Electron. Commun. Probab. **19** (2014), no. 46, 1–6.
DOI: 10.1214/ECP.v19-3157
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Mixing under monotone censoring*

Jian Ding† Elchanan Mossel‡

Abstract

Consider critical percolation on a rhombus in the hexagonal lattice, and let A the event of a left to right crossing (by 1's). Suppose we sample configurations from A by starting from some fixed configuration in A, and then at each step a uniformly random hexagon is resampled as long as the resulting configuration is in A. It is easy to see that this sampling procedure converges to the uniform distribution on A. How long would it take? In this short note we will analyze the mixing properties of this chain and, more generally, initiate the study of mixing times of Markov chains under monotone censoring. A number of open problems are presented.

Keywords: Mixing time; testing monotonicity.
AMS MSC 2010: 60J10.
Submitted to ECP on November 25, 2013, final version accepted on May 12, 2014.

1 Introduction

1.1 A motivating example

Consider critical percolation on a rhombus in the hexagonal lattice. Formally this is given by the probability space $\{0,1\}^{H_n}$ with the uniform distribution, where we denote by H_n the hexagons in the hexagonal lattice in the rhombus. It is trivial to sample a configuration from this model by sampling each hexagon independently. Let A be the event of a left to right crossing (by 1's). It is well known, by duality, that $\mathbb{P}[A] = 0.5$. Suppose we want to sample a configuration of A. One natural way to do so is by rejection sampling: sampling a random configuration and accepting it if and only if it is in A. A different natural way to sample is to start with a particular left to right crossing configuration and then repeatedly re-sample hexagons uniformly at random as long as the resulting configuration is in A. It is not hard to see that the second procedure will also converge to the uniform distribution on A. However, how long would it take to converge?

1.2 Monotone sampling in $\{0,1\}^n$

We will study a more general question. Consider the partial order on $\{0,1\}^n$ where $x \leqslant y$ if and only if $x_i \leqslant y_i$ for all $i \in [1,n]$. We say a set $A \subset \{0,1\}^n$ is monotone if $x \in A$ and $x \leqslant y$ imply that $y \in A$. For a monotone set A and $x_0 \in A$, let $M_A^{x_0}$ denote the following Markov chain started at x_0.

*Support: JD: NSF DMS-1313596; EM: NSF DMS-1106999, CCF 1320105, DOD ONR N000141110140.
†University of Chicago, USA. E-mail: jianding@galton.uchicago.edu
‡UC Berkeley, USA. E-mail: mossel@stat.berkeley.edu

- Given the current state x, pick a coordinate i uniformly at random and re-randomize x_i to obtain y.
- Let the next state of the chain be y if $y \in A$. Otherwise let it be x.

The monotonicity of A implies that the chain is irreducible, and therefore the chain converges to the uniform distribution on A. We aim to analyze the mixing time for the chain $M_A^{x_0}$. We recall that the mixing time for a Markov chain $X = (X_t)_{t=0,1,\ldots}$ on a countable space Ω with stationary distribution π is defined to be

$$\tau_{mix}(X) = \max_{x\in\Omega} \min\{t : \|\mathbb{P}(X_t \in \cdot \mid X_0 = x) - \pi(\cdot)\|_{\mathrm{TV}} \leqslant 1/4\}\,,$$

where $\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|$ is the total variation distance between two probability measures μ and ν (on space Ω). For more background on mixing times, see, e.g., [1, 9].

In order to bound the mixing time, we will use a standard geometric bound on the mixing time given by the *conductance* of the underlying graph for a Markov chain. Given a graph $G = G(V, E)$, the conductance $\phi(G)$ is defined to be

$$\phi(G) = \min_{S\subseteq V:|\mathrm{vol}(S)|\leqslant|E|} \Phi(S)\,, \text{where} \Phi(S) \triangleq \frac{|\partial_E S|}{\mathrm{vol}(S)}\,, \tag{1.1}$$

where $\mathrm{vol}(S)$ is the sum of degrees over vertices in S and $\partial_E(S) = \{(x, y) \in E : x \in S, y \notin S\}$ denotes the edge boundary set of S. For a subset A of the hypercube, we let G_A be the graph corresponding to the Markov chain $M_A^{\cdot}$. More precisely, G_A is the induced subgraph of the hypercube $\{0, 1\}^n$ restricted to A with a self-loop added to vertex x for each edge $(x, y) \in E \cap A \times A^c$ (we use the convention that each self-loop is counted as degree 1).

In what follows, we denote by $\mathbb{P}$ the uniform probability measure on $\{0, 1\}^n$.

