CS294 Markov Chain Monte Carlo: Foundations & Applications	Fall 2009

Lecture 8: October 1

Lecturer: Prof. Alistair Sinclair — *Scribes: Alistair Sinclair*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

8.1 Lattice Tilings

Consider a simply connected region S of the two-dimensional Cartesian lattice (e.g. a $2n \times 2n$ chessboard). A *tiling* of S is a covering of all its squares by non-overlapping dominoes, each of which occupies two adjacent squares of the lattice. We want to generate u.a.r. a tiling of S. Such problems arise in statistical physics, where the tilings correspond to configurations of a so-called *dimer system* on S. Various physical properties of the system are related to the expected value, over the uniform distribution, of some function defined over configurations, such as the number of horizontal dominoes or the correlation between the orientation of dominoes at two given squares. The Markov chain most commonly used to generate u.a.r. a tiling of S picks a 2×2 square in S u.a.r. and rotates it if it contains exactly two dominoes. (See, e.g., Figure 8.1.) Thurston [Th90] proved that the set of all tilings is connected by such moves (and we shall see this below).

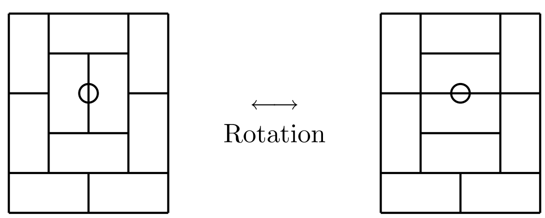

Figure 8.1: Domino Tilings

8.2 Lozenge Tilings

For the purpose of this lecture, we will analyze the mixing time of an analogous Markov chain for *lozenge* tilings on the triangular lattice as shown in Figure 8.2. (A *lozenge* is the analogue of a domino in the Cartesian lattice). This analysis is due to Luby, Randall and Sinclair [LRS01], and extends with a little more work to the Cartesian lattice. We use the following Markov chain to sample lozenge tilings of a simply connected region S of the triangular lattice.

- Pick a point in the interior of region S u.a.r.
- If the hexagon surrounding the chosen point contains exactly three lozenges, rotate it by $60°$ (as shown in Figure 8.2), else do nothing.

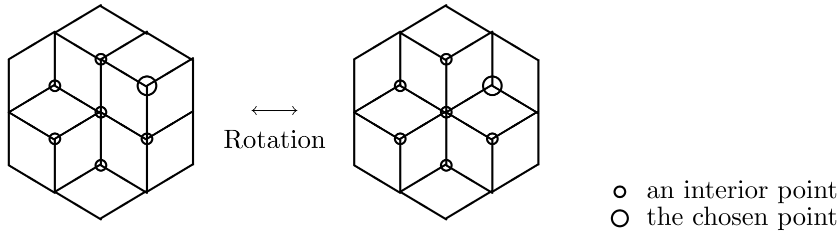

Figure 8.2: Lozenge Tilings

It turns out that it is easier to analyze the above process when viewed over so-called *routings*, which are defined below.

8.2.1 Correspondence With Routings

The lozenge tilings of a region of the triangular lattice are in 1-1 correspondence with *routings* on an associated Cartesian lattice. Given a simply connected region S of the triangular lattice, we can define a corresponding region S' of the Cartesian lattice as follows. There is a vertex in region S' corresponding to the midpoint of a vertical edge in region S. Two vertices in region S' are connected iff the corresponding points in region S lie on adjacent triangles. The vertices in S' that correspond to the vertical edges on the left (or right) sided boundary of S are called sources (or sinks). If a lozenge tiling exists for S, then the number of sources in S' is equal to the number of sinks. For example, in Figure 8.3, there are two sources s_1, s_2 and two sinks t_1, t_2. A *routing* in S' is a set of non-intersecting shortest paths from each source to the corresponding sink. Note that every routing contains the same number of points of S'; indeed, the path for each source-sink pair contains the same number of points in every routing. We call the interior points (i.e., those that are not sources or sinks) the *points along* the routing.

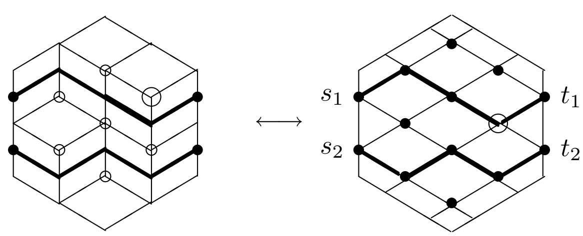

Figure 8.3: Lozenge Tiling ≡ Routing

What does the above Markov chain look like in the routings world? Well, rotation of a hexagon containing three lozenges in a tiling corresponds to flipping a "valley" to a "peak" or vice-versa in the corresponding routing (see Figure 8.3). This leads us to the following description of a Markov chain on routings:

- Pick a point $p \in \{1, \ldots, n\}$ u.a.r. along the routing, and a direction $d \in \{\uparrow, \downarrow\}$ u.a.r.
- Move point p in direction d if possible.

