Sampling Adsorbing Staircase Walks Using a New Markov Chain Decomposition Method

Russell A. Martin*
School of Mathematics
Georgia Institute of Technology
martin@math.gatech.edu

Dana Randall*
College of Computing
and School of Mathematics
Georgia Institute of Technology
randall@cc.gatech.edu

Abstract

Staircase walks are lattice paths from $(0,0)$ to $(2n,0)$ which take diagonal steps and which never fall below the x-axis. A path hitting the x-axis k times is assigned a weight of λ^k, where $\lambda > 0$. A simple local Markov chain which connects the state space and converges to the Gibbs measure (which normalizes these weights) is known to be rapidly mixing when $\lambda = 1$, and can easily be shown to be rapidly mixing when $\lambda < 1$. We give the first proof that this Markov chain is also mixing in the more interesting case of $\lambda > 1$, known in the statistical physics community as adsorbing staircase walks. The main new ingredient is a decomposition technique which allows us to analyze the Markov chain in pieces, applying different arguments to analyze each piece.

1. Introduction

1.1. The model

Staircase walks (also called *Dyck paths*) are walks in $\mathbb{Z}^2$ from $(0,0)$ to (n,n) which stay above the diagonal $x = y$. Rotating by 45^o, they correspond to walks from $(0,0)$ to $(2n,0)$ which take diagonal steps by adding $(1,1)$ or $(1,-1)$ at each step and which never fall below the x-axis (see figure 1). The number of staircase walks is exactly $C(n)$, the nth Catalan number, which can be calculated exactly so sampling can be done recursively without a Markov chain. However, there is also a simple Markov chain on the set of staircase walks which has been very useful for sampling other combinatorial objects including triangulations [8] and planar matchings [5, 14]. The Markov chain consists of "mountain/valley" flips by choosing $i \in [2n]$ and if the ith step of the walk is a local optimum (a mountain), inverting it so that it is a local minimum (a valley), or vice-versa.

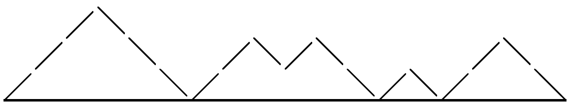

Figure 1: A staircase walk ($n = 9$)

A natural generalization, studied in the statistical physics community, is to weight the set of staircase walks according to the number of times they hit the x-axis. We assign a weight λ^k to a walk which hits the x-axis k times. The Gibbs measure normalizes this so that the probability of a walk w is $\pi(w) = \frac{\lambda^{k_w}}{\sum_v \lambda^{k_v}}$, where the sum is taken over all walks v and k_v is the number of times v hits the x-axis. When $\lambda = 1$ this is just the uniform probability where $\pi(w) = 1/C(n)$. Taking $\lambda < 1$ favors walks which stay away from the x-axis, and taking $\lambda > 1$, called *adsorbing walks*, favors walks which hit the x-axis many times. It was shown by van Rensburg [10] that there is a phase transition at $\lambda = 2$: when $\lambda < 2$, the walks wander $O(\sqrt{n})$ away from the x-axis, whereas when $\lambda > 2$ the walks never wander more than $o(\sqrt{n})$ away.

Adsorbing staircase walks are closely related to "returning" walks on an infinite d-ary tree which start and end at the root. We label the edges of the complete d-ary tree as a Cayley graph so that each vertex is adjacent to one edge with each of d labels. A returning walk of length $2n$, starting at the root, has n edges

* Supported in part by NSF Grant No. CCR-9703206.

leading away from the root and n returning. Whenever we are at the root we have d choices of labeled edges; whenever we are away from the root we have $d-1$ edges which moves us farther away and a unique edge which will bring us closer to the root. Hence, there are $d^k(d-1)^{n-k}$ walks of length $2n$ that hit the root k times (including the initial time, but not the final). Using adsorbing staircase walks, sampling is easy: (1) Select a staircase walk of length $2n$ according to the Gibbs measure with $\lambda = \frac{d}{d-1}$. A staircase walk hitting the x-axis k times appears with probability proportional to $(\frac{d}{d-1})^k = \frac{d^k(d-1)^{(n-k)}}{(d-1)^n}$. The up edges in this walk correspond to steps in the tree that move away from the root, and the down edges are those that move back towards the root. (2) Assign labels to the up edges uniformly at random (from the set of d labels for edges starting from the x-axis, and from a suitable set of $d-1$ labels for edges above the x-axis), assigning labels to the down edges that equal the label of the most recent unpaired up edge preceding it. This gives a sequence of labeled edges corresponding to a labeled walk of length $2n$ in the d-ary tree.

The mountain/valley Markov chain can be modified to incorporate the Gibbs weights. It is straightforward to show that it is rapidly mixing when $\lambda \leq 1$. Wilson [14] gives a tight bound of $O(n^3(\log n + \log \frac{1}{\varepsilon}))$ when $\lambda = 1$, which provides an upper bound in the case $\lambda < 1$.

When $\lambda > 1$ a simple coupling argument is insufficient. Informally, for coupling to succeed we need to construct a coupled Markov chain so that close configurations tend to come closer together. However, in the adsorbing case, pairs of walks that differ near the x-axis will tend to diverge initially. We note that there are alternative, recursive methods for sampling adsorbing staircase walks based on generating functions, but our goal in this paper is to understand how these new weightings affect the mixing time of the mountain/valley Markov chain.

A natural approach in trying to circumvent this difficulty is to introduce a new Markov chain based on a *heat bath* algorithm. A heat bath works by erasing a larger piece of the current configuration (larger than the mountain/valley walk which erases only two edges of the walk) and moving to a new configuration, consistent with the current (remaining) boundary information, according to the conditional probability. As with many such algorithms, this more elaborate Markov chain appears to be too difficult to analyze.

1.2. Our results

In this paper, we show that the mountain/valley Markov chain is rapidly mixing even in the case $\lambda > 1$ (and hence all values of λ). The main new ingredient is a general decomposition technique which we believe has many other applications. This new technique is similar to a decomposition theorem of Madras and Randall [6,7], but is much more natural and much simpler to apply. Suppose the state space can be naturally partitioned into sets $\{A_i\}$ (subject to certain constraints, outlined in section 4.2). Further, suppose that the Markov chain is rapidly mixing when restricted to any of the A_i. Finally, suppose that a *projection* (also defined later) of these sets is rapidly mixing, suggesting that it is easy to travel from any of the A_i to any other A_j. Then we can conclude that the original Markov chain is rapidly mixing as well. This is quite similar in spirit to the Madras/Randall result, however their decomposition theorem requires that the $\{A_i\}$ form a *cover* and must in fact have considerable overlaps. We have found that the new theorem is far more natural for several applications.

