each of the $U_1^*(j)$ follows approximately the distribution $\pi(U_1\,|\,V_1, V_2)$. When a new observation V_3 is considered, the trial density $\pi(U_1, U_2\,|\,V_1, V_2)$ associated with the rejuvenated procedure is closer to the target density $\pi(U_1, U_2\,|\,V_1, V_2, V_3)$ than is the trial density $\pi(U_1\,|\,V_1)\pi(U_2\,|\,U_1, V_1, V_2)$ induced by the plain sequential imputation. In addition, if the expectation of a function of all the imputed U's, say $E_\pi(\theta\,|\,V_1, \ldots, V_n)$, in a Bayesian problem is estimated by using the imputations $U_1(j), \ldots, U_n(j)$, then rejuvenation becomes a good strategy when $C^2(W)$ is large, because the loss is minimal and the potential gain is large.

Remark. An alternative method of rejuvenation is to select the rejuvenated imputation sequences by deterministic-plus-residual sampling. Specifically, let the current N independent imputations be $\{\mathbf{X}_t(1), \ldots, \mathbf{X}_t(N)\}$. To rejuvenate after observing y_{t+1}, we compute the normalized weights $\{w_{t+1}(1), \ldots, w_{t+1}(N)\}$. Let $n_i = [Nw_{t+1}(i)]$, the greatest integer less than or equal to $Nw_{t+1}(i)$. Then n_i copies of $X_t(i)$ are retained after rejuvenation. This is done for $i = 1, \ldots, N$ to obtain $\sum_{i=1}^{N} n_i$ rejuvenated imputations. The remaining $N - \sum_{i=1}^{N} n_i$ rejuvenated imputations are selected by sampling from $\{\mathbf{X}_t(1), \ldots, \mathbf{X}_t(N)\}$, with replacement with probabilities proportional to $w_{t+1}(i) - n_i/N$. This scheme can effectively bring down the extra variation introduced by resampling. For example, in the ideal case when all the $Nw_{t+1}(i)$ are integers, there is no loss from rejuvenation in the first stage; that is, $\hat{\mu}_1$ and $\tilde{\mu}_1$ are identical. On the other hand, the first part of formula (18) is reduced to zero. By writing $w_2(j)$ as n_j/N, we immediately observe $\operatorname{var}(\tilde{\mu}_2) \le \operatorname{var}(\hat{\mu}_2)$, which implies that in this case the rejuvenation is always beneficial. A theoretical analysis for the general case is difficult, however.

5. OTHER MODIFICATIONS

5.1 Early-Stage Imputation With Gibbs Sampling

Sequential imputation is most useful when the data are collected sequentially. But a practical drawback is that a large N is required when the system is large; that is, when m and q are large. This occurs because the early imputations are primarily based on our priors and are very likely to lose predictability for the future observations. To retain a reasonable number of effective imputations while information accumulates, one would have to start with a large number of imputation sequences. An alternative is to observe several output values from sequence before starting sequential imputation. With enough information, "good" starting imputations for the unobserved signal can be found, so that N can be kept small. In other words, instead of starting sequential imputation early on, we observe the first batch of signals $y_1, \ldots, y_{n_1}$ and try to impute $x_{1-q}(j), \ldots, x_{n_1}(j)$ from the partial *posterior* distribution $\pi(\mathbf{X}_{n_1}\,|\,\mathbf{Y}_{n_1})$. The Gibbs sampler is a natural choice for this purpose.

