F. Leisen, A. Mira

Coalescence time and second largest eigenvalue modulus in the monotone reversible case

2006/11

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell'Università dell'Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

Printed in Italy in July 2006
Università degli Studi dell'Insubria
Via Monte Generoso, 71, 21100 Varese, Italy

COALESCENCE TIME AND SECOND LARGEST EIGENVALUE MODULUS IN THE MONOTONE REVERSIBLE CASE

Fabrizio Leisen*and Antonietta Mira†

July 27, 2006

Abstract

If T is the coalescence time of the Propp and Wilson [15], perfect simulation algorithm, the aim of this paper is to show that T depends on the second largest eigenvalue modulus of the transition matrix of the underlying Markov chain. This gives a relationship between the ordering based on the speed of convergence to stationarity in total variation distance and the ordering defined in terms of speed of coalescence in perfect simulation.

1 Introduction

In [15], Propp and Wilson define an algorithm to sample "exactly" from a probability distribution π defined on a finite set E. To this aim, we need a Markov chain that has π as its unique stationary distribution. The updating rule of the Markov chain can be considered as a function:

$$\begin{array}{rcl} \phi : E \times [0,1] & \rightarrow & E \\ (x, U) & \rightarrow & \phi(x, U) \end{array}$$

where U is a uniform random variable on the interval $[0, 1]$. Let $P = \{p_{ij}\}_{i,j \in E}$ be the transition matrix of the Markov chain, then $p_{ij} = P(\phi(i, U) = j)$. Clearly, given a distribution of interest π, the choice of the Markov chain is not unique, and an open question is how this choice can be made "optimally" in some sense. For Markov chains reversible with respect to π, we can define an ordering, called "speed of convergence", that is based on the second largest eigenvalue modulus (SLEM throughout the paper) of the transition matrix (see [12] for a survey on orderings for Markov chains). In this ordering convergence is measured in terms of total variation distance. The purpose of this paper

*Università di Modena e Reggio Emilia, Dipartimento di Matematica, Via Campi 213/b Modena, E-mail: leisen.fabrizio@unimore.it

†Università dell'Insubria, Dipartimento di Economia, Via Monte Generoso 71, 21100 Varese, E-mail: antonietta.mira@uninsubria.it

is to show that, in monotone reversible cases, the speed of convergence ordering implies an ordering on the coalescence times. In Section 2, we review the Propp and Wilson perfect simulation algorithm and set the notation. In Section 3 we give preliminary theorems and in Section 4 we state the main result that will be used, in Section 5, to define an ordering for the coalescence times.

2 Propp and Wilson exact algorithm

In their seminal paper, [15], Propp and Wilson device a way to turn a Markov chain Monte Carlo algorithm (MCMC) into a perfect simulation algorithm. It is well know that in MCMC the crucial problem is how to detect when the Markov chain has reached its stationary regime. Many convergence diagnostics have been designed that aim at detecting failure of convergence based on the simulated path of the Markov chain. All these convergence diagnostics can at most give negative answers but will never reassure the user on the fact that their Markov chains have reached stationarity. Propp and Wilson propose a clever way to simulate a Markov chain so that at the end of the perfect simulation algorithm an exact sample from the stationary distribution, π, is obtained.

2.1 The exact simulation algorithm

Let :

- $E = \{x_1, x_2, \cdots, x_N\}$ be a finite states space
- $U_{-1}, U_{-2}, U_{-3}, \cdots$ be a sequence of i.i.d. random variables
- $\phi = \phi(x, U)$ $x \in E$ be the updating rule of an ergodic Markov chain with stationary distribution π with the property that for all $i, j \in E$ $P(\phi(i, U_t) = j) = p_{ij}$.