Given this new decomposition technique, we can state our strategy for analyzing the mountain/valley chain. We first decompose the state space Ω into $\bigcup \mathcal{S}_k$, where $\mathcal{S}_k$ is the set of staircase walks which hit the x-axis exactly k times. First we show that $|\mathcal{S}_k|$ is log-concave in k. This immediately implies that our first projection (according to the decomposition theorem) is mixing in polynomial time, so it suffices to show that the Markov chain restricted to $\mathcal{S}_k$, $\mathcal{M}_{\mathcal{S}_k}$, is rapidly mixing for each k.

To show that $\mathcal{M}_{\mathcal{S}_k}$ is rapidly mixing, we apply the decomposition theorem a second time. This time we partition the state space (i.e., the set of staircase walks which hit the x-axis exactly k times) into $\binom{n-1}{k}$ sets according to which k points on the x-axis the paths hit. Showing that $\mathcal{M}$ is rapidly mixing when restricted to any of these sets is straightforward and follows the unbiased case when $\lambda = 1$, which has been previously analyzed.

The projection arising from the second decomposition can be viewed as an interesting particle process: we want to sample from the $\binom{n-1}{k}$ ways to place k particles on the x-axis between 0 and n so that each configuration occurs with probability $\Pi_i C(x_i)$, where x_i is the length of the gap between the ith and $i+1$st particles, and $C(n)$ is the nth Catalan number. The Markov chain arising from this projection is quite natural: choose $(i, d) \in [k] \times \{\ell, r\}$ and move the ith particle in the left or right direction according to d, if possible.

This particle process isolates the difficulty in the original mountain/valley Markov chain on weighted staircase walks; indeed a simple path coupling argument fails for an analogous reason. The final step of our analysis is noticing that a heat bath algorithm is easy to define for this particle system which does have the desirable properties. Namely, it converges to the correct stationary distribution and, moreover, we can show it mixes in polynomial time. By a standard comparison argument we can show that the mixing rate of the heat bath algorithm is close to the mixing rate of the original particle process. Hence, we can apply the decomposition theorem for a second time, thereby establishing the polynomial-time mixing rate of the original mountain/valley chain on the entire state space of staircase walks, as desired.

2. A Markov chain on $\mathcal{S}$

We define a natural Markov chain $\mathcal{M}$ on $\mathcal{S}$, the set of staircase walks with $2n$ edges, for a fixed $n \geq 1$. This mountain/valley Markov chain has previously appeared in [5, 8, 14]. The transitions of the chain are *inversions* which replace local maxima with local minima, or vice-versa, by interchanging two edges along the walk. If the cth point on the path is $v_c = (x_c, y_c)$, we call it a mountain if $y_{c-1} = y_c - 1 = y_{c+1}$ and inverting it consists of setting $y'_c = y_c - 2$. Likewise, inverting a valley where $y_{c-1} = y_c + 1 = y_{c+1}$ consists of setting $y'_c = y_c + 2$. The Markov chain $\mathcal{M}$ iterates the following steps.

One step of Markov chain $\mathcal{M}$:

1. Pick c uniformly at random from $\{2, \ldots, 2n-2\}$, and let v denote the point on the path whose x-coordinate is c.

2.
- If v is the bottom of a valley lying on the x-axis, with probability $\frac{1}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v. Otherwise, set $X_{t+1} = X_t$.
- If v is the top of a mountain, and inverting it will put it on the x-axis, with probability $\frac{\lambda}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.
- If v is the bottom of a valley not lying on the x-axis, or if v is the top of a mountain and inverting at v does not put it on the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t inverted at v.
- In all other cases, set $X_{t+1} = X_t$.

Note that this Markov chain is aperiodic, reversible, and the stationary distribution is the Gibbs distribution, namely $\pi(\sigma) = \frac{\lambda^k}{Z}$, where k is the number of times the walk σ touches the x-axis. We show in the next section that the mixing time for this chain is polynomially bounded when $\lambda \leq 1$ using path coupling.

3. Mixing machinery

In what follows, we assume that $\mathcal{M}$ is an ergodic (i.e. irreducible and aperiodic), reversible Markov chain with finite state space Ω, transition probability matrix P, and stationary distribution π.

The time a Markov chain takes to converge to its stationary distribution, the mixing time of the chain, is measured in terms of the distance between the distribution at time t and the stationary distribution. Letting $P^t(x, y)$ denote the t-step probability of going from x to y, the *total variation distance* at time t is

$$\|P^t, \pi\|_{tv} = \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|.$$

For $\varepsilon > 0$, the *mixing time* $\tau(\varepsilon)$ is

$$\tau(\varepsilon) = \min\{t : \|P^{t'}, \pi\|_{tv} \leq \varepsilon, \forall t' \geq t\}.$$

We say a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log \frac{1}{\varepsilon}$, where n is the size of each configuration in the state space.

It is well known that the mixing rate is related to the *spectral gap* of the transition matrix. For the transition matrix P, we let $Gap(P) = \lambda_0 - |\lambda_1|$ denote its spectral gap, where $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ are the eigenvalues of P and $1 = \lambda_0 > |\lambda_1| \geq |\lambda_i|$ for all $i \geq 2$. The following result relates the spectral gap with the mixing time of the chain (see, e.g., [12]):

Theorem 3.1 *Let $\pi_* = \min_{x \in \Omega} \pi(x)$. For all $\varepsilon > 0$ we have*

1. $\tau(\varepsilon) \leq \frac{1}{Gap(P)} \log(\frac{1}{\pi_* \varepsilon})$
2. $\tau(\varepsilon) \geq \frac{|\lambda_1|}{Gap(P)} \log(\frac{1}{2\varepsilon})$.