More precisely, Gibbs sampling iterations are used to process the first batch of signals $y_1, \ldots, y_{n_1}$ to produce N imputed complete data $(\mathbf{X}_{n_1}(j), \mathbf{Y}_{n_1})$, $j = 1, \ldots, N$, each as-

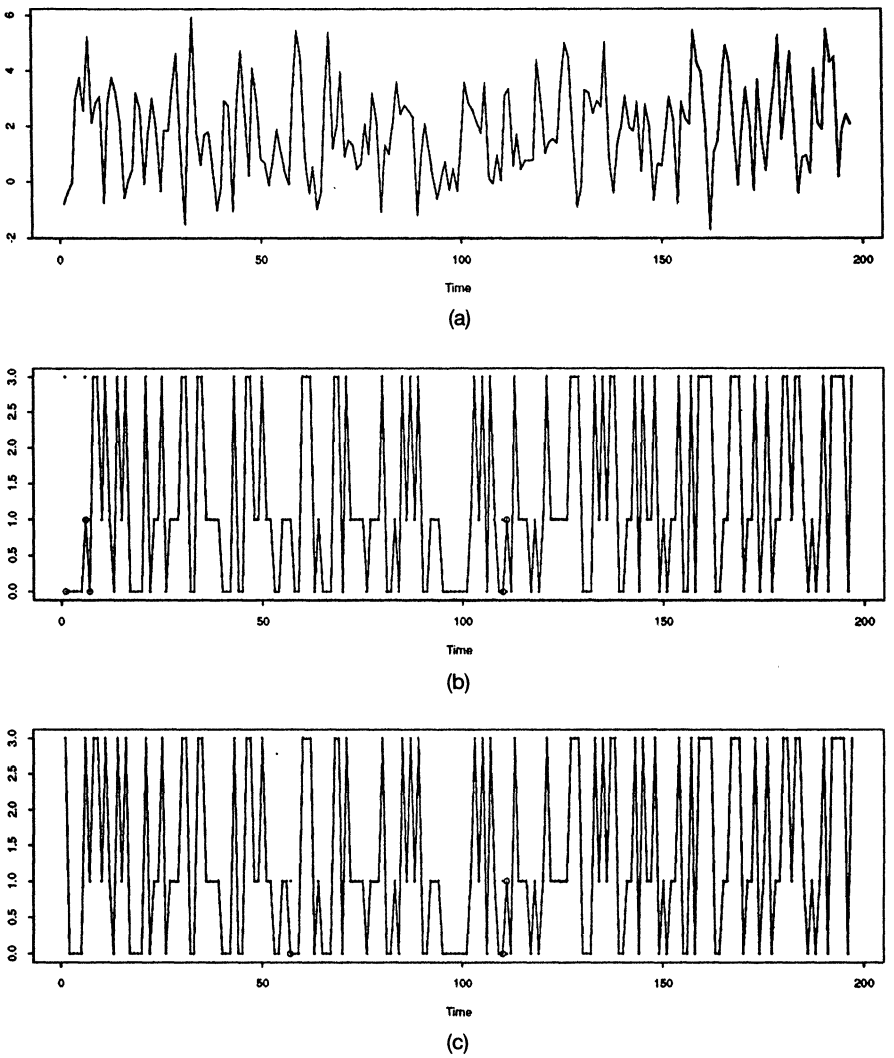

Figure 1. Example I: (a) Output Signal, First 200 Observations; (b) Estimates of the Input Signals Using Real Time Restoration; (c) Estimates of the Input Signals Using Final Weights.

sociated with equal weight $1/N$. (See Chen and Li 1993 for details on implementing the Gibbs sampler.) When new signals $y_{n_1+1}, y_{n_1+2}, \ldots$ are collected, $x_{n_1+1}(j), x_{n_1+2}(j), \ldots, j = 1, \ldots, N$, are imputed by using the sequential imputation method described in Section 3.2, treating the estimated system variance as known.

Our experiences show that this approach works well when the system variance is unknown or the number of possible levels m of the input signal is large. In addition, we observe that the combination strategy is able, with the help of importance weight, to discriminate good imputations from bad ones resulting from the Gibbs sampler and, consequently, avoids being trapped in a local mode of the posterior distribution. Chen and Li (1993) reported that the Gibbs sampler is often trapped in a shift mode. A further implication of this observation is that in our use of the Gibbs sampler, the global convergence is not required for each independent Gibbs sampling chain.

