The Annals of Applied Probability
1996, Vol. 6, No. 2, 444–465

COORDINATE SELECTION RULES FOR GIBBS SAMPLING[1]

BY GEORGE S. FISHMAN

University of North Carolina

This paper studies several different plans for selecting coordinates for updating via Gibbs sampling. It exploits the inherent features of the Gibbs sampling formulation, most notably its neighborhood structure, to characterize and compare the plans with regard to convergence to equilibrium and variance of the sample mean. Some of the plans rely completely or almost completely on random coordinate selection. Others use completely or almost completely deterministic coordinate selection rules. We show that neighborhood structure induces idempotency for the individual coordinate transition matrices and commutativity among subsets of these matrices. These properties lead to bounds on eigenvalues for the Gibbs sampling transition matrices corresponding to several of the plans. For a frequently encountered neighborhood structure, we give necessary and sufficient conditions for a commonly employed deterministic coordinate selection plan to induce faster convergence to equilibrium than the random coordinate selection plans. When these conditions hold, we also show that this deterministic selection rule achieves the same worst-case bound on the variance of the sample mean as that arising from the random selection rules when the number of coordinates grows without bound. This last result encourages the belief that faster convergence for the deterministic selection rule may also imply a variance of the sample mean no larger than that arising for random selection rules.

Introduction. Given its relatively simple algorithmic formulation, the appeal of Gibbs sampling is understandable as a method for generating sample paths in Monte Carlo experiments. Recent interest in the method has generated a considerable literature on its convergence properties [e.g., Amit and Grenander (1989), Barone and Frigessi (1990), Frigessi, Hwang, Sheu and di Stefano (1993), Liu, Wong and Kong (1994, 1995), Roberts and Polson (1991), Roberts and Smith (1992) and Schervish and Carlin (1993)]. While this literature exploits the positivity properties of the Gibbs sampler, a comparable exploitation of its intrinsic neighborhood properties has been less common. A notable exception is Amit (1991). The present paper describes five plans for performing Gibbs sampling and uses neighborhood properties to characterize and compare them with regard to convergence to equilibrium and variance of the sample mean. Although the term Gibbs sampling usually connotes the iterative updating of coordinates one at a time in a prescribed *deterministic*

Received December 1994; revised January 1996.

[1]Research partially supported by NSF Grant DDM-89-13344. The Government has certain rights in this material. Any opinions, findings and conclusions or recommendations in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

AMS 1991 *subject classifications*. 60J05, 62E25.

Key words and phrases. Coordinate selection, Gibbs sampling, Markov random field, Monte Carlo.

order, the present paper includes *random* coordinate selection as part of Gibbs sampling.

Plan 1 allows random coordinate selection on each step. Plan 2 restricts Plan 1 so that no coordinate can be repeatedly updated on successive steps. Plan 3 describes a strictly deterministic coordinate selection plan often suggested to ensure reversibility. Plan 4 describes an alternative deterministic selection strategy that, before updating begins, randomly decides whether to update coordinates in the order $l_1, \ldots, l_m$ or $l_m, \ldots, l_1$. Plan 5 is a special case of Plan 4 that demonstrates the benefit of batching coordinates into two groups based on specified neighborhood properties. Before updating begins, it randomly chooses the order in which the groups are to be updated on each pass through the coordinates.

Section 1 formulates the problem as one of sampling from a Markov random field on an m-dimensional countably finite product space. For random coordinate selection rules, Section 2 uses the reversibility, idempotency and reducibility of the Gibbs sampler induced transition matrices $\mathbf{P}_1, \ldots, \mathbf{P}_m$ for the m coordinates to derive lower and upper bounds on the eigenvalues of the one-step expected transition matrix for Plan 1. For alternative sampling plans, Section 3 introduces criteria for comparing convergence rates and for comparing variances of sample means. It illustrates these criteria when comparing Plan 1 based on randomly selecting coordinates with equal probability on each step and Plan 2 based on barring the coordinate selected on step $j-1$ from being selected on step j.

Section 4 begins the discussion of deterministic coordinate selection rules. It shows that one can expect Plan 4 to induce convergence to equilibrium at least as fast as Plan 3 does. This property encourages deeper study in Section 5 of strategies of the general form of Plan 4.

When the neighborhood condition assumes a special but commonly encountered form, Section 5 shows how to partition the coordinates into two subsets that facilitate analysis. In particular, transition matrices corresponding to coordinates within a subset commute with each other but not with transition matrices in the other subset. Section 5 exploits this property in Plan 5 and derives necessary and sufficient conditions for the plan to induce faster convergence than Plans 1 and 2.

Section 6 concentrates on variances. In particular, its analysis encourages the conjecture that Plan 5 induces a smaller variance for the sample mean than Plans 1 and 2 whenever Plan 5 induces faster convergence.

In what follows, we repeatedly make use of the inequalities $\operatorname{rank}(\mathbf{A}+\mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B})$ and $\operatorname{rank}(\mathbf{AB}) = \min[\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})]$ for any two matrices $\mathbf{A}$ and $\mathbf{B}$, and on a theorem of Weyl:

THEOREM 0. *Let $\lambda_0(\mathbf{C}) \geq \lambda_1(\mathbf{C}) \geq \cdots \geq \lambda_{v-1}(\mathbf{C})$ denote the ordered eigenvalues of the $v \times v$ symmetric matrix $\mathbf{C}$. If $\mathbf{A}$ and $\mathbf{B}$ are $v \times v$ symmetric matrices, then for all $0 \leq l \leq v-1$,*

$$\max[\lambda_l(\mathbf{A}) + \lambda_{v-1}(\mathbf{B}), \lambda_{v-1}(\mathbf{A}) + \lambda_l(\mathbf{B})] \\ \leq \lambda_l(\mathbf{A}+\mathbf{B}) \leq \min[\lambda_l(\mathbf{A}) + \lambda_0(\mathbf{B}), \lambda_0(\mathbf{A}) + \lambda_l(\mathbf{B})].$$

See, for example, Horn and Johnson [(1985), Theorem 4.3.1, page 181] or Marcus and Minc (1964).

1. Basic notation. Let $\{\mathbf{S}_j = (S_{1j}, \dots, S_{mj}),\ j \ge 0\}$ denote a stochastic process taking values in the countably finite state space $\mathscr{S} = \mathscr{S}_1 \times \cdots \times \mathscr{S}_m$ of size $v = |\mathscr{S}|$ and let $\boldsymbol{\pi} = \{\pi(\mathbf{x}),\ \mathbf{x} = (x_1, \dots, x_m) \in \mathscr{S}\}$ denote a probability mass function such that for each $\mathbf{x} \in \mathscr{S}$ and all $j \ge 0$,

$$\mathrm{pr}(\mathbf{S}_j = \mathbf{x}) = \pi(\mathbf{x}).$$

For each $x \in \mathscr{S}_i$ and $1 \le i \le m$,

$$\mathrm{pr}(S_{ij} = x | S_{lj} = x_l,\ \forall\ l \ne i) = \pi_i(x|x_l,\ \forall\ l \ne i) := \frac{\pi(\mathbf{x}_i(x))}{\sum_{y \in \mathscr{S}_i} \pi(\mathbf{x}_i(y))},$$

where

$$\mathbf{x}_i(y) := (x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_m).$$

Let $\mathscr{M} = \{1, \dots, m\}$ denote the collection of all sites or coordinates in the state vector and let $\mathscr{G}_1, \dots, \mathscr{G}_m$ denote subsets of $\mathscr{M}$ which for each i and each $l \in \mathscr{M}$ have the properties $i \notin \mathscr{G}_i,\ l \notin \mathscr{G}_l$ and $i \in \mathscr{G}_l \Leftrightarrow l \in \mathscr{G}_i$. Assume that for each $\mathbf{x} \in \mathscr{S}$,

$$\pi(\mathbf{x}) > 0 \quad \text{(positivity condition)}$$

and

$$\pi_i(x|x_l,\ \forall\ l \ne i) = \pi_i(x|x_l,\ l \in \mathscr{G}_i) \qquad \forall\ x \in \mathscr{S}_i,\ 1 \le i \le m$$
$$\text{(neighborhood condition).}$$

The function $\boldsymbol{\pi}$ is called a *Markov random field* with *neighborhood system* $\{\mathscr{G}_1, \dots, \mathscr{G}_m\}$, where $\mathscr{G}_i$ denotes the neighborhood of coordinate i.

Let $\{u(\mathbf{x}),\ \mathbf{x} \in \mathscr{S}\}$ denote a one-to-one mapping of the states $\mathbf{x}$ in $\mathscr{S}$ onto the integers $\mathscr{V} = \{0, 1, \dots, v-1\}$ and let $\pi_{u(\mathbf{x})} := \pi(\mathbf{x})$, for each $\mathbf{x} \in \mathscr{S}$. For each coordinate $l = 1, \dots, m$, let $\mathbf{P}_l := \| p_{ijl} \|_{i,\,j=0}^{v-1}$ denote a $v \times v$ Markov transition matrix with

$$p_{u(\mathbf{x})u(\mathbf{x}_l(y))l} := \pi_l(y|x_i,\ i \in \mathscr{G}_l) \qquad \forall\ y \in \mathscr{S}_l \text{ and } \forall\ \mathbf{x} \in \mathscr{S}.$$

Since

$$\pi_l(y|x_i,\ i \in \mathscr{G}_l)\, \pi(\mathbf{x}) = \pi_l(y|x_i,\ i \in \mathscr{G}_l)\, \pi_l(x_l|x_i,\ i \in \mathscr{G}_l)\, \pi(\mathbf{x}_{-l}),$$

where

$$\mathbf{x}_{-l} := (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_m)$$

and since

$$\pi(\mathbf{x}(y)) = \pi_l(y|x_i,\ i \in \mathscr{G}_l)\, \pi(\mathbf{x}_{-l}),$$

then

$$\pi_l(y|x_i,\ i \in \mathscr{G}_l)\, \pi(\mathbf{x}) = \pi_l(x_l|x_i,\ i \in \mathscr{G}_l)\, \pi(\mathbf{x}_l(y)),$$

so that $\mathbf{P}_l$ is reversible w.r.t. $\boldsymbol{\pi}$. That is, $\mathbf{D}\mathbf{P}_l = \mathbf{P}_l^T\mathbf{D}$ for $\mathbf{D} := \mathbf{diag}(\pi_0, \pi_1, \dots, \pi_{v-1})$.