Consider the following exact simulation algorithm:

step -1 Compute :
$\phi_1 = \phi(x_1, U_{-1})$
$\phi_2 = \phi(x_2, U_{-1})$
...
...
$\phi_n = \phi(x_n, U_{-1})$
If $\phi_1 = \phi_2 = \cdots = \phi_n = k$ for some $k \in E$ then return k.
else go to step -2

step -2 Compute :
$\phi_1 = \phi(\phi(x_1, U_{-2}), U_{-1})$
$\phi_2 = \phi(\phi(x_2, U_{-2}), U_{-1})$
...
...
$\phi_n = \phi(\phi(x_n, U_{-2}), U_{-1})$
If $\phi_1 = \phi_2 = \cdots = \phi_n = k$ then return k.
else go to step -3

.....
.....
.....

step $-t$ Compute :
$\phi_1 = \phi(\phi(\cdots\phi(\phi(x_1, U_{-t}), U_{-t+1})\cdots, U_{-2}), U_{-1})$
$\phi_2 = \phi(\phi(\cdots\phi(\phi(x_2, U_{-t}), U_{-t+1})\cdots, U_{-2}), U_{-1})$
...
...
$\phi_n = \phi(\phi(\cdots\phi(\phi(x_n, U_{-t}), U_{-t+1})\cdots, U_{-2}), U_{-1})$
If $\phi_1 = \phi_2 = \cdots = \phi_n = k$ then return k.
else go to step $-t-1$

.....

Propp and Wilson, [15], proved that this algorithm ends with probability one and that the value k that is returned is distributed as π. The time step, T, where $\phi_1 = \phi_2 = \cdots = \phi_n = k$ is called "coalescence time" and it depends on $U_{-1}, U_{-2}, U_{-3}, \cdots$, so T is a random variable.

2.2 Monotone case and Coupling inequality

Let $\leq_E$ be a partial ordering on the states space E and assume that E admits maximum and minimum with respect to this ordering. Throughout the paper we denote the maximal and minimal state by $\hat{1}$ and $\hat{0}$.
If the updating rule of an ergodic Markov chain with stationary distribution π satisfies the property:

$$\forall x, y \in E \quad \text{such that} \quad x \leq_E y \quad \text{we have} \quad \phi(x, U) \leq_E \phi(y, U)$$

then the chain is called monotone. This property is very important for the Propp and Wilson algorithm because it guaranties that it is sufficient to run only two chains, starting at $\hat{0}$ (minimal chain) and $\hat{1}$ (maximal chain) respectively, and to verify coalescence only for these two chains. The monotonicity property ensures that if the maximal and minimal chain

have coalesced, then all other chains (started from any other possible value in the state space) will also coalesce because they will be "trapped" between the maximal and the minimal chain. Computationally, running only the maximal and minimal chain this is more convenient then running all possible chains.
The question is: if ϕ_1, ϕ_2 are two monotone updating rules of two ergodic Markov chains with stationary distribution π and transitions matrix $\mathbf{P}$ and $\mathbf{Q}$ respectively, which one coalesces before? In this paper the above question will be addressed. For monotone Markov chains we have two inequalities called "coupling inequalities":

$$P(T > t) \leq l\overline{d(t)}$$

$$P(T > t) \geq \overline{d(t)} \geq d(t)$$

where l is the length of the longest chain in the totally ordered space E and :

$$\overline{d(t)} = \max_{x,y \in E} ||\mathbf{P}^t(x,\cdot) - \mathbf{P}^t(y,\cdot)||_{TV};$$

$$d(t) = \max_{x \in E} ||\mathbf{P}^t(x,\cdot) - \pi(\cdot)||_{TV}.$$

3 Preliminaries

For each fixed $M \in \mathbb{N}$, we define two Markov chains on the state space E in the following way :

$$X_0^M = \hat{0} \quad X_1^M = \phi(\hat{0}, U_{-M}) \quad X_2^M = \phi(X_1^M, U_{-M+1}) \cdots X_M^M = \phi(X_{M-1}^M, U_{-1});$$

$$Y_0^M = \hat{1} \quad Y_1^M = \phi(\hat{1}, U_{-M}) \quad Y_2^M = \phi(Y_1^M, U_{-M+1}) \cdots Y_M^M = \phi(Y_{M-1}^M, U_{-1}).$$