5.2 Time Delay Output in Real Time Restoration

It is most desirable to estimate the input signals by using the final weighted average as indicated in (12) or (13). To carry out real time signal restoration, however, the x_t must be estimated during the process. One obvious solution to this problem is to instantaneously estimate the signals by computing the weighted average of the $x_t(j)$ using the current weights $w_t(j)$ for time t. But this is usually very inaccurate at early stages of processing. A better alternative is to allow for a d-lag delay, because future observations $y_{t+1}, \ldots, y_{t+d}$ contain substantial information about the current signal x_t. Note that in general, $\pi(\mathbf{X}_t | \mathbf{Y}_{t+d})$ is closer to the target distribution $\pi(\mathbf{X}_t | \mathbf{Y}_n)$ than $\pi(\mathbf{X}_t | \mathbf{Y}_t)$ is. Specifically, if we are allowed to have d-lag delay, the signal x_t is estimated using $x_t(j)$ with importance weights at time $t + d$. Thus at time $t + d$, the posterior expectation of the input signal x_t (d-lag backwards) is estimated by using the *current* importance weights $w_{t+d}(j), j = 1, \ldots, N$; that is, $\hat{x}_t = \{\sum_{j=1}^{N} w_{t+d}(j)x_t(j)\}/W_{t+d}$, where $W_{t+d} = \sum_{j=1}^{N} w_{t+d}(j)$. The MAP estimate $\hat{x}_t$ is approximated by s_{i_0}, a maximizer of the approximate marginal posterior probability function $\pi(x_t = s_i | Y_{t+d})$ that can be computed by (13) with the $w(j)$ replaced by the $w_{t+d}(j)$. The delay time was chosen to be $q + 1$ in our examples that follow.

Table 1. Comparing Different Thresholds for Rejuvenation

ESS	1	2	3	4	5	10	20	38	74	130	230	333	500	667	883	1,000
Real time	31	5	5	3	3	5	4	6	7	6	6	5	5	5	4	6
Final	19	3	3	1	4	4	4	1	1	4	2	3	2	2	1	3
# restarts	0	3	3	4	4	5	5	7	9	13	14	17	22	31	44	1,000

6. SIMULATION STUDY

6.1 Direct Application of Sequential Imputation

This example is taken from Chen and Li (1993), who assumed the states of signals to be 0, 1, and 3. The blurring equation is

$$y_t = x_t + .8x_{t-1} - .4x_{t-2} + \varepsilon_t. \tag{19}$$

The input signals are iid uniform from $\{0, 1, 3\}$. The standard deviation of the white noise, which is roughly .3, was chosen to make the signal-to-noise ratio fixed at 15 dB.

One thousand observations were simulated from this system. We carried out the sequential imputation procedure using $N = 1{,}000$ imputation sequences. The prior distribution for the system coefficient $\mathbf{H}$ was assumed to be a product of independent $N(0, 1{,}000)$, and the distribution of the x_t were assumed known. The threshold for rejuvenation was set to ESS = 3, where ESS is defined in (14). For real time signal restoration, the delay time was chosen to be 3. We used the MAP estimate for the input signals. Note also that the weighted average estimate $\hat{x}_t$ in (12) works well in most situations, because the posterior probability $\pi(x_t = s_i \mid y)$ often concentrates on one value.

A signal is said to be *misclassified* if the estimated signal differs from the true signal. We tested our algorithm on the first 50, 100, 200, 500, and 1,000 of the simulated observations. The real time restoration misclassified 3, 3, 5, 7, and 12 signals, whereas the estimation using the final weights computed by (13) misclassified 0, 0, 3, 5, and 12 signals. Figure 1a displays the first 200 observations of the output series. Figure 1b shows a real time restoration for the first 200 observations. The dots are the true signals, and the lines are the estimates. Figure 1c illustrates the estimates using the final weights.