2. Random coordinate selection. One method of generating a sample path $\{\mathbf{S}_j\}$ randomly selects a coordinate on each step and updates it. Let $w_i > 0,\ 1 \le i \le m$ and $w_1 + \cdots + w_m = 1$. Plan 1 describes the approach:

PLAN 1.
Randomly generate $\mathbf{S}_0$ from an initializing p.m.f. $\boldsymbol{\pi}_0$ on $\mathscr{S}$.
Randomly generate I from $\mathscr{M}$ with p.m.f. $\{w_l,\ 1 \le l \le m\}$.
Randomly generate S_{Ij} from $\mathscr{S}_I$ using row $u(\mathbf{S}_{j-1})$ of $\mathbf{P}_I$.
$(*)$ $S_{lj} \leftarrow S_{l,j-1},\ \forall\, l \ne I.$
$j \leftarrow j+1.$

Step $(*)$ is included for completeness here and in subsequent plans. In practice, most implementations are able to avoid it, thereby reducing cost.

On each step, Plan 1 has the one-step expected transition matrix

$$\mathbf{L}(\mathbf{w}) = w_1\mathbf{P}_1 + \cdots + w_m\mathbf{P}_m$$

and k-step transition matrix

$$\mathbf{Q}_{1k}(\mathbf{w}) = \mathbf{L}^k(\mathbf{w}).$$

These relatively meager specifications for $\mathbf{P}_l$ suffice to characterize several properties of $\mathbf{L}(\mathbf{w})$.

PROPOSITION 1. *For each $l \in \mathscr{M}$, let $v_l := |\mathscr{S}_l|$. Then $\mathbf{P}_l$ has v/v_l unit eigenvalues and $v - v/v_l$ zero eigenvalues.*

PROOF. Since $l \notin \mathscr{G}_l$, $\mathbf{P}_l$ has v/v_l linearly independent rows and, therefore, has $v - v/v_l$ zero eigenvalues and v/v_l nonzero eigenvalues. Also, $\mathbf{P}_l^2 = \mathbf{P}_l$ reveals $\mathbf{P}_l$ and $\boldsymbol{\Delta}_l := \mathbf{D}^{1/2}\mathbf{P}_l\mathbf{D}^{-1/2}$ to be idempotent matrices. The reversibility of $\mathbf{P}_l$ implies that $\boldsymbol{\Delta}_l$ is symmetric and that $\mathbf{P}_l$ and $\boldsymbol{\Delta}_l$ have the same spectrum. Since a symmetric idempotent matrix has all its eigenvalues in $\{0, 1\}$ [e.g., Horn and Johnson (1985), page 37], $\mathbf{P}_l$ has exactly v/v_l unit eigenvalues. □

PROPOSITION 2. *The matrix $\mathbf{L}(\mathbf{w})$ is aperiodic and irreducible.*

PROOF. Since the neighborhood condition implies that for each $\mathbf{x}_{-l} \in \mathscr{S}\backslash\mathscr{S}_l$ and for all $x, y \in S_l$, $p_{u(\mathbf{x}_l(y))u(\mathbf{x}_l(x))l} = p_{u(\mathbf{x}_l(x))u(\mathbf{x}_l(x))l}$, there must be some state i such that $p_{iil} > 0$. Therefore, $\mathbf{L}(\mathbf{w})$ is aperiodic. For each $\mathbf{l} := (l_1, \dots, l_m) \in \mathscr{M}^m$, let $\mathbf{Q}(\mathbf{l}) := \mathbf{P}_{l_1} \times \mathbf{P}_{l_2} \times \cdots \times \mathbf{P}_{l_m}$, so that

$$\mathbf{L}^m(\mathbf{w}) = m^{-m} \sum_{\mathbf{l}\in\mathscr{M}^m} \left(\prod_{i=1}^m w_{l_i}\right)\mathbf{Q}(\mathbf{l}).$$

Since there exists at least one $\mathbf{Q}(\mathbf{l}) > \mathbf{0}$, $\mathbf{L}^m(\mathbf{w}) > \mathbf{0}$ and, therefore, $\mathbf{L}(\mathbf{w})$ is irreducible. □

A like result appears in Liu, Wong and Kong (1995).

With regard to eigenvalues, the positive semidefiniteness of each $\mathbf{P}_l$ implies that $\mathbf{L}(\mathbf{w})$ is positive semidefinite. Also, the reversibility of each $\mathbf{P}_l$ implies the reversibility of $\mathbf{L}(\mathbf{w})$. Therefore, $\mathbf{L}(\mathbf{w})$ has all real nonnegative eigenvalues. Let $\lambda_0(\mathbf{A}) \geq \lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_{v-1}(\mathbf{A})$ denote the eigenvalues of a symmetric $v \times v$ matrix $\mathbf{A}$. Since $\lambda_i(\mathbf{L}(\mathbf{w})) = \lambda_i(\sum_{l=1}^m w_l \Delta_l)$, $\mathbf{L}(\mathbf{w})$ has the spectrum $\lambda_0(\mathbf{L}(\mathbf{w})) = 1 > \lambda_1(\mathbf{L}(\mathbf{w})) \geq \cdots \geq \lambda_{v-1}(\mathbf{L}(\mathbf{w})) \geq 0$.

THEOREM 3. *For the matrix* $\mathbf{L}(\mathbf{w})$,

$$\lambda_1(\mathbf{L}(\mathbf{w})) \geq \max(w_1, \dots, w_m; 1 - w_1, \dots, 1 - w_m).$$

PROOF. From Theorem 0, $\lambda_1(\mathbf{L}(\mathbf{w})) \geq \lambda_1(\sum_{l=2}^m w_l \mathbf{P}_l) + \lambda_{v-1}(w_1 \mathbf{P}_1)$. Since $\lambda_{v-1}(w_1 \mathbf{P}_1) = 0$ by Proposition 1, $\lambda_1(\mathbf{L}(\mathbf{w})) \geq \lambda_1(\sum_{l=2}^m w_l \mathbf{P}_l)$. From Theorem 0, $\lambda_1(\sum_{l=2}^m w_l \mathbf{P}_l) \geq \lambda_1(\sum_{l=3}^m w_l \mathbf{P}_l) + \lambda_{v-1}(w_2 \mathbf{P}_2)$. Since $\lambda_{v-1}(w_2 \mathbf{P}_2) = 0$, it follows that $\lambda_1(\mathbf{L}(\mathbf{w})) \geq \lambda_1(\sum_{l=3}^m w_l \mathbf{P}_l)$. Iteratively applying Theorem 0 $m - 3$ more times leads to $\lambda_1(\mathbf{L}(\mathbf{w})) \geq \lambda_1(w_m \mathbf{P}_m) + \lambda_{v-1}(w_{m-1} \mathbf{P}_{m-1}) = \lambda_1(w_m \mathbf{P}_m)$. Because $\lambda_1(\mathbf{P}_l) = 1$ for all $l \in \mathscr{M}$, $\lambda(\mathbf{L}(\mathbf{w})) \geq w_m$. Since a like result must hold for each permutation $\mathbf{l}$ of the coordinates $1, \dots, m$, $\lambda_1(\mathbf{L}(\mathbf{w})) \geq \max(w_1, \dots, w_m)$.

Because $\sum_{l=2}^m w_l \mathbf{P}_l / \sum_{l=2}^m w_l$ is a reducible stochastic matrix with v/v_1 closed sets, $\lambda_1(\sum_{l=2}^m w_l \mathbf{P}_l / \sum_{l=2}^m w_l) = 1$. Therefore,

$$\sum_{l=2}^m w_l = 1 - w_1 \leq \lambda_1\left(\sum_{l=2}^m w_l \mathbf{P}_l\right) + \lambda_{v-1}(w_1 \mathbf{P}_1) \leq \lambda_1(\mathbf{L}(\mathbf{w})).$$

Again, a like inequality holds for each coordinate, so that

$$\lambda_1(\mathbf{L}(\mathbf{w})) \geq \max(1 - w_1, \dots, 1 - w_m),$$

which completes the proof. □

As an immediate consequence of Theorem 3, the distribution $w_1 = \cdots = w_m = 1/m$ leads to the smallest lower bound $\lambda_1(\mathbf{L}(\mathbf{w})) \geq 1 - 1/m$.

To derive a lower bound for each eigenvalue of $\mathbf{L}(\mathbf{w})$, Theorem 4 relies on Theorem 0 and knowledge of the number of closed sets in a convex linear combination of the $\mathbf{P}_l$. To derive upper bounds, Theorem 5 relies on the same theorem and knowledge of upper bounds on the ranks of these convex combinations. As a consequence, Corollary 6 reveals that the smallest lower and upper bounds obtain for $w_1 = \cdots = w_m = 1/m$.

THEOREM 4. *Let* $\mathscr{L}$ *denote the set of all permutations of the integers* $1, \dots, m$ *and for each* $\mathbf{l} \in \mathscr{L}$, *let* $s_0(\mathbf{l}) := 0$ *and* $s_j(\mathbf{l}) := \sum_{i=1}^j w_{l_i}$ *for* $1 \leq j \leq m$. *Assume* $v_1 \leq v_2 \leq \cdots \leq v_m$ *and let* $u_0 := 1$ *and* $u_j := u_{j-1} v_j$ *for* $1 \leq j \leq m$. *Then*

$$\lambda_l(\mathbf{L}(\mathbf{w})) \geq \max_{\mathbf{l} \in \mathscr{L}} [1 - s_j(\mathbf{l})], \qquad 1 \leq l \leq u_j - 1,\ 1 \leq j \leq m. \tag{1}$$

The Appendix contains the proof.

THEOREM 5. *Assume* $v_1 \le v_2 \le \cdots \le v_m$. *Then*

$$\lambda_l(\mathbf{L}(\mathbf{w})) \le \begin{cases} \max_{\mathbf{l}\in\mathscr{L}}[1 - s_{j-1}(\mathbf{l})], & l \ge v \sum_{i=m-j+2}^{m} (1/v_i),\ 2 \le j \le m, \\ 1, & l \ge 0. \end{cases} \tag{2}$$

The Appendix contains the proof. Note that $\lambda_l(\mathbf{L}(\mathbf{w})) := 0$ for $l \ge v$.