If T is the coalescence time, we have:

$$X_T^T, Y_T^T \sim \pi.$$

We now construct a new Markov chain on the enlarged state space $E \times E$, in the following way:

$$\begin{array}{lllllll}
Z_0^M & = & (\hat{0}, \hat{1}) & & & & \\
Z_1^M & = & \hat{\phi}((\hat{0}, \hat{1}), U_{-M}) & = & (\phi(\hat{0}, U_{-M}), \phi(\hat{1}, U_{-M})) & = & (X_1^M, Y_1^M) \\
Z_2^M & = & \hat{\phi}(Z_1^M, U_{-M+1}) & = & (\phi(X_1^M, U_{-M+1}), \phi(Y_1^M, U_{-M+1})) & = & (X_2^M, Y_2^M) \\
 & \cdot & & \cdot & & \cdot & \\
 & \cdot & & \cdot & & \cdot & \\
 & \cdot & & \cdot & & \cdot & \\
Z_M^M & = & \hat{\phi}(Z_{M-1}^M, U_{-1}) & = & (\phi(X_{M-1}^M, U_{-1}), \phi(Y_{M-1}^M, U_{-M})) & = & (X_M^M, Y_M^M)
\end{array}$$

The transition matrix of the Markov chain Z defined above is:

$$\tilde{P} = \{p_{(x_{i_1}, x_{i_2}) \to (x_{j_1}, x_{j_2})}\}_{(x_{i_1}, x_{i_2}), (x_{j_1}, x_{j_2}) \in E \times E}$$

where :

$$p_{(x_{i_1}, x_{i_2}) \to (x_{j_1}, x_{j_2})} = p_{x_{i_1} x_{j_1}} p_{x_{i_2} x_{j_2}}.$$

It is easy to check that this is indeed a stochastic matrix in fact, for any (x_{i_1}, x_{i_2}) we have that:

$$\sum_{(x_{j_1},x_{j_2})\in E\times E} p_{(x_{i_1},x_{i_2})\to(x_{j_1},x_{j_2})} = \sum_{x_{j_1}\in E}\sum_{x_{j_2}\in E} p_{x_{i_1}x_{j_1}} p_{x_{i_2}x_{j_2}} = 1.$$

We note that the matrix $\tilde{P}$ can be written in the convenient form:

$$\tilde{P} \quad = \quad \begin{bmatrix} p_{x_1x_1}P & p_{x_1x_2}P & \cdots & p_{x_1x_N}P \\ p_{x_2x_1}P & p_{x_2x_2}P & \cdots & p_{x_2x_N}P \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{x_Nx_1}P & p_{x_Nx_2}P & \cdots & p_{x_Nx_N}P \end{bmatrix}$$

and it's easy to prove that :

$$\tilde{P}^n \quad = \quad \begin{bmatrix} p^{(n)}_{x_1x_1}P^n & p^{(n)}_{x_1x_2}P^n & \cdots & p^{(n)}_{x_1x_N}P^n \\ p^{(n)}_{x_2x_1}P^n & p^{(n)}_{x_2x_2}P^n & \cdots & p^{(n)}_{x_2x_N}P^n \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p^{(n)}_{x_Nx_1}P^n & p^{(n)}_{x_Nx_2}P^n & \cdots & p^{(n)}_{x_Nx_N}P^n \end{bmatrix}$$

Theorem 3.1 *If P is reversible w.r.t. π then $\tilde{P}$ is also reversible w.r.t. $\tilde{\pi}$, where $\tilde{\pi}$ is defined, for all $i, j \in E \times E$, as:*

$$\tilde{\pi}(i,j) = \pi(i)\pi(j).$$

Proof:
We show that for all $(x_{i_1}, x_{i_2}), (x_{j_1}, x_{j_2}) \in E \times E$ the following identity holds:

$$\tilde{\pi}(x_{i_1}, x_{i_2}) p_{(x_{i_1},x_{i_2})\to(x_{j_1},x_{j_2})} = \tilde{\pi}(x_{j_1}, x_{j_2}) p_{(x_{j_1},x_{j_2})\to(x_{i_1},x_{i_2})}.$$