Theorem 1.1. *For any monotone set $A \subset \{0, 1\}^n$, we have*

$$\phi(G_A) \geqslant \frac{\mathbb{P}[A]}{16n}\,.$$

Combined with standard results in the theory of Markov chains [8, Theorem 2.1] and [12, Lemma 3.3] (see also [9, Theorem 13.14]), Theorem 1.1 yields the following corollary on the mixing time of M_A.

Corollary 1.2. *For any monotone set $A \subset \{0, 1\}^n$, the mixing time for the chain M_A satisfies*

$$\tau_{mix}(M_A) \leqslant 2(\tfrac{16n}{\mathbb{P}[A]})^2 \log(4 \cdot 2^n \mathbb{P}(A))\,.$$

Note that this implies that the mixing time is polynomial in n as long as A is large (of measure at least inverse polynomial in n). In particular, our motivating example of sampling a critical percolation configuration with a left to right crossing has mixing time at most $O(n^3)$. Our conductance bound in Theorem 1.1 is tight up to polynomial factors in n as the following example shows:

Example 1.3. *Assume $n \geqslant 2m$ and let $A = \{x : x_1 = x_2 \ldots = x_m = 1\} \cup \{x : x_{m+1} = \ldots = x_{2m} = 1\}$. Clearly A is monotone and $\mathbb{P}[A] = 2^{-m+1} - 2^{-2m}$. Considering $\Phi(B)$ for $B = \{x : x_1 = x_2 \ldots = x_m = 1\} \subset A$, we see that $\phi(G_A) \leqslant 2^{-m}$. Similarly starting from the point $x = (x_i)$ for $x_i = \mathbf{1}_{i\in A}$ it is easy to see that the mixing time is lower bounded by the time to hit $y = (y_i)$ for $y_i = \mathbf{1}_{i\in A\cup B}$ with probability at least $1/4$, which is lower bounded by 2^{m-4}.*

1.3 Property Testing

Our proof uses a new ingredient in the context of mixing of Markov chain, a result from the theory of *property testing*. Property testing, explicitly defined in [11], plays a central role in probabilistically checkable proofs. However, it was extended and extensively studied on its own right for checking properties such as graph properties with fascinating connections to many areas of combinatorics including, in particular, regularity lemmas. One concrete question on this topic is to test the monotonicity of a boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ (this is equivalent to testing the monotonicity of a subset of a hypercube). A natural algorithm for testing monotonicity is to sample a number of random neighboring pairs (x,y) in the hypercube, and to reject the monotonicity if and only if $f(x) > f(y)$ while $x < y$ for at least one of the sampled pairs. It was proved in [6] that this natural algorithm works well with high probability. The key to the success of this algorithm, which is also the key to our proof of Theorem 1.1, is the following structural theorem on approximately monotone sets.

Theorem 1.4. *[6, Theorem 2] For any set $S \subset \{0,1\}^n$, define*

$$\delta(S) = (n2^n)^{-1}|\{(x,y) \in \{0,1\}^n \times \{0,1\}^n : |x-y| = 1, x \leqslant y, x \in S, y \notin S\}|\,,$$
$$\varepsilon(S) = \min\{\mathbb{P}(S \oplus A) : A \text{ is monotone }\}\,.$$

where $\oplus$ denotes the symmetric difference of two sets. Then we have $\delta(S) \geqslant \varepsilon(S)/n$.

1.4 Proof of Theorem 1.1

Recall that $\mathbb{P}$ is the uniform measure on $\{0,1\}^n$. In light of definition (1.1), it suffices to prove that

$$\Phi(B) \geqslant \frac{\mathbb{P}(A)}{16n} \text{ for all } B \subset A \text{ such that } \mathbb{P}(B) \leqslant \mathbb{P}(A)/2\,. \tag{1.2}$$

It is clear that (1.2) holds if $\mathbb{P}(A)\mathbb{P}(B) < 8 \cdot 2^{-n}$, since in this case, by connectivity of G_A, we have that

$$\Phi(B) \geqslant \frac{1}{\mathrm{vol}(B)} = \frac{1}{\mathbb{P}(B)n2^n} \geqslant \frac{\mathbb{P}(A)}{8n}\,.$$

It remains to consider the case when $\mathbb{P}(A)\mathbb{P}(B) \geqslant 8 \cdot 2^{-n}$. Denote by $C = A \setminus B$ and by Ω the collection of monotone sets in the hypercube $\{0,1\}^n$. We claim that

$$\textbf{either } \mathbb{P}(B \oplus F) \geqslant \frac{\mathbb{P}(A)\mathbb{P}(B)}{16},\ \text{for all } F \in \Omega\,, \textbf{or } \mathbb{P}(C \oplus F) \geqslant \frac{\mathbb{P}(A)\mathbb{P}(B)}{16}\,,\ \text{for all } F \in \Omega\,. \tag{1.3}$$