Remark. For simplicity, we may add self-loops with probability $\frac{1}{2}$ to each point in the state space to ensure that $\lambda_1 \geq 0$, also ensuring aperiodicity of the Markov chain. $\square$

We give a brief review of some of the techniques that are used to bound the mixing time (or spectral gap) of a Markov chain, before introducing our new method in section 4.2.

3.1. Path coupling

A *coupling* is a Markov chain on $\Omega \times \Omega$ with the following properties: Instead up updating the pair of configurations independently, the coupling updates them so that the two processes will tend to correlate, or "move together" under some measure of distance, but each process, viewed in isolation, is just performing transitions of the original Markov chain. Also, once the pair of configurations agree at some time, the coupling guarantees they agree from that time forward. The mixing time can be bounded by the expected time for configurations to coalesce under any valid coupling.

More simply, *path coupling* lets us bound the mixing time by analyzing a subset of $\Omega \times \Omega$. The method of path coupling is described in the next theorem, adapted from [3]:

Theorem 3.2 **[3]** *Let d be an integer valued metric defined on $\Omega \times \Omega$ taking values in $\{0, \ldots, D\}$. Let U be a subset of $\Omega \times \Omega$ such that for all $(x_t, y_t) \in \Omega \times \Omega$ there exists a path $x_t = z_0, z_1, \ldots, z_r = y_t$ between x_t and y_t such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and*

$$\sum_{i=0}^{r-1} d(z_i, z_{i+1}) = d(x_t, y_t).$$

Define a coupling $(x_t, y_t) \to (x_{t+1}, y_{t+1})$ of $\mathcal{M}$ on all pairs $(x_t, y_t) \in U$. If $E(\Delta d(x_t, y_t)) \leq 0$ for all $(x_t, y_t) \in U$, and there exists $\alpha > 0$ such that $Pr[d(x_{t+1}, y_{t+1}) \neq d(x_t, y_t)] \geq \alpha$ for all t, then the mixing time satisfies

$$\tau(\varepsilon) \leq \lceil \frac{eD^2}{\alpha} \rceil \lceil \log(\frac{1}{\varepsilon}) \rceil.$$

3.2. Staircase walks with $\lambda \leq 1$

We demonstrate the method of path coupling to show that the mountain/valley chain on staircase walks is rapidly mixing when $\lambda \leq 1$.

We define our distance measure Φ to be one-half of the area between the configurations, i.e., drawing the configurations on the same set of axis bounds rectangular regions between the pair of walks. The distance between the two walks is one-half of the sum of the areas of these rectangular regions. For the coupling, we take the point c in step 1 of $\mathcal{M}$, and attempt to perform the same transition in each walk. To use path coupling, we must examine a pair of walks that differ solely by a single transition of the chain (a single square).

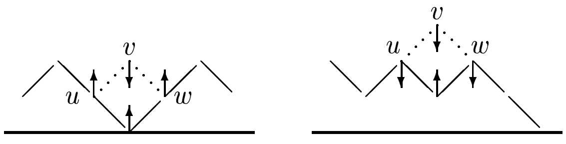

Figure 2: Typical situations for path coupling

Lemma 3.3 *Let $X, Y \in \mathcal{S}$ with $\Phi(X,Y) = 1$. After one step of $\mathcal{M}$ we have $E(\Delta\Phi(X,Y)) \leq 0$, provided $\lambda \leq 1$.*

Proof. Consider the configurations in figure 2, which show pieces of walks that agree everywhere except at a single square, and let Y denote the "upper" walk. If the square is adjacent to the x-axis as in figure 2.A, then there are two transitions that decrease the distance by one, inverting v in one of the walks so they now agree everywhere. Inverting u or w in Y increases the distance by one. Every other transition not involving u, v, or w does not change the distance between X and Y. Therefore, in this case we find $E(\Delta\Phi(X,Y)) \leq \frac{1}{2n}(\frac{1}{4} + \frac{1}{4} - \frac{1}{2(1+\lambda)} - \frac{\lambda}{2(1+\lambda)}) = 0$. (This is an inequality since one or both of the moves which increase the distance might not be valid moves.) The second case is if the differences between the two walks occur a unit distance from the x-axis, as in figure 2.B. There are also two good inversions at v, each with probability $\frac{1}{4}$, and those at u and w increase the distance between X and Y by one; all other moves preserve the distance between the pair. In this case, we have $E(\Delta\Phi(X,Y)) \leq \frac{1}{2n}(\frac{\lambda}{2(1+\lambda)} + \frac{\lambda}{2(1+\lambda)} - \frac{1}{4} - \frac{1}{4})$. This last expression is non-positive if $\lambda \leq 1$. Other situations where X and Y differ by a square that is far away from the x-axis are neutral; two good moves decrease the distance by one, and (at most) two bad moves increase the distance by one. Each of these moves occurs with equal probability, so in these cases, we also have $E(\Delta\Phi(X,Y)) \leq 0$. $\square$

An application of theorem 3.2 gives a polynomial bound on the mixing rate. These details are left to the reader. We note that in the case demonstrated in figure 2.B, the distance will increase in expectation if $\lambda > 1$.

3.3. The decomposition method

The Madras/Randall *decomposition method* [7] offers a different approach for bounding the mixing time of a Markov chain and will be the main motivation behind our analysis in this paper. The intuition behind this method is that we look at subsets of the state space and show that the Markov chain restricted to each subset is mixing. Then, if the sets overlap enough (and cover all of Ω), we can deduce a bound on the mixing rate of the original chain on the entire state space.

Following [7], let $A_1, \ldots, A_m$ be subsets of Ω such that $\cup_i A_i = \Omega$. We are interested in two classes of induced Markov chains. The first is a set of *restricted* Markov chains, obtained by restricting $\mathcal{M}$ to each subset A_i, i.e., any move of $\mathcal{M}$ that would take us from an element $x \in A_i$ to some $y \notin A_i$, $j \neq i$, is rejected. In particular, the restriction to A_i is a Markov chain, $\mathcal{M}_i$, where the transition matrix P_{A_i} is defined as follows: If $x \neq y$ and $x, y \in A_i$ then $P_{A_i}(x,y) = P(x,y)$; if $x \in A_i$ then $P_{A_i}(x,x) = 1 - \sum_{y \in A_i, y \neq x} P_{A_i}(x,y)$.