The above identity follows from:

$$\begin{aligned} \tilde{\pi}(x_{i_1}, x_{i_2}) p_{(x_{i_1},x_{i_2})\to(x_{j_1},x_{j_2})} &= \pi(x_{i_1})\pi(x_{i_2}) p_{x_{i_1}x_{j_1}} p_{x_{i_2}x_{j_2}} \\ &= \pi(x_{i_1}) p_{x_{i_1}x_{j_1}} \pi(x_{i_2}) p_{x_{i_2}x_{j_2}} \\ &= p_{x_{j_1}x_{i_1}} \pi(x_{j_1}) p_{x_{j_2}x_{i_2}} \pi(x_{j_2}) \\ &= p_{x_{j_1}x_{i_1}} p_{x_{j_2}x_{i_2}} \pi(x_{j_1})\pi(x_{j_2}) \\ &= \tilde{\pi}(x_{j_1}, x_{j_2}) p_{(x_{j_1},x_{j_2})\to(x_{i_1},x_{i_2})} \ \square \end{aligned}$$

The next theorem establishes a connection between the Z chain and the two chains that start from $\hat{0}$ and $\hat{1}$. We use this notation: $\delta_{\hat{0}}$, $\delta_{\hat{1}}$ for the measures on E, concentrated, respectively, on $\hat{0}$, $\hat{1}$, and $\delta_{(\hat{0},\hat{1})}$ for the

measure on $E \times E$ concentrated on $(\hat{0}, \hat{1})$. Note that,
$\delta_{(\hat{0},\hat{1})}(x, y) = \delta_{\hat{0}}(x)\delta_{\hat{1}}(y)$, so we can write $\delta_{(\hat{0},\hat{1})}$ in this compact form:

$$\delta_{(\hat{0},\hat{1})} \quad = \quad \begin{bmatrix} \delta_{\hat{0}}(x_1)\delta_{\hat{1}} \\ \delta_{\hat{0}}(x_2)\delta_{\hat{1}} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{\hat{0}}(x_N)\delta_{\hat{1}} \end{bmatrix}$$

Vectors are considered as column and we use the superscript t to denote the transpose.

Theorem 3.2 *Let T be a random variable with values in $\{1, 2, \cdots\}$. The following two statements are equivalent:*

1. *$X_T^T \sim \pi$ and $Y_T^T \sim \pi$*
2. *$Z_T^T \sim \tilde{\pi}$*

Proof:
"$(1 \Rightarrow 2)$" If $X_T^T \sim \pi$,$Y_T^T \sim \pi$ then

$$P(X_T^T = i) = \pi(i) = (\delta_{\hat{0}} P^T)(i) = p_{\hat{0}i}^{(T)}$$

$$P(Y_T^T = i) = \pi(i) = (\delta_{\hat{1}} P^T)(i) = p_{\hat{1}i}^{(T)}$$

So

$$\begin{aligned} \delta_{(\hat{0},\hat{1})}^t \tilde{P}^T &= [\delta_{\hat{0}}(\hat{0})\delta_{\hat{1}}^t p_{\hat{0}x_1}^{(T)} \tilde{P}^T, \delta_{\hat{0}}(\hat{0})\delta_{\hat{1}}^t p_{\hat{0}x_2}^{(T)} \tilde{P}^T, \cdots, \delta_{\hat{0}}(\hat{0})\delta_{\hat{1}}^t p_{\hat{0}x_N}^{(T)} \tilde{P}^T] \\ &= [\pi(x_1)\pi^t, \pi(x_2)\pi^t, \cdots, \pi(x_N)\pi^t] \\ &= \tilde{\pi} \end{aligned}$$