To study the effect of rejuvenation, we applied the sequential imputation procedure with different restarting thresholds for the first 200 simulated observations. Table 1 summarizes the result of using different threshold values (i.e., ESS) for rejuvenation found in the first row. The second row contains the numbers of misclassified signals for real time restoration with delay time 3. The numbers in the third row are the misclassified signals using final weight. The fourth row records the number of times that the algorithm rejuvenates for each threshold value.

Here a threshold of ESS = 1 implies that we never rejuvenate. From Table 1, it is obvious that rejuvenation is important. But the effect of the threshold value (other than 1) is very small regarding the final results. Because rejuvenation requires computing time, a small ESS threshold is preferred.

To illustrate the effect of time delay in the real time restoration, we simulated 200 signal sequences, each with 200 observations, from the system (19). Figure 2 is the boxplot of the number of misclassifications for real time restorations with delay time 0 to 4 and the restoration using the final weights, where $N = 1{,}000$ for all the simulations. We can see that a certain delay time is needed to get acceptable results. For the estimates using the final weights, the sequential imputation procedure misclassified more than 10 input signals 22 times in the 200 simulated data sets.

6.2 Combining Gibbs Sampling and Sequential Imputations

In this section we study a large system,

$$y_t = .9162x_t - .1833x_{t-1} + .4812x_{t-2} - .1987x_{t-3} + \varepsilon_t,$$

where the x_t were randomly drawn from 16 possible levels $\{-15, -13, \ldots, -1, 1, \ldots, 13, 15\}$ with equal probabilities. The noise ε_t were iid from a normal distribution with mean zero and standard deviation .3144, with a signal-to-noise ratio of 30 dB. A similar example has also been used by Hatzinakos and Nikias (1991) and Chen and Li (1993). Figure 3(a) shows the first 200 output signals y_t, $(t = 1, \ldots, 198)$. Because of the large number of possible levels that x can take, the information at early stages is poor. Plain sequential imputation is not practical in this situation for the reasons discussed in Section 5.2. Note that in assigning the presignal values x_{-2}, x_{-1}, x_0, there are $16^3 = 4{,}096$ possible combinations. Experience shows that 5,000 imputation se-

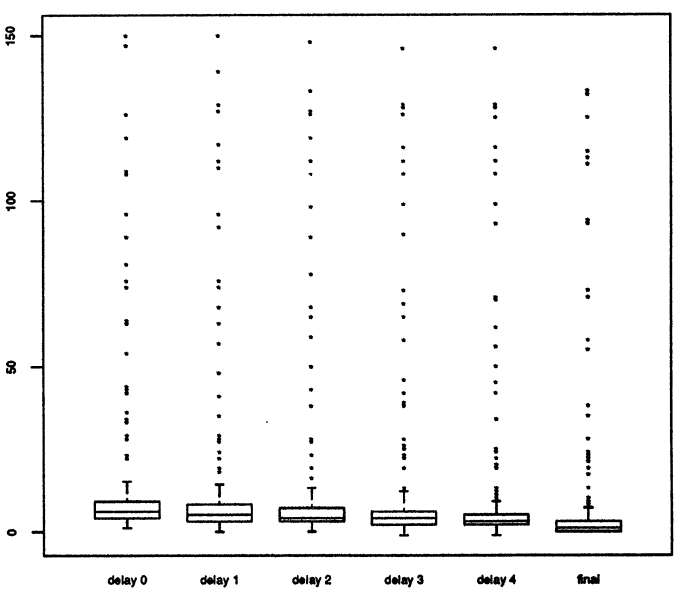

Figure 2. The Boxplot of the Number of Signals Misspecified by Real Time Restoration With Delay Time 0 to 4 and Using the Final Weight in 200 Simulated Sequences.