The second Markov chain is the *projection* $\mathcal{M}_H$ of the cover $\{A_1, \ldots, A_m\}$, defined on the set $[m]$, where each point i is associated with the set A_i. Let $\Theta = \max_{x \in \Omega} |\{i : x \in A_i\}|$. The transition matrix P_H for Markov chain $\mathcal{M}_H$ is defined by letting $P_H(i,j) = \frac{\pi(A_i \cap A_j)}{\Theta \pi(A_i)}$ for $i \neq j$, and $P_H(i,i) = 1 - \sum_{j \neq i} P_H(i,j)$. The limiting distribution ρ of this chain is given by $\rho(i) = \pi(A_i)/\hat{Z}$, where $\hat{Z} = \sum_i \sum_{x \in A_i} \pi(x) \leq \Theta$. From [7] we have

Theorem 3.4 [7] *In the preceding framework,*

$$Gap(P) \geq \frac{1}{\Theta^2} Gap(P_H) \min_{i=1,\ldots,m} Gap(P_{A_i}).$$

4. A new decomposition result

Our goal is to give a method, analogous to that in section 3.3, but using a partition of Ω into *disjoint* pieces. We relate the spectral gap of the original chain to the spectral gap of the restriction to each set in the partition, and that of a new projection of this partition.

We first briefly introduce the framework that Caracciolo, Pelisetto, and Sokal (CPS) use in the context of simulated tempering (see [7] for further details).

4.1. The CPS tempering method

Let P denote a transition matrix of a Markov chain on the finite state space Ω that is reversible with respect to the probability distribution π. Suppose that the state space is partitioned into m disjoint pieces $\Omega_1, \ldots, \Omega_m$. For each $i = 1, \ldots, m$, define P_{Ω_i}, the restriction of P to Ω_i, by rejecting jumps that leave Ω_i (as in section 3.3). Let π_i be the normalized restriction of π to Ω_i, i.e., $\pi_i(A) = \frac{\pi(A \cap \Omega_i)}{b_i}$ where $b_i = \pi(\Omega_i)$. Let Q be another transition matrix that is also reversible with respect to π. Define $\overline{Q}$ to be the following aggregated transition matrix on the state space $\{1, \ldots, m\}$:

$$\overline{Q}(i,j) = \frac{1}{b_i} \sum_{\substack{x \in \Omega_i, \\ y \in \Omega_j}} \pi(x) Q(x,y).$$

We note that

$$b_i \overline{Q}(i,j) = b_j \overline{Q}(j,i)$$

so $\overline{Q}$ is reversible with respect to the probability measure $b = (b_1, \ldots, b_m)$ on $\{1, \ldots, m\}$.

Theorem 4.1 ([7], Thm A.1) *Assume Q is positive semidefinite. Let $Q^{1/2}$ denote the nonnegative square root of Q. Then*

$$Gap(Q^{1/2} P Q^{1/2}) \geq Gap(\overline{Q}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

4.2. Disjoint decomposition

We use theorem 4.1 to derive a bound on the spectral gap of P.

Theorem 4.2 *Let P_{Ω_i} be as above, and let $\overline{P}$ be defined as above with P in place of Q. Then*

$$Gap(P) \geq \frac{1}{2} Gap(\overline{P}) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

Proof. Take $Q = P$ in theorem 4.1 above, and use that the eigenvalues of P^2 are the squares of the eigenvalues of P, so that $Gap(P) = 1 - \sqrt{1 - Gap(P^2)} \geq \frac{1}{2} Gap(P^2)$. (The inequality follows from the Taylor series of $1 - \sqrt{1-x}$.) $\square$

We also derive a useful corollary. Suppose we replace the matrix $\overline{P}$ with a transition matrix P_M on the set $\{1,\ldots,m\}$, with Metropolis transitions, i.e., $P_M(i,j) = \min\{1, \frac{\pi(\Omega_j)}{\pi(\Omega_i)}\}$. Let $\partial_i(\Omega_j) = \{y \in \Omega_j : \exists x \in \Omega_i \text{ with } P(x,y) > 0\}$. First we state a useful lemma that follows immediately from the "functional definition" of the spectral gap (see [7, Eq. (7)]):

Lemma 4.3 *Suppose R and T are Markov chains on the same state space, each reversible with respect to the distribution π. Suppose there are constants c_1 and c_2 such that $c_1 R(x,y) \leq T(x,y) \leq c_2 R(x,y)$ for all $x \neq y$. Then $c_1 Gap(R) \leq Gap(T) \leq c_2 Gap(R)$.*

Corollary 4.4 *With P_M as above, suppose there exists $\alpha > 0$ and $\beta > 0$ such that*

1. *$P(x,y) \geq \alpha$ for all $x \sim y$ in P;*
2. *$\pi(\partial_i(\Omega_j)) \geq \beta\pi(\Omega_j)$ for all pairs $i \sim j$ in the Markov chain defined by $\overline{P}$.*

Then

$$Gap(P) \geq \frac{1}{2}\alpha\beta\, Gap(P_M) \min_{i=1,\ldots,m} Gap(P_{\Omega_i}).$$

Proof. Note that

$$\begin{aligned}
\sum_{\substack{x\in\Omega_i,\\ y\in\Omega_j}} \pi(x)P(x,y) &= \sum_{\substack{x\in\partial_j(\Omega_i),\\ y\in\partial_i(\Omega_j)}} \pi(x)P(x,y) \\
&= \sum_{\substack{x\in\partial_j(\Omega_i),\\ y\in\partial_i(\Omega_j)}} \pi(y)P(y,x) \\
&\geq \sum_{y\in\partial_i(\Omega_j)} \pi(y)\alpha \\
&\geq \alpha\beta\pi(\Omega_j),
\end{aligned}$$

where the second equality follows from reversibility and the inequalities follow from conditions 1 and 2. Multiplying by $\frac{1}{\pi(\Omega_i)}$, we see $\overline{P}(i,j) \geq \alpha\beta P_M(i,j)$, so $Gap(\overline{P}) \geq \alpha\beta\, Gap(P_M)$ by lemma 4.3. $\square$

We illustrate the use of this new decomposition theorem on the problem of sampling from the set of weighted staircase walks. Our goal is not to find an optimal bound on the mixing time, but rather to demonstrate the applicability of this new method.