Otherwise, there exist monotone sets B' and C' such that

$$\mathbb{P}(B \oplus B') < \frac{\mathbb{P}(A)\mathbb{P}(B)}{16} \text{ and } \mathbb{P}(C \oplus C') < \frac{\mathbb{P}(A)\mathbb{P}(B)}{16}\,. \tag{1.4}$$

In particular, we have $\mathbb{P}(B') \geqslant \mathbb{P}(B)/2$ and $\mathbb{P}(C') \geqslant \mathbb{P}(C) - \frac{\mathbb{P}(A)\mathbb{P}(B)}{16} \geqslant \frac{\mathbb{P}(A)}{2} - \frac{\mathbb{P}(A)}{16} = \frac{7}{16}\mathbb{P}(A)$. An application of FKG inequality [5] gives that

$$\mathbb{P}(B' \cap C') \geqslant \mathbb{P}(B') \cdot \mathbb{P}(C') \geqslant \frac{7}{32}\mathbb{P}(A)\mathbb{P}(B)\,.$$

Combined with (1.4), it follows that

$$\mathbb{P}(B \cap C) \geqslant \mathbb{P}(B' \cap C') - \mathbb{P}(B \oplus B') - \mathbb{P}(C \oplus C') \geqslant \frac{7}{32}\mathbb{P}(A)\mathbb{P}(B) - 2\frac{1}{16}\mathbb{P}(A)\mathbb{P}(B) > 0\,,$$

contradicting with the fact that $B \cap C = \emptyset$. Thus, we completed verification of (1.3).

Without loss of generality we assume now that $\mathbb{P}(B \oplus F) \geqslant \frac{\mathbb{P}(A)\mathbb{P}(B)}{16}$, for all $F \in \Omega$ (if the same holds for C, we just apply the following analysis to C in the same manner, with the observation that $\partial_E B = \partial_E C$). By Theorem 1.4, we get that

$$|\Psi(B)| \geqslant \frac{2^n \mathbb{P}(A)\mathbb{P}(B)}{16}$$

where

$$\Psi(B) \triangleq \{(x,y) \in \{0,1\}^n \times \{0,1\}^n : |x-y| = 1, x \leqslant y, x \in B, y \notin B\}.$$

For $(x,y) \in \Psi(B)$, we have $x \in B$ and $x \leqslant y$, and thus $y \in A$ since A is a monotone set. Therefore, we get $(x,y) \in \partial_E B$, yielding that $\Psi(B) \subseteq \partial_E B$. This implies that $|\partial_E B| \geqslant \frac{2^n \mathbb{P}(A)\mathbb{P}(B)}{16}$. Combined with the fact that $\mathrm{vol}(B) = n2^n \mathbb{P}(B)$, it completes the proof of (1.2) and thus the proof of the theorem.

1.5 Discussions and open problems

It seems plausible that the bound on the mixing time obtained in Corollary 1.2 is not sharp. A case of particular interest is when $\mathbb{P}(A) \geqslant 1/2$. Indeed, we ask the following open question.

Question 1.5. *Suppose that there exists a constant $c > 0$ such that a monotone subset $A \subset \{0,1\}^n$ has measure $\mathbb{P}(A) \geqslant c$. Is it true that $\tau_{mix}(M_A) \leqslant Cn \log n$, where $C > 0$ is a constant depending only on c?*

In a different direction, our results suggest testing non-product measures. For example, suppose we wish to reproduce Theorem 1.1 for the Ising model on some graph G, where we denote by μ the stationary measure. For this to work, we will need an analogue of the testing result. In this setup it is natural to define for a set $S \subset \{0,1\}^n$ (identifying 0 with $-$ and 1 with $+$)

$$\delta(S) = \sum_{(x,y) \in \Psi(S)} \frac{\mu(x)}{n}$$

where

$$\Psi(S) = \{(x,y) \in \{0,1\}^n \times \{0,1\}^n : |x-y| = 1, x \leqslant y, x \in S, y \notin S\},$$

and

$$\varepsilon(S) = \min\{\mu(S \oplus A) : A \text{ is monotone}\}.$$

We then ask

Question 1.6. *Consider the ferromagnetic Ising model on a graph $G = (V,E)$. Under what assumptions is it the case that $\delta(S) \geqslant (\varepsilon(S)/n)^a$ for all $S \subset \{0,1\}^n$ and a fixed constant $a > 0$?*