Note that if the chosen point p is not a peak or a valley, it will not move in either direction d. Moreover, not all peaks and valleys can move: some are "blocked" because of the constraint that the paths be non-intersecting.

The above Markov chain is aperiodic since there is at most one direction in which a point along a routing can be moved, and hence choosing direction d u.a.r. implies a self-loop probability of at least $\frac{1}{2}$ at every state. The Markov chain is irreducible since it is possible to get from any routing to the maximum routing (i.e., the routing in which every path is "as high as possible") by iteratively converting valleys into peaks by rotation, processing the paths top-to-bottom. Similarly, it is possible to go from the maximum routing to any other routing as the Markov chain is symmetric. The Markov chain thus converges to the uniform distribution over the set of all routings (or equivalently tilings) Ω.

We will use a coupling argument to bound the mixing time of this Markov chain. We will exploit the fact that the chain admits a *complete* coupling that is *monotone* w.r.t. a natural partial order on Ω, as defined in the previous lecture.

The partial order is the following.

Definition 8.1 *For two routings X and Y, we say that $X \preceq Y$ iff X does not lie above Y at any corresponding point along their paths.*

Thus the unique maximum routing $\top$ is the one in which all paths are "as high as possible", and the minimum routing $\perp$ is the one in which all paths are "as low as possible."

The complete coupling is defined by the following distribution over functions $f : \Omega \to \Omega$. The function f is specified by choosing a point p and a direction d, both u.a.r. Then, for any routing $X \in \Omega$, the image $f(X)$ is obtained by moving point p in direction d if possible. It is clear that this distribution is consistent with the Markov chain, so we have a valid complete coupling.

It is easy to verify [**exercise!**] that this coupling is monotone w.r.t. $\preceq$. As discussed in the previous lecture, this implies that the coupling time T_{xy} for any pair of initial states x, y is bounded by $T_{\top\perp}$, the coupling time for the two extremal states $\top$ and $\perp$. This gives us an *algorithmic* method for estimating the mixing time, even if we have no analytical bound: namely, simulate the coupled evolution starting from the states $\top$ and $\perp$ and observe the time for them to meet.

However, we now go further and show how to analyze the coupling time rigorously. Note that monotonicity also simplifies this analysis: first, we need only consider initial states $X_0 = \top$ and $Y_0 = \perp$; second, we can assume that at all future times the two copies satisfy $X_t \succeq Y_t$.

8.2.2 Coupling Analysis of Single Path Routings

As described earlier, the coupling between two routings X and Y is the following. We choose the same point p and direction d in both X and Y. Then we make the corresponding move in both (if possible). For the purpose of simplicity, we will initially restrict attention to routings in which there is only one path: this is simpler since no moves are ever blocked by the constraint that the paths be non-intersecting. By monotonicity, we may assume that one path always lies above the other; see Figure 8.4. We define the *distance* between the two paths as the area enclosed between them.

Now consider the possible moves under the coupling. We can mark some points on the two paths as either good (G) or bad (B). A point is marked G if by selecting that point and making the corresponding move, we decrease the area between the two paths (by one unit). It is marked B if the corresponding move increases the area (by one unit). The following claim is not hard to prove and is left as an exercise.

Claim 8.2 *$\#G \geq \#B$.*

This claim implies that the expected change in distance is non-positive. Also, it is easy to see that when the

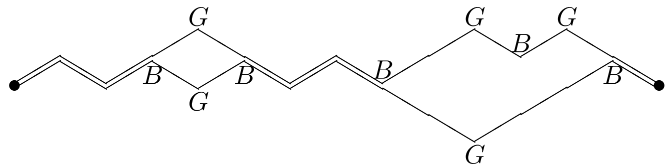

Figure 8.4: Distance = area between the two paths = 6

paths differ there is always at least one good move. Hence, the probability of a change in the distance is at least $\frac{1}{2N}$, where N is the number of points along a routing. Taking as a natural measure of problem size the area n of S, we have clearly that $N \leq n$, and hence the coupling time is stochastically dominated by the hitting time to 0 of a random walk on the integers $[0, n]$ which has zero drift and probability at least $\frac{1}{2n}$ of making a move. Thus the mixing time of our process is $O(n^3)$.