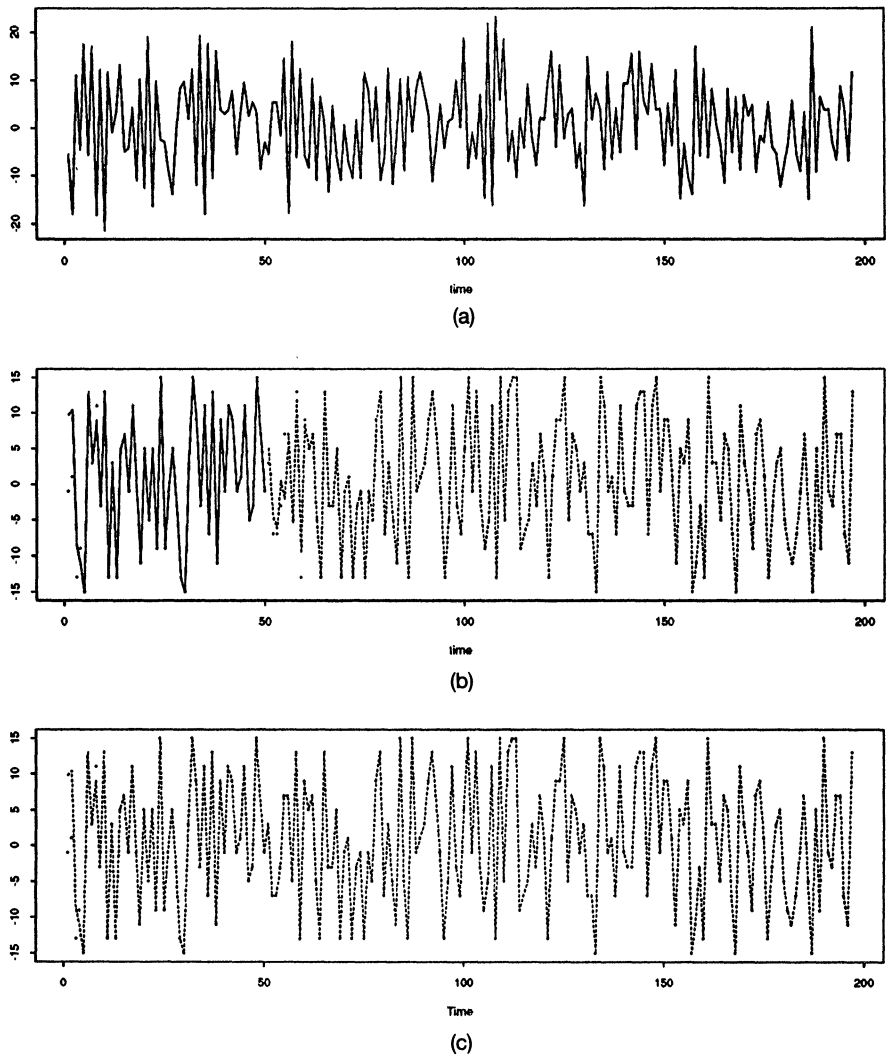

Figure 3. Example II: (a) Output Signal, First 200 Observations; (b) Estimates of the Input Signals Using Real Time Restoration, With the First 50 Signals Estimated Using the Gibbs Sampler With Sequential Imputation Weights; (c) Estimates of the Input Signals Using Final Weight.

quences are not enough for such a system. The procedure proposed in Section 5.2 is more appropriate. For this example, 10 independent Gibbs sampling chains were applied to the first 55 observations, each with 900 iterations. The first 500 iterations for each chain were discarded, and 20 imputations were saved from the rest of the 400 iterations by choosing 1 in every 20 iterations. The imputations of the last 5 signals were discarded, because the Gibbs sampler is usually not very accurate in estimating the last several signals (Chen and Li 1993). Thus at the end of the Gibbs sampling procedure, 200 imputed signal sequences, each 50 signals long, were constructed. The posterior mean vector and covariance matrix of $\mathbf{H}$ were then computed and saved for each of the 200 imputed sequences. Sequential imputations (i.e., Steps 1 and 2) were performed to process the rest of the data.