COROLLARY 6. *The assignment* $w_i = 1/m$, $1 \le i \le m$, *minimizes* $\max_{\mathbf{l}\in\mathscr{L}}[1 - s_j(\mathbf{l})]$ *for each* $1 \le j \le m$.

The proof is immediate.

While the assignment of Corollary 6 does not imply minimization of each $\lambda_l(\mathbf{L}(\mathbf{w}))$, it does suggest a reasonable choice when no additional information is available. Accordingly, we hereafter adopt it exclusively and write $\mathbf{L} := m^{-1} \sum_{l=1}^{m} \mathbf{P}_l$ and $\beta_l := \lambda_l(\mathbf{L})$ for $0 \le l \le v-1$. This case offers the additional advantage of the tighter lower bounds $\lambda_l(\mathbf{L}) \ge 1 - j/m$ for $l \le \prod_{i=m-j+1}^{m} v_i$ and $1 \le j \le m$.

3. Figures of merit. The relative desirabilities of alternative sampling plans depend on how well they perform with regard to convergence from an arbitrarily selected initial state to the steady state and with regard to the variation they induce. For each Plan i and all $x, y \in \mathscr{V}$, let

$$q_{ixy}^{(k)} := \text{probability of moving from state } x \text{ to state } y \text{ in } k \text{ steps} \tag{3}$$

and let $\mathbf{Q}_{ik} := \| q_{ixy}^{(k)} \|_{x,\,y=0}^{v-1}$, which we call the k-step transition matrix for Plan i. If Plans i and j each have irreducible and aperiodic transition matrices, then $\lim_{k\to\infty} q_{ixy}^{(k)} = \pi_y$, $\lim_{k\to\infty} q_{jxy}^{(k)} = \pi_y$, $\forall\ x, y \in \mathscr{V}$, and

$$W_{ijxy}^{(k)} := \left| \frac{q_{ixy}^{(k)} - \pi_y}{q_{jxy}^{(k)} - \pi_y} \right| \tag{4}$$

measures their relative speeds of convergence. If for all $x, y \in \mathscr{V}$, $W_{ijxy}^{(k)}$ converges to zero as $k \to \infty$, then Plan i has greater appeal than Plan j has according to this criterion. If for all $x, y \in \mathscr{V}$, $W_{ijxy}^{(k)}$ diverges as $k \to \infty$, then Plan j has the greater appeal.

With regard to variation, let $\{g_i,\ 0 \le i \le v-1\}$ denote a bounded function, let

$$\mu := \sum_{i=0}^{v-1} g_i \pi_i$$

and assume μ is unknown and to be estimated. Let $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_n$ denote $n+1$ successive states generated by Plan i and let $X_j := g_{u(\mathbf{S}_j)}$. Consider

$$\overline{\mu}_{in} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

as an estimator of μ. For irreducible, aperiodic transition matrices,

$$\lim_{n\to\infty}[\mathrm{E}(\overline{\mu}_{in}|\mathbf{S}_0 = \mathbf{x}) - \mu] = 0$$

and

$$\lim_{n\to\infty} n \operatorname{var}(\overline{\mu}_{in}|\mathbf{S}_0 = \mathbf{x}) = T_i := \lim_{n\to\infty} n \operatorname{var}\overline{\mu}_{in},$$

where $\operatorname{var}\overline{\mu}_{in}$ denotes the variance that obtains when starting in the steady state. Comparing T_i and T_j for Plans i and j, respectively, provides a basis for determining which induces a smaller asymptotic variance. In particular, $T_i < T_j$ favors Plan i.

Let $\mathbf{f}_l := (f_{0l}, f_{1l}, \ldots, f_{v-1,l})^T$ denote the left eigenvector of $\mathbf{L}$ corresponding to β_l with $\mathbf{f}_l^T\mathbf{f}_l = 1$ and $\mathbf{f}_j^T\mathbf{f}_l = 0,\ j \neq l$. Then

$$q_{1xy}^{(k)} = \pi_y + \sqrt{\pi_y/\pi_x}\sum_{l=1}^{v-1} f_{xl} f_{yl}\beta_l^k, \qquad k \geq 0, \tag{5}$$

and for $c_l := \sum_{x=0}^{v-1} g_x\sqrt{\pi_x}\, f_{xl},\ 1 \leq l \leq v-1$ [e.g., Kemeny and Snell (1960)],

$$T_1 = \sum_{l=1}^{v-1} c_l^2(1+\beta_l)/(1-\beta_l), \tag{6}$$

where $\operatorname{var} X_j = \sum_{l=1}^{v-1} c_l^2$. In terms of matrices, $T_1 = \mathbf{g}^T\mathbf{D}(\mathbf{I}+\mathbf{L}-2\mathbf{\Sigma})(\mathbf{I}-\mathbf{L}+\mathbf{\Sigma})^{-1}\mathbf{g}$, where $\mathbf{g} := (g_0, g_1, \ldots, g_{v-1})^T$ and $\mathbf{\Sigma}$ denotes a $v\times v$ matrix with $(\pi_0, \pi_1, \ldots, \pi_{v-1})$ in each row. Hereafter, we assume $\operatorname{var} X_j = 1$. In what follows, expressions (5) and (6) establish the baselines against which we compare the merits of other sampling plans.

As a consequence of Theorems 4 and 5 and Corollary 6, for $m > 2$,

$$\begin{aligned}&\sum_{j=1}^{m}\left(\sum_{l=u_{j-1}}^{u_j-1} c_l^2\right)\left(\frac{2m-j}{j}\right)\\ &\qquad\leq T_1 \leq \frac{1+\beta_1}{1-\beta_1}\sum_{l=1}^{t_m-t_{m-1}-1} c_l^2 + \sum_{j=2}^{m-1}\left(\sum_{l=t_m-t_{m-j+1}}^{t_m-t_{m-j}-1} c_l^2\right)\left(\frac{2m-j}{j}\right) + m\sum_{l=t_m-t_1}^{v-1} c_l^2,\end{aligned}$$

where $t_1, \ldots, t_m$ are defined in the Appendix and $c_l := 0$ for all $l \geq v$. Two examples illustrate the significance of these bounds. If $c_l = 0$ for all $v_1 \leq l \leq v-1$, then

$$2m-1 \leq T_1 \leq (1+\beta_1)/(1-\beta_1).$$

Recall that $\beta_1 \geq 1 - 1/m$. Alternatively, if $v_1 = \cdots = v_m$ and $c_1^2 = \cdots = c_{v-1}^2 = 1/(v-1)$, then

$$\sum_{l=u_{j-1}}^{u_j - 1} c_l^2 = \frac{v_1^{j-1}(v_1 - 1)}{v_1^m - 1}$$

and

$$\sum_{l=t_m - t_{m-j+1}}^{t_m - t_{m-i} - 1} c_l^2 \leq \frac{(j-i)v_1^{m-1}}{(v_1^m - 1)}, \qquad 1 \leq i \leq j \leq m,$$

where equality holds if and only if $m \leq v_1$. More importantly,

$$\begin{aligned} 2m\left(1 - \frac{1}{v_1}\right) &\leq \frac{2m(v_1 - 1)}{1 - v_1^{-m}} \sum_{j=1}^{m} \frac{1}{jv_1^{m-j+1}} \\ &\leq T_1 \leq \frac{1}{v_1(1 - v_1^{-m})}\left[\frac{1+\beta_1}{1-\beta_1} + O(m \ln m)\right]. \end{aligned}$$

Note that both examples have lower bounds that grow linearly with the number of coordinates m. However, the second example leads to an upper bound that grows, at least, as $m \ln m$.

Since $\mathbf{P}_i$ is idempotent, repeatedly selecting i, as may occur in Plan 1, contributes nothing to convergence. To reduce this limitation, consider the strategy in Plan 2 which, after updating coordinate I, randomly selects the next coordinate uniformly from $\mathscr{M} \setminus \{I\}$ [e.g., Amit and Grenander (1989)].

PLAN 2.
Randomly generate $\mathbf{S}_0$ from an initializing p.m.f. $\boldsymbol{\pi}_0$ on $\mathscr{S}$.
Randomly generate I from $\mathscr{M}$ with probability $1/m$.
Randomly generate S_{I1} from $\mathscr{S}_I$ using row $u(\mathbf{S}_0)$ of $\mathbf{P}_I$.
$S_{i1} \leftarrow S_{i0}, \ \forall\, i \neq I$.
$j \leftarrow 1$ and $J \leftarrow I$.
On each step $j \geq 2$:
  Randomly generate I from $\mathscr{M} \setminus \{J\}$ with probability $1/(m-1)$.
  Randomly generate S_{Ij} from $\mathscr{S}_I$ using row $u(\mathbf{S}_{j-1})$ of $\mathbf{P}_I$.
  $S_{ij} \leftarrow S_{i,j-1}, \ \forall\, i \neq I$.
  $J \leftarrow I$.
  $j \leftarrow j + 1$.

Proposition 7 gives an immediate consequence of this plan.

PROPOSITION 7. *Plan 2 has the k-step expected transition matrix*

$$\mathbf{Q}_{2k} = \mathbf{L}\left(\frac{m\mathbf{L} - \mathbf{I}}{m-1}\right)^{k-1}, \qquad k \geq 1. \tag{7}$$

PROOF. Let $I_1, \ldots, I_k$ denote the randomly selected coordinates for sampling on steps 1 through k, let $\mathbf{V}_k := \mathbf{P}_{I_1} \times \cdots \times \mathbf{P}_{I_k}$ and observe that $\mathbf{Q}_{2k} = \mathrm{E}_{I_1,\ldots,I_k}\mathbf{V}_k = \mathrm{E}_{I_1}\mathrm{E}_{I_2,\ldots,I_k|I_1}\mathbf{V}_k$. Then

$$\begin{aligned}
\mathbf{Q}_{2k} &= \mathrm{E}_{I_1}\mathrm{E}_{I_2,\ldots,I_k|I_1}\mathbf{V}_k \\
&= \mathrm{E}_{I_1}\mathrm{E}_{I_2,\ldots,I_{k-1}|I_1}\mathbf{V}_{k-1}\mathrm{E}_{I_k|I_{k-1}}\mathbf{P}_{I_k} \\
&= \mathrm{E}_{I_1}\mathrm{E}_{I_2,\ldots,I_{k-1}|I_1}\mathbf{V}_{k-1}\left(\frac{m\mathbf{L}-\mathbf{P}_{I_{k-1}}}{m-1}\right) \\
&= \mathrm{E}_{I_1}\mathrm{E}_{I_2,\ldots,I_{k-2}|I_1}\mathbf{V}_{k-2}\left(\frac{m\mathbf{L}-\mathbf{P}_{I_{k-2}}}{m-1}\right)\left(\frac{m\mathbf{L}-\mathbf{I}}{m-1}\right) \\
&\ \ \vdots \\
&= \mathrm{E}_{I_1}\mathbf{P}_{I_1}\left(\frac{m\mathbf{L}-\mathbf{I}}{m-1}\right)^{k-1} \\
&= \mathbf{L}\left(\frac{m\mathbf{L}-\mathbf{I}}{m-1}\right)^{k-1}. \qquad \square
\end{aligned}$$

Theorem 8 compares the performances of Plans 1 and 2.