"$(2 \Rightarrow 1)$" If $Z_T^T \sim \tilde{\pi}$ then

$$||\delta_{(\hat{0},\hat{1})} P^T - \tilde{\pi}||_{TV} = 0$$

Since:

$$\begin{aligned} ||\delta_{(\hat{0},\hat{1})}^t P^T - \tilde{\pi}||_{TV} &= \tfrac{1}{2} \textstyle\sum_{i=1}^N \sum_{j=1}^N |(\delta_{(\hat{0},\hat{1})}^t \tilde{P}^T)(x_i, x_j) - \tilde{\pi}(x_i, x_j)| \\ &= \tfrac{1}{2} \textstyle\sum_{i=1}^N \sum_{j=1}^N |p_{\hat{0}x_i}^{(T)} p_{\hat{1}x_j}^{(T)} - \pi(x_i)\pi(x_j)| \\ &= \text{(we apply the triangular inequality)} \\ &\geq \tfrac{1}{2} \textstyle\sum_{i=1}^N |p_{\hat{0}x_i}^{(T)} - \pi(x_i)| \\ &= ||\delta_{\hat{0}}^t P^T - \pi||_{TV} \end{aligned}$$

it follows:

$$0 \leq ||\delta_{\hat{0}}^t P^T - \pi||_{TV} \leq ||\delta_{(\hat{0},\hat{1})}^t P^T - \tilde{\pi}||_{TV} = 0.$$

Analogously, we can show that:

$$||\delta_1^t P^T - \pi||_{TV} = 0$$

We can thus conclude that:

$$X_T^T, Y_T^T \sim \pi$$

□

We now want to derive the eigenvalues of the matrix $\tilde{P}$ from the eigenvalues of the matrix P. We recall that a reversible matrix is always diagonalizable. In particular let R be the matrix:

$$R = \text{diag}(\pi(x_1), \pi(x_2), \cdots, \pi(x_N))$$

then the matrix

$$S = R^{\frac{1}{2}} P R^{-\frac{1}{2}}$$

is symmetric. So there exists an orthogonal matrix O such that:

$$D = O^t S O$$

where D is a diagonal matrix that has the same eigenvalues of P. Thus, if $\lambda_1, \lambda_2, \cdots, \lambda_N$ are the eigenvalues of P, then

$$D = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_N).$$

Theorem 3.3 *If $\lambda_1, \lambda_2, \cdots, \lambda_N$ are the eigenvalues of P then the set:*

$$\{\lambda_i \lambda_j : i, j = 1, \cdots, N\}$$

is the set of all the eigenvalues of $\tilde{P}$.

Proof: We use the preceding technique. Let

$$\tilde{R} = \text{diag}(\tilde{\pi}(x_1, x_1), \tilde{\pi}(x_1, x_2), \cdots, \tilde{\pi}(x_N, x_N))$$

If $S = \{s_{x_i x_j}\}_{x_i, x_j \in E}$ from a direct calculation it follows:

$$\tilde{S} = \begin{bmatrix} s_{x_1 x_1} S & s_{x_1 x_2} S & \cdots & s_{x_1 x_N} S \\ s_{x_2 x_1} S & s_{x_2 x_2} S & \cdots & s_{x_2 x_N} S \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ s_{x_N x_1} S & s_{x_N x_2} S & \cdots & s_{x_N x_N} S \end{bmatrix}$$

And if $O = \{o_{x_i x_j}\}_{x_i, x_j \in E}$ then we can choose $\tilde{O}$:

$$\tilde{O} = \begin{bmatrix} o_{x_1 x_1} O & o_{x_1 x_2} O & \cdots & o_{x_1 x_N} O \\ o_{x_2 x_1} O & o_{x_2 x_2} O & \cdots & o_{x_2 x_N} O \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ o_{x_N x_1} O & o_{x_N x_2} O & \cdots & o_{x_N x_N} O \end{bmatrix}$$

This is an orthogonal matrix and $\tilde{D} = \tilde{O}^t \tilde{S} \tilde{O}$ is a diagonal matrix with the same eigenvalues of $\tilde{P}$. This matrix has the form:

$$\tilde{D} = \begin{bmatrix} \lambda_1 D & 0 & \cdots & 0 \\ 0 & \lambda_2 D & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \lambda_N D \end{bmatrix}$$

□

The following two technical lemmas will be used in Section 4.

