

Convergence Analyses and Comparisons of Markov Chain Monte Carlo Algorithms in Digital Communications

Rong Chen, *Member, IEEE*, Jun S. Liu, and Xiaodong Wang, *Member, IEEE*

Abstract—Recently, Markov chain Monte Carlo (MCMC) methods have been applied to the design of blind Bayesian receivers in a number of digital communications applications. The salient features of these MCMC receivers include the following: a) They are optimal in the sense of achieving minimum symbol error rate; b) they do not require the knowledge of the channel states, nor do they explicitly estimate the channel by employing training signals or decision-feedback; and c) they are well suited for iterative (turbo) processing in coded systems. In this paper, we investigate the convergence behaviors of several MCMC algorithms (both existing and new ones) in digital communication applications. The geometric convergence property of these algorithms is established by considering only the chains or the marginal chains corresponding to the transmitted digital symbols, which take values from a finite discrete set. We then focus on three specific applications, namely, the MCMC decoders in AWGN channels, ISI channels, and CDMA channels. The convergence rates for these algorithms are computed for small simulated datasets. Different convergence behaviors are observed. It is seen that differential encoding, parameter constraining, collapsing, and grouping are efficient ways of accelerating the convergence of the MCMC algorithms, especially in the presence of channel phase ambiguity.

Index Terms—Bayesian receiver, duality, eigenvalue, forward-backward algorithm, Gibbs sampler, Metropolis algorithm, transition function.

I. INTRODUCTION

IN DIGITAL communications, the transmitter transmits a sequence of symbols $\mathbf{X} = (x_1, x_2, \dots, x_n)$, where $x_i \in \mathcal{A}$ and \mathcal{A} is a finite set of all possible transmitted symbols. At the receiver end, the received signal denoted as $\mathbf{Y} = r(\tau; \mathbf{X}, \Theta)$ is a random continuous function of time τ , which is dependent of the transmitted symbol sequence \mathbf{X} and some unknown parameters Θ . Typically, Θ is a multidimensional continuous-valued quantity used to describe the channel characteristic. In this context, the following two statistical inference problems based on the maximum-likelihood (ML) principle are of interest and have

been the subject of intense research in the fields of communications and signal processing.

- *Sequence Detection:* Assuming Θ is known, find the ML estimate of \mathbf{X}

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X} \in \mathcal{A}^n} p(\mathbf{Y} | \mathbf{X}). \quad (1)$$

- *Channel Estimation:* Assuming \mathbf{X} is known, find the ML estimate of Θ

$$\hat{\Theta} = \arg \max_{\Theta} p(\mathbf{Y} | \Theta). \quad (2)$$

Typically, the transmitter first transmits a sequence of known symbols, based on which channel estimation is performed at the receiver. The information symbol sequence is then transmitted, and sequence detection is performed at the receiver using the estimated channel. Such an “estimate-then-plug-in” approach, although popular in engineering practice, is *ad hoc* and bears no theoretical optimality. In fact, the objective of digital communication is to reliably transmit information between the transmitter and the receiver, and channel estimation is not necessarily essential for achieving this. Indeed, the channel parameters Θ can be viewed as “hidden” random data, and the ultimate symbol sequence detection problem becomes finding

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X} \in \mathcal{A}^n} E\{p(\mathbf{Y} | \Theta, \mathbf{X}) | \mathbf{X}\} \quad (3)$$

where $E\{\cdot\}$ denotes the expectation taken with respect to the distribution of Θ . The inference problem (3) can be solved using the expectation-maximization (EM) algorithm [11]. This approach to receiver design in digital communications, although theoretically appealing, suffers from a spectral efficiency problem. This is because that in order for the EM algorithm to converge to the ML solution, the initial estimate of the symbol sequence must be close to the ML estimate. Hence, a training symbol sequence has to be employed for this purpose, which results in a loss of spectral efficiency.

An entirely different approach to the design of optimal digital communication receiver is based on the Bayesian formulation of the problem. In this approach, all unknowns \mathbf{X} and Θ are treated as random quantities with some independent prior distributions $p(\mathbf{X})$ and $p(\Theta)$, respectively. We are then interested

Manuscript received January 31, 2001; revised October 3, 2001. This work was supported in part by the National Science Foundation under Grants DMS-9803649, CCR-9980599, DMS-0073601, DMS-0094613, and DMS-0073651. The associate editor coordinating the review of this paper and approving it for publication was Prof. Simon J. Godsill.

R. Chen is with the Department of Information and Decision Science, University of Illinois at Chicago, Chicago, IL 6060 USA.

J. S. Liu is with the Department of Statistics, Harvard University, Cambridge, MA 02138 USA (e-mail: jliu@bioinfo.stat.harvard.edu).

X. Wang is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA.

Publisher Item Identifier S 1053-587X(02)00552-4.

in computing the *a posteriori* probability distribution of each transmitted symbol

$$P(x_t = a_j | \mathbf{Y}) = \sum_{\mathbf{X}_{[-t]} \in \mathcal{A}^{n-1}} p(\mathbf{X} | \mathbf{Y}) \quad (4)$$

$$= \sum_{\mathbf{X}_{[-t]} \in \mathcal{A}^{n-1}} \int p(\mathbf{X}, \Theta | \mathbf{Y}) d\Theta \quad (5)$$

$\forall a_j \in \mathcal{A}$, where $\mathbf{X}_{[-t]} \triangleq \mathbf{X} \setminus x_t$. Note that the joint posterior density in (5) can be written as

$$p(\mathbf{X}, \Theta | \mathbf{Y}) \propto p(\mathbf{Y} | \mathbf{X}, \Theta) p(\mathbf{X}) p(\Theta). \quad (6)$$

The computation in (5) is clearly prohibitive, and we therefore resort to Monte Carlo method. Suppose we can generate random samples (either independent or dependent)

$$\left(\mathbf{X}^{(1)}, \Theta^{(1)} \right), \left(\mathbf{X}^{(2)}, \Theta^{(2)} \right), \dots, \left(\mathbf{X}^{(m)}, \Theta^{(m)} \right)$$

from the joint posterior distribution (6) or samples $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(m)}$ from the marginal distribution

$$p(\mathbf{X} | \mathbf{Y}) = \int p(\mathbf{X}, \Theta | \mathbf{Y}) d\Theta. \quad (7)$$

Then, we can approximate the marginal posterior $p(x_i | \mathbf{Y})$ by the empirical distribution based on the corresponding component in the Monte Carlo sample, i.e., $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}$, and approximate the inference (5) by

$$P(x_i = a_j | \mathbf{Y}) \cong \frac{1}{m} \sum_{\ell=1}^m \delta(x_i^{(\ell)} = a_j) \quad (8)$$

where $\delta(\cdot)$ is an indicator function such that $\delta(x = a) = 1$ if $x = a$, and $\delta(x = a) = 0$ if $x \neq a$. Markov chain Monte Carlo (MCMC) methods such as the Gibbs sampler and the Metropolis algorithm [14] are effective means for achieving the sampling from the posterior distributions (6) or (7).

The first application of the Gibbs sampler to the design of Bayesian receiver in digital communications appeared in [6], where a blind equalizer based on the Gibbs sampler was proposed for symbol recovery in an intersymbol-interference (ISI) channel. Recently, this approach has been applied to treat more complicated communication systems, such as the coded ISI channel [23] and the coded code-division multiple-access (CDMA) channel [22], with Gaussian or impulsive ambient noise. A salient feature of these Gibbs-sampler-based receivers is that they can incorporate the *a priori* symbol probabilities, and they produce as output the *a posteriori* symbol probabilities—that is, they are *soft-input soft-output* algorithms. Hence, these methods are well suited for iterative processing in a coded system, which allows the symbol detector to refine its processing based on the decoding stage, and vice versa [22], [23].

Our recent studies [22], [23] have focused on developing Bayesian receiver algorithms based on the Gibbs sampler. However, the convergence behavior of the Gibbs sampler in these digital communication applications has not been thoroughly examined. In this paper, we first establish the general convergence properties of the MCMC algorithms used in these Bayesian receivers. We then discuss a few specific MCMC detectors in AWGN, ISI, and the CDMA channels. The convergence rates of these algorithms are computed for small data sizes, in which different convergence behaviors are observed.

For a different type of deconvolution problems, [1] proposed a fancier MCMC algorithm that, in addition to the regular Gibbs sampler, uses also a reversible-jump proposal. The geometric convergence of their algorithm was proved by using a method pioneered by [17]. A similar technique has also been used in [18], [19] to bound the convergence rate of a MCMC sampler. In comparison with their methods, the arguments in this paper are simpler and more intuitive. Besides proposing a few new MCMC algorithms and establishing the qualitative convergence properties of these and some Gibbs sampling algorithms proposed earlier in the literature, this paper emphasizes the understanding of the nature of these algorithms: What causes their slow convergence, and how we can improve them.

The rest of this paper is organized as follows. In Section II, we provide some general convergence results for MCMC samplers and, in particular, those applicable to the Bayesian digital communication receivers. In Sections III–V, we study the convergence properties of several specific MCMC receivers based on the Gibbs sampler in AWGN channels, ISI channels, and CDMA channels, respectively. Section VI contains the conclusion.

II. CONVERGENCE OF MCMC SAMPLERS

In all MCMC algorithms, a Markov transition rule (or kernel) is first constructed so that its limiting distribution is the desired posterior distribution. Then, a realization of the Markov chain based on this transition rule is generated, with an arbitrary initial value. If the Markov chain converges to its equilibrium distribution, then after certain burn-in period, the realization of the Markov chain can be used as *correlated samples* from the target distribution. The convergence rate of the Markov chain affects not only the necessary length of the burn-in period but the efficiency of the posterior estimation based on the Monte Carlo samples as well. When the convergence rate is low, the auto-correlations among the realized Monte Carlo draws are high, resulting in an inaccurate estimate by using (8). In this section, we provide some simple theoretical results that are directly related to the algorithms to be studied. More details can be found in [14].

A. Basics of Markov Chain

Suppose $X^{(1)}, X^{(2)}, \dots, X^{(n)}, \dots$ are the realizations of a Markov chain defined on the state space \mathbf{E} . Let \mathcal{E} be the a countably generated σ -algebra on \mathbf{E} . Let $P(x, \cdot)$ denote the transition function for the chain. That is, $P(x, A) = P(X^{(t+1)} \in A | X^{(t)} = x)$ for all $A \in \mathcal{E}$ and $x \in \mathbf{E}$. To conform with the

general notations in this paper, we also use $P(\cdot|x)$ to denote the same function and use $P^{(n)}(x, \cdot)$ to denote the n -step transition function, i.e.,

$$P^{(n)}(x, A) = P\left(X^{(n)} \in A | X^{(0)} = x\right). \quad (9)$$

A transition kernel is said to be ϕ -irreducible for a measure ϕ on $(\mathbf{E}, \mathcal{E})$ if $\forall x \in \mathbf{E}$ and $A \in \mathcal{E}$ with $\phi(A) > 0$ there exists an integer $n \geq 1$ such that $P^{(n)}(x, A) > 0$. When an invariant distribution π exists for the Markov chain, it is most relevant to consider whether the chain is π -irreducible. Intuitively, *irreducibility* reflects the connectivity of the space under the transition rule P .

For a discrete state space, a state $x \in \mathbf{E}$ is said to be *aperiodic* if the greatest common divisor of $\{n: P^{(n)}(x, x) > 0\}$ is one. This definition can also be extended to a general state space with a bit more measure-theoretic dressings. Clearly, if a state x is aperiodic and the chain is irreducible, then every state in \mathbf{E} must be aperiodic. Last, we define the total variation distance between probability measures P and Q as

$$\begin{aligned} \|P - Q\|_{\text{var}} &= \sup_{A \in \mathcal{E}} |P(A) - Q(A)| \\ &= \frac{1}{2} \int |P(dx) - Q(dx)| \equiv \frac{1}{2} \|P - Q\|_{L^1}. \end{aligned} \quad (10)$$

The following two convergence results are stated without proofs. See a proof of [3, Th. 2], [14] and the proof of [21, Th. 2].