THEOREM 8. *For $m > 2$, $W_{12xy}^{(k)} \to \infty$ as $k \to \infty$. For $m = 2$, a necessary and sufficient condition for Plan 2 to converge more rapidly is $1 - \beta_1 < 2\beta_{\min}$, where $\beta_{\min}$ denotes the smallest positive eigenvalue of $\mathbf{L}$. Also,*

$$\begin{aligned}
T_2 &= \mathbf{g}^T\mathbf{D}\left[\mathbf{I}+\mathbf{L}-2\boldsymbol{\Sigma}-\frac{2}{m}(\mathbf{L}-\boldsymbol{\Sigma})\right](\mathbf{I}-\mathbf{L}+\boldsymbol{\Sigma})^{-1}\mathbf{g} \\
&= \sum_{l=1}^{v-1} c_l^2\left(\frac{1+\beta_l-2\beta_l/m}{1-\beta_l}\right)
\end{aligned}$$

and

$$1-\frac{1}{m} \le \frac{T_2}{T_1} = 1-\frac{2}{m}\frac{\sum_{i=1}^{v-1} c_i^2(\beta_i/(1-\beta_i))}{\sum_{i=1}^{v-1} c_i^2(1+\beta_i/(1-\beta_i))} \le 1.$$

PROOF. From (5) and Proposition 7,

$$q_{2xy}^{(k)} = \pi_y + \sqrt{\frac{\pi_y}{\pi_x}}\sum_{l=1}^{v-1} f_{xl}f_{yl}\beta_l\left(\frac{m\beta_l-1}{m-1}\right)^{k-1}, \qquad k \ge 1,$$

so that

$$W_{12xy}^{(k)} = \left|\frac{\sum_{l=1}^{v-1} f_{xl}f_{yl}\beta_l^k}{\sum_{l=1}^{v-1} f_{xl}f_{yl}\beta_l((m\beta_l-1)/(m-1))^{k-1}}\right|, \qquad k \ge 1.$$

Therefore, a necessary and sufficient condition for $W_{12xy}^{(k)} \to \infty$ as $k \to \infty$ is that for all positive β_l,

$$\frac{1-(m-1)\beta_1}{m} < \beta_l < \frac{1+(m-1)\beta_1}{m}.$$

Since $\beta_l \le \beta_1$, $\beta_1 \ge (m-1)/m$ and $(1-(m-1)\beta_1)/m \le -(m^2-3m+1)/m^2 < 0$ for all $m > 2$, $W_{12xy}^{(k)} \to \infty$ as $k \to \infty$ for all $m > 2$. Moreover, $(1-\beta_1)/2 < \beta_l$ is necessary and sufficient for $m = 2$.

The quantity $\bar{\mu}_{2n}$ has asymptotic variance

$$T_2 = \lim_{n\to\infty} n \operatorname{var} \bar{\mu}_{2n} = \mathbf{g}^T \left[\mathbf{I} - \Sigma + 2 \sum_{k=1}^{\infty} (\mathbf{Q}_{2k} - \mathbf{\Sigma}) \right] \mathbf{g}$$

and since

$$\mathbf{Q}_{2k} - \mathbf{\Sigma} = (\mathbf{L} - \mathbf{\Sigma}) \left(\frac{m\mathbf{L} - \mathbf{I} - m\mathbf{\Sigma}}{m-1} \right)^{k-1} \quad \text{for } k \ge 1,$$

then

$$T_2 = \mathbf{g}^T \mathbf{D} \left[\mathbf{I} + \mathbf{L} - 2\mathbf{\Sigma} - \frac{2}{m} (\mathbf{L} - \mathbf{\Sigma}) \right] (\mathbf{I} - \mathbf{L} + \mathbf{\Sigma})^{-1} \mathbf{g}.$$

The lower bound follows from the observation that

$$1 - \frac{2}{m} \frac{\sum_{i=1}^{v-1} c_i^2 (\beta_i/(1-\beta_i))}{\sum_{i=1}^{v-1} 2c_i^2 (\beta_i/(1-\beta_i))} \le \frac{T_2}{T_1}. \qquad \square$$

The result for $m = 2$ calls for a bit more study. Let $h(\beta, k, m) := \beta(1-m\beta)^{k-1}$ for $0 \le \beta \le 1/m$ and $k, m \ge 2$. It is easily seen that

$$h(\beta, k, m) \le \frac{1}{(k-1)m} \left(1 - \frac{1}{k}\right)^k \le \frac{1}{(k-1)m\mathrm{e}}. \tag{8}$$

This property is reassuring for Plan 2 with $m = 2$. If $2\beta_{\min} - 1 < -\beta_1$, then $\beta_{\min}|2\beta_{\min} - 1|^{k-1}$ controls the convergence rate for $q_{2xy}^{(k)}$. However, (8) implies

$$\beta_1^k < \beta_{\min}|2\beta_{\min} - 1|^{k-1} \le \frac{1}{2\mathrm{e}(k-1)} \le \frac{0.1840}{k-1}$$

so that, while Plan 1 offers faster convergence, the convergence rate for Plan 2 is also rapid. $\square$

4. Deterministic coordinate selection. Let $\mathbf{l} = (l_1, \ldots, l_m)$ denote a permutation of the integers $\{1, \ldots, m\}$. Then iteratively updating coordinates in the order $l_1, \ldots, l_m$ induces the km-step transition matrix $\mathbf{Q}^k(\mathbf{l})$, where $\mathbf{Q}(\mathbf{l})$ is not reversible. At least two options exist for recovering reversibility. Plan 3 describes the first [e.g., Johnson (1989)]. On each iteration it updates coordinates in the order $l_1, \ldots, l_{m-1}, l_m, l_{m-1}, \ldots, l_1$.

PLAN 3.
Given: $\mathbf{l} = (l_1, \dots, l_m)$.
$i \leftarrow 1$, $j \leftarrow 1$ and $t \leftarrow 1$.
While $i < m$: $l_{m+i} \leftarrow l_{m-i}$ and $i \leftarrow i + 1$.
Randomly select $\mathbf{S}_0$ from the initializing p.m.f. $\boldsymbol{\pi}_0$ on $\mathscr{S}$.
On each step $j \geq 1$:
Sample $\mathbf{S}_j$ from row $u(\mathbf{S}_{j-1})$ of $\mathbf{P}_{l_t}$.
$S_{ij} \leftarrow S_{i,j-1}\ \forall\ i \neq l_t$.
$t \leftarrow t \pmod{2m-1} + 1$.

The idempotency of $\mathbf{P}_m$ implies that $\mathbf{P}_{l_1} \times \cdots \times \mathbf{P}_{l_{m-1}} \times \mathbf{P}_{l_m} \times \mathbf{P}_{l_{m-1}} \times \cdots \times \mathbf{P}_{l_1} = \mathbf{R}(\mathbf{l})$, where $\mathbf{R}(\mathbf{l}) := \mathbf{Q}(l_1, \dots, l_m)\mathbf{Q}(l_m, \dots, l_1)$. The corresponding $k(2m-1)$-step transition matrix is $\mathbf{Q}_{3,\,k(2m-1)}(\mathbf{l}) = \mathbf{R}^k(\mathbf{l})$. Since $\mathbf{DR}(\mathbf{l}) = \mathbf{R}^T(\mathbf{l})\mathbf{D}$, $\mathbf{R}(\mathbf{l})$ is reversible and since $[\mathbf{D}^{1/2}\mathbf{Q}(l_1, \dots, l_m)\mathbf{D}^{-1/2}]^T = \mathbf{D}^{1/2}\mathbf{Q}(l_m, \dots, l_1)\mathbf{D}^{-1/2}$, $\mathbf{D}^{1/2}\mathbf{R}(\mathbf{l})\mathbf{D}^{-1/2}$ is symmetric positive semidefinite. Therefore, $\mathbf{Q}_{3,\,k(2m-1)}(\mathbf{l})$ has spectrum

$$\lambda_i(\mathbf{Q}_{3,\,k(2m-1)}(\mathbf{l})) = \lambda_i^k(\mathbf{R}(\mathbf{l})) = \lambda_i^k(\mathbf{D}^{1/2}\mathbf{R}(\mathbf{l})\mathbf{D}^{-1/2}) \in [0, 1], \quad 0 \leq i \leq v-1. \tag{9}$$

Plan 4 describes a second option for inducing reversibility. At the beginning of the sampling experiment, one randomly chooses the coordinate permutation $(l_1, \dots, l_m)$ or the reverse permutation $(l_m, \dots, l_1)$ with equal probabilities. Thereafter, on each iteration it updates all coordinates in that order.

PLAN 4.
Given: $\mathbf{l} = (l_1, \dots, l_m)$.
$j \leftarrow 1$ and $t \leftarrow 1$.
Sample J from $\{0, 1\}$ with probabilities $\{1/2, 1/2\}$.
If $J = 0$: while $t \leq m$, $i \leftarrow l_t$ and $t \leftarrow t + 1$.
Otherwise: while $t \leq m$, $i_t \leftarrow l_{m-t+1}$ and $t \leftarrow t + 1$.
Randomly select $\mathbf{S}_0$ from the initializing p.m.f. $\boldsymbol{\pi}_0$ on $\mathscr{S}$.
$t \leftarrow 1$.
For each step $j \geq 1$:
Sample $\mathbf{S}_j$ from row $u(\mathbf{S}_{j-1})$ of $\mathbf{P}_{i_t}$.
$S_{ij} \leftarrow S_{i,j-1}\ \forall\ i \neq i_t$.
$t \leftarrow t \pmod{m} + 1$.