8.2.3 Coupling Analysis of Multi-Path Routings

Let us now extend the above analysis to routings with multiple paths. The difficulty here lies in taking care of blocked moves: because some moves are blocked, we can no longer argue that $\#G \geq \#B$. We shall get around this with a trick: we shall modify the Markov chain so that no moves are blocked! I.e., whenever the chosen point p is a peak or a valley, we will always make a move. To describe this, note that any peak (valley) defines a *tower* of some height $h \geq 1$, which is the maximal contiguous sequence of peaks (valleys) above (below) it. The tower of an unblocked peak/valley has height 1; if it is blocked the height will be greater than 1. Note that a tower of any height can be rotated in an analogous fashion to the simple rotations (of towers of height 1) considered earlier; see Figure 8.5.

Here is the modified Markov chain, with tower moves:

- Pick a point p and direction d as before.
- Rotate the tower at p in direction d if possible (i.e., if p is a peak and $d = \downarrow$, or p is a valley and $d = \uparrow$) with probability $1/h$, where h is the height of the tower.

Note that we rotate a tower only with probability inversely proportional to its height. This is because rotating a tower of height h changes the distance by $\pm h$.

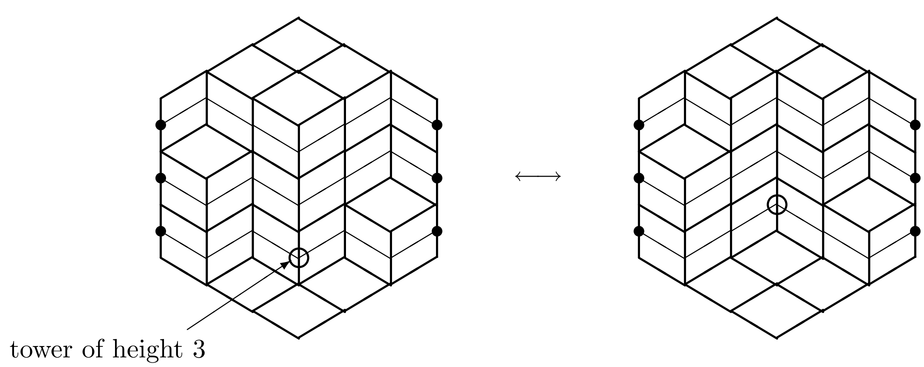

Figure 8.5: Tower Rotations

It is easy to check that the same complete coupling f as before (modified to allow tower moves) is still monotone w.r.t. $\preceq$. Hence, we can work with the coupling time for the min ($\perp$) and max ($\top$) routings, and can assume that the two routings at any future time lie one above the other. We define the *distance* between any two routings as the sum of the areas between their corresponding paths; it is not hard to check [**exercise**] that the maximum possible distance (i.e., that between $\perp$ and $\top$) is $O(n^{3/2})$, where again n is the area of S.

As before, identify a peak or valley as good (G) or bad (B) according to whether it decreases or increases the distance when it moves. Let us mark all bad nodes with a tower of height h above/below them as B_h, and all good nodes with a tower of height h above/below them as G_h. We can now analyze the expected change in distance as follows:

$$\begin{aligned} E[\Delta(Distance)] &= \sum_h \frac{1}{2Nh}(\#B_h - \#G_h)h \\ &= (\frac{1}{2N})\sum_h(\#B_h - \#G_h) \\ &\leq 0, \end{aligned}$$

where $N \leq n$ is the total number of points along a routing. In the first line here, we are using the facts that a tower of height h is rotated with probability $1/2Nh$, and that the resulting change in area is $\pm h$. In the last line we are using the combinatorial fact seen earlier that the number of good moves is at least as large as the number of bad moves; this property is inherited from the single-path case as no moves are blocked.

Thus, this process is stochastically dominated by the hitting time to 0 of a random walk on the integers $[0, n^{1.5}]$ which has zero drift, and whose variance in one step is at least $\frac{1}{2n}$.[1] Hence the mixing time is $O(n^4)$.

Note: Wilson [W04] subsequently obtained a tight bound of $O(n^3 \log n)$ for the mixing time on a hexagonal region, using a more technical argument.

References

[LRS01] M. Luby and D. Randall and A. Sinclair, "Markov chain algorithms for planar lattice structures," *SIAM Journal on Computing* **31** (2001), pp. 167–192. Preliminary version in *ACM STOC*, 1995.

[Th90] W. Thurston, "Conway's tiling groups," *American Mathematical Monthly* **97** (1990), pp. 757–773.

[W04] D.B. Wilson, "Mixing times of lozenge tiling and card shuffling Markov chains," *The Annals of Applied Probability*, 14(1):274–325, 2004.

[1] Note that the process can take jumps of arbitrary size h, not just 1 as in simple random walk. However, standard martingale techniques still imply that the expected hitting time to 0 for such a process on $[0, D]$ with zero drift is D^2/V, where V is a lower bound on the variance of a move (i.e., the expected squared distance moved in one step). In our case, we can take $V \geq \frac{1}{2Nh} \times h^2 = \frac{h}{2N} \geq \frac{1}{2n}$.