5. Decomposition of $\mathcal{S}$

Noticing that path coupling fails to show rapid mixing of $\mathcal{M}$ when $\lambda > 1$, it is natural to try to sample from a subset of $\mathcal{S}$, say $\mathcal{S}_k$, the set of staircase walks that hit the x-axis exactly k times between the endpoints. We further break this down by decomposing $\mathcal{S}_k$ into sets $\mathcal{S}_{k,T}$, based on the location of the x-axis hits, where sampling from these subsets is easy. In the final two sections we show how to formalize this approach using the decomposition method of section 4.2.

The mountain/valley Markov chain $\mathcal{M}$ is insufficient for sampling from $\mathcal{S}_k$ since we will never be able to alter the places that a path hits the x-axis. For this reason, we need to introduce a slight variant $\widehat{\mathcal{M}}$ on $\mathcal{S}$ for the purposes of the analysis; the rapid mixing of the simpler chain $\mathcal{M}$ follows from the rapid mixing of $\widehat{\mathcal{M}}$ by a very simple application of the comparison method (see [2, 9]).

In this new Markov chain $\widehat{\mathcal{M}}$ there are two basic types of moves. The first type of moves are inversions. The second type of move consists of changing one "propeller-like" structure into its mirror image. Letting D denote a "down" edge and U an "up" edge, if there is a sequence of four edges $DUUD$, we can change it to the sequence $UDDU$, or vice-versa. These moves are only allowed when one point of the propeller touches the boundary. See figure 3 for a pictorial depiction of this move. We call such a change a propeller move (centered) at v.

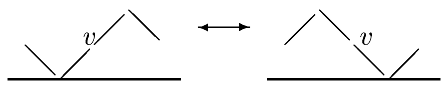

Figure 3: The propeller move

More formally, $\widehat{\mathcal{M}}$ iterates these steps.

One step of Markov chain $\widehat{\mathcal{M}}$:

1. Pick c uniformly at random from $\{2,\ldots,2n-2\}$, and let v denote the point on the path whose x-coordinate is c.

2.
- If $c = 2$ and v is the bottom of a valley on the x-axis, with probability $\frac{1}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.
- If $c = 2$ and v is the top of a peak and inverting it will put it on the x-axis, with probability $\frac{\lambda}{2(1+\lambda)}$ set X_{t+1} equal to X_t inverted at v.

- If $c \in \{3, 4, \ldots, 2n-2\}$, and v is the bottom of a valley not lying on the x-axis, or if v is the top of a peak and inverting at v does not put it onto the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t inverted at v.
- If v is the central vertex of a propeller structure where the lowest point lies on the x-axis, with probability $\frac{1}{4}$ set X_{t+1} equal to X_t after performing a propeller move at v.
- In all other cases, set $X_{t+1} = X_t$.

First note that $\widehat{\mathcal{M}}$ is aperiodic ($X_{t+1} = X_t$ with probability at least $\frac{1}{2}$). Second, the only time a transition is possible from a path hitting the boundary k times to one hitting the boundary $k+1$ (respectively $k-1$) times is when there is a peak (respectively valley) at the beginning of the walk, and we select that vertex in step 1 of the chain. All other moves of $\widehat{\mathcal{M}}$ preserve the weight of the walk.

Having described the Markov chain, we use it to define a metric d on $\mathcal{S}$. For any pair of states $X, Y \in \mathcal{S}$, if $P(X,Y) > 0$ (so X and Y are nearest neighbors), we define $d(X,Y)$ to equal one-half of the area of the symmetric difference of the two staircase walks. If $P(X,Y) = 0$ (i.e., moving from X to Y requires more than one move), first consider a *path of states* $X = Z_0, Z_1, \ldots, Z_r = Y$ between X and Y, where $P(Z_i, Z_{i+1}) > 0$ for each $i = 0, \ldots r-1$; then define $d(X,Y) = \min \sum_{i=0}^{r-1} d(Z_i, Z_{i+1})$ where the minimum is taken over all paths joining X and Y. We call d the *transition metric*.

We apply the decomposition method of section 4.2 to show rapid mixing of $\widehat{\mathcal{M}}$ on $\mathcal{S}$. To do so we need to examine the projection of the partition $\{\mathcal{S}_k\}$ and bound the spectral gap for the restriction to each of the subsets $\mathcal{S}_k$. We do this is the next sections following a brief combinatorial excursion.

5.1. A combinatorial look at $\mathcal{S}$

For this subsection, we let S^n denote the set of staircase walks with $2n$ edges and let $\mathcal{S}_k^n$ denote the subset of S^n containing those walks with k internal x-axis hits. The cardinalities of $\mathcal{S}_k^n$ can be shown to be log-concave, i.e., the sequence of numbers $|\mathcal{S}_0^n|, |\mathcal{S}_1^n|, |\mathcal{S}_2^n|, \ldots, |\mathcal{S}_{n-1}^n|$ is a log-concave sequence. This follows from two simple lemmas.

Lemma 5.1 *For $n \geq 3$ and $1 \leq k \leq n-2$,* $|\mathcal{S}_k^n| = |\mathcal{S}_{k-1}^{n-1}| + |\mathcal{S}_{k+1}^n|$.

Lemma 5.2 *For $n \geq 3$ and $1 \leq k \leq n-2$,*

$$|\mathcal{S}_k^n| = |\mathcal{S}_{k-1}^{n-1}| + |\mathcal{S}_k^{n-1}| + \cdots + |\mathcal{S}_{n-2}^{n-1}|.$$

Also, for $n \geq 2$,

$$|\mathcal{S}_0^n| = |\mathcal{S}^{n-1}| = |\mathcal{S}_0^{n-1}| + \cdots + |\mathcal{S}_{n-2}^{n-1}|.$$

Proof. For the first part, we use lemma 5.1 iteratively, so

$$\begin{aligned} |\mathcal{S}_k^n| &= |\mathcal{S}_{k-1}^{n-1}| + |\mathcal{S}_{k+1}^n| \\ &= |\mathcal{S}_{k-1}^{n-1}| + |\mathcal{S}_k^{n-1}| + |\mathcal{S}_{k+2}^n| = \cdots \end{aligned}$$

For the second, there is a bijection between $\mathcal{S}_0^n$ and $\mathcal{S}^{n-1}$ by taking $\sigma \in \mathcal{S}_0^n$, deleting the initial and terminal edges of the walk, and shifting the walk down and to the left by one to obtain a staircase walk joining $(0,0)$ to $(2n-2, 0)$. □

Theorem 5.3 *For a fixed $n \geq 3$, $|\mathcal{S}_k^n|$ is log-concave. In particular, for $1 \leq k \leq n-2$,*

$$|\mathcal{S}_{k-1}^n| \cdot |\mathcal{S}_{k+1}^n| \leq |\mathcal{S}_k^n|^2. \tag{1}$$

Proof. We use induction on n. For notational simplicity we suppress the cardinality symbols.