Plan 4 has km-step transition matrix $\mathbf{Q}_{4,\,km}(\mathbf{l}) = \mathbf{N}_k(\mathbf{l})$, where $\mathbf{N}_k(\mathbf{l}) := \frac{1}{2}[\mathbf{Q}^k(l_1, \dots, l_m) + \mathbf{Q}^k(l_m, \dots, l_1)]$ is clearly reversible. Moreover, Theorem 9 gives a motivation for preferring Plan 4 over Plan 3.

THEOREM 9. *For Plans* 3 *and* 4 *and* $k \geq 1$,

$$|\lambda_i(\mathbf{N}_{2k}(\mathbf{l}))| \leq \lambda_i^k(\mathbf{R}(\mathbf{l})), \qquad 0 \leq i \leq v-1. \tag{10}$$

PROOF. For a $v \times v$ real matrix $\mathbf{A}$, it is known that [Fan and Hoffman (1955), Marshall and Olkin (1979), Theorem 9.F.4]

$$\lambda_i^2((\mathbf{A}+\mathbf{A}^T)/2) \le \lambda_i(\mathbf{A}\mathbf{A}^T), \qquad 0 \le i \le v-1,$$

and that for all integers $l \ge 1$ [Fan (1949), Marshall and Olkin (1979), Theorem 9.E.4],

$$|\lambda_i(\mathbf{A}^l(\mathbf{A}^T)^l)| \le \lambda_i^l(\mathbf{A}\mathbf{A}^T).$$

Therefore,

$$\begin{aligned}\lambda_i^2((\mathbf{A}^l+(\mathbf{A}^T)^l)/2) &\le \lambda_i(\mathbf{A}^l(\mathbf{A}^l)^T)\\ &= \lambda_i(\mathbf{A}^l(\mathbf{A}^T)^l)\\ &\le \lambda_i^l(\mathbf{A}\mathbf{A}^T).\end{aligned}$$

The theorem is proved for $\mathbf{A} := \mathbf{D}^{1/2}\mathbf{Q}(\mathbf{l})\mathbf{D}^{-1/2}$ and $l = 2k$. $\square$

Theorem 9 reveals that for $2k$ iterations each of m updates, Plan 4 induces convergence at a rate at least as fast as k iterations of $2m-1$ updates each of Plan 3. Plan 3 induces at least one additional property that deserves attention. Since $\mathbf{P}_{l_1}^2 = \mathbf{P}_{l_1}$, updating coordinate l_1 at the end of iteration $j-1$ contributes no benefit to convergence. Accordingly, one may choose to update $2m-2$ coordinates in the order $l_1, \dots, l_{m-1}, l_m, l_{m-1}, \dots, l_2$ on each of $(k-1)$ iterations and update coordinates $l_1, \dots, l_{m-1}, l_m, l_{m-1}, \dots, l_1$ on the last iteration for a total of $k(2m-2)$ updates. This compares to $2km$ coordinate updates under Plan 3. Since k iterations with Plan 3 have convergence rate $\lambda_1^k(\mathbf{R}(\mathbf{l}))$, whereas $2km$ steps with Plan 1 have convergence rate β_1^{2km}, it is clear that Plan 3 converges more rapidly if and only if $\beta_1^m > \lambda_1^{1/2}(\mathbf{R}(\mathbf{l}))$.

5. Partitioning the coordinates. Although Section 4 demonstrates the advantage of Plan 4 for convergence, it tells us nothing about how a particular coordinate permutation $\mathbf{l}$ affects convergence. To address this issue, we focus on a special representation that occurs frequently in practice and that exploits the neighborhood concept introduced in Section 1. Let $\mathscr{M}_1$ and $\mathscr{M}_2$ denote mutually exclusive and exhaustive subsets of $\mathscr{M}$ and assume they satisfy $\mathscr{M}_1 = \bigcup_{i\in\mathscr{M}_2}\mathscr{G}_i$ and $\mathscr{M}_2 = \bigcup_{i\in\mathscr{M}_1}\mathscr{G}_i$. Suppose that for all $\mathbf{x} \in \mathscr{S}$,

$$\pi_i(x_i|x_l,\ i \ne l) = \begin{cases}\pi_i(x_i|x_l,\ l \in \mathscr{G}_i \subseteq \mathscr{M}_2), & i \in \mathscr{M}_1,\\ \pi_i(x_i|x_l,\ l \in \mathscr{G}_i \subseteq \mathscr{M}_1), & i \in \mathscr{M}_2.\end{cases}$$

That is, given $\mathbf{S}_{j-1}$, the random variables $(S_{ij},\ i \in \mathscr{M}_1)$ are independent of $(S_{i,j-1},\ i \in \mathscr{M}_1)$, and $(S_{ij},\ i \in \mathscr{M}_2)$ are independent of $(S_{i,j-1},\ i \in \mathscr{M}_2)$. As an immediate consequence, $\{\mathbf{P}_i,\ i \in \mathscr{M}_1\}$ forms a commuting family of idempotent matrices, and $\{\mathbf{P}_i,\ i \in \mathscr{M}_2\}$ does likewise. As an illustration, assume m is even and let $m_1 = m_2 = m/2$, $\mathscr{M}_1 := \{1, 3, \dots, m-3, m-1\}$ and

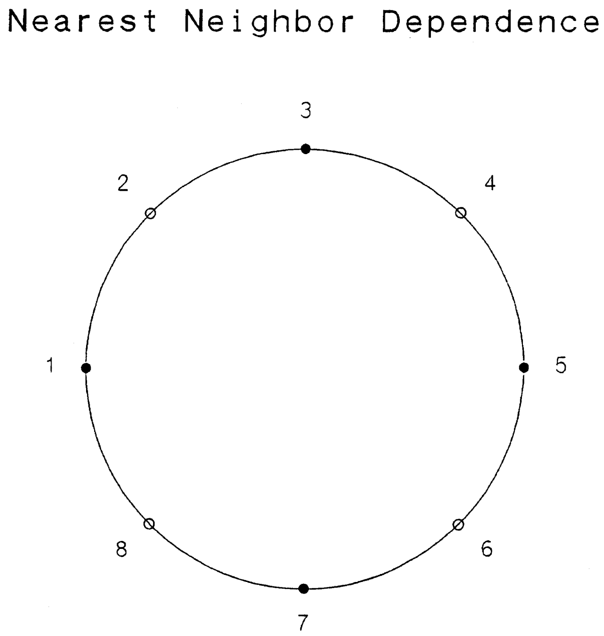

FIG. 1.

$\mathscr{M}_2 := \{2, 4, \ldots, m-2, m\}$. Conceptually, one can think of the sites as m consecutively numbered points on a circle (Figure 1), where for a *nearest-neighbor* specification point i has the neighborhood

$$\mathscr{G}_i = \begin{cases} \{2, m\}, & \text{if } i = 1, \\ \{i-1, i+1\}, & \text{if } i = 2, \ldots, m-1, \\ \{1, m-1\}, & \text{if } i = m. \end{cases}$$

This odd/even representation also fits a commonly encountered two-dimensional site model. Consider a square $d \times d$ lattice as in Figure 2 with sites labeled $\mathbf{t} = (t_1, t_2),\ t_1, t_2 \in \{0, 1, \ldots, d-1\}$. Let site $\mathbf{t}$ have the nearest-neighbor neighborhood

$$\mathscr{G}_{\mathbf{t}} = \{(t_1 - 1, t_2), (t_1 + 1, t_2), (t_1, t_2 - 1), (t_1, t_2 + 1)\},$$

where each coordinate argument is taken modulo d. Then $\mathscr{M}$ denotes the set of all $m = d^2$ sites and there exist two disjoint and exhaustive subsets $\mathscr{M}_1$ and $\mathscr{M}_2$ of $\mathscr{M}$ such that for every site $\mathbf{t} \in \mathscr{M}_1$, $\mathscr{G}_{\mathbf{t}} \subseteq \mathscr{M}_2$, and for every site $\mathbf{t} \in \mathscr{M}_2$, $\mathscr{G}_{\mathbf{t}} \subseteq \mathscr{M}_1$. In the case of Figure 2, these subsets are

$$\mathscr{M}_1 = \{(0,0), (0,2), (1,1), (1,3), (2,0), (2,2), (3,1), (3,3)\}$$

and

$$\mathscr{M}_2 = \{(0,1), (0,3), (1,0), (1,2), (2,1), (2,3), (3,0), (3,2)\}.$$

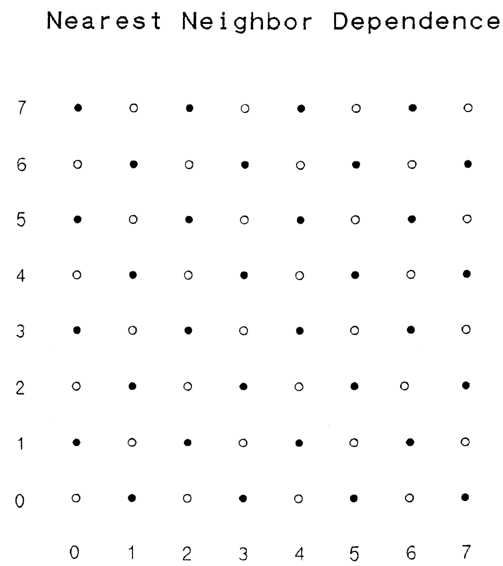

FIG. 2.

For an example of a partition with one coordinate in $\mathscr{M}_1$ and $m-1$ in $\mathscr{M}_2$, see the Gamma posterior distribution illustration in Gelfand and Smith (1990).

Let $m_1 = |\mathscr{M}_1|$ and $m_2 = |\mathscr{M}_2|$. Plan 5 offers a special, but important, case of Plan 4, where the coordinate permutation $\mathbf{l} = (l_1, \ldots, l_m) \in \mathscr{L}$ is chosen so that $\mathscr{M}_1 = \{l_1, \ldots, l_{m_1}\}$ and $\mathscr{M}_2 = \{l_{m_1+1}, \ldots, l_m\}$. If $J = 0$, the plan first updates all coordinates in $\mathscr{M}_1$ in the order $l_1, \ldots, l_{m_1}$ and then all coordinates in $\mathscr{M}_2$ in the order $l_{m_1+1}, \ldots, l_m$ on each iteration. If $J = 1$, it first updates all coordinates in $\mathscr{M}_2$ in the order $l_{m_1+1}, \ldots, l_m$ and then all coordinates in $\mathscr{M}_1$ in the order $l_1, \ldots, l_m$. Here $(l_1, \ldots, l_{m_1})$ can be any one of the $m_1!$ permutations of the elements of $\mathscr{M}_1$ and $(l_{m_1+1}, \ldots, l_m)$ can be any one of the $m_2!$ permutations of $\mathscr{M}_2$. Note that a commonly encountered version of the Ising model relies on a nearest-neighbor concept that fits this formulation.