Theorem 1: Suppose the state space \mathbf{E} of a Markov chain is discrete and finite. If the transition kernel of this chain is irreducible and aperiodic, then $P^{(n)}(x, \cdot)$ as a probability measure on \mathbf{E} converges to the invariant distribution $\pi(\cdot)$ geometrically in total variation distance, that is, there exists $0 < r < 1$ and $c > 0$ such that

$$\left\| P^{(n)}(x, \cdot) - \pi \right\|_{\text{var}} \leq cr^n. \quad (11)$$

Theorem 2 [21]: Suppose P is π -irreducible and $\pi P = \pi$. Then, P is positive recurrent, and π is the unique invariant distribution of P . If P is also aperiodic, then, for all x but a subset whose measure under π is zero (i.e., π -almost all x),

$$\left\| P^{(n)}(x, \cdot) - \pi \right\|_{\text{var}} \rightarrow 0. \quad (12)$$

Clearly, all of the MCMC algorithms we designed for the signal reception problems in digital communications have (5) as their invariant distribution. It is also not difficult to see that they are irreducible and aperiodic; thus, their convergence to stationary distribution is guaranteed by Tierney's theorem.

B. Convergence Rate and Autocorrelation

Once convergence is established, the following arguments can be used to understand the rate of convergence.

Consider a Markov chain on a finite discrete state space. Suppose the number of states is κ . We write the transition kernel as

a $\kappa \times \kappa$ matrix $\mathbf{P} = [a_{ij}]$, where its entry a_{ij} is equal to the transition probability from state i to state j .

Theory for non-negative matrices [4, (Perron–Frobenius theorem)] indicates that \mathbf{P} 's largest eigenvalue in absolute value is 1 and that it is a simple eigenvalue if \mathbf{P} is irreducible (all its eigenvalues are simple and real if \mathbf{P} also corresponds to a reversible Markov chain). Thus, we can express \mathbf{P} as

$$\mathbf{P} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^{-1} \quad (13)$$

where $\mathbf{\Lambda} = \text{diag}\{1, J_2, \dots, J_l\}$ with $l \leq \kappa$ and $1 \geq |\lambda_2| \geq \dots \geq |\lambda_l|$. Here, J_k is the Jordan form corresponding to λ_k . As $n \rightarrow \infty$

$$\begin{aligned} \mathbf{P}^n &= \mathbf{B} \begin{pmatrix} 1 & 0 & & 0 \\ 0 & J_2^n & & \\ & & \ddots & \\ 0 & & & J_l^n \end{pmatrix} \mathbf{B}^{-1} \\ &\xrightarrow{n \rightarrow \infty} \mathbf{B} \begin{pmatrix} 1 & 0 & & 0 \\ 0 & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix} \mathbf{B}^{-1} \end{aligned} \quad (14)$$

if and only if $|\lambda_2| < 1$, which holds true if \mathbf{P} is aperiodic. Because $\pi \mathbf{P}^n = \pi$ and the limit of \mathbf{P}^n as $n \rightarrow \infty$ is of rank 1, every row of the limiting matrix \mathbf{P}^∞ must be the same as π . Hence, the rate of convergence is exactly $r = |\lambda_2|$. As noted in [16], the above argument can be extended to uncountable state spaces via elementary functional analysis. It can be shown that when the conditional expectation operator is compact and has a spectral radius ρ on the mean-zero functional space, then $\|P^{(n)}(x, \cdot) - \pi(\cdot)\|_{L^2}$ converges to zero with rate ρ .

The convergence rate is also closely related to the sample autocorrelations. Suppose $X^{(0)}, X^{(1)}, \dots$ is a realization of the Markov chain starting from stationary distribution π . Let $h(X)$ be a square-integrable function with respect to π . Without loss of generality, we set $E_\pi(h(X)) = 0$ and $\text{var}_\pi(h(X)) = 1$. Define

$$\rho_i(h) = \text{corr}\left(h\left(X^{(0)}\right), h\left(X^{(i)}\right)\right). \quad (15)$$

For simplicity, we consider only the case of a discrete state space. Let $\mathbf{v}_1, \mathbf{v}_2, \dots$ be the orthonormal right-eigenvectors of \mathbf{P} (i.e., $\mathbf{P}\mathbf{v}_i = \lambda_i\mathbf{v}_i$). Then, the spectral expansion of $\mathbf{h} \equiv h(X)$ is

$$\mathbf{h} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots$$

Since $E_\pi h(X) = 0$, $a_1 = 0$. Hence, $\rho_i(h) = a_2\lambda_2^i + a_3\lambda_3^i + \dots$. In the case when \mathbf{h} is chosen as \mathbf{v}_2 , the first autocorrelation $|\rho_1(h)| = |\lambda_2|$ is exactly the convergence rate. Otherwise, $|\rho_i(h)|$ is dominated by $|\lambda_2|^i$ when i is large. In most applications, it is impossible to observe directly on λ_2 . A useful way to gain insight on the convergence rate is to plot $\rho_i(h)$ for a number of testing functions h .

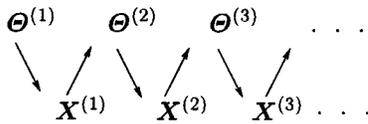


Fig. 1. Graphical illustration of the two-component Gibbs sampler.

The sample autocorrelations directly influence the estimation efficiency of a MCMC algorithm. When estimating $E_{\pi}(h)$ by using the sample average, we have

$$\lim_{n \rightarrow \infty} n \text{var}(\bar{h}) \approx 1 + 2(\rho_1(h) + \rho_2(h) + \dots) \equiv 2T_h$$

which is large when $\rho_i(h)$ decreases to zero slowly. In physics, T_h is called the *autocorrelation time*.

C. Duality of the Two-Component Gibbs Sampler

Suppose we run the Gibbs sampler on a two-component random vector (\mathbf{X}, Θ) with the target distribution $\pi(\mathbf{X}, \Theta)$, e.g., the posterior distribution (6). A Gibbs sampler for this problem iterates between two steps: i) Generate Θ from the conditional sampling distribution $\pi(\Theta | \mathbf{X})$, and ii) generate \mathbf{X} from the conditional distribution $\pi(\mathbf{X} | \Theta)$. More precisely, given the sample $(\mathbf{X}^{(k)}, \Theta^{(k)})$ generated at the k th step, the $(k+1)$ th sample is generated as follows:

$$\begin{aligned} \Theta^{(k+1)} &\sim \pi(\Theta | \mathbf{X}^{(k)}) \\ \mathbf{X}^{(k+1)} &\sim \pi(\mathbf{X} | \Theta^{(k+1)}). \end{aligned} \quad (16)$$

Fig. 1 shows graphically the relationship among these random draws.

It is easy to check that the transition rule determined by the above iterations has π as its invariant distribution [15]. If we look at the component \mathbf{X} marginally, $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots$ also form a Markov chain (see Fig. 1). Moreover, this marginal chain has the same convergence rate as that of the original chain [15]. Note that this is also true for the Θ component when the \mathbf{X} can be sampled in one step as in (16). To summarize, we have the following theorem due to [15].

Theorem 3: Suppose the chain $(\Theta^{(0)}, \mathbf{X}^{(0)})$, $(\Theta^{(1)}, \mathbf{X}^{(1)})$, \dots resulting from a two-component Gibbs sampler is in stationarity. Then, we have the following.

a) For an arbitrary function h ,

$$\text{cov} \left\{ h(\mathbf{X}^{(0)}), h(\mathbf{X}^{(1)}) \right\} = \text{var}_{\pi} \{ E_{\pi} \{ h(\mathbf{X}) | \Theta \} \}. \quad (17)$$

b) The *marginal chains* $\{\mathbf{X}^{(t)}, t = 1, 2, \dots\}$ and $\{\Theta^{(t)}, t = 1, 2, \dots\}$ are both reversible Markov chains. In particular, the transition function for, say, the first chain is

$$\begin{aligned} P_1(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}) \\ = \int \pi(\mathbf{X}^{(1)} | \Theta^{(0)}) \pi(\Theta^{(0)} | \mathbf{X}^{(0)}) d\Theta^{(0)}. \end{aligned} \quad (18)$$

c) The convergence rate of the joint chain is equal to that of the marginal chain, and they are equal to the maximal correlation between Θ and \mathbf{X} , which is defined as

$$r \triangleq \sup_g \frac{\text{var} \{ E \{ g(\mathbf{X}) | \Theta \} \}}{\text{var}_{\pi} \{ g(\mathbf{X}) \}} \quad (19)$$

where the supremum is taken over all function g such that $0 < \text{var}_{\pi}(g) < \infty$.

In the Bayesian digital communication receiver design problem, $\mathbf{X} = (x_1, \dots, x_n)$, where each symbol x_i takes value from a finite set \mathcal{A} . Thus, the state space of the \mathbf{X} -chain is the finite space \mathcal{A}^n . For the three communication channels considered in this paper, it is easily established in later sections that the Markov chain corresponding to the Gibbs receiver is irreducible and aperiodic. Hence, the Gibbs samplers in these receivers converge geometrically. In practice, however, it is their convergence rate that matters the most. Note that (19) is usually difficult to evaluate. However, because of Theorem 3(c), we can obtain the convergence rate of the joint chain by studying the convergence rate of the marginal \mathbf{X} -chain. When \mathbf{X} is on a finite state space, results in Section II-B can be used.

Most of the time, sampling from $\pi(\mathbf{X} | \Theta)$ in one step is difficult, in which case one may iteratively generate individual x_t from the conditional distributions $\pi(x_t | \Theta, \mathbf{X}_{[-t]})$, $t = 1, \dots, n$. In other words, given $(\mathbf{X}^{(k)}, \Theta^{(k)})$, the sample at the next step is generated as follows:

$$\begin{aligned} \Theta^{(k+1)} &\sim \pi(\Theta | \mathbf{X}^{(k)}) \\ x_t^{(k+1)} &\sim \pi(x_t | \Theta^{(k+1)}, \mathbf{X}_{[-t]}^{(k)}), \quad t = 1, \dots, n \end{aligned} \quad (20)$$

where

$$\mathbf{X}_{[-t]}^{(k)} \triangleq (x_1^{(k+1)}, \dots, x_{t-1}^{(k+1)}, x_{t+1}^{(k)}, \dots, x_n^{(k)}).$$

In this case, the marginal chain $\mathbf{X}^{(k)}$ is still a Markov chain (but is no longer reversible), and its rate of convergence is still the same as the joint chain. In comparison with the previous two-component sampler, the convergence of this full Gibbs sampler is slower [14], [15].

D. Remarks

In practice, computing the exact convergence rate of an MCMC sampler is formidable even for very simple problems. Some recent research has focused on finding a bound for $|\lambda_2|$. One major class of work is built on the geometric techniques invented for differential geometry problems, e.g., the Poincaré inequality, Cheeger's inequality, and log-Sobolev inequality [8], [12]. Another class of work uses the coupling technique and follows the approach of [17] to establish geometric convergence or to bound convergence rate [1], [18], [19].

In what follows, we study the convergence behavior of the Bayesian MCMC detectors in three digital communication channels, namely, the additive Gaussian noise (AWGN) channel, the ISI channel, and the CDMA channel, and we compute the exact convergence rates of various MCMC

samplers in these applications. One distinctive feature of many digital communication problems is that the involved unknown random quantities can be divided into two classes: the transmitted digital symbols, which take values from a finite set, and the channel parameters, which take continuous values. For Gibbs samplers that iterate between these two components, our strategy is to establish convergence properties based only on the discrete component. For other algorithms, such as the one-component Gibbs decoder and the Metropolis decoder, their geometric convergence can be directly established by Theorem 1 because they operate on the discrete symbol space.

III. MCMC DETECTORS IN AWGN CHANNELS

A. Target Distributions and the Algorithm

We start with the simplest channel model in digital communications: the additive white Gaussian noise (AWGN) channel. After filtering and sampling of the continuous-time received waveform, the discrete-time received signal in such a channel is given by

$$y_t = \phi x_t + v_t, \quad t = 1, 2, \dots, n \quad (21)$$

where

y_t	received signal at time t ;
$x_t \in \{+1, -1\}$	transmitted binary symbol at time t ;
$\phi \in \mathbb{R}$	received signal amplitude;
v_t	independent Gaussian noise with zero-mean and variance σ^2 , i.e., $v_t \sim \mathcal{N}(0, \sigma^2)$.

Denote $\mathbf{X} \triangleq [x_1, \dots, x_n]$ and $\mathbf{Y} \triangleq [y_1, \dots, y_n]$. Our problem is to estimate the *a posteriori* probability distribution of each symbol based on the received signal \mathbf{Y} without knowing the channel parameters (ϕ, σ^2) . The solution to this problem based on the Gibbs sampler is as follows. Assuming a uniform prior for ϕ , a uniform prior for \mathbf{X} (on $\{-1, +1\}^n$), and an inverse χ^2 prior for σ^2 , $\sigma^2 \sim \chi^{-2}(\nu, \lambda)$ ($\nu = 1$ and $\lambda = 0.1$ were used in [22], [23]), the complete posterior distribution is given by

$$p(\mathbf{X}, \phi, \sigma^2 | \mathbf{Y}) \propto p(\mathbf{Y} | \mathbf{X}, \phi, \sigma^2) p(\phi) p(\sigma^2) p(\mathbf{X}). \quad (22)$$

The Gibbs sampler starts with arbitrary initial values of $\mathbf{X}^{(0)}$ and, for $k = 0, 1, \dots$, iterates between the following two steps.