For $n = 3$, by a simple enumeration of the possibilities, we find that $\mathcal{S}_0^3 = \mathcal{S}_1^3 = 2$ and $\mathcal{S}_2^3 = 1$, so that $\mathcal{S}_0^3 \cdot \mathcal{S}_2^3 \leq (\mathcal{S}_1^3)^2$.

Now assume that for some $n-1$ that $\mathcal{S}_k^{n-1}$ is log-concave. Also, assume first that $k \geq 2$. We want to show that (1) holds. To do this, it suffices to show the inequality

$$\begin{aligned} &\mathcal{S}_{k-2}^{n-1}[\mathcal{S}_k^{n-1} + \mathcal{S}_{k+1}^{n-1} + \cdots + \mathcal{S}_{n-2}^{n-1}] \leq \\ &\qquad \mathcal{S}_{k-1}^{n-1}[\mathcal{S}_{k-1}^{n-1} + \mathcal{S}_k^{n-1} + \cdots + \mathcal{S}_{n-2}^{n-1}], \end{aligned} \tag{2}$$

since lemma 5.2 implies that (2) is equivalent to

$$\mathcal{S}_{k-2}^{n-1} \cdot \mathcal{S}_{k+1}^n \leq \mathcal{S}_{k-1}^{n-1} \cdot \mathcal{S}_k^n. \tag{3}$$

By adding $\mathcal{S}_k^n \cdot \mathcal{S}_{k+1}^n$ to both sides of (3), factoring, and applying lemma 5.1, we get (1).

To show (2), it suffices to show the set of inequalities

$$\mathcal{S}_{k-2}^{n-1} \cdot \mathcal{S}_{k-1+i}^{n-1} \leq \mathcal{S}_{k-1}^{n-1} \cdot \mathcal{S}_{k-2+i}^{n-1}$$

for all $i \in [n-k-1]$. These inequalities all hold by our induction hypothesis that $\mathcal{S}_k^{n-1}$ is log-concave. Adding them, and the extra term $\mathcal{S}_{k-1}^{n-1} \cdot \mathcal{S}_{n-2}^{n-1}$ to the right hand side, gives us (2).

All that remains is the case $\mathcal{S}_0^n \cdot \mathcal{S}_2^n \leq (\mathcal{S}_1^n)^2$ (when $k = 1$). We use that $\mathcal{S}_0^n = \mathcal{S}_1^n$ and, from lemma 5.1, we see $\mathcal{S}_2^n \leq \mathcal{S}_1^n$. Therefore, $\mathcal{S}_0^n \cdot \mathcal{S}_2^n = \mathcal{S}_1^n \cdot \mathcal{S}_2^n \leq \mathcal{S}_1^n \cdot \mathcal{S}_1^n$. □

5.2. Projection 1: $\mathcal{S} = \dot{\cup}\mathcal{S}_k$

We bound the mixing rate of the projection by appealing to theorem 5.3. The (disjoint) projection P_M of the partition $\mathcal{S} = \dot{\cup}_{k=0}^{n-1}\mathcal{S}_k$ is a random walk on $\{0, \ldots, n-1\}$ with stationary probabilities $\frac{\lambda^k|\mathcal{S}_k|}{Z}$, where $Z = \sum_{k=0}^{n-1}\lambda^k|\mathcal{S}_k|$, and Metropolis transition probabilities. Path coupling yields the following theorem.

Lemma 5.4 *The mixing time of P_M satisfies*

$$\tau_{P_M}(\varepsilon) = O(n^2 \log n \log(\frac{1}{\varepsilon})).$$

5.3. Restriction 1: Mixing on S_k

By the disjoint decomposition theorem, it suffices to show that the restricted Markov chains ($\widehat{\mathcal{M}}$ restricted to S_k) are rapidly mixing in order to conclude that $\widehat{\mathcal{M}}$ is mixing on the whole state space $\mathcal{S}$. We show this in the next section.

6. Decomposition of S_k

In this section we show that $\widehat{\mathcal{M}}_{S_k}$, the Markov chain restricted to $\mathcal{S}_k$, is rapidly mixing. To do this we apply the decomposition method a second time. First we partition $\mathcal{S}_k$ and show that $\widehat{\mathcal{M}}_{S_k}$ is mixing when restricted to each of set of this partition. Following that, we show the projection is mixing using heat bath dynamics and the comparison theorem, setting the stage for corollary 4.4.

Let T denote a subset of $\mathcal{S}_k$ where each walk touches the x-axis in the same k locations. For example, (in the case that $n \geq 6$) we can consider the set of walk that hit the x-axis at the points with x-coordinates $2, 6$, and 10 in the interior between the two endpoints. There are $\binom{n-1}{k}$ ways to specify the location of k internal hits, as the x-coordinate of each hit must be an even number. We write $\mathcal{S}_k = \dot{\cup}_T S_{k,T}$, where this union is over all $\binom{n-1}{k}$ ways of specifying the hits on the x-axis.