PLAN 5.
Randomly select $\mathbf{S}_0$ from the initializing p.m.f. $\boldsymbol{\pi}_0$ on $\mathscr{S}$.
Randomly select J from $\{0, 1\}$ with probabilities $\{1/2, 1/2\}$.
$j \leftarrow 1$ and $k \leftarrow 1$.
On each iteration $k \geq 1$:
  For $t = 1$ to m:
    $l \leftarrow l_{(Jm_1+t-1)(\mathrm{mod}\, m)+1}$.
    Randomly generate S_{lj} from $\mathscr{S}_l$ using row $u(\mathbf{S}_{j-1})$ of $\mathbf{P}_l$.
    $S_{ij} \leftarrow S_{i,j-1} \; \forall \; i \neq l$.
    $j \leftarrow j+1$.
  $k \leftarrow k+1$.

THEOREM 10. *The matrix* $\mathbf{M}_i := \Pi_{l\in\mathscr{M}_i}\mathbf{P}_l$ *is idempotent and reversible with* $v/\Pi_{l\in\mathscr{M}_i}v_l$ *unit eigenvalues and* $v - v/\Pi_{l\in\mathscr{M}_i}v_l$ *zero eigenvalues. Also* $\mathbf{M} := \frac{1}{2}(\mathbf{M}_1 + \mathbf{M}_2)$ *is reversible, nonnegative definite, irreducible and aperiodic.*

PROOF. Since the matrices $\{\mathbf{P}_l,\ l \in \mathscr{M}_i\}$ commute, there exists a nonsingular orthogonal matrix $\mathbf{G}_i$ that simultaneously diagonalizes them. Then $\mathbf{P}_l = \mathbf{G}_i\,\boldsymbol{\lambda}_l\mathbf{G}_i^T$, where $\boldsymbol{\lambda}_l$ is a $v \times v$ diagonal matrix of the eigenvalues of $\mathbf{P}_l$. Therefore, $\mathbf{M}_i = \mathbf{G}_i\Pi_{l\in\mathscr{M}_i}\boldsymbol{\lambda}_l\mathbf{G}_i^T$. Since each $\boldsymbol{\lambda}_l$ has only zero and unit entries on its main diagonal, $\mathbf{M}_i^2 = \mathbf{M}_i$ so that $\mathbf{M}_i$ is idempotent.

Since $\mathbf{M}_1 = \Pi_{i=1}^{m_1}\mathbf{P}_{l_i}$, $\mathbf{D}\mathbf{M}_1 = \mathbf{D}\mathbf{P}_{l_1}\Pi_{i=2}^{m_1}\mathbf{P}_l = \mathbf{P}_{l_1}^T\mathbf{D}\Pi_{i=2}^{m_1}\mathbf{P}_{l_i} = (\mathbf{P}_{l_1}^T \times \cdots \times \mathbf{P}_{l_{m_1}}^T)\mathbf{D} = \mathbf{M}_1^T\mathbf{D}$ so that $\mathbf{M}_1$ is reversible. An analogous result holds for $\mathbf{M}_2$. Since $\mathbf{M}_i$ has $v/\Pi_{l\in\mathscr{M}_i}v_l$ closed sets and is idempotent, it has $v/\Pi_{l\in\mathscr{M}_i}v_l$ units and $v - v/\Pi_{l\in\mathscr{M}_i}v_l$ eigenvalues. $\square$

Theorem 10 reveals that $\mathbf{M}_1$ and $\mathbf{M}_2$ in Plan 5 have analogous properties to those of $\mathbf{P}_1, \ldots, \mathbf{P}_m$ in Plans 1 and 2. In particular, idempotency implies that repeatedly updating coordinates in $\mathscr{M}_1$ before updating the coordinates in $\mathscr{M}_2$ contributes nothing to convergence. Theorems 11 and 12 develop additional properties of Plan 5 and Theorem 13 shows that its expected k-iteration transition matrix $\mathbf{Q}_{5,\,km}$ has a form analogous to the expected k-step transition matrix $\mathbf{Q}_{2k}$ in (7) for Plan 2.

THEOREM 11. *Let* $\mathbf{M} = \frac{1}{2}(\mathbf{M}_1 + \mathbf{M}_2)$. *Then* $\mathbf{M}$ *is a reversible, nonnegative definite, irreducible aperiodic matrix.*

PROOF. Nonnegative definiteness for $\mathbf{M}$ follows from the property that the sum of two nonnegative definite matrices is nonnegative definite. Since $\mathbf{M}_1$ and $\mathbf{M}_2$ are reversible, $\mathbf{M}$ is clearly reversible. Since $\mathbf{M}_1\mathbf{M}_2 > \mathbf{0}$, $\mathbf{M}^2 > \mathbf{0}$ so that $\mathbf{M}$ is irreducible. Since the diagonal elements of $\mathbf{M}_1$ and $\mathbf{M}_2$ are positive, $\mathbf{M}$ is aperiodic. $\square$

THEOREM 12. *Plan* 5 *has expected km-step transition matrix*

$$\mathbf{Q}_{5,\,km} = \tfrac{1}{2}[(\mathbf{M}_1\mathbf{M}_2)^k + (\mathbf{M}_2\mathbf{M}_1)^k] = \mathbf{M}(2\mathbf{M} - \mathbf{I})^{2k-1}, \qquad k \geq l. \tag{11}$$

PROOF. If $J = 0$ in Plan 5, then the km-step transition matrix is $(\mathbf{M}_1\mathbf{M}_2)^k$. If $J = 1$, it is $(\mathbf{M}_2\mathbf{M}_1)^k$. Therefore, $\mathbf{Q}_{5,\,km} = \frac{1}{2}[(\mathbf{M}_1\mathbf{M}_2)^k + (\mathbf{M}_2\mathbf{M}_1)^k]$. By idempotency $4\mathbf{M}^2 = \mathbf{M}_1\mathbf{M}_2 + \mathbf{M}_1\mathbf{M}_2 + \mathbf{M}_1 + \mathbf{M}_2$ so that $\frac{1}{2}(\mathbf{M}_1\mathbf{M}_2 + \mathbf{M}_2\mathbf{M}_1) = \mathbf{M}(2\mathbf{M} - \mathbf{I})$. By induction on k (≥ 2) it is easily seen that

$$(\mathbf{M}_1\mathbf{M}_2)^k + (\mathbf{M}_2\mathbf{M}_1)^k = [(\mathbf{M}_1\mathbf{M}_2)^{k-1} + (\mathbf{M}_2\mathbf{M}_1)^{k-1}](2\mathbf{M} - \mathbf{I})^2$$

so that $\mathbf{Q}_{5,\,km} = \mathbf{M}(2\mathbf{M} - \mathbf{I})^{2k-1}$ for $k \geq 1$. $\square$

As an immediate consequence of Theorems 11 and 12, $\mathbf{Q}_{5,\,km}$ has spectrum $\{\gamma_i(2\gamma_i-1)^{2k-1},\ 1\le i\le v-1\}$. Therefore it is of interest to learn more about $\{\gamma_i,\ 1\le i\le v-1\}$.

THEOREM 13. *Let* $r_i := \Pi_{l\in\mathscr{M}_i} v_l$ *for* $i=1$ *and* 2, $v_* := \min(r_1, r_2)$ *and* $v^* = \max(r_1, r_2)$. *Then*

$$
\begin{aligned}
\gamma_0 &= 1, & \\
\tfrac{1}{2} \le \gamma_l &< 1, & 1 \le l \le v_* - 1,\\
\gamma_l &= \tfrac{1}{2}, & v_* \le l \le v^* - 1,\\
0 \le \gamma_l &\le \tfrac{1}{2}, & v^* \le l \le v-1,
\end{aligned}
\tag{12}
$$

and at least $v-2v_*$ *of the* γ_l*'s take values in* $\{0, 1/2\}$.

PROOF. Since $\mathbf{M}_i$ is idempotent, it has $\operatorname{rank}(\mathbf{M}_i) = r_i$ unit eigenvalues and $v-r_i$ zero eigenvalues. Therefore,

$$\tfrac{1}{2}\max[\lambda_l(\mathbf{M}_1)+\lambda_{v-1}(\mathbf{M}_2), \lambda_{v-1}(\mathbf{M}_1)+\lambda_l(\mathbf{M}_2)] = \begin{cases} \tfrac{1}{2}, & 0\le l\le v^*-1,\\ 0, & v^*\le l\le v-1,\end{cases}$$

and

$$\tfrac{1}{2}\min[\lambda_l(\mathbf{M}_1)+\lambda_0(\mathbf{M}_2), \lambda_0(\mathbf{M}_1)+\lambda_l(\mathbf{M}_2)] = \begin{cases} 1, & 0\le l\le v_*-1,\\ \tfrac{1}{2}, & v_*\le l\le v-1,\end{cases}$$

which, together with Theorem 0, establish (12).

Whereas $\operatorname{rank}(\mathbf{M}) \le \operatorname{rank}(\mathbf{M}_1)+\operatorname{rank}(\mathbf{M}_2) = r_1+r_2$, $\operatorname{rank}(\frac{1}{2}(\mathbf{M}_1\mathbf{M}_2+\mathbf{M}_2\mathbf{M}_1)) \le \operatorname{rank}(\mathbf{M}_1\mathbf{M}_2)+\operatorname{rank}(\mathbf{M}_2\mathbf{M}_1) \le 2\min[\operatorname{rank}(\mathbf{M}_1), \operatorname{rank}(\mathbf{M}_2)] = 2v_*$. Since $\frac{1}{2}(\mathbf{M}_1\mathbf{M}_2+\mathbf{M}_2\mathbf{M}_1) = \mathbf{M}(2\mathbf{M}-\mathbf{I})$ has the spectrum $\{\gamma_l(2\gamma_l-1),\ 0\le l\le v-1\}$ and is of rank no greater than $2v_*$, it has at least $v-2v_*$ of the γ_l's in $\{0, \frac{1}{2}\}$. □

Since Theorem 13 implies that no more than $2v_*-1$ of the γ_l's are in $(0,\frac{1}{2})\cup(\frac{1}{2},1)$, this result becomes significant for small v_*. For example, if $m_1=1$ and $r_1=2$, then no more than three γ_l's lie in $(0,\frac{1}{2})\cup(\frac{1}{2},1)$.