The following example suggests that some assumptions are needed. Consider the Curie-Weiss model (Ising model on the complete graph) at low temperature (so the stationary measure admits double wells, see [2, 3]) with n sites. For convenience, suppose that n is odd. Let $A = \{x : \sum_{i=1}^n x_i \leqslant \frac{n-1}{2}\}$. We claim that $\varepsilon(A) \geqslant 1/6$. In order to see this, let $A_k = \{x : \sum_{i=1}^n x_i = k\}$ and $A'_k = \{x : \sum_{i=1}^n x_i = n-k\}$ for $k \leqslant \frac{n-1}{2}$. For $x \in A_k$ and $y \in A^c$, define

$$a(x,y) = \frac{\mathbf{1}_{y \in A'_k, y \geqslant x} \mu(x)}{|\{z : z \in A'_k, z \geqslant x\}|}.$$

Thus, $\sum_{y \in A^c} a(x,y) = \mu(x)$ for all $x \in A$. In addition, by symmetry we see that

$$a(x,y) = \frac{\mathbf{1}_{y \in A'_k, y \geqslant x} \mu(y)}{|\{z : z \in A_k, y \geqslant z\}|}.$$

Thus for every $y \in A^c$ we have

$$\sum_{x\in A} a(x,y) = \mu(y)$$

(so $a(\cdot,\cdot)$ is a mass transportation from A to A^c with respect to measure μ). Therefore, for any monontone set B we have

$$\mu(B\cap A^c) = \sum_{x\in A}\sum_{y\in B\cap A^c} a(x,y) \geqslant \sum_{x\in B\cap A}\sum_{y\in A^c} a(x,y)$$

where the last inequality follows from the fact that $\mathbf{1}_{y\in B}a(x,y) \geqslant \mathbf{1}_{x\in B}a(x,y)$ for all $x\in A$ and $y \in A^c$, due to the monotonicity of B. Thus, we obtain

$$\mu(B\cap A^c) \geqslant \sum_{x\in B\cap A} \mu(x) = \mu(A\cap B)\,.$$

This implies that $\mu(B\cap A^c) \geqslant \mu(B)/2$. Combined with the simple fact that $\mu(A) = 1/2$, it follows that

$$\mu(A\oplus B) \geqslant \max(\mu(A)-\mu(B), \mu(B)/2) \geqslant 1/6\,,$$

as desired. However, it is clear that

$$\delta(A) \leqslant \mu(A_{(n-1)/2})\,,$$

which is exponentially small in n at low temperature [2, 3].

Finally, we note that the effect of censoring on mixing times was studied in [10], where it was shown that the mixing can only be delayed for Glauber dynamics on monotone spin systems by censoring some updates (the censoring is prescribed without information on what is the proposed update). In [7], an example was given to demonstrate that censoring can indeed speed up the mixing for proper coloring. This question was then studied in [4] in much more general settings, which introduced a certain partial order on the class of stochastically monotone Markov kernels and proved that the monotonicity of Markov chains implies monotonicity of mixing times. These results are different from ours in at least the following two senses: (1) They focus on Markov chains with the same stationary measure while the censoring considered here changes the state space of the Markov chain; (2) They aim at qualitative results which ensure monotonicity for mixing times of Markov chains under consideration, while ours aims to give a quantitative bound on the mixing time for the censored Markov chain.

Acknowledgement. We than two anonymous referees for helpful suggestions that lead to a significant improvement on exposition.

References

[1] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. In preparation, available at `http://www.stat.berkeley.edu/ aldous/RWG/book.html`.

[2] R. S. Ellis and C. M. Newman. Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Z. Wahrsch. Verw. Gebiete*, 44(2):117–139, 1978. MR-0503333

[3] R. S. Ellis, C. M. Newman, and J. S. Rosen. Limit theorems for sums of dependent random variables occurring in statistical mechanics. II. Conditioning, multiple phases, and metastability. *Z. Wahrsch. Verw. Gebiete*, 51(2):153–169, 1980. MR-0566313

[4] J. Fill and J. Kahn. Comparison inequalities and fastest-mixing Markov chains. *Ann. Appl. Probab.*, 23(5):1778–1816, 2013. MR-3114917

[5] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971. MR-0309498

[6] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samorodnitsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000. MR-1774842

[7] A. E. Holroyd. Some circumstances where extra updates can delay mixing. *J. Stat. Phys.*, 145(6):1649–1652, 2011. MR-2863724

[8] G. F. Lawler and A. D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988. MR-0930082

[9] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson. MR-2466937

[10] Y. Peres and P. Winkler. Can Extra Updates Delay Mixing? *Comm. Math. Phys.*, 323(3):1007–1016, 2013. MR-3106501

[11] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. MR-1379300

[12] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989. MR-1003059