Algorithm III.1 [Two-Component Gibbs Detector in AWGN Channel]:

- Draw a sample $(\phi^{(k+1)}, \sigma^{2(k+1)})$ from the conditional distribution (given $\mathbf{X}^{(k)}$)

$$p(\phi, \sigma^2 | \mathbf{X}^{(k)}, \mathbf{Y}) \propto (\sigma^2)^{-(n/2)} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \phi x_t^{(k)})^2\right] \times (\sigma^2)^{-((\nu+2)/2)} \exp\left(-\frac{\nu\lambda}{2\sigma^2}\right) \quad (23)$$

$$\propto \pi^{(k+1)}(\phi | \sigma^2) \pi^{(k+1)}(\sigma^2) \quad (24)$$

where

$$\pi^{(k+1)}(\sigma^2) \sim \chi^{-2}\left(\nu + n - 1, \frac{1}{\nu + n - 1} \times \left[\nu\lambda + \sum_{t=1}^n y_t^2 - \frac{1}{n} \left(\sum_{t=1}^n y_t x_t^{(k)}\right)^2\right]\right) \quad (25)$$

and

$$\pi^{(k+1)}(\phi | \sigma^2) \sim \mathcal{N}\left(\frac{1}{n} \sum_{t=1}^n y_t x_t^{(k)}, \frac{\sigma^2}{n}\right). \quad (26)$$

- Draw a sample $\mathbf{X}^{(k+1)}$ from the following conditional distribution, given $(\phi^{(k+1)}, \sigma^{2(k+1)})$

$$p(\mathbf{X} | \phi^{(k+1)}, \sigma^{2(k+1)}, \mathbf{Y}) = \prod_{t=1}^n p(x_t | y_t, \phi^{(k+1)}, \sigma^{2(k+1)}) \quad (27)$$

$$\propto \prod_{t=1}^n \exp\left[-\frac{1}{2\sigma^{2(k+1)}} (y_t - \phi^{(k+1)} x_t)^2\right]. \quad (28)$$

That is, for $t = 1, \dots, n$ and $b \in \{+1, -1\}$, draw $x_t^{(k+1)}$ from

$$P(x_t^{(k+1)} = b) = \left[1 + \exp\left(-\frac{2b\phi^{(k+1)}y_t}{\sigma^{2(k+1)}}\right)\right]^{-1}. \quad (29)$$

In the derivations for the first step, we used the fact that $\sum_{t=1}^n x_t^2 = n$ for all possible \mathbf{X} . Note that this step is slightly different from but more efficient than that in [22] and [23] in which they used separate steps for ϕ and σ^2 .

It is worthwhile to note that one can integrate out ϕ and σ^2 in (22) analytically to get the marginal target distribution of \mathbf{X} , which can provide some further insight. More precisely, we have

$$\pi(\mathbf{X}) \propto \left[\nu\lambda + \sum_{t=1}^n y_t^2 - \frac{1}{n} \left(\sum_{t=1}^n x_t y_t\right)^2\right]^{-(n+\nu)/2}. \quad (30)$$

This defines a distribution on the space of an n -dimensional cube. The mode of this distribution is clearly at $\tilde{\mathbf{X}}$ and $-\tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}} = \text{sign}(\mathbf{Y})$. Intuitively, this is the “obvious solution” in this simple setting but is not easy to generalize. As will be discussed below, the difficulty in dealing with $\pi(\mathbf{X})$ is its bimodality and the strong separation of these two modes when the signal-to-noise ratio (SNR) (which is defined as ϕ^2/σ^2) is large. Based on (30), we can derive another Gibbs sampling algorithm:

Algorithm III.2 [One-Component Gibbs Detector in AWGN Channel]:

- Choose t from $1, \dots, n$ by either the random scan (i.e., the t is chosen at random) or the deterministic scan (i.e.,

one cycles t from 1 to n systematically). Update $\mathbf{X}^{(k)}$ to $\mathbf{X}^{(k+1)}$, where $x_s^{(k+1)} = x_s^{(k)}$ for $s \neq t$ and $x_t^{(k+1)}$ is drawn from the conditional distribution

$$\begin{aligned} \pi(x_t = b | \mathbf{X}_{[-t]}^{(k)}) \\ = \frac{\pi(x_t = b, \mathbf{X}_{[-t]}^{(k)})}{\pi(x_t = b, \mathbf{X}_{[-t]}^{(k)}) + \pi(x_t = -b, \mathbf{X}_{[-t]}^{(k)})} \end{aligned} \quad (31)$$

where $\pi(\mathbf{X})$ is as in (30). When the variance σ^2 is known

$$\pi(\mathbf{X}) \propto \exp \left\{ \frac{1}{2n\sigma^2} \left(\sum_{t=1}^n x_t y_t \right)^2 \right\}. \quad (32)$$

For the deterministic scan, it is conventional to consider a complete cycle from 1 to n as one iteration, and we will follow this convention. Besides the two Gibbs samplers just described, an attractive alternative is the Metropolis algorithm applied directly to (30). Suppose $\mathbf{X}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$. At step $k+1$, the Metropolis algorithm proceeds as follows.

Algorithm III.3 [Metropolis Detector in AWGN Channel]:

- Choose $t \in \{1, \dots, n\}$ either by the random scan or by the deterministic scan. Define $\mathbf{Z} = (z_1, \dots, z_n)$, where $z_t = -x_t^{(k)}$ and $z_s = x_s^{(k)}$ for $s \neq t$. Generate independent $U \sim \text{unif}(0, 1)$. Let $\mathbf{X}^{(k+1)} = \mathbf{Z}$ if

$$U \leq \min \left\{ 1, \frac{\pi(\mathbf{Z})}{\pi(\mathbf{X}^{(k)})} \right\} \quad (33)$$

and let $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)}$ otherwise.

This Metropolis algorithm differs from the one-component Gibbs detector only slightly in the way of updating $x_t^{(k)}$ to $x_t^{(k+1)}$. That is, the Metropolis algorithm always forces the change (to $-x_t^{(k)}$) unless it is rejected, whereas the Gibbs sampler “voluntarily” selects whether to make the change so that no rejection is incurred. It is shown in [13] that when the random scan is used, the Metropolis rule always results in a smaller second-largest eigenvalue (not in absolute value) than the corresponding Gibbs sampler. Thus, when the target distribution is relatively peaked (high SNR), the Metropolis algorithm is slightly preferable. However, as observed by [10] in the Ising model, the Metropolis algorithm may have a large (in absolute value) negative eigenvalue when the target distribution is flatter (low SNR). This is consistent with our simulation studies. In practice, however, the large negative eigenvalue is not a serious concern. No clear theory is available when a deterministic scan is used for updating. Our simulation suggests that a similar result to that of the random scan samplers seems to hold well.

B. Analysis of the Gibbs Chain

The two-component Gibbs detector (see *Algorithm 1*) generates a Markov chain

$$\mathbf{X}^{(0)}, (\phi^{(1)}, \sigma^{2(1)}), \mathbf{X}^{(1)}, \dots, (\phi^{(k)}, \sigma^{2(k)}), \mathbf{X}^{(k)}, \dots$$

as described in Section II-C. Now, consider only the \mathbf{X} part of the above chain, which takes values on the finite set $\{+1, -1\}^n$. Algorithm 1 dictates that the transition probability from $\mathbf{X}^{(k)}$ to $\mathbf{X}^{(k+1)}$ given by (18) is

$$\begin{aligned} P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) \\ \propto \int p(\mathbf{X}^{(k+1)} | \phi, \sigma^2, \mathbf{Y}) p(\phi, \sigma^2 | \mathbf{X}^{(k)}, \mathbf{Y}) d\phi d\sigma^2 \\ = \int \prod_{t=1}^n p(x_t^{(k+1)} | \phi, \sigma^2, y_t) \\ \times p(\phi, \sigma^2 | \mathbf{X}^{(k)}, \mathbf{Y}) d\phi d\sigma^2 \\ \propto \int \prod_{t=1}^n \left[1 + \exp\left(-\frac{2x_t^{(k+1)} \phi y_t}{\sigma^2}\right) \right]^{-1} (\sigma^2)^{-((n+\mu+2)/2)} \\ \times \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \phi x_t^{(k)})^2 - \frac{\nu\lambda}{2\sigma^2}\right] d\phi d\sigma^2. \end{aligned} \quad (34)$$

It is easily seen that $P(\cdot | \cdot) > 0$ when the true variance of v_t in (21) is not zero. This means that the Markov chain is irreducible and aperiodic. As shown in Section II, the second largest eigenvalue (in absolute value) of the $2^n \times 2^n$ transition matrix is the geometric convergence rate. The smaller it is, the faster the Gibbs sampler converges. A rate of 0.9 is usually considered to be fast for most of the problems. Note that this rate depends on the actual value of \mathbf{Y} ; hence, it is random.

The integral (34) is difficult to evaluate. In order to gain some insights on the convergence properties, we consider a simpler problem that assumes σ^2 known (this is a reasonable assumption in practice since it is the noise power of the channel and can be measured at the receiver before the communication session starts). Then, (34) becomes

$$\begin{aligned} P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) \\ \propto \int \prod_{t=1}^n \left[1 + \exp\left(-\frac{2x_t^{(k+1)} \phi y_t}{\sigma^2}\right) \right]^{-1} \\ \times \psi\left(\phi; \frac{1}{n} \sum_{t=1}^n y_t x_t^{(k)}, \frac{\sigma^2}{n}\right) d\phi \end{aligned} \quad (35)$$

where $\psi(x; \mu, s^2)$ is the density function of a Gaussian $\mathcal{N}(\mu, s^2)$ random variable. The integral (35) can be easily evaluated using Monte Carlo integration by sampling ϕ from

$$\mathcal{N}\left(\frac{1}{n} \sum_{t=1}^n y_t x_t^{(k)}, \frac{\sigma^2}{n}\right)$$

and averaging the corresponding values of

$$\prod_{t=1}^n \left[1 + \exp\left(-\frac{2x_t^{(k+1)} \phi y_t}{\sigma^2}\right) \right]^{-1}.$$

The transition matrices $P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)})$ for the one-component Gibbs detector and Metropolis detector are easy to obtain.

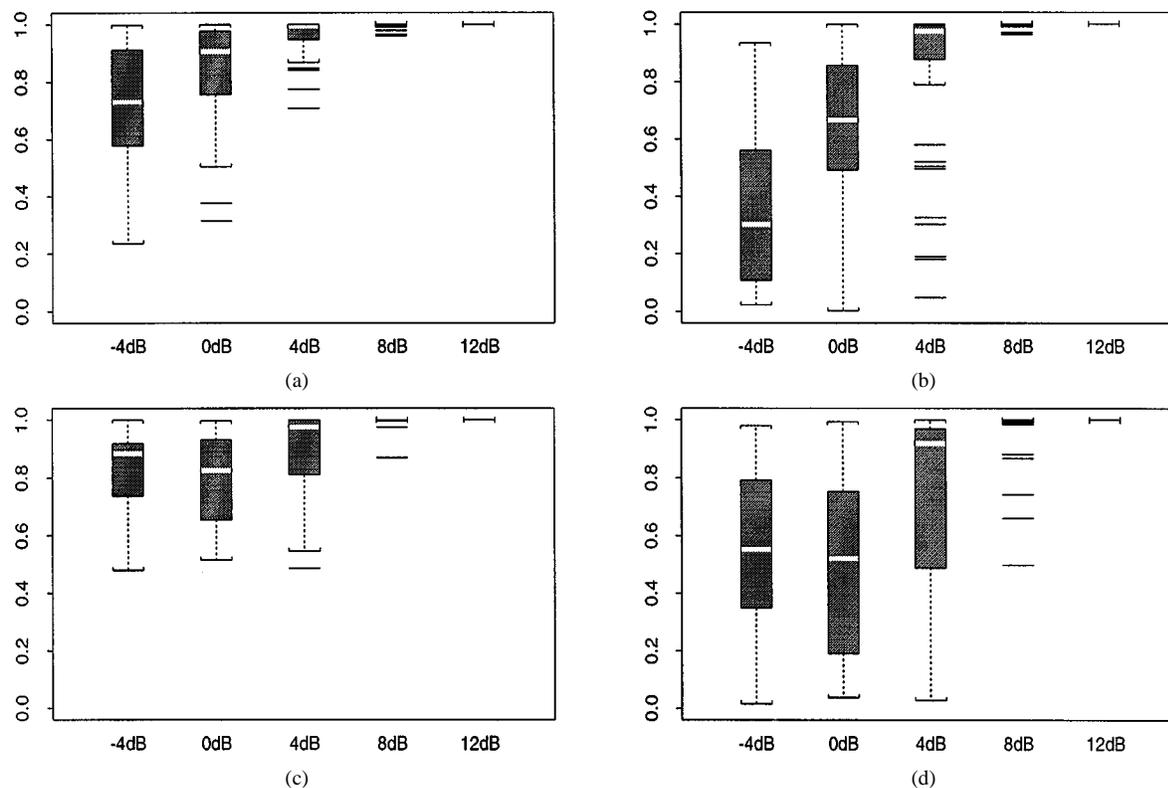


Fig. 2. Boxplots of convergence rates in AWGN channels (without differential coding) with $n = 5$ and $\phi = 1$. Two-component Gibbs detector, one-component deterministic scan Gibbs detector, one-component deterministic scan Metropolis detector, and five-step one-component random scan Metropolis detector are shown in (a)–(d), respectively.