Lemma 3.1 *Let P be a reversible irreducible transition matrix on the finite state space E with stationary distribution π. Then, for $n \geq 1$,all $i \in E$, and all $A \subset E$*

$$|\delta_i^t P^n(A) - \pi^t(A)| \leq (\frac{1-\pi(i)}{\pi(i)})^{\frac{1}{2}} \min(\pi(A)^{\frac{1}{2}}, \frac{1}{2}) \rho^n$$

where ρ is the SLEM of P.

An immediate consequence of the Lemma (3.1) is that if $A = \{i\}$ then

$$|\delta_i^t P^n(i) - \pi^t(i)| \leq \rho^n$$

For a proof of Lemma (3.1), see Theorem 3.3, p.209 of [3].
If $\mathbf{Q} = \{q_{ij}\}_{i,j \in E}$ is a stochastic matrix then we can define the "ergodicity coefficient of Dobrushin, τ" as:

$$\tau(\mathbf{Q}) = \max_{x,y} \frac{1}{2} \sum_{i \in E} |q_{xi} - q_{yi}|.$$

So obviously:

$$\overline{d(t)} = \tau(\mathbf{P}^t)$$

where $\overline{d(t)}$ is the norm defined in Section 2.2. We recall a result on stochastic matrices and the ergodicity coefficient from [17] pp. 81-82:

Theorem 3.4 *Let $\mathbf{w} = \{w_i\}$ be an arbitrary vector and $\mathbf{P} = \{p_{ij}\}$ a stochastic matrix indexed by E. If $\mathbf{z} = \mathbf{P}\mathbf{w}$, $\mathbf{z} = \{z_i\}$, then, for any two indexes h, h':*

$$z_h - z_{h'} \leq \frac{1}{2} \sum_{j \in E} |p_{hj} - p_{h'j}| \{\max_{j \in E} w_j - \min_{j \in E} w_j\}$$

and

$$\{\max_{j \in E} z_j - \min_{j \in E} z_j\} \leq \tau(\mathbf{P}) \{\max_{j \in E} w_j - \min_{j \in E} w_j\}$$

or equivalently:

$$\max_{h,h' \in E} |z_h - z_{h'}| \leq \tau(\mathbf{P}) \{\max_{j,j' \in E} |w_j - w_{j'}|\}. \quad (1)$$

An immediate consequence of this Theorem is the following Lemma.

Lemma 3.2 *If P is a stochastic matrix with ρ as its second largest eigenvalues modulus, then*

$$\overline{d(t)} \geq \rho^t$$

Proof: If $\mathbf{w} = \{w_i\}_{i \in E}$ is an eigenvector with corresponding eigenvalue $\beta, |\beta| < 1$. Then:

$$\mathbf{P}\mathbf{w} = \beta\mathbf{w}$$
$$\mathbf{P}^t\mathbf{w} = \beta^t\mathbf{w}.$$

If $\mathbf{z} = \mathbf{P}^t\mathbf{w}$ then

$$z_h = \quad \mathbf{P}^t\mathbf{w}(h) = \beta^t w_h$$

$$z_{h'} = \quad \mathbf{P}^t\mathbf{w}(h') = \beta^t w_{h'}$$

$$\max_{h,h' \in E} |z_h - z_{h'}| = \quad \max_{h,h' \in E} |\beta^t w_h - \beta^t w_{h'}| = |\beta|^t \max_{h,h' \in E} |w_h - w_{h'}|$$

So by applying (1) we obtain:

$$|\beta|^t \max_{h,h' \in E} |w_h - w_{h'}| \leq \tau(\mathbf{P}^t)\{\max_{j,j' \in E} |w_j - w_{j'}|\}$$