6.1. Restriction 2: Mixing of $\mathcal{S}_{k,T}$

Let $\widehat{\mathcal{M}}_{S_{k,T}}$ denote the restriction of $\widehat{\mathcal{M}}_{S_k}$ to the set T. We have the following result, whose proof is a simple application of path coupling:

Lemma 6.1 *Let T be a subset of $\mathcal{S}_k$ as above, and let $X, Y \in T$ with $d(X,Y) = 1$. After one step of the Markov chain, $\widehat{\mathcal{M}}_{S_k,T}$, on the set T, we have $E(\Delta d(X,Y)) \leq 0$.*

Proof. This proof is similar to lemma 3.3, except that we need only consider the situation when the paths differ by a square that is at least distance one from the x-axis as in figure 4. There are two good inversions at v that decrease the distance by one, and at most two inversions increasing the distance by one. Each of these inversions happens with equal probability, so $E(\Delta d(X,Y) \leq 0$. □

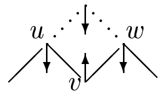

Figure 4: Possible transitions inside $\mathcal{S}_{k,T}$

Lemma 6.1 gives the first piece for the path coupling theorem. If $d(X,Y) > 0$, the probability of the distance changing in one step of $\widehat{\mathcal{M}}_{S_k,T}$ is at least $\frac{1}{4(n-k)}$, since when we select a vertex on the walk we may avoid choosing one that lies on the x-axis, and its immediate neighbor to the right, as these vertices will never move. For pairs of walks in $\mathcal{S}_k$, we have $0 \leq d(X,Y) \leq (n-k-1)^2 - (n-k-1)$. For $X, Y \in \mathcal{S}_{k,T}$ there is a sequence of $d(X,Y)$ inversions that will transform one walk into the other.

By a straightforward application of theorem 3.2, we have

Lemma 6.2 *The mixing time of $\widehat{\mathcal{M}}_{S_k,T}$ satisfies*

$$\tau_{\widehat{\mathcal{M}}_{S_k,T}}(\varepsilon) \leq \lceil \frac{e((n-k-1)(n-k-2))^2}{\frac{1}{4(n-k)}} \log(\frac{1}{\varepsilon}) \rceil$$
$$= \mathrm{O}(n^5 \log(\frac{1}{\varepsilon})).$$

6.2. Projection 2: $\mathcal{S} = \dot{\cup}_T\mathcal{S}_{k,T}$

The projection of $\mathcal{S}_{k,T}$ can be viewed as a particle process on $[n]$. The particles represent the places that a path hits the x-axis. The projection of the Markov chain can be viewed as single-site dynamics on the set of particles which moves one particle to the left or right in each step. Interestingly, analyzing the mixing rate of this particle process cannot be done using a simple path coupling argument, which seems to isolate the difficulty with using path coupling on the original mountain/valley chain. However, the particle process is sufficiently simple that we can analyze it indirectly. We formalize this approach in this subsection.

- **Particle process with single-site dynamics**

Consider a set of $n-1$ sites in a linear arrangement, k of which contain particles, and let $\mathcal{P}_k$ denote the set of all $\binom{n-1}{k}$ such configurations. An element $\sigma \in \mathcal{P}_k$ will correspond to the set, $\mathcal{S}_{k,T(\sigma)}$, of all staircase walks in $\mathcal{S}_k$ that have x-axis hits at the locations determined by the particles. For example, if $k=3$ and σ is the configuration with particles at sites 2, 3, and 8, then $\mathcal{S}_{3,T(\sigma)}$ consists of all walks in $\mathcal{S}_3$ that hit the x-axis at coordinates 4, 6, and 16 (recalling that walks only hit the x-axis at even coordinates).

The transitions on this set, commonly referred to as *single-site* (or *Glauber*) dynamics, will consist of selecting a particle at random and moving it one space to the left or right (provided the destination is unoccupied). More formally, if T and T' are two configurations, each consisting of k particles, which differ by a single particle (at distance 1), then $P_k(T,T') = \min\{1, \frac{\rho(T')}{\rho(T)}\}$. The stationary probability $\rho(T)$ of a configuration T is proportional to $C(x_1)C(x_2)\cdots C(x_{k+1})$, where x_j is the distance between particles j and $j+1$ in T. A direct coupling argument on this particle process will fail because under any coupling there will be configurations T and T' which will tend to move farther apart in expectation. Instead we consider heat bath dynamics which allows particles to move greater distances in one move.

- **Particle process with heat bath dynamics**

We examine the set $\mathcal{P}_k$ using heat bath dynamics, with a final goal of deriving a bound on the mixing time of the single-site dynamics on this set. For convenience, we add two particles at sites 0 and n (corresponding to the endpoints of the walks in S_k) and consider instead a particle model with $n+1$ sites (with coordinates 0 through n) and $k+2$ particles. Only particles in sites $1,\ldots,n-1$ are allowed to move. We still refer to this set as $\mathcal{P}_k$. We use $\sigma(j)$ to denote the location (coordinate value) of particle j in σ. So $\sigma(1)=0$ and $\sigma(k+2)=n+1$ for every $\sigma \in \mathcal{P}_k$.

The Markov chain $\tilde{P}_k$ on $\mathcal{P}_k$ has heat bath dynamics with transitions as follows: Choose a particle at random (excluding the two fixed particles). Remove this particle, and reinsert it in the interval between its two neighbors with correct conditional probabilities. That is, if σ_1, σ_2 differ solely in the location of particle j, then $\tilde{P}_k(\sigma_1,\sigma_2) = C(x_{j-1})C(x_j)/Z$, where $x_j = \sigma_2(j+1) - \sigma_2(j)$ is the distance between particles j and $j+1$. The normalizing constant $Z = \sum_{t=1}^{m-1} C(t)C(m-t)$, where $m = x_{j-1} + x_j$. If σ_1 and σ_2 differ in the position of two or more particles, then $\tilde{P}_k(\sigma_1,\sigma_2) = 0$.

Markov chain $\tilde{P}_k$ is reversible with stationary probabilities $\tilde{\rho}(\sigma) = C(x_1)C(x_2)\cdots C(x_{k+1})/\tilde{Z}_k$, where $\tilde{Z}_k = \sum C(y_1)C(y_2)\cdots C(y_{k+1})$ is the normalizing constant and the sum is over all positive solutions to $y_1 + y_2 + \ldots + y_{k+1} = n$.

Define a distance metric δ on the set of particle configurations, with $\delta(\sigma_1,\sigma_2) = \sum_{j=1}^{k} |\sigma_1(j) - \sigma_2(j)|$, the sum of distances between corresponding particles. Note that $0 \le \delta(\sigma_1,\sigma_2) \le k(n-k-1)$. We examine the heat bath dynamics using path coupling, so we must consider elements differing by unit distance.