By analogy with the development for Plan 1, $\mathbf{M}$ is the expected one-cycle transition matrix corresponding to a plan that on each cycle randomly selects a subset $\mathscr{M}_i$ and updates all of its coordinates. Then $2k$ cycles have the expected transition matrix $\mathbf{M}^{2k}$ and km expected coordinate updates. In a manner analogous to the relationship between Plans 1 and 2, Plan 5 induces faster convergence than this alternative plan if and only if $\gamma_1+2\gamma_{\min}>1$, where $\gamma_{\min}$ denotes the smallest positive eigenvalue of $\mathbf{M}$. Since $\gamma_1\ge\frac{1}{2}$, $\gamma_{\min}>\frac{1}{4}$ is sufficient for this faster convergence. Since this alternative plan allows repeated updating of the same coordinate subset, we focus on Plan 5, which guarantees exhaustive updating before repetition.

We now compare convergence rates to the equilibrium distribution $\boldsymbol{\pi}$ for Plan 5, an essentially deterministic coordinate selection plan, with Plan 2, an essentially random coordinate selection plan.

THEOREM 14. *Let $m \geq 1$. If $\gamma_1+\gamma_{\min} \geq 1$, then for all $x, y \in \mathscr{S}$, $W_{25xy}^{(km)} \to \infty$ as $k \to \infty$ if and only if*

$$\gamma_1 < \frac{1}{2}\left[1+\left(\frac{m\beta_1-1}{m-1}\right)^{m/2}\right]. \tag{13}$$

If $\gamma_1+\gamma_{\min} < 1$, then for all $x, y \in \mathscr{S}$, $W_{25xy}^{(km)} \to \infty$ as $k \to \infty$ if and only if

$$\gamma_{\min} > \frac{1}{2}\left[1-\left(\frac{m\beta_1-1}{m-1}\right)^{m/2}\right]. \tag{14}$$

PROOF. The expected km-step transition matrix $\mathbf{Q}_{2,\,km}$ has elements

$$q_{2xy}^{(km)} = \pi_y + \sqrt{\frac{\pi_y}{\pi_x}} \sum_{l=1}^{v-1} f_{xl} f_{yl} \beta_l \left(\frac{m\beta_l-1}{m-1}\right)^{km-1},$$

whereas the expected k-iteration transition matrix $\mathbf{Q}_{5,\,km}$ has elements

$$q_{5xy}^{(km)} = \pi_y + \sqrt{\frac{\pi_y}{\pi_x}} \sum_{l=1}^{v-1} f'_{xl} f'_{yl} \gamma_l (2\gamma_l-1)^{2k-1}, \tag{15}$$

where $\{f'_{xl}\}$ are elements of the left eigenvector of $\mathbf{D}^{1/2}\mathbf{M}\mathbf{D}^{-1/2}$ corresponding to γ_l. Then $W_{25xy}^{(km)} \to \infty$ as $k \to \infty$ if and only if

$$\max\left[\left(\frac{m\beta_1-1}{m-1}\right)^m, \left|\frac{m\beta_{\min}-1}{m-1}\right|^m\right] > \max[(2\gamma_1-1)^2, (2\gamma_{\min}-1)^2].$$

Since $\beta_1 \geq 1-1/m$, $m\beta_1-1 \geq |m\beta_{\min}-1|$ for all $m \geq 3$. If $\gamma_1+\gamma_{\min} \geq 1$, then

$$\left(\frac{m\beta_1-1}{m-1}\right)^m > (2\gamma_1-1)^2 \tag{16}$$

is necessary and sufficient for convergence. However, this is equivalent to (13).

If $\gamma_1+\gamma_{\min} \leq 1$, then

$$\left(\frac{m\beta_1-1}{m-1}\right)^{m/2} > 1-2\gamma_{\min} \tag{17}$$

is necessary and sufficient and this is equivalent to (14). □

It is of special interest to assess convergence as the number of coordinates m grows. Again the bound $\beta_1 \geq 1-1/m$ provides insight. If $\gamma_1+\gamma_{\min} > 1$, then a sufficient condition for Plan 5 to converge faster is

$$\gamma_1 < \frac{1+\mathrm{e}^{-1/2}}{2} = 0.803265\ldots \quad \text{as } m \to \infty.$$

If $\gamma_1 + \gamma_{\min} \le 1$, then a sufficient condition is

$$\gamma_{\min} > \frac{1 - \mathrm{e}^{-1/2}}{2} = 0.196735\ldots \quad \text{as } m \to \infty.$$

Since Theorem 8 already has shown that Plan 2 converges faster than Plan 1 for $m > 2$, Plan 5 also converges faster than Plan 1 whenever the conditions in Theorem 13 are met.

It is easily seen that the following corollary holds.

COROLLARY 15. *If either*

$$\gamma_1 > \frac{1}{2}\left[1 + \left(\frac{m\beta_1 - 1}{m-1}\right)^{m/2}\right] \quad \textit{or} \quad \gamma_{\min} < \frac{1}{2}\left[1 - \left(\frac{m\beta_1 - 1}{m-1}\right)^{m/2}\right],$$

then $W_{25xy}^{(km)} \to 0$ *as* $k \to \infty$ *so that Plan* 2 *converges more rapidly than Plan* 5.

These conditions become $\gamma_1 > 0.803265\ldots$ and $\gamma_{\min} < 0.196735\ldots$ as $m \to \infty$.

6. Variance considerations. For $m_1 = M_2 = 1$, Plan 5 reduces to Plan 2. For $m_1 = m_2 = m/2$ for even m, an expression for $T_5 := \lim_{n\to\infty} n \operatorname{var} \bar{\mu}_{5n}$ is derived in Fishman (1994). However, the expression for T_5 does not lend itself to meaningful comparison with T_1 so that it is not possible to state conditions under which $T_5 < T_1$ and $T_5 > T_1$. However, deterministic coordinate selection in Gibbs sampling often takes one observation X_{lm} on each iteration l for $l = 1, \ldots, k$ and uses

$$\tilde{\mu}_{5k} = \frac{1}{k} \sum_{l=1}^{k} X_{lm} \tag{18}$$

to estimate μ. For (18), Theorem 16 gives a limiting result for variance that provides some insight into the benefit of deterministic versus random coordinate selection.

THEOREM 16. *Let* $\mathscr{V}_* := \{l \in \mathscr{V}\colon\ \gamma_l \notin (0, \frac{1}{2})\}$. *Then*

$$\begin{aligned} \tilde{T}_5 := \lim_{k\to\infty} k \operatorname{var} \tilde{\mu}_{5k} &= 1 + \sum_{l\in\mathscr{V}_*} \frac{d_l^2(2\gamma_l - 1)}{2(1-\gamma_l)} \\ &= \sum_{l\in\mathscr{V}_*} \frac{d_l^2}{2(1-\gamma_l)} + \sum_{l\in\mathscr{V}\setminus\mathscr{V}_*} d_l^2, \end{aligned} \tag{19}$$

where $d_l^2 > 0$ *for all* $l \in \mathscr{V}_*$ *and* $\sum_{l=1}^{v-1} d_l^2 = 1$.

PROOF. As before, we have var $X_{im} = 1$. Based on (15),

$$\begin{aligned} \operatorname{cov}(X_{im}, X_{jm}) &= \sum_{x=0}^{v-1}\sum_{y=0}^{v-1} g_x g_y \pi_x (q_{5xy}^{(|i-j|m)} - \pi_y) \\ &= \sum_{l=1}^{v-1}\left(\sum_{x=0}^{v-1}\sqrt{\pi_x}\, g_x f'_{xl}\right)^2 \gamma_l (2\gamma_l - 1)^{2|i-j|-1} \qquad (20)\\ &= \sum_{l=1}^{v-1} d_l^2 \gamma_l (2\gamma_l - 1)^{2|i-j|-1}, \end{aligned}$$

where $d_l := \sum_{x=0}^{v-1} g_x \sqrt{\pi_x}\, f'_{xl}$. Recall that $\{f'_{xl},\ 0 \le x \le v-1\}$ is the left eigenvector of $\mathbf{D}^{1/2}\mathbf{M}\mathbf{D}^{-1/2}$ corresponding to eigenvalue γ_l. Since

$$\sum_{l=1}^{v-1} f'_{xl} f'_{yl} = \begin{cases} -\sqrt{\pi_x \pi_y}, & \text{if } x \ne y, \\ 1 - \pi_x, & \text{if } x = y, \end{cases}$$

$\sum_{l=1}^{v-1} d_l^2 = 1$. Then

$$\begin{aligned} \tilde{T}_5 &= 1 + 2\sum_{j=1}^{\infty} \operatorname{cov}(X_0, X_{jm}) \\ &= 1 + 2\sum_{j=1}^{\infty}\sum_{l=1}^{v-1} d_l^2 \gamma_l (2\gamma_l - 1)^{2j-1} \\ &= 1 + \sum_{l \in \mathscr{V}_*} \frac{d_l^2 (2\gamma_l - 1)}{2(1-\gamma_l)}. \end{aligned}$$ □

Observe that $\tilde{T}_5 \le \frac{1}{2}(1-\gamma_1)$, whereas (6) implies $T_1 \le (1+\beta_1)/(1-\beta_1)$. Moreover, the nonnegativity of the eigenvalues implies var $\bar{\mu}_{1n} \le T_1/n \le (1+\beta_1)/(1-\beta_1)n$. If Plan 5 induces faster convergence than Plan 1 does, then $\gamma_1 \le \frac{1}{2}[1 + ((m\beta_1 - 1)/m - 1)^{m/2}] \le \frac{1}{2}(1 + \beta_1^{m/2})$, by Theorem 14, so that $\tilde{T}_5 \le \frac{1}{2}(1-\gamma_1) \le \frac{1}{2}(1 - \beta_1^{m/2})$. As a consequence, for large k and $n := km$,

$$\operatorname{var} \tilde{\mu}_{5k} \le \frac{1}{(1-\beta_1^{m/2})k} = \frac{m}{(1-\beta_1^{m/2})n}. \qquad (21)$$

If there exists an $\varepsilon > 0$ such that $\beta_1 \ge 1 - 1/m^{1+\varepsilon}$, then

$$\lim_{m\to\infty} \frac{m}{(1-\beta_1^{m/2})} \cdot \frac{1-\beta_1}{1+\beta_1} = 1.$$

Expression (21) implies that for large m the uppermost bound $m/((1-\beta_1^{m/2})n)$ for var $\tilde{\mu}_{5k}$ for k observations under Plan 5 is close to the uppermost bound for var $\bar{\mu}_{1n}$ for n observations under Plan 1 and, therefore, close to the uppermost bound for var $\bar{\mu}_{2n}$ for n observations under Plan 2. While not a substitute for an exact comparison of variances, this result does suggest that little may be lost in statistical efficiency by using $\tilde{\mu}_{5k}$ under Plan 5 in place of $\bar{\mu}_{2n}$ under Plan 2. Moreover, since it is known that $\operatorname{var}\bar{\mu}_{5n} \le \operatorname{var}\tilde{\mu}_{5k}$ [e.g., MacEachern and Berliner (1994)], this analysis encourages the conjecture that $\operatorname{var}\bar{\mu}_{5n} \le \operatorname{var}\bar{\mu}_{2n}$ for large n when Plan 5 induces faster convergence than Plan 2 does and $\beta_1 \ge 1 - 1/m^{1+\varepsilon}$.

