Mixing Times of the Schelling Segregation Model and Biased Permutations

Prateek Bhakta (Advisor: Dana Randall)

Markov chains are fundamental tools used throughout the sciences and engineering; the design and analysis of Markov Chains has been a focus of theoretical computer science for the last 20 years. A Markov Chain takes a random walk in a large state space Ω, converging to a target stationary distribution π over Ω. The number of steps needed for the random walk to have distribution close to π is the *mixing time*. For the Markov Chain to be useful, the mixing time should be polynomial in the input to the problem, typically logarithmic in $|\Omega|$. Such a chain is called *rapidly mixing*.

I am interested in applying Markov chain techniques to applied problems, and have been working on two primary problems: the Schelling Segregation Model, a well studied model in economics, and a biased permutation problem arising from the Move Ahead One list update paradigm in caching. In both cases, we leverage intuition gained from studying models in statistical physics.

1 Mixing of the Schelling Segregation Model

The Schelling Segregation Model was proposed by Nobel Prize winning economist Thomas Schelling as a means to understand causes of racial segregation.[9] He empirically showed that a local preference that one's neighbors be of the same "type" as oneself can lead to *segregation*, the global presence of large connected neighborhoods predominately of one type. We analyze generalizations of this model in the Markov Chain framework, with the goal of rigorously establishing bounds on the mixing time at various parameters.

More formally, we consider a tiling model on the n x n grid G where there are three tile types, A,B, and 0 for empty. We say that each non-empty tile is *influenced* by tiles within distance r, and each tile prefers that influencing tiles be of the same type as itself. Although Schelling proposed using dynamics based on a binary threshold function to determine if a particular person was "happy" or "unhappy", we feel that it is more natural to analyze a geometric bias function, where a person can be more or less happy depending on the number of similar neighbors. In our model, we are given parameters $\lambda \geq \mu \geq 1$, and the weight π of a particular board configuration σ is $\pi(\sigma) = \lambda^{\#\text{A-A edges} + \#\text{B-B edges}} \cdot \mu^{\#\text{AB edges}}$. More simply, each matched $A-A$ or $B-B$ influence edge contributes λ to the weight of the configuration, and each $A-B$ influence contributes μ. This scheme corresponds to Non-Saturated Ising model on the graph G^r, where each node is connected to all nodes within distance r on the grid G.

The Markov chain M_r starts at any initial board configuration σ_0. It then, for each time t, iteratively chooses a tile and a color $c \in \{A, B, 0\}$ to potentially replace this tile with. We call this proposed state τ. Finally, it makes this transition according to the metropolis distribution, that is $\sigma_{t+1} = \tau$ with probability $\min(1, \frac{\pi(\tau)}{\pi(\sigma_t)}$, and $\sigma_{t+1} = \sigma_t$ otherwise. It is known that for the $r = 1$ case, there are constants $c < d$ for which the chain is fast when $\mu \leq \lambda \leq c$, and the chain is slow when $\lambda/\mu > d$.[5] In our model, we note that each tile is influenced by $O(r^2)$ tiles instead of only the 4 nearest neighbors, and we are led to the following conjecture.

Conjecture: *The Markov Chain M_r is rapidly mixing for all $\mu < \lambda \leq 1 + c/r^2$ and is slowly mixing for all $\lambda/\mu \geq 1 + d/r^2$ for some constants c and d.*

1.1 Previous work: The related Ising model of statistical physics has been well studied in the Markov Chain framework. [7][8]. The case where $r = 1$, the Non-Saturated Ising model on the grid, was analyzed by Greenburg and Randall.[5] They show that for constants $c < d$, the chain is fast for $\lambda < c$ using a coupling argument, and that the chain is slow for $\lambda > d$ by showing that a cut exists in Ω with exponentially small weight. The first real progress on the Schelling model itself was done very recently by Brandt et al.[3]. Their analysis provides some of the only rigorous analysis of the model, but only in the one-dimensional setting.

1.2 Our Progress: My advisor, Dana Randall, and I have recently made a great deal of progress towards resolving our conjecture and its equivalent on related models. In particular, we make the first real progress towards the analysis of the 2-dimensional Schelling model. First, we extend the results of Greenburg and Randall, using a new fault line argument to exhibit an exponentially small cut in the state space. The following outlines our intended approach.

Claim: The Markov Chain M_r is rapidly mixing for all $\mu \leq \lambda \leq 1 + c/r^2$ and is slowly mixing for all $\lambda/\mu \geq 1 + d \log r/r^2$ for some constants c and d.

We define a fault component to be a maximally connected set of edges between neighboring *nonmatched* tiles or between empty tiles. A *fault line* is a fault component that spans either from the top to the bottom of the grid, or from the left to the right. Let Ω be the set of nxn board configurations, and $\Omega_F \subset \Omega$ be the subset that contains a fault line. It is known that Ω_F forms a cut in the state space Ω, and our goal is to show that this set has exponentially small weight. The existence of a small cut in Ω places a bound on the conductance Φ, which is a well known measure of the mixing time.[6]

We introduce the concept of an *extended* fault line, which is a maximal collection of fault components beginning with a fault line that iteratively adds nearby fault components within distance r. We believe that we can create a mapping $f : \Omega_F \to \Omega$ that reverses all the colors in fault components of the input σ to obtain a configuration τ with exponentially more weight. We can also bound the number of preimages of any $\tau \in \Omega$ by encoding the extended fault line in each mapping as a depth-first search path. This can be used to show that this cut will be exponentially small when $\lambda/\mu > 1 + \frac{d \log r}{r^2}$ for some d. Finally, we believe that we can remove the $\log r$ term by careful pruning of fault components that are expensive to encode.