For example, for a deterministic scan one-component Gibbs or Metropolis sampler, we use

$$P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) = \prod_{t=1}^n q(x_t^{(k+1)} | x_1^{(k+1)}, \dots, x_{t-1}^{(k+1)}, x_{t+1}^{(k)}, \dots, x_n^{(k)}) \quad (36)$$

where $q(\cdot)$ is implied by (31) for Gibbs and by (33) for Metropolis. For the random-scan Metropolis decoder, the transition for $\mathbf{X}^{(k+1)} \neq \mathbf{X}^{(k)}$ is

$$P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) = \frac{1}{n} \min \left\{ 1, \pi(\mathbf{X}^{(k+1)}) / \pi(\mathbf{X}^{(k)}) \right\} \quad (37)$$

where $\mathbf{X}^{(k+1)}$ differs from $\mathbf{X}^{(k)}$ only at one coordinate, and π is as in (30) or (32).

Fig. 2 shows the result of a numerical study. For data size $n = 5$ and five different SNR values (which are defined as ϕ^2/σ^2) at $-4, 0, 4, 8,$ and 12 dB, 50 samples of (\mathbf{X}, \mathbf{Y}) are simulated from the system with $\phi = 1$ and $v_t \sim \mathcal{N}(0, \sigma^2)$ with σ^2 determined by the SNR. Then, the $2^5 \times 2^5$ transition matrix is constructed according to (31) and (35)–(37). The second largest eigenvalue in absolute value of each matrix is computed. Fig. 2

shows the boxplots¹ of the convergence rates. Since the computation cost for n random scans is the same as one iteration of the deterministic scan, the convergence rate for the random-scan Metropolis algorithm corresponds to the n th power of the rate of the one-step random scan Metropolis chain.

From Fig. 2, we can see that the one-component Gibbs sampler performed consistently better than the two-component one. This is consistent with the comparison theorems in [2], [14], and [15] because the former corresponds to integrating out the θ component in the latter algorithm (i.e., collapsing). The Gibbs performed better than the Metropolis decoders when SNR is low, whereas the Metropolis ones were slightly better when SNR is high. These results are in good agreement with the theory in [10] and [13]. It is also noted that all the algorithms performed poorly when SNR is high, which is due to the ambiguity of the system. Note that

$$p(\mathbf{X}, \phi, \sigma^2 | \mathbf{Y}) = p(-\mathbf{X}, -\phi, \sigma^2 | \mathbf{Y}) \quad (38)$$

for all \mathbf{X} , ϕ , and σ^2 . Hence, there are two modes of equal size in the posterior distribution. As SNR increases, the “energy” gap separating the two modes becomes deeper, making it more difficult for the chain to travel from one mode to the other. In fact, when the SNR is infinite ($\sigma^2 = 0$), the Markov chain is no longer irreducible; hence, it will be trapped in one mode forever.

¹A boxplot is a simple way to show the distribution of a data set. The two ends of the box represents the upper and the lower quartiles; hence, 50% of the middle range data is in the box, 25% is above the box, and 25% is below the box. The middle bar represents the median, and the points beyond the upper and the lower bounds are regarded as outliers.

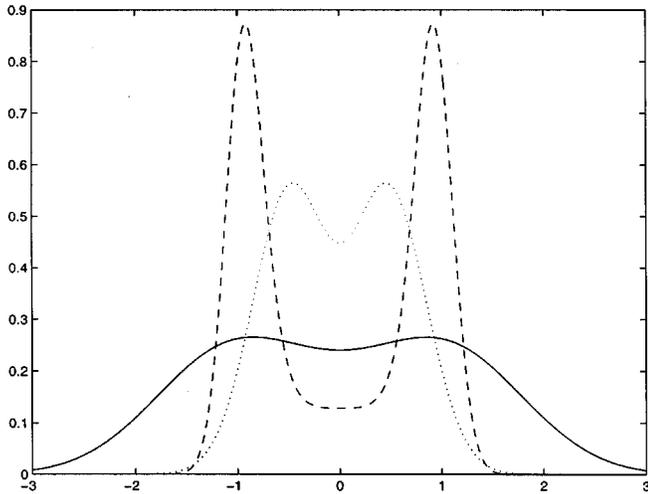


Fig. 3. Marginal distribution of $p(\phi|Y)$ in AWGN channels with $n = 5$, $\phi = 1$, and SNR = -4, 2, and 8 dB corresponding to solid, dotted, and dashed lines, respectively.

The bimodality feature can be seen from the 1-D marginal distribution of ϕ , as in Fig. 3. When σ^2 is known, the posterior marginal distribution of ϕ is

$$p(\phi|Y) = \sum_{\mathbf{X} \in \{-1, 1\}^n} p(\phi|Y, \mathbf{X}) p(\mathbf{X}|Y) \quad (39)$$

where $p(\phi|Y, \mathbf{X})$ is the Gaussian distribution in (26). Hence, the marginal distribution is a mixture Gaussian distribution with 2^n components. The effect of SNR on the energy gap between the two modes is evident.

C. Effect of Differential Encoding

To overcome the phase ambiguity, one can either restrict ϕ to be positive or, alternatively, use differential encoding. Let the information sequence be $s_t \in \{+1, -1\}$, $t = 2, \dots, n$. In differential coding, we construct the transmitted sequence $x_t \in \{+1, -1\}$, $t = 1, \dots, n$ such that $x_t = x_{t-1}s_t$. To obtain Monte Carlo draws from the posterior distribution of $p(\mathbf{S}, \phi, \sigma^2|Y)$, we use one of the MCMC algorithms in the previous subsection to generate a Markov chain on $(\mathbf{X}, \phi, \sigma^2)$ and then convert the samples of \mathbf{X} to \mathbf{S} using $s_t^{(k)} = x_t^{(k)} x_{t-1}^{(k)}$, $t = 2, \dots, n$. Note that in this way, \mathbf{X} and $-\mathbf{X}$ result in the same \mathbf{S} . Since $\{\mathbf{X}^{(k)}\}$ is a Markov chain, so is $\{\mathbf{S}^{(k)}\}$. The transition probability from $\mathbf{S}^{(k)}$ to $\mathbf{S}^{(k+1)}$ is given by

$$P(\mathbf{S}^{(k+1)}|\mathbf{S}^{(k)}) = P(\mathbf{X}^{(k+1)}|\mathbf{X}^{(k)}) + P(-\mathbf{X}^{(k+1)}|\mathbf{X}^{(k)}) \quad (40)$$

where both $\mathbf{X}^{(k+1)}$ and $-\mathbf{X}^{(k+1)}$ result in $\mathbf{S}^{(k+1)}$, and $\mathbf{X}^{(k)}$ results in $\mathbf{S}^{(k)}$. Note that both $\mathbf{X}^{(k)}$ and $-\mathbf{X}^{(k)}$ result in $\mathbf{S}^{(k)}$, but since $P(-\mathbf{X}^{(k+1)}|-\mathbf{X}^{(k)}) = P(\mathbf{X}^{(k+1)}|\mathbf{X}^{(k)})$, either one can be used. Again, $P(\mathbf{X}^{(k+1)}|\mathbf{X}^{(k)})$ and $P(-\mathbf{X}^{(k+1)}|\mathbf{X}^{(k)})$ can be obtained from (31), (33), (35), or (36) according to the sampling scheme used.

By denoting $s_1 = x_1$ and $\mathbf{S} \triangleq [s_1, s_2, \dots, s_n]$, we can modify (30) to give rise to the marginal target distribution for the s_t :

$$\pi(s_1, \dots, s_n) \propto \left\{ \nu\lambda + \sum_{t=1}^n y_t^2 - \frac{1}{n} \left(\sum_{t=1}^n y_t \prod_{i=1}^t s_i \right)^2 \right\}^{-(n+\nu)/2} \quad (41)$$

Clearly, s_1 is independent of all the other s and has a uniform marginal distribution.

It is trickier to implement an efficient Gibbs sampler or Metropolis algorithm based on (41). For example, the single-site update method (i.e., changing one s_t at a time) may be inefficient because when we propose to change s_t to $-s_t$, all the signs on y_t, y_{t+1}, \dots have to be changed. This may result in a very small acceptance rate. Since a single update from x_t to $-x_t$ corresponds to changing (s_t, s_{t+1}) ($-s_t, -s_{t+1}$), we can employ proposals

$$(s_t, s_{t+1}) \rightarrow (-s_t, -s_{t+1}), \quad t < n$$

and $s_n \rightarrow -s_n$ for distribution (41).

Fig. 4 shows the boxplots similar to those in Fig. 2 for SNR = -4, 0, 4, 8, 12, and 16 dB. It is clear that with differential encoding, the convergence rate becomes faster as SNR increases (the bimodality problem is eliminated). The one-component Gibbs is, again, among the best performers. It is also interesting to note that the random-scan Metropolis algorithm converges to a fixed nonzero rate as SNR increases, whereas the corresponding deterministic-scan samplers converge to a near-zero rate. This is because that it takes an average of $n(1 + 2^{-1} + \dots + n^{-1})$ steps of the single random-scan to update all the n components (the coupon collecting problem), which incurs an additional $\log(n)$ factor than the deterministic-scan samplers in the high SNR cases.

IV. GIBBS EQUALIZER IN ISI CHANNELS

In this section, we consider the convergence of the Gibbs sampler for blind equalization in an intersymbol interference (ISI) channel [6], [23]. After filtering and sampling the continuous-time received waveform, the discrete-time received signal in such a channel is given by

$$y_t = \sum_{s=0}^q \phi_s x_{t-s} + v_t, \quad t = 1, 2, \dots, n \quad (42)$$

where

- $(q+1)$ channel order;
- $\phi_i \in \mathbb{R}$ value of the i th channel tap $i = 0, \dots, q$;
- $x_t \in \{+1, -1\}$ transmitted binary symbol at time t ;
- $v_t \sim \mathcal{N}(0, \sigma^2)$ independent Gaussian noise at time t .

Let $\mathbf{X} \triangleq [x_{1-q}, \dots, x_n]$, $\mathbf{Y} \triangleq [y_1, \dots, y_n]$, $\boldsymbol{\phi} \triangleq [\phi_0, \dots, \phi_q]^T$. With a uniform prior for $\boldsymbol{\phi}$, a uniform prior for \mathbf{X} , and an inverse χ^2 prior for σ^2 (e.g., $\sigma^2 \sim \chi_{\nu}^{-2, \lambda}$), the complete posterior distribution is

$$p(\mathbf{X}, \boldsymbol{\phi}, \sigma^2|Y) \sim p(Y|\mathbf{X}, \boldsymbol{\phi}, \sigma^2) p(\boldsymbol{\phi}) p(\sigma^2) p(\mathbf{X}). \quad (43)$$

The Gibbs sampler approach to this problem starts with an arbitrary initial value of $\mathbf{X}^{(0)}$ and iterates between the following two steps.