But:

$$\max_{h,h' \in E} |w_h - w_{h'}| \quad = \quad \max_{j,j' \in E} |w_j - w_{j'}|$$

$$\overline{d(t)} \quad = \quad \tau(\mathbf{P}^t)$$

So if $\rho = |\beta|$ is the SLEM of $\mathbf{P}$ we obtain:

$$\overline{d(t)} \geq \rho^t$$

□

Now recall that

$$d(t) \leq \overline{d(t)} \leq 2d(t)$$

so

$$2d(t) \geq \rho^t.$$

We note that, given two transition matrices, P_1 and P_2, such that $\rho_1 < \rho_2$, then $\tilde{\rho}_1 < \tilde{\rho}_2$, where $\tilde{\rho}_1$ and $\tilde{\rho}_2$ are the SLEM of the enlarged chain, $\tilde{P}_1$ and $\tilde{P}_2$ respectively.

4 Main result

Let us begin with a remark. If T is a coalescence time we have that:

$$||\delta^t_{(\hat{0},\hat{1})}\tilde{P}^T - \tilde{\pi}||_{TV}$$

Furthermore, for monotone Markov chains, we have that:

$$d(T) = \max_{(x_i,x_j) \in (E \times E)} ||\delta^t_{(x_i,x_j)}\tilde{P}^T - \tilde{\pi}||_{TV} = 0.$$

This is due to the fact that, if the minimal and maximal chains (i.e. the chains starting from $\hat{0}$ and $\hat{1}$ respectively), coalesce then, thanks to the monotonicity property, chains starting from any other initial state also coalesce. We are now ready to state the main result.

Theorem 4.1 *Let ϕ_1, ϕ_2 be two monotone updating rules of Markov chains with transitions matrix P_1 and P_2 reversible with respect to π. If T_1, T_2 are the coalescence times respectively of ϕ_1, ϕ_2 and*

$$\rho_1 < \rho_2$$

where $\rho_1 = SLEM\ (P_1), \rho_2 = SLEM\ (P_2)$,then

$$T_1 < T_2,\ a.s.\ \pi.$$

Proof: If $\rho_1 < \rho_2$ then

$$\tilde{\rho_1} < \tilde{\rho_2}$$

where $\tilde{\rho_1} = \text{SLEM } (\tilde{P_1}), \tilde{\rho_2} = \text{SLEM } (\tilde{P_2})$.
If T_1,T_2 are coalescence times then:

$$||\delta^t_{(\hat{0},\hat{1})}\tilde{P_1}^{T_1} - \tilde{\pi}||_{TV} = 0;$$

$$||\delta^t_{(\hat{0},\hat{1})}\tilde{P_2}^{T_2} - \tilde{\pi}||_{TV} = 0.$$

From the first we obtain:

$$\frac{1}{2}\sum_{(x_i,x_j)\in E\times E} |(\delta^t_{(\hat{0},\hat{1})}\tilde{P_1}^{T_1})(x_i,x_j) - \tilde{\pi}(x_i,x_j)| = 0$$

But this is a sum of positive terms, so every terms must be equal to zero. In particular:

$$|(\delta^t_{(\hat{0},\hat{1})}\tilde{P_1}^{T_1})(\hat{0},\hat{1}) - \tilde{\pi}(\hat{0},\hat{1})| = 0. \tag{2}$$

Suppose that $T_1 \geq T_2$, then

$$||\delta^t_{(\hat{0},\hat{1})}\tilde{P_2}^{T_1} - \tilde{\pi}||_{TV} = 0$$

but from the remark at the beginning of the section:

$$\max_{(x_i,x_j)\in(E\times E)} ||\delta^t_{(x_i,x_j)}\tilde{P_2}^{T_1} - \tilde{\pi}||_{TV} = 0. \tag{3}$$

By combining (2) with (3) and using the lemmas of the previous section, we obtain:

$$0 = (2) - 2\cdot(3) \leq \tilde{\rho_1}^{T_1} - \tilde{\rho_2}^{T_1}.$$

We thus obtain:

$$\tilde{\rho_2} \leq \tilde{\rho_1},$$

which contradicts the hypothesis.
□

5 An ordering for coalescence time

Consider two transition matrices, P_1 and P_2, with the same stationary distribution, that are monotone with respect to some partial ordering defined on the state space. Let ρ_1, ρ_2 be the SLEMs and T_1, T_2 be the coalescence times of P_1 and P_2 respectively.