APPENDIX

PROOF OF THEOREM 4. Let

$$\mathbf{A}_j(\mathbf{l}) := \frac{1}{s_j(\mathbf{l})}\sum_{i=1}^{j} w_{l_i}\mathbf{P}_{l_i}$$

and

$$\mathbf{B}_{m-j}(\mathbf{l}) := \frac{1}{1-s_j(\mathbf{l})}\sum_{i=j+1}^{m} w_{l_i}\mathbf{P}_{l_i}$$

so that $\mathbf{L}(\mathbf{w}) = s_j(\mathbf{l})\mathbf{A}_j(\mathbf{l}) + [1 - s_j(\mathbf{l})]\mathbf{B}_{m-j}(\mathbf{l})$. From Theorem 0, for each $l = 1, \dots, v-1$,

$$\begin{aligned}\lambda_l(\mathbf{L}(\mathbf{w})) \ge \max\{&s_j(\mathbf{l})\lambda_l(\mathbf{A}_j(\mathbf{l})) + [1-s_j(\mathbf{l})]\lambda_{v-1}(\mathbf{B}_{m-j}(\mathbf{l})),\\ &s_j(\mathbf{l})\lambda_{v-1}(\mathbf{A}_j(\mathbf{l})) + [1-s_j(\mathbf{l})]\lambda_l(\mathbf{B}_{m-j}(\mathbf{l}))\}.\end{aligned}$$

Let $u_0(\mathbf{l}) := 1$ and $u_j(\mathbf{l}) = u_{j-1}(\mathbf{l})v_{l_j}$ for $1 \le j \le m$. Each $\mathbf{A}_j(\mathbf{l})$ and $\mathbf{B}_{m-j}(\mathbf{l})$ is positive semidefinite. Also, $\mathbf{A}_j(\mathbf{l})$ and $\mathbf{B}_{m-j}(\mathbf{l})$ have exactly $v/u_j(\mathbf{l})$ and $u_j(\mathbf{l})$ closed sets, respectively. Since the number of unit eigenvalues corresponds to the number of closed sets in a stochastic matrix,

$$\lambda_l(\mathbf{L}(\mathbf{w})) \ge s_j(\mathbf{l}), \qquad 1 \le l \le v/u_j(\mathbf{l}) - 1,$$

and

$$\lambda_l(\mathbf{L}(\mathbf{w})) \ge 1 - s_j(\mathbf{l}), \qquad 1 \le l \le u_j(\mathbf{l}) - 1,\ 1 \le j \le m.$$

For each $\mathbf{l} \in \mathscr{L}$ there exists another permutation $\mathbf{l}' \in \mathscr{L}$ with the coordinate sequence reversed so that $s_j(\mathbf{l}) = 1 - s_{m-j+1}(\mathbf{l}')$, $s_{m-j+1}(\mathbf{l}) = 1 - s_j(\mathbf{l}')$ and $v/u_{m-j+1}(\mathbf{l}) = u_j(\mathbf{l}') \ge u_j$. Therefore, for each $\mathbf{l} \in \mathscr{L}$, $\lambda_l(\mathbf{L}(\mathbf{w})) \ge 1 - s_j(\mathbf{l})$, $1 \le l \le u_j - 1$, $1 \le j \le m$, which establishes (1). □

PROOF OF THEOREM 5. From Theorem 0,

$$\begin{aligned}\lambda_l(\mathbf{L}(\mathbf{w})) \le \min\{&s_j(\mathbf{l})\lambda_l(\mathbf{A}_j(\mathbf{l})) + [1-s_j(\mathbf{l})]\lambda_0(\mathbf{B}_{m-j}(\mathbf{l})),\\ &s_j(\mathbf{l})\lambda_0(\mathbf{A}_j(\mathbf{l})) + [1-s_j(\mathbf{l})]\lambda_l(\mathbf{B}_{m-j}(\mathbf{l}))\},\end{aligned}$$

where $\lambda_0(\mathbf{A}_j(\mathbf{l})) = \lambda_0(\mathbf{B}_{m-j}(\mathbf{l})) = 1$. Let $t_0 := 0$, $t_j := t_{j-1} + v/v_j$, $t_0(\mathbf{l}) := 0$ and $t_j(\mathbf{l}) := t_{j-1}(\mathbf{l}) + v/v_{l_j}$ for $1 \le j \le m$ and observe that

$$\operatorname{rank}(\mathbf{A}_j(\mathbf{l})) \le \sum_{i=1}^{j}\operatorname{rank}(\mathbf{P}_{l_i}) = t_j(\mathbf{l})$$

and

$$\text{rank}(\mathbf{B}_{m-j}(\mathbf{l})) \le \sum_{i=j+1}^{m} \text{rank}(\mathbf{P}_{l_i}) = t_m(\mathbf{l}) - t_j(\mathbf{l}).$$

Therefore, $\mathbf{A}_j(\mathbf{l})$ and $\mathbf{B}_{m-j}(\mathbf{l})$, respectively, have no more than $t_j(\mathbf{l})$ and $t_m(\mathbf{l}) - t_j(\mathbf{l})$ positive eigenvalues, implying

$$\lambda_l(\mathbf{L}(\mathbf{w})) \le s_j(\mathbf{l}), \qquad l \ge t_m(\mathbf{l}) - t_j(\mathbf{l}),$$

and

$$\lambda_l(\mathbf{L}(\mathbf{w})) \le 1 - s_{j-1}(\mathbf{l}), \qquad l \ge t_{j-1}(\mathbf{l}). \tag{22}$$

For all $\mathbf{l} \in \mathscr{L}$ and $1 \le j \le m$, observe that $t_m - t_{m-j} \le t_j(\mathbf{l}) \le t_j$. Also, for each $\mathbf{l} \in \mathscr{L}$ there exists an $\mathbf{l}'$ such that $t_j(\mathbf{l}) = t_m(\mathbf{l}') - t_j(\mathbf{l}')$, $s_j(\mathbf{l}) = 1 - s_{m-j}(\mathbf{l}')$ and $s_{m-j}(\mathbf{l}) = 1 - s_j(\mathbf{l}')$. Therefore,

$$\lambda_l(\mathbf{L}(\mathbf{w})) \le \max_{\mathbf{l} \in \mathscr{L}} [1 - s_{j-1}(\mathbf{l})], \qquad l \ge t_m - t_{m-j+1},\ 1 \le j \le m,$$

which establishes (22). □

Acknowledgment. I am grateful to my colleague Mark Hartmann for several fruitful discussions on early aspects of this research.

REFERENCES

AMIT, Y. (1991). On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.* **38** 82–89.

AMIT, Y. and GRENANDER, U. (1989). Comparing sweeping strategies for stochastic relaxation. *J. Multivariate Anal.* **37** 197–222.

BARONE, P. and FRIGESSI, A. (1990). Improving stochastic relaxation for Gaussian random fields. *Problems in the Informational Sciences* **4** 369–389.

FAN, K. (1949). On a theorem of Weyl concerning eigenvalues of linear transformations. I. *Proc. Nat. Acad. Sci. U.S.A.* **35** 652–655.

FAN, K. and HOFFMAN, A. (1955). Some metric inequalities in the space of matrices. *Proc. Amer. Math. Soc.* **6** 111–116.

FISHMAN, G. S. (1994). Coordinate selection rules for Gibbs sampling. Report UNC/OR TR-92/15, Operations Research Dept., Univ. North Carolina at Chapel Hill.

FRIGESSI, A., HWANG, C.-R., SHEU, S.-J. and DI STEFANO, P. (1993). Convergence rates of the Gibbs sampler, Metropolis algorithm and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B* **55** 205–219.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sample-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

HORN, R. A. and JOHNSON, C. A. (1985). *Matrix Analysis*. Cambridge Univ. Press.

JOHNSON, V. (1989). On statistical image reconstruction. Ph.D. dissertation, Dept. Statistics, Univ. Chicago.

KEMENY, J. G. and SNELL, J. L. (1960). *Finite Markov Chains*. Van Nostrand, New York.

LIU, J., WONG, W. H. and KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. *Biometrika* **81** 27–40.

LIU, J., WONG, W. H. and KONG, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B* **57** 157–169.

MACEACHERN, S. and BERLINER, L. M. (1994). Subsampling the Gibbs sampler. *Amer. Statist.* **48** 188–190.

MARCUS, M. and MINC, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

ROBERTS, G. O. and POLSON, N. G. (1991). A note on the geometric convergence of the Gibbs sampler. Dept. Mathematics, Univ. Nottingham.

ROBERTS, G. O. and SMITH, A. F. M. (1992). Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. Research Report 92-30, Statistical Laboratory, Univ. Cambridge.

SCHERVISH, M. J. and CARLIN, B. P. (1993). On the convergence of successive substitution sampling. Dept. Statistics, Carnegie Mellon Univ.

DEPARTMENT OF OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599
E-MAIL: gfish@fish.or.unc.edu