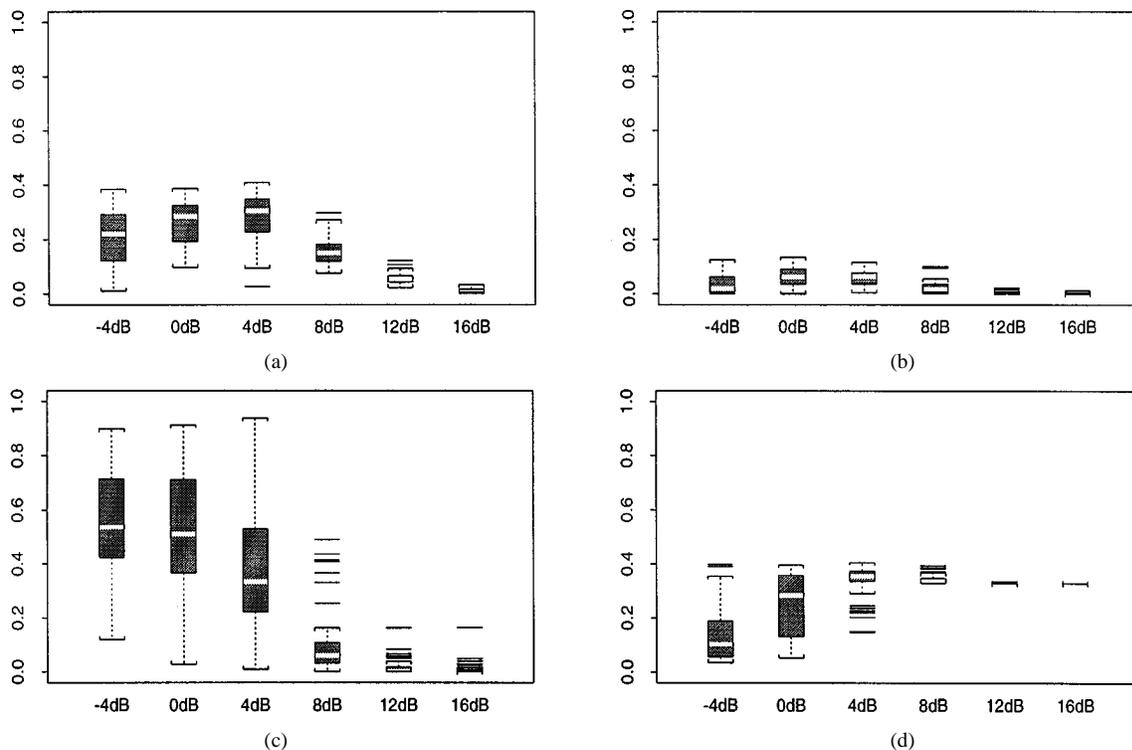


Fig. 4. Boxplots of convergence rates in AWGN channels (with differential coding) with $n = 5$, $\phi = 1$. Two-component Gibbs detector, one-component deterministic scan Gibbs detector, one-component deterministic scan Metropolis detector, and five-step one-component random scan M-detector are shown in (a)–(d), respectively.

Algorithm IV.1 [Two-Component Gibbs Equalizer in ISI Channel]:

- Draw a sample $(\phi^{(k+1)}, \sigma^{2(k+1)})$ from the conditional distribution [given $\mathbf{X}^{(k)}$]

$$\begin{aligned}
 & p(\phi, \sigma^2 | \mathbf{X}^{(k)}, \mathbf{Y}) \\
 & \propto (\sigma^2)^{-(n/2)} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \phi^T \mathbf{x}_t^{(k)})^2\right] \\
 & \quad \times (\sigma^2)^{-((\nu+2)/2)} \exp\left(-\frac{\nu\lambda}{2\sigma^2}\right) \\
 & \propto \pi^{(k+1)}(\phi | \sigma^2) \pi^{(k+1)}(\sigma^2) \quad (44)
 \end{aligned}$$

where $\mathbf{x}_t^{(k)} \triangleq [x_t^{(k)}, \dots, x_{t-q}^{(k)}]^T$ for $k = 0, 1, \dots$, and

$$\pi^{(k+1)}(\sigma^2) \sim \chi^{-2}\left(\nu + n - 1, \frac{\nu\lambda + W^{(k+1)}}{\nu + n - 1}\right) \quad (45)$$

$$\pi^{(k+1)}(\phi | \sigma^2) \sim \mathcal{N}(\boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}) \quad (46)$$

$$\begin{aligned}
 W^{(k+1)} &= \sum_{t=1}^n y_t^2 - \left[\sum_{t=1}^n \mathbf{x}_t^{(k)} y_t \right]^T \\
 & \quad \times \left[\sum_{t=1}^n \mathbf{x}_t^{(k)} \mathbf{x}_t^{(k)T} \right]^{-1} \left[\sum_{t=1}^n \mathbf{x}_t^{(k)} y_t \right] \quad (47)
 \end{aligned}$$

$$\boldsymbol{\Sigma}^{(k+1)} = \left[\frac{1}{\sigma^2} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t^T \right]^{-1} \quad (48)$$

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\Sigma}^{(k+1)} \left(\frac{1}{\sigma^2} \sum_{t=1}^n \mathbf{x}_t^{(k)} y_t \right). \quad (49)$$

- Draw a sample $\mathbf{X}^{(k+1)}$ from the conditional distribution, given $(\phi^{(k+1)}, \sigma^{2(k+1)})$ through the following iterations. For $t = 1 - q, \dots, n$, generate $x_t^{(k+1)}$ from

$$\begin{aligned}
 & p\left(x_t | \phi^{(k+1)}, \sigma^{2(k+1)}, \mathbf{Y}, \mathbf{X}_{[-t]}^{(k)}\right) \\
 & \propto \exp\left[-\frac{1}{2\sigma^{2(k+1)}} \sum_{j=1}^n (y_j - \phi^{(k+1)T} \mathbf{x}_j^{(k)})^2\right] \quad (50)
 \end{aligned}$$

where

$$\mathbf{X}_{[-t]}^{(k)} \triangleq [x_{1-q}^{(k)}, \dots, x_{t-1}^{(k)}, x_{t+1}^{(k)}, \dots, x_M^{(k)}]$$

and

$$\mathbf{x}_j^{(k)} \triangleq [\mathbf{X}_{[-t]}^{(k)}]_{j-q:j}.$$

The above algorithm generates a Markov chain

$$\mathbf{X}^{(0)}, (\phi^{(1)}, \sigma^{2(1)}), \mathbf{X}^{(1)}, \dots, (\phi^{(k)}, \sigma^{2(k)}), \mathbf{X}^{(k)}, \dots$$

Now, consider only the \mathbf{X} part of the above chain. It is again a first-order Markov chain taking values on a finite set $\{+1, -1\}^{n+q+1}$ with transition probability from $\mathbf{X}^{(k)}$ to $\mathbf{X}^{(k+1)}$ being

$$\begin{aligned}
 & P\left(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}\right) \\
 & \propto \int p\left(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}, \phi, \sigma^2, \mathbf{Y}\right) \\
 & \quad \times p\left(\phi, \sigma^2 | \mathbf{X}^{(k)}, \mathbf{Y}\right) d\phi d\sigma^2 \\
 & \propto \int \prod_{t=-q}^n p\left(x_t^{(k+1)} | \mathbf{X}_{[-t]}^{(k)}, \phi, \sigma^2, \mathbf{Y}\right) \\
 & \quad \times p\left(\phi, \sigma^2 | \mathbf{X}^{(k)}, \mathbf{Y}\right) d\phi d\sigma^2. \quad (51)
 \end{aligned}$$

With σ^2 known, (51) is simplified to

$$P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) \propto \int \prod_{t=-q}^n p(x_t^{(k+1)} | \phi, \mathbf{Y}, \mathbf{X}_{[-t]}^{(k)}) \times \psi(\phi; \boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}) d\phi \quad (52)$$

where $\psi(z; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density function of a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The integral (52) can be easily evaluated using Monte Carlo integration by sampling ϕ from $\mathcal{N}(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{2(k+1)})$ and average the corresponding values of

$$\prod_{t=-q}^n p(x_t^{(k+1)} | \phi, \mathbf{Y}, \mathbf{X}_{[-t]}^{(k)}). \quad (53)$$

Another interesting Gibbs sampling scheme, which can help alleviate the channel memory problem, is based on the ‘‘grouping’’ idea [5], [7], [9], [15]. In particular, Carter and Kohn [5] utilize a forward–backward algorithm to sample \mathbf{X} jointly, conditional on \mathbf{Y} and the parameters. This scheme is shown to be effective when the \mathbf{X} forms a Gaussian Markov model or a Markov chain whose state variable takes on only a few values. In the ISI channel decoding problem, the x_t are i.i.d. symbols *a priori*, but they are correlated *a posteriori* because of the observed signal \mathbf{Y} and the relationship (42). The induced correlation among the x_t vanishes after lag q . More precisely, instead of using (50) to sample \mathbf{X} iteratively, one can draw \mathbf{X} altogether.

Algorithm IV.2 [Grouping-Gibbs Equalizer in ISI Channel]:

- The first few steps are identical to the previous Gibbs equalizer.
- The last step is replaced by the forward–backward scheme. Conditional on ϕ and σ (we suppress the superscript for iteration numbers), we have the joint distribution of \mathbf{X} :

$$p(\mathbf{X} | \phi, \sigma, \mathbf{Y}) \propto \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - \phi^T \mathbf{x}_j)^2\right] \equiv \exp\{g_1(\mathbf{x}_1) + \dots + g_n(\mathbf{x}_n)\} \quad (54)$$

where $\mathbf{x}_j = (x_{j-q}, \dots, x_j)$. Thus, each \mathbf{x}_j can take 2^{q+1} possible values. The following two steps produce a sample \mathbf{X} from $p(\mathbf{X} | \phi, \sigma, \mathbf{Y})$.

- *Forward Summation:* Define $f_1(\mathbf{x}_1) = \exp\{g_1(\mathbf{x}_1)\}$ and compute recursively

$$f_{j+1}(\mathbf{x}_{j+1}) = \sum_{x_{j-q}=-1}^1 [f_j(\mathbf{x}_j) \exp\{g_{j+1}(\mathbf{x}_{j+1})\}]. \quad (55)$$

- *Backward Sampling:* First, draw $\mathbf{x}_n = (x_{n-q}, \dots, x_n)$ from distribution $P(\mathbf{x}_n) \propto f_n(\mathbf{x}_n)$. Then, for $j = n - q - 1, \dots, 1$, draw $P(x_j | x_{j+1}, \dots, x_n) \propto f_{j+q}(x_j, \dots, x_{j+q})$.

Although the grouping idea is attractive for overcoming the channel memory problem, the additional computation cost may offset its advantages. More precisely, the forward–backward procedure needs about 2^q times more memory and about 2^q times more basic operations.

Similar to the previous section, we can integrate out the continuous parameters and write down the marginal target distribution of \mathbf{X} :

$$\pi(\mathbf{X}) \propto [\nu\lambda + W]^{-(n+\nu)/2} \quad (56)$$

where

$$W = \sum y_t^2 - \left[\sum \mathbf{x}_t y_t \right]^T \left[\sum \mathbf{x}_t \mathbf{x}_t^T \right]^{-1} \left[\sum \mathbf{x}_t y_t \right]. \quad (57)$$

A similar idea in the ISI channel framework also appeared in [2]. We can then derive the one-component Gibbs and Metropolis algorithms accordingly. The phase ambiguity (i.e., likelihood unchanged when \mathbf{X} is changed to $-\mathbf{X}$) can be clearly seen from this joint distribution.

Algorithm IV.3 [One-Component Gibb/Metropolis Equalizer in ISI Channel]:

- Choose t from $1, \dots, n$ by either the random scan or the systematic scan. Let $\mathbf{X}^{(k+1)} = \mathbf{Z}$, where $z_s = x_s^{(k)}$ for $s \neq t$ and $z_t = -x_t^{(k)}$ with probability

$$\frac{\pi(\mathbf{Z})}{\pi(\mathbf{X}^{(k)}) + \pi(\mathbf{Z})} \quad (58)$$

for the Gibbs equalizer or with probability

$$\min\left\{1, \frac{\pi(\mathbf{Z})}{\pi(\mathbf{X}^{(k)})}\right\} \quad (59)$$

for the Metropolis equalizer, where $\pi(\mathbf{X})$ is as in (56). Otherwise, let $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)}$. When the variance σ^2 is known

$$\pi(\mathbf{X}) \propto \left\{ \frac{1}{|\sum \mathbf{x}_t \mathbf{x}_t^T|^{q/2}} \right\} \times \exp\left\{ \frac{1}{2\sigma^2} \left[\sum \mathbf{x}_t y_t \right]^T \left[\sum \mathbf{x}_t \mathbf{x}_t^T \right]^{-1} \left[\sum \mathbf{x}_t y_t \right] \right\}. \quad (60)$$

To overcome the phase ambiguity, we use differential coding in all of our algorithms. Denote $\mathbf{S} \triangleq [s_2, \dots, s_n]$ as the information bits. Let $s_t^{(k)} = x_t^{(k)} x_{t-1}^{(k)}$, $t = 2, \dots, n$. Since $\mathbf{X}^{(k)}$ forms a Markov chain, $\mathbf{S}^{(k)}$ is a Markov chain as well. The transition probability from $\mathbf{S}^{(k)}$ to $\mathbf{S}^{(k+1)}$ is

$$P(\mathbf{S}^{(k+1)} | \mathbf{S}^{(k)}) = P(\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) + P(-\mathbf{X}^{(k+1)} | \mathbf{X}^{(k)}) \quad (61)$$

where both $\mathbf{X}^{(k+1)}$ and $-\mathbf{X}^{(k+1)}$ result in $\mathbf{S}^{(k+1)}$, and $\mathbf{X}^{(k)}$ results in $\mathbf{S}^{(k)}$.