Definition 5.1 *We say that P_1 dominates P_2 in terms of* speed of convergence *and write $P_1 \geq_S P_2$, if*

$$\rho_1 \leq \rho_2.$$

This ordering was already defined in [12].

Definition 5.2 *We say that P_1 dominates P_2 in terms of* coalescence time *and write $P_1 \geq_C P_2$, if*

$$T_1 \leq T_2.$$

Theorem 5.1 *If $P_1 \geq_S P_2$ then $P_1 \geq_C P_2$.*

The result follows from Theorem 4.1 in Section 4.

6 Conclusions

We give a sufficient condition for selecting an updating rule in perfect simulation that is optimal in the sense of minimizing the coalescence time. The results holds for finite state spaces and monotone chains. We plan to investigate perfect sampling algorithm with continuous state spaces and non-monotone updating mechanism in further research.

References

[1] Aldous D. and Diaconis P. (2000) *Preliminary version of a book on finite Markov chains* available electronically at http://www.stat.berkeley.edu/users/aldous

[2] Bellman R. (1972) *Introduction to Matrix Analysis* McGraw-hill.

[3] Brémaud P. (1999) *Markov Chains,Gibbs Fields,Monte Carlo Simulation, and Queues* Springer.

[4] Diaconis P. (1988) *Group Representation in Probability and statistics* Institute of mathematical statistics Lecture notes-Monograph series

[5] Diaconis P. and Stroock D. (1991) *Geometric Bounds for eigenvalues of Markov Chains* Appl. Prob. 1 36-61

[6] Dimakos X. (1999) *A Guide to Exact Simulation* International Statistical Review vol. 69 No. 1 27-48

[7] Fill J. (1998) *An interruptible algorithm for perfect sampling via Markov chains*, Annals of Applied Probability, 7

[8] Frigessi A., Hwang C., and Younes L. (1992) *Optimal spectral structure of reversible stochastic matrices,Monte Carlo Methods and the simulation of Markov Random Fields* Ann. Appl. Prob. 2 610-628

[9] Frigessi A., Di Stefano P., Hwang C. and Sheu A. (1993) *Convergence Rates of the Gibbs Sampler,the Metropolis Algorithm and the other single site updating dynamics* J.R. Statist. Soc. Ser. B 55 205-219

[10] Geman S. and Geman D. (1984) *Stochastic relaxation,Gibbs distributions and the Bayesian restoration of images* IEEE Transactions on Pattern Analysis and Machine Intelligence 6 721-741

[11] Hastings W.K. (1970) *Monte Carlo Sampling Methods using Markov chains and their applications* Biometrika 57 97t09

[12] Mira A. (2001) *Ordering and Improving MCMC performances* Statistical Science,Vol 16 No. 4 340-350.

[13] Mira A. and Geyer C.J. (1999) *Ordering Monte Carlo Markov Chains* School of Statistics University of Minnesota Tech. Report 632.

[14] Peskun P.H. (1973) *Optimum Monte Carlo sampling using Markov Chains* Biometrika 60 607-612

[15] Propp J.G. and Wilson D.B. (1998) *Exact Sampling with Coupled Markov Chains and applications to Statistical Mechanics* Random Structures and Algorithms 9 223-252.

[16] Saloff-Coste L. (2004) *Total variation Lower bound for finite Markov chains:Wilson's Lemma* Random walks and geometry: proceedings of a workshop at the Erwin Schrodinger Institute, Vienna, June 18 - July 13, 2001, Walter de Gruyter

[17] Seneta E. (1981) *Nonnegative Matrices and Markov Chains* Springer

[18] Tierney L. (1998) *A Note on Metropolis Hastings Kernels for General state spaces* Ann. of App. Prob. 8 1-9

[19] Wilson D. (2004) *Mixing Times of lozenge tiling and card shuffling Markov chains* Ann. Appl. Probab. 14 (2004) 274-325