Figure 3 shows some of the results. The dots are the true signals, and the solid line in Figure 3b indicates the signals recovered using the Gibbs sampler imputations with the final sequential imputation weights. The dashed lines are the real-time signal restoration when performed with sequential imputations. The total number of misclassified signals was 14. Figure 3c plots the estimated input signals using the final weight. It misclassified 6 signals for the entire sequence of 204 signals.

7. SUMMARY

In this article we proposed a sequential imputation algorithm for blind deconvolution. The basic idea of the method is to treat the unobserved signals as missing data and to impute them sequentially. Rubin (1987b) suggested multiple imputations to incorporate the extra variations caused by missing data. In our case, however, multiple imputed sequences are needed to enhance a competition among them in the predictability of the future observables, so as to direct the signal recovery. A restarting strategy termed *rejuvenation* is introduced to reduce the Monte Carlo variations when the coefficient of variation in the sequential imputation weight is large. We expect the method to be useful in many other situations.

The advantage of the algorithm is its ability to recover system coefficients sequentially without a training sequence.

It also requires less of the system than methods of other investigators. Our simulations show that the procedure works well in complicated systems with small sample sizes. Furthermore, while the Gibbs sampler can become trapped in a local mode (Chen and Li 1993), sequential imputation does not. This is partly because N independent sequences are simulated and competing. Those sequences trapped in a local mode will lose their predictabilities gradually, and the ones with better predictabilities will eventually dominate.

[Received July 1993. Revised June 1994.]

REFERENCES

Chen, R., and Li, T. H. (1993), "Blind Restoration of Linearly Degraded Discrete Signals Via Gibbs Sampler," technical report, Texas A&M University, Dept. of Statistics.

Cheng, Q. (1990), "Maximum Standardized Cumulant Deconvolution of Non-Gaussian Linear Processes," *The Annals of Statistics,* 18, 1745–1783.

Donoho, D. (1981), "On Minimum Entropy Deconvolution," in *Applied Time Series Analysis, II,* ed. D. F. Findley, New York: Academic Press.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* 85, 398–409.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrika,* 57, 1317–1339.

Giannakis, G. B., and Mendel, J. M. (1989), "Identification of Nonminimum Phase Systems Using Higher Order Statistics," *IEEE Transactions on Acoustic, Speech, and Signal Processing,* ASSP-37, 665–674.

Godard, D. N. (1980), "Self-Covering Equalization and Carrier Tracking in Two-Dimensional Data Communication System," *IEEE Transactions on Communication,* COM-28, 1867–1875.

Hatzinakos, D., and Nikias, C. L. (1991), "Blind Equalization Using a Tricepstrum-Based Algorithm," *IEEE Transactions on Communication,* COM-39, 669–682.

Kong, A., Liu, J. S., and Wong, W. H. (1994), "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association,* 89, 278–288.

Li, T. H. (1992), "Blind Identification and Deconvolution of Linear Systems Driven By Binary Random Sequences," *IEEE Transactions on Information Theory,* 38, 26–38.

——— (1993), "Estimation and Blind Deconvolution of AR Systems With Nonstationary Binary Inputs," *Journal of Time Series Analysis,* 14, 575–588.

Lii, K. S., and Rosenblatt, M. (1982), "Deconvolution and Estimation of Transfer Function Phase and Coefficients for Non-Gaussian Linear Processes," *The Annals of Statistics,* 10, 1195–1208.

Liu, J. S. (1992), "Metropolized Independent Sampling With Comparisons to Rejection Sampling and Importance Sampling," Technical Report R-427, Harvard University, Dept. of Statistics.

Rubin, D. B. (1987a), "A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information are Modest: The SIR Algorithm," *Journal of the American Statistical Association,* 52, 543–546.

——— (1987b), *Multiple Imputation for Nonresponse in Surveys,* New York: John Wiley.

Sato, Y. (1975), "A Method for Self-Recovering Equalization for Multilevel Amplitude-Modulation System," *IEEE Transactions on Communication,* COM-23, 679–682.