Fig. 5 shows the result of a simulation study for sample size $n = 5$, $\phi = [1, 0.5]$, and five different SNR values (which are defined as $\|\phi\|^2/\sigma^2$) at $-2, 0, 2, 4$ and 6 dB. Fifty samples of the \mathbf{X}, \mathbf{Y} are simulated with $v_t \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is determined by the SNR. Then, the transition matrix is constructed according to Algorithms IV.1–IV.3 (for the one-component deterministic scan Gibbs equalizer). Then, second-largest eigenvalues in absolute value are obtained. It can be seen that the one-component deterministic-scan Gibbs [see Fig. 5(b)] outperformed all other algorithms. The grouping-Gibbs equalizer [see Fig. 5(c)] is better than the two-component Gibbs [see Fig. 5(a)]. However, its computation is about twice (2^q with $q = 1$) that of the two-component Gibbs. Fig. 5(d) plots the results for the two-step transition of the two-component Gibbs, which requires about the same computation as the grouping-Gibbs. It shows that the standard two-component Gibbs equalizer is comparable with the grouping-Gibbs in this example. We can also

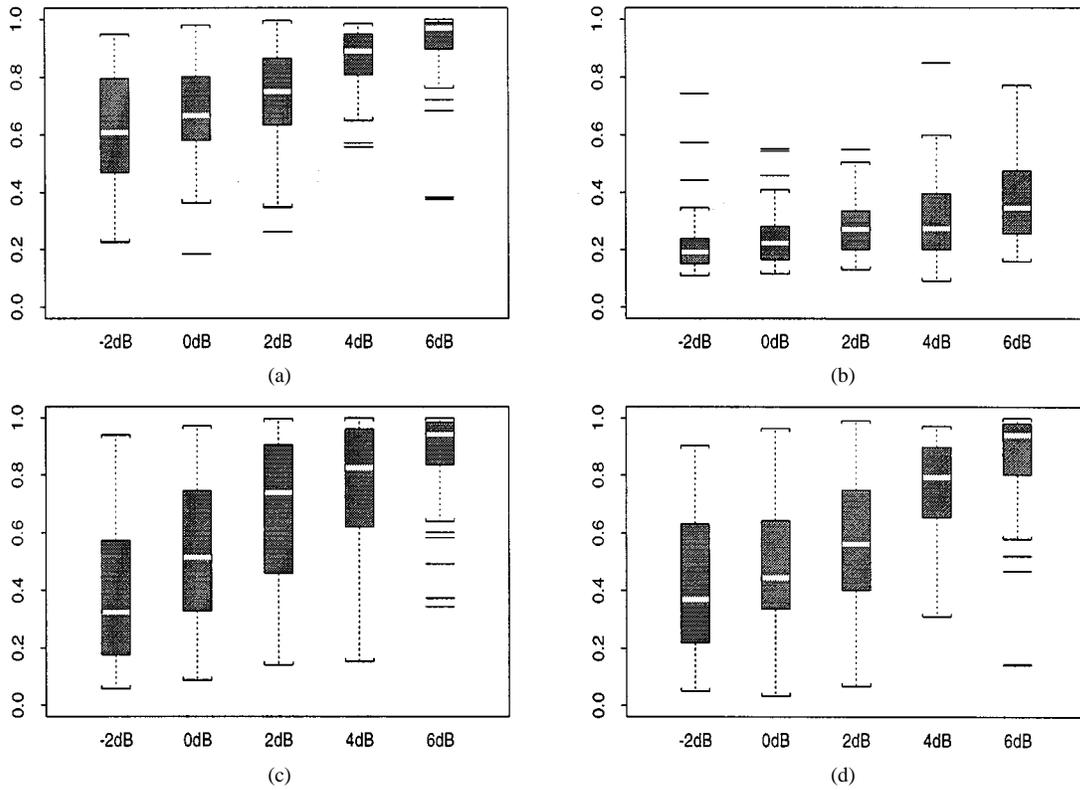


Fig. 5. Boxplots of convergence rates in ISI channels with $q = 1$: $n = 5$, $\phi = [1, 0.5]$. Two-component Gibbs detector, one-component deterministic scan Gibbs detector, grouping-Gibbs equalizer, and two-step two-component Gibbs detector are shown in (a)–(d), respectively.

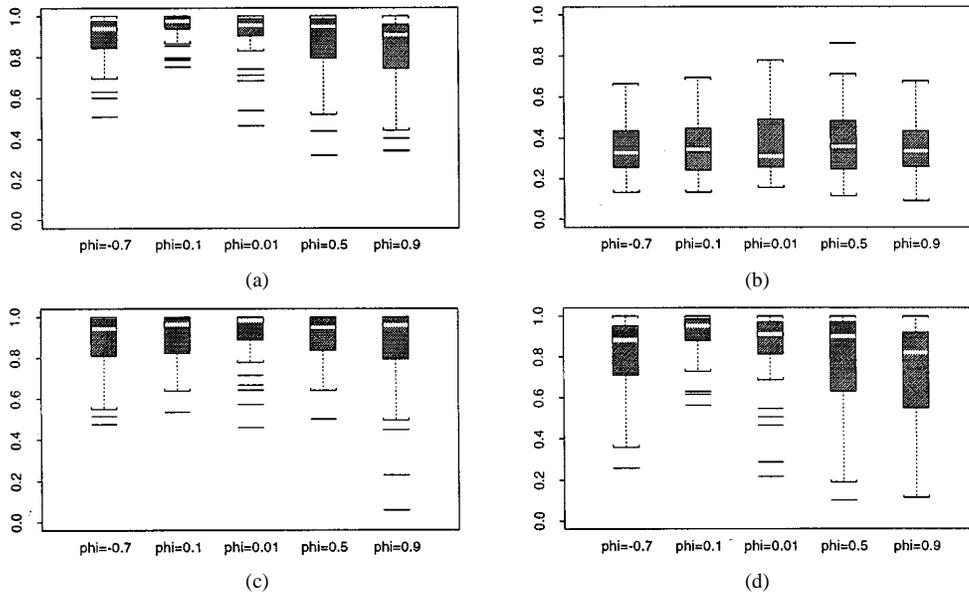


Fig. 6. Boxplots of convergence rates in ISI systems with $q = 1$, $n = 5$, SNR = 6 dB, and $\phi = [1, \phi_1]$. From (a) Two-component Gibbs. (b) One-component deterministic scan Gibbs. (c) Grouping-Gibbs. (d) Two-step two-component Gibbs detectors.

see that as the SNR increases, the convergence becomes slower. This is contrary to that in the AWGN channel (Fig. 4), where higher SNR results in faster convergence. The key reason is the “channel memory” problem and the shifting ambiguity to be discussed later.

Fig. 6 shows the boxplots of the same setting as that in Fig. 5 for SNR = 6 dB and $\phi = [1, \phi_1]$. The boxplots show the convergence rates for $\phi_1 = -0.7, 0.01, 0.1, 0.5, 0.9$ for Algorithms IV.1–IV.3. We can see that the Gibbs sampler converges slower

when ϕ_1 is close to 0. This is due to the fact that the shift ambiguity increases when ϕ_1 is close to 0. Note that the posterior distribution evaluated at $\phi = [r, 0]$ and $\phi = [0, r]$ have the same value. When data is generated from $\phi = [1, 0]$, the posterior modes for ϕ will be around both $\phi = [1, 0]$ and $\phi = [0, 1]$ with the same height, corresponding to (x_1, x_2, \dots, x_n) and its shifted version $(x_2, \dots, x_n, x_{n+1})$, respectively. If SNR is large, the Gibbs sample will have difficulty moving from one mode to the other, resulting in slow convergence rate. Wang and

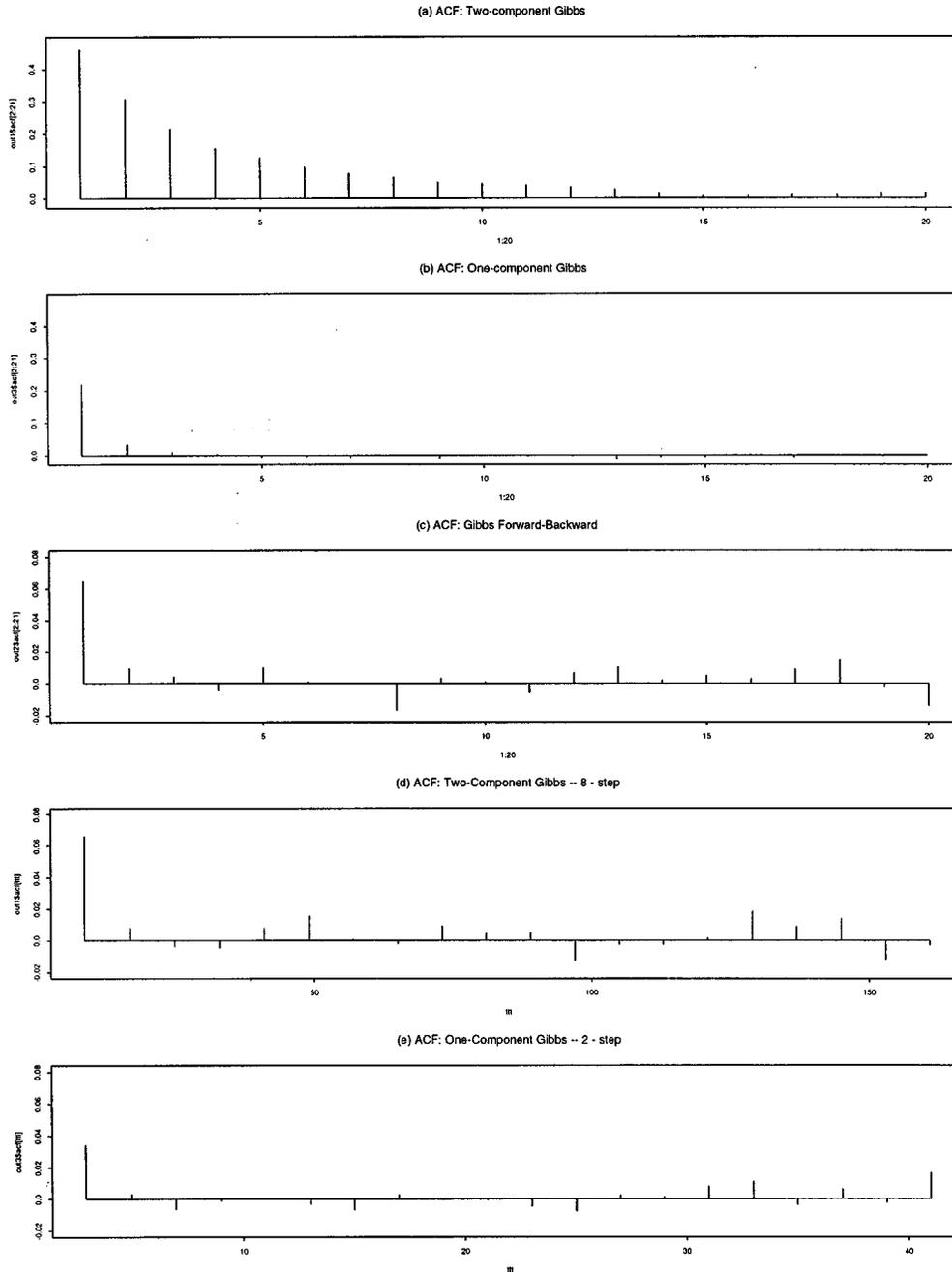


Fig. 7. Autocorrelations of the MCMC equalizers for an ISI system with $q = 4$, $n = 50$, and SNR = 8 dB. (a) Two-component Gibbs. (b) One-component Gibbs. (c) Group-Gibbs. (d) Eight-step transition for two-component Gibbs. (e) Two-step one-component Gibbs.