1.3 Exponentially Decaying Influence Model: We also show similar progress for the variant of the Schelling model where the influence between two tiles decays exponentially as a function of the distance between them. This can be viewed as a more realistic model for influence from distant neighbors.

We introduce a parameter $\alpha < 1$, and let the influence between two tiles at distance k to be α^{k-1}. As before, if two tiles match, they contribute $\lambda^{\alpha^{k-1}}$ to the weight, and if they mis-match, they contribute $\mu^{\alpha^{k-1}}$. The weight of a board configuration is again the product of all influences. An important quantity is the sum S of all influences on a particular tile, which is bounded by $\frac{4}{(1-\alpha)^2}$. We make the following conjecture, based on intuition gained from the analysis of M_r.

Conjecture: *The Markov Chain M_r is rapidly mixing for all $\mu < \lambda \leq 1 + c/S$ and is slowly mixing for all $\lambda/\mu \geq 1 + d/S$ for some constants c and d.*

We have also recently made similar progress to resolving this conjecture.

Claim: The Markov Chain M_α is rapidly mixing for all $\mu \leq \lambda \leq 1 + c/S$ and is slowly mixing for all $\lambda/\mu \geq 1 + d \log(-\log \alpha)/S$ for some constants c and d.

Our main argument chooses $r = f(\alpha)$ and extends the analysis of M_r to obtain these bounds. In this case, we believe that we can remove the $\log(-\log \alpha)$ term by only including fault components that have large gain relative to the cost to encode them.

2 The Mixing Time of Biased Permutations

A very different setting I am currently considering is the mixing of various Markov Chains in the context of biased permutations. I considered the following Markov chain on the space Ω of permutations of $\{1, \ldots, n\}$.

Given $p_{i,j}$ for $i, j \in \{1, \ldots, n\}$ for $i \neq j$ such that for $i < j$, $p_{i,j} \geq \frac{1}{2}$ and $p_{i,j} + p_{j,i} = 1$, the Markov chain M_{SWP} starts at any initial permutation σ, then iteratively chooses a position $i \in 1, \ldots, n-1$ uniformly, and swaps the elements $\sigma(i), \sigma(i+1)$ with probability $p_{\sigma(i+1),\sigma(i)}$; else it does nothing. This very simple Markov chain has defied analysis for many years, and was only known to mix rapidly for constant $p_{i,j} = c$[1] or the case where $p_{i,j} = 1/2$ *or* 1 [4]. Ten years ago, Jim Fill first studied this model under a convexity condition that requires $p_{i,j} \leq p_{i,j+1}$ for $1 \leq i < j \leq n-1$ and $p_{i,j} \leq p_{i-1,j}$ for $2 \leq i < j \leq n$.

Conjecture: *M_{SWP} is rapidly mixing with the convexity condition.*

We've recently done exiting work to prove that the chain is rapidly mixing for two broad classes of $p_{i,j}$, and disproved the conjecture that M_{SWP} is rapidly mixing when all $p_{i,j} \geq 1/2$ for $i < j$. This work was recently submitted for publication[2]. I hope to extend these results to the more general setting assuming only the convexity condition on the $p_{i,j}$. One strong approach under consideration involves showing that the conductance of any subset of Ω cannot be too small by iteratively modifying one value of $p_{i,j}$ at a time, preserving convexity at each step.

References

[1] Itai Benjamini, Noam Berger, Christopher Hoffman, and Elchanan Mossel. Mixing times of the biased card shuffling and the asymmetric exclusion process. *Trans. Amer. Math. Soc*, 2005.

[2] Prateek Bhakta, Sarah Miracle, Dana Randall, and Amanda Pascoe Streib. Mixing times of self-organizing lists and biased permutations. *CoRR*, 2012.

[3] Christina Brandt, Nicole Immorlica, Gautam Kamath, and Robert D. Kleinberg. An analysis of one-dimensional schelling segregation. *CoRR*, 2012.

[4] Russ Bubley and Martin Dyer. Faster random generation of linear extensions. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, 1998.

[5] Sam Greenberg and Dana Randall. Convergence rates of markov chains for some self-assembly and non-saturated ising models. *Theoretical Computer Science*, 2009.

[6] Mark Jerrum and Alastair Sinclair. Approximate counting, uniform generation and rapidly mixing markov chains. In *Information and Computation*, 1989.

[7] Mark Jerrum and Alastair Sinclair. Poly-time approximation algorithms for the ising model. In *Society for Industrial and Applied Mathematics*, 1993.

[8] Dana Randall and David Wilson. Sampling spin configurations of an ising system. In *Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '99, 1999.

[9] Thomas Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1971.