Chen [23] showed that an effective method is to impose constraints on the channel parameters ϕ and to employ rejection sampling in combination with shift operation to generate Monte Carlo samples. In the MCMC framework, one can also incorporate a shift proposal and use the general Metropolis–Hastings rule [14] to guide the Markov transitions.

To compare the algorithms more realistically, we simulated 50 observations from model (42) with

$$\phi = [-0.1987, 0.4812, 0.9162, -0.1833]^T.$$

Algorithms IV.1–IV.3 were performed under the constraint that ϕ_3 is positive and has the largest magnitude. Each algorithm was run 20 000 iterations, and the statistic W in (57) were recorded

for each iteration. We observed that all the algorithms were trapped in a shift mode in 20 000 iterations if no constraint was imposed.

As discussed in Section II, sample autocorrelations of a test function can be used as an alternative measure of convergence rate. The autocorrelation of the sequence W , after discarding the first 2000 iterations, were calculated and shown in Fig. 7(a)–(e). [Note that the scales for Fig. 7(a)–(e) are different.] Fig. 7(c) shows the eight-step autocorrelation of the two-component Gibbs, which is comparable with that of Fig. 7(b). Since the computation cost for the grouping-Gibbs is about eight times as much as the two-component Gibbs, we conclude that for this example, the two-component Gibbs

is as efficient as the Grouping-Gibbs. It is also seen that the one-component deterministic-scan Gibbs sampler again performed the best (which is about four times more efficient than the two-component one) in this example.

V. GIBBS MULTIUSER DETECTOR IN CDMA CHANNELS

Consider a CDMA system with K users, employing normalized modulation waveforms $\mathbf{h}_1, \dots, \mathbf{h}_K$ and signaling in the presence of additive white Gaussian noise. After filtering and sampling the continuous-time received waveform, the received signal at time t is given by

$$\mathbf{r}_t = \mathbf{H}\Phi\mathbf{x}_t + \mathbf{v}_t \quad (62)$$

where $\mathbf{H} \triangleq [\mathbf{h}_1 \ \mathbf{h}_2 \ \dots \ \mathbf{h}_K]$ contains the spreading waveforms of all users, $\Phi \triangleq \text{diag}\{\phi_1, \phi_2, \dots, \phi_K\}$ contains real amplitude of each user, $\mathbf{x}_t \triangleq [x_{1,t} \ x_{2,t} \ \dots \ x_{K,t}]^T$ contains the binary symbols sent by all users at time t , and $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ is the ambient Gaussian noise vector at time t . Clearly, the CDMA system is a generalization of the AWGN in that the multiuser symbol can be viewed as a ‘‘supersymbol.’’

At the receiver, the received signal \mathbf{r}_t is correlated with each of the spreading waveform to obtain the following sufficient statistic:

$$\mathbf{y}_t \triangleq \mathbf{H}^T \mathbf{r}_t = \mathbf{R}\Phi\mathbf{x}_t + \mathbf{u}_t \quad (63)$$

where

$$\begin{aligned} \mathbf{R} &\triangleq \mathbf{H}^T \mathbf{H} \\ \mathbf{u}_t &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}). \end{aligned}$$

The multiuser waveform correlation matrix \mathbf{R} is known at the receiver. For equiponderated signals, it is given by

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}_{K \times K}$$

where ρ ($0 \leq \rho < 1$) is the correlation between the signaling waveform of any two users.

Denote $\boldsymbol{\phi} \triangleq [\phi_1 \ \dots \ \phi_K]^T$ and $\mathbf{X}_t = \text{diag}\{x_{1,t}, \dots, x_{K,t}\}$. Denote also $\boldsymbol{\mathcal{X}} \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_n]$ and $\mathbf{Y} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_n]$. With a flat prior for $\boldsymbol{\phi}$ and known σ^2 , a Gibbs sampler starts with an initial value $\boldsymbol{\mathcal{X}}^{(0)}$ and iterates between the following steps for $j = 0, 1, \dots$

Algorithm V.1 [Gibbs Multiuser Detector in CDMA Channel]:

- Draw a sample $\boldsymbol{\phi}^{(j+1)}$ from the following conditional distribution [given $\boldsymbol{\mathcal{X}}^{(j)}$]:

$$p(\boldsymbol{\phi} | \boldsymbol{\mathcal{X}}^{(j)}, \mathbf{Y}) \sim \mathcal{N}(\boldsymbol{\mu}^{(j+1)}, \boldsymbol{\Sigma}^{(j+1)}) \quad (64)$$

where

$$\boldsymbol{\Sigma}^{(j+1)} \triangleq \left[\frac{1}{\sigma^2} \sum_{t=1}^n \mathbf{X}_t^{(j)} \mathbf{R} \mathbf{X}_t^{(j)} \right]^{-1} \quad (65)$$

$$\boldsymbol{\mu}^{(j+1)} \triangleq \boldsymbol{\Sigma}^{(j+1)} \left(\frac{1}{\sigma^2} \sum_{t=1}^n \mathbf{X}_t^{(j)} \mathbf{y}_t \right). \quad (66)$$

- For $t = 1, \dots, n$ and $k = 1, \dots, K$, draw a sample $x_{k,t}^{(j+1)}$ from the following:

$$\begin{aligned} & \frac{p[x_{k,t} = +1 | \mathbf{Y}, \boldsymbol{\phi}^{(j+1)}, \boldsymbol{\mathcal{X}}_{[-k,-t]}^{(j)}]}{p[x_{k,t} = -1 | \mathbf{Y}, \boldsymbol{\phi}^{(j+1)}, \boldsymbol{\mathcal{X}}_{[-k,-t]}^{(j)}]} \\ &= \exp \left\{ \frac{2\phi_k^{(j+1)}}{\sigma^2} \left(y_{k,t} - \mathbf{e}_k^T \mathbf{R} \Phi^{(j+1)} \mathbf{x}_{k,t}^{(j)} \right) \right\} \end{aligned} \quad (67)$$

where

$$\mathbf{x}_{k,t}^{(j)} \triangleq [x_{1,t}^{(j+1)} \ \dots \ x_{k-1,t}^{(j+1)} \ 0 \ x_{k+1,t}^{(j)} \ \dots \ x_{K,t}^{(j)}]^T$$

and \mathbf{e}_k denotes the k th unit basis vector of \mathbb{R}^K .

The one-component Gibbs detector can be derived easily as in the AWGN cases and is omitted here. Again, to cope with the phase ambiguity, we use differential encoding for each user. Let the information sequence of the k th user be $s_{k,t} \in \{+1, -1\}$, $t = 2, \dots, n$; then, the transmitted sequence for this user is given by $x_{k,t} = x_{k,t-1} s_{k,t}$, $t = 1, \dots, n$. Denote $\mathbf{s}_t \triangleq [s_{1,t} \ \dots \ s_{K,t}]^T$ and $\boldsymbol{\mathcal{S}} \triangleq [\mathbf{s}_2, \dots, \mathbf{s}_n]$, and let $\text{diff}(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{S}}$.

The foregoing Gibbs iteration generates a Markov chain

$$\boldsymbol{\mathcal{S}}^{(0)}, \boldsymbol{\phi}^{(1)}, \boldsymbol{\mathcal{S}}^{(1)}, \dots, \boldsymbol{\phi}^{(n)}, \boldsymbol{\mathcal{S}}^{(n)}, \dots$$

The transition matrix is

$$\begin{aligned} P(\boldsymbol{\mathcal{S}}^{(j+1)} | \boldsymbol{\mathcal{S}}^{(j)}) &= P(\boldsymbol{\mathcal{X}}^{(j+1)} | \boldsymbol{\mathcal{X}}^{(j)}) \\ &\quad + P(-\boldsymbol{\mathcal{X}}^{(j+1)} | \boldsymbol{\mathcal{X}}^{(j)}) \end{aligned} \quad (68)$$

where $\text{diff}(\boldsymbol{\mathcal{X}}^{(j)}) = \boldsymbol{\mathcal{S}}^{(j)}$, $\text{diff}(\boldsymbol{\mathcal{X}}^{(j+1)}) = \text{diff}(-\boldsymbol{\mathcal{X}}^{(j+1)}) = \boldsymbol{\mathcal{S}}^{(j+1)}$, and

$$\begin{aligned} & P(\boldsymbol{\mathcal{X}}^{(j+1)} | \boldsymbol{\mathcal{X}}^{(j)}) \\ &= \int \left\{ \prod_{t=1}^n \prod_{k=1}^K p(x_{k,t}^{(j+1)} | \mathbf{Y}, \boldsymbol{\phi}, \boldsymbol{\mathcal{X}}_{[-k,-t]}^{(j)}) \right. \\ &\quad \left. \times \psi(\boldsymbol{\phi}; \boldsymbol{\mu}^{(j+1)}, \boldsymbol{\Sigma}^{(j+1)}) \right\} d\boldsymbol{\phi}. \end{aligned} \quad (69)$$

To investigate the convergence property of this algorithm, we carried out a small simulation study. For $k = 2$, $n = 4$, $\rho = 0.5$, $\phi_1 = \phi_2 = 1$, and four different levels of SNR (defined as ϕ_1^2/σ^2) (SNR = 2, 4, 6, 8, 10, 12 dB), we simulated fifty sequences each and calculated their convergence rate. Fig. 8(a) shows the boxplots of the convergence rate for different SNRs.

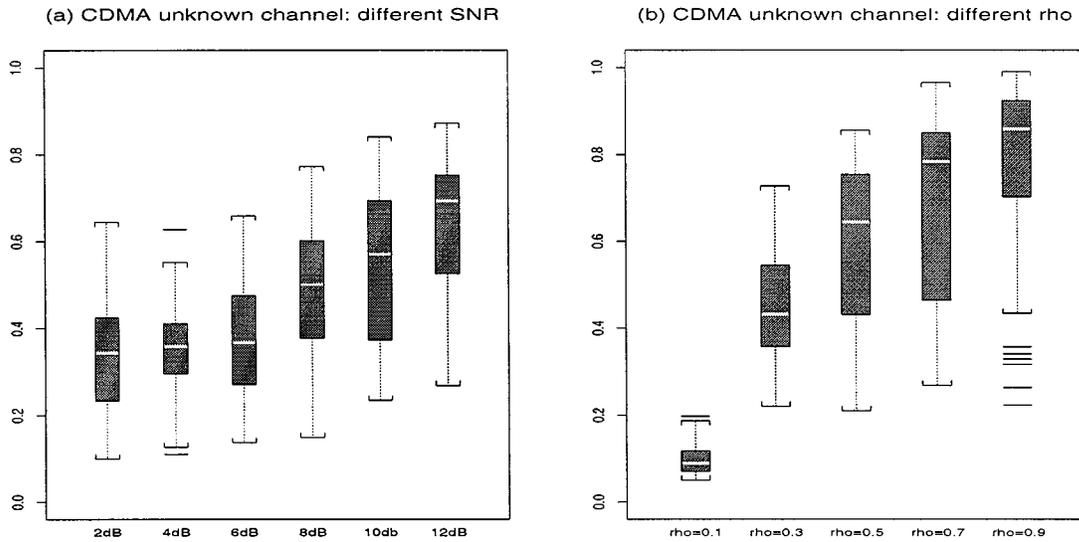


Fig. 8. Boxplots of the convergence rates in CDMA channels. (a) $n = 4$, $K = 2$, $\rho = 0.5$. (b) $n = 4$, $K = 2$, and SNR = 0 dB.

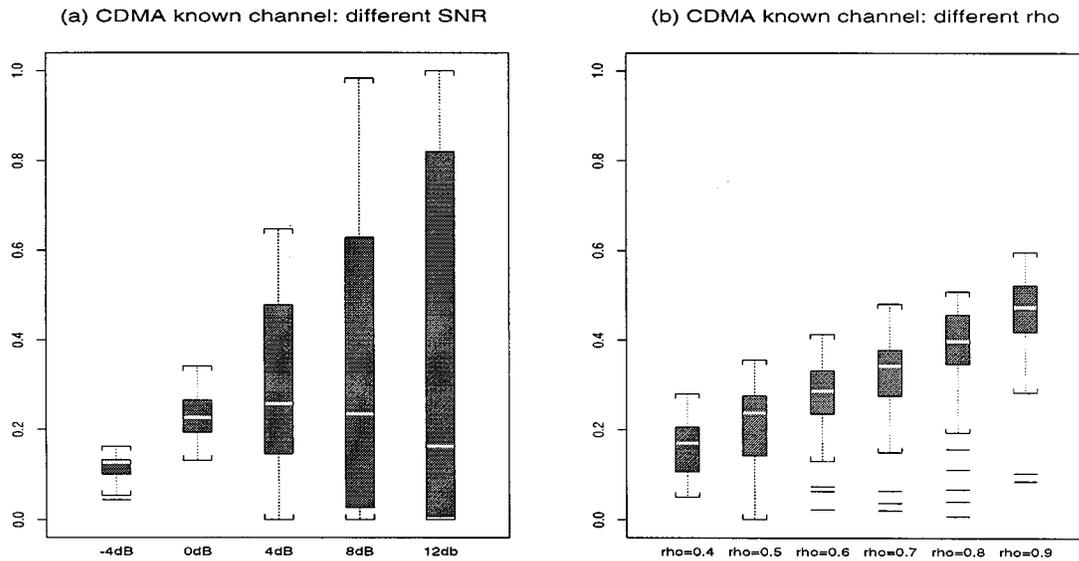


Fig. 9. Boxplots of convergence rates for multiuser systems. (a) $n = 1$, $K = 5$, $\rho = 0.5$. (b) $n = 1$, $K = 5$, and SNR = 0 dB.

We can see that as SNR becomes larger, the Gibbs sampler becomes stickier, with slower convergence rate, due to the fact that x_t 's are sampled conditioning on the other x 's as that in the ISI channel (see Fig. 5).

The next simulation is designed to see the effect of the correlation ρ . For $K = 2$, $n = 4$, $\phi_1 = \phi_2 = 1$, and SNR = 2 dB and four different values of $\rho = [0.1, 0.3, 0.5, 0.7, 0.9]$, we simulated 50 samples each and calculated their convergence rate. Fig. 8(b) shows the boxplots of the convergence rate for each ρ . It is seen that the high correlation of user signals results in a slower convergence of the Gibbs multiuser detector.

Another interesting situation is the detection of the multiuser symbols when the channel conditions (i.e., ϕ and σ^2) are known to the receiver [20]. In this case, the system (63) becomes independent for different t . That is

$$p(\mathcal{X}|\mathbf{Y}) = \prod_{t=1}^n p(\mathbf{x}_t | \mathbf{y}_t).$$

Hence, we can simply consider the case that $n = 1$. The Gibbs algorithms for interference cancellation becomes the following. Given initial values $\mathbf{x}^{(0)}$, a Gibbs sampler iterates between the following steps for $j = 1, 2, \dots$

Algorithm V.2 [Gibbs Interference Canceller in CDMA Channel]:

- For $k = 1, \dots, K$, draw a sample $x_k^{(j+1)}$ from the following.

$$\frac{p \left[x_k = +1 \mid \mathbf{Y}, \mathbf{x}_{[-k]}^{(j)} \right]}{p \left[x_k = -1 \mid \mathbf{Y}, \mathbf{x}_{[-k]}^{(j)} \right]} = \exp \left\{ \frac{2\phi_k^{(j+1)}}{\sigma^2} \left(y_k - \mathbf{e}_k^T \mathbf{R} \Phi \mathbf{x}_k^{0(j)} \right) \right\} \quad (70)$$

where

$$\mathbf{x}_k^{0(j)} \triangleq \left[x_1^{(j+1)} \dots x_{k-1}^{(j+1)}, 0, x_{k+1}^{(j)} \dots x_K^{(j)} \right]^T$$

and \mathbf{e}_k denotes the k th unit basis vector of \mathbb{R}^K .

The transition matrix then becomes

$$P(\mathbf{x}^{(j+1)} | \mathbf{x}^{(j)}) = \prod_{k=1}^K p(x_k^{(j+1)} | Y, \mathbf{x}_{[-k]}^{(j)}). \quad (71)$$

Differential coding is not necessary since there is no phase ambiguity when Φ is known. Note that the above Gibbs sampler is defined on the $\{-1, 1\}^K$ cube and is, thus, geometrically convergent. To get some insights on its convergence rate, we simulated 50 samples from the system with $K = 5$, $\phi_1 = \dots = \phi_5 = 1$, and $\rho = 0.5$ for SNR = -4, 0, 4, 8, 12 dB, respectively. The corresponding convergence rates were calculated and shown in Fig. 9(a). Again, it can be seen that for small SNR, the convergence is faster, whereas for large SNR, the range of the convergence rate becomes larger. It depends significantly on the observed signals, with many rates close to 1.

To obtain some insights on the effect of the correlation ρ , we simulated 50 samples for $K = 5$, $\phi_1 = \dots = \phi_5 = 1$, SNR = 0 dB, and $\rho = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Fig. 9(b) shows the boxplots. We can see clearly that as ρ increases, the convergence becomes slower.

VI. CONCLUSIONS

In this paper, we investigated the convergence properties of Bayesian MCMC receivers in AWGN channels, ISI channels, and multiuser CDMA channels. We have shown that these MCMC receivers converge geometrically. The rate of the convergence depends on various characteristics of the channel, particularly the SNR, the correlation between the observations, and various ambiguities. Among many possible MCMC designs, we found that the idea of integrating out continuous parameters [2], [15] provides a consistent performance improvement for the standard two-component Gibbs sampler in all examples we have tested. The grouping technique based on the forward-backward method [5] can be useful for decoding the ISI channels. However, its additional computation cost may be more than offset its benefits when the number of channel taps are large.

The differential coding is efficient in removing the phase ambiguity and can be used to accelerate the convergence of the Gibbs sampler. When applied to more complicated problems such as ISI and CDMA channels, however, the effect of differential coding diminishes, and some other tricks, such as using a shift proposal or imposing appropriate constraints, are needed. By implementing various tactics to accelerate the convergence, the Bayesian receivers can be more efficient and accurate. Our study also shows that although the exact convergence rate of an MCMC sampler or even a reasonable bound of it is difficult to obtain, results based on small and simple systems and on autocorrelation plots often yield important insights on the convergence properties.

REFERENCES

[1] C. Andrieu, E. Barat, and A. Doucet, "Bayesian deconvolution of noisy filtered point processes," *IEEE Trans. Signal Processing*, vol. 49, pp. 134–146, Jan. 1999.

[2] C. Andrieu, A. Doucet, and S. J. Godsill, "Blind deconvolution of convolutive discrete source mixture," in *Proc. Workshop IEEE NNSP*, Cambridge, MA, 1998.

[3] R. N. Bhattacharya and E. C. Waymire, *Stochastic Processes with Applications*. New York: Wiley, 1990.

[4] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA: SIAM, 1994.

[5] C. K. Carter and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, vol. 81, no. 3, pp. 541–553, 1994.

[6] R. Chen and T.-H. Li, "Blind restoration of linearly degraded discrete signals by Gibbs sampler," *IEEE Trans. Signal Processing*, vol. 43, pp. 2410–2413, Sept. 1995.

[7] T. C. Clapp and S. J. Godsill, "Bayesian blind deconvolution for mobile communications," in *Proc. Inst. Elect. Eng. Colloq. Adaptive Signal Process. Mobile Commun. Syst.*, London, U.K., 1997.

[8] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues for Markov chains," *Ann. Appl. Probab.*, vol. 1, pp. 36–61, 1991.

[9] A. Doucet and P. Duvaut, "Fully Bayesian analysis of hidden Markov models," in *Proc. Euro. Signal Process. Conf.*, Trieste, 1996.

[10] A. Frigessi, P. Distefano, C. R. Hwang, and S. J. Sheu, "Convergence rates of the Gibbs sampler; the Metropolis algorithm and other single-site updating dynamics," *J. R. Statist. Soc. B*, vol. 55, no. 1, pp. 205–219, 1993.

[11] C. N. Georghiadis and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 300–308, Mar. 1997.

[12] M. Jerrum and A. Sinclair, "Approximating the permanent," *SIAM J. Comput.*, vol. 18, no. 6, pp. 1149–1178, 1989.

[13] J. S. Liu, "Peskun's theorem and a modified discrete-state Gibbs sampler," *Biometrika*, vol. 83, no. 3, pp. 681–682, 1996.

[14] —, *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag, 2001.

[15] 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.

[16] J. S. Liu, W. H. Wong, and A. Kong, "Covariance structure and convergence rate of the Gibbs sampler with various scans," *J. R. Statist. Soc. B*, vol. 57, no. 1, pp. 157–169, 1995.

[17] E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge, U.K.: Cambridge Univ. Press, 1984.

[18] G. O. Roberts and J. S. Rosenthal, "Markov-chain Monte Carlo: Some practical implications of theoretical results," *Can. J. Statist.*, vol. 26, no. 1, pp. 5–20, 1998.

[19] J. S. Rosenthal, "Minorization conditions and convergence-rates for Markov-chain Monte-Carlo," *J. Amer. Statist. Assoc.*, vol. 90, no. 430, pp. 558–566, 1995.

[20] T. M. Schmidl, A. Gatherer, X. Wang, and R. Chen, "Interference cancellation using the Gibbs sampler," in *Proc. 2000 IEEE Semi Annual Veh. Technol. Conf.*, Boston, MA, Sept. 27–29, 2000.

[21] L. Tierney, "Markov-chains for exploring posterior distributions," *Ann. Statist.*, vol. 22, no. 4, pp. 1701–1728, 1994.

[22] X. Wang and R. Chen, "Adaptive Bayesian multiuser detection for synchronous CDMA in Gaussian and impulsive noise," *IEEE Trans. Signal Processing*, vol. 48, pp. 2013–2028, July 2000.

[23] —, "Blind turbo equalization in Gaussian and impulsive noise," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 1092–1105, July 2001.



Rong Chen (M'92) received the B.S. degree in mathematics from Peking University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in statistics from Carnegie Mellon University, Pittsburgh, PA, in 1987 and 1990, respectively.

He is a Professor with the Department of Information and Decision Sciences, College of Business Administration, University of Illinois at Chicago (UIC). Before joining UIC in 1999, he was with the Department of Statistics, Texas A&M University, College Station. His main research interests are in time series analysis, statistical computing and Monte Carlo methods in dynamic systems, and statistical applications in engineering and business. He is an Associate Editor for the *Journal of American Statistical Association*, the *Journal of Business and Economic Statistics*, *Statistica Sinica*, and *Computational Statistics*.



Jun S. Liu received the B.S. degree in mathematics in 1985 from Peking University, Beijing, China, and the Ph.D. degree in statistics in 1991 from the University of Chicago, Chicago, IL.

He was Assistant Professor of Statistics at Harvard University, Cambridge, MA, from 1991 to 1994 and joined the Statistics Department of Stanford University, Stanford, CA, in 1994, as Assistant Professor. In 2000, he returned to Harvard Statistics Department as Professor. His main research interests are Bayesian modeling and computation, Monte

Carlo methods, bioinformatics, dynamic systems, and nonlinear filtering. He has worked with collaborators on designing novel Gibbs sampling-based algorithms to find subtle repetitive patterns in biopolymer sequences. These algorithms have been successfully applied to predict protein functions and to understand gene regulation mechanisms. He has also worked on theoretical analyses of Markov chain Monte Carlo algorithms, such as efficiency comparison of different algorithms, novel algorithmic designs, and convergence rate studies. His research recently focuses on designing more efficient sequential Monte Carlo-based filtering methods for signal processing and developing statistical tools for genomic data analysis.

Dr. Liu received the CAREER Award from the National Science Foundation in 1995 and the Mitchell Prize from the International Society of Bayesian Analysis in 2000.



Xiaodong Wang (M'98) received the B.S. degree in electrical engineering and applied mathematics (with the highest honor) from Shanghai Jiao Tong University, Shanghai, China, in 1992, the M.S. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1995, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1998.

From July 1998 to December 2001, he was with the Department of Electrical Engineering, Texas A&M University, College Station, as an Assistant Professor.

He is now with the Department of Electrical Engineering, Columbia University, New York, NY, as an Assistant Professor. His research interests fall in the general areas of computing, signal processing, and communications. He has worked in the areas of digital communications, digital signal processing, parallel and distributed computing, nanoelectronics, and quantum computing. His current research interests include multiuser communications theory and advanced signal processing for wireless communications. He was with AT&T Labs—Research, Red Bank, NJ, during the summer of 1997.

Dr. Wang is a member of the American Association for the Advancement of Science. He received the 1999 NSF CAREER Award and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.