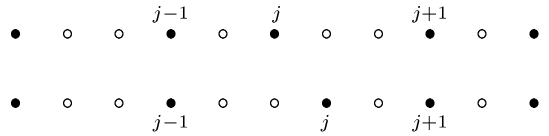

Figure 5: Typical situation for path coupling

Figure 5 shows parts of two configurations with k particles, differing only at particle j. In our coupling, if particle j is chosen in the first step of the move, we can reinsert it at the same position in each configuration, decreasing the distance by one. Also, choosing any other particle except $j-1$ or $j+1$ allows us to reinsert it at the same position in each configuration with identical probabilities, leaving the distance unchanged. So we need to consider how to couple the moves if we choose particle $j-1$ or $j+1$. As the other case is similar, consider the case when particle $j-1$ is selected. Figure 6 shows the situation with particle $j-1$ removed in each configuration, with an indication of how the moves are coupled.

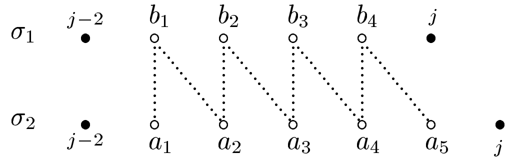

Figure 6: The coupling for the particle system

Let $m = \sigma_1(j) - \sigma_1(j-2)$. For $i \in [m-1]$ let $b_i = C(i)C(m-i)/Z_1$ be the probability to insert at position i between $j-2$ and j in the upper configuration, and for $i \in [m]$, let $a_i = C(i)C(m-i+1)/Z_2$ be the probability to insert the particle in the lower configuration σ_2, where $Z_1 = \sum_{i=1}^{m-1} C(i)C(m-i)$ is the normalizing constant for σ_1 and $Z_2 = \sum_{i=1}^{m} C(i)C(m-i+1)$ is the normalizing constant for σ_2. We have the following surprising combinatorial lemmas:

Lemma 6.3 *With a_i, b_i as above, these probabilities satisfy $a_i \le b_i$ for $i = 1, \ldots, m-1$.*

Proof. First we note that $Z_1 = \frac{2m-2}{m+2}C(m)$ and $Z_2 = \frac{2m}{m+3}C(m+1)$, obtained using the defining recurrence relation for the Catalan numbers (see, e.g., [13]). Second, observe that $\frac{C(j+1)}{C(j)} = \frac{2(2j+1)}{j+2}$ for $j \ge 0$, so this ratio increases as $j \to \infty$. We have $a_i \le b_i$ if and only if

$$\frac{C(i)C(m-i+1)}{Z_2} \le \frac{C(i)C(m-i)}{Z_1}$$
$$\Leftrightarrow \frac{C(m-i+1)}{C(m-i)} \le \frac{2m(m+2)C(m+1)}{(m+3)(2m-2)C(m)}.$$

This last inequality follows from our second observation above (i.e., $\frac{C(m-i+1)}{C(m-i)} \le \frac{C(m+1)}{C(m)}$) and from the fact that $\frac{2m(m+2)}{(m+3)(2m-2)} > 1$. Hence $a_i \le b_i$. □

Lemma 6.4 *With a_i, b_i as above, for all $k \in [m-1]$, we have*

$$\sum_{i=1}^{k} a_i \le \sum_{i=1}^{k} b_i \le \sum_{i=1}^{k+1} a_i.$$

Proof. Half of these inequalities come directly from lemma 6.3, i.e. $a_1 \le b_1$ and $a_2 \le b_2$ imply $a_1 + a_2 \le b_1 + b_2$. Similarly, $a_1 + a_2 + a_3 \le b_1 + b_2 + b_3$, etc. Then start with the final inequality of lemma 6.3, which is actually an equality, $b_1 + \cdots + b_{m-1} = a_1 + \cdots + a_m$. By the symmetry in the definition of a_i and b_i, we see that $a_{m-i+1} = a_i$ (for $i = 1, \ldots, m$), and $b_{m-i} = b_i$ (for $i = 1, \ldots, m-1$), so from lemma 6.3, $a_m \le b_{m-1}$. This implies that $b_1 + \cdots + b_{m-2} \le a_1 + \cdots + a_{m-1}$. Since $a_{m-1} \le b_{m-2}$, then $b_1 + \cdots + b_{m-3} \le a_1 + \cdots + a_{m-2}$, and so forth to give the remaining inequalities. □

Lemmas 6.3 and 6.4 allow us to couple moves in the "zig-zag" manner shown in figure 6. We reinsert particle $j-1$ at position 1 in both σ_1 and σ_2 with probability a_1. We place it at position 2 in σ_2 and at position 1 in σ_1 with probability $b_1 - a_1$, and so forth. In general, we place particle $j-1$ at position i in both with probability $\sum_{r=1}^{i} a_r - \sum_{r=1}^{i-1} b_r$, whereupon the distance between the configurations remains unchanged. We place particle $j-1$ at position i in σ_1 and $i+1$ in σ_2 with probability $\sum_{r=1}^{i} b_r - \sum_{r=1}^{i} a_r$, increasing the distance by one.

Lemma 6.5 *Let $\sigma_1, \sigma_2 \in \mathcal{P}_k$ with $\delta(\sigma_1, \sigma_2) = 1$. After one step of the coupled Markov chain, $\tilde{P}_k$, we have $E(\Delta\delta(\sigma_1, \sigma_2)) \le 0$.*

Proof. We have noted that selecting a particle other than $j-1$, j, or $j+1$ does not change the distance, and that by choosing particle j the distance decreases by one. The bad moves that increase the distance are those in which we insert $j-1$ (or $j+1$) at different positions in σ_1 and σ_2. If we select either of these two particles, say $j-1$, the expected change in distance is

$$\begin{aligned}\beta &= \sum_{i=1}^{m-1}\left(\sum_{r=1}^{i} b_r - \sum_{r=1}^{i} a_r\right)\\ &= (m-1)b_1 + (m-2)b_2 + \cdots + b_{m-1}\\ &- (m-1)a_1 - (m-2)a_2 - \cdots - a_{m-1}.\end{aligned}$$

Recalling that $b_i = b_{m-i}$ and $a_i = a_{m-i+1}$, we can also write

$$\begin{aligned}\beta &= (m-1)b_{m-1} + (m-2)b_{m-2} + \cdots + b_1\\ &- (m-1)a_m - (m-2)a_{m-1} - \cdots - a_2.\end{aligned}$$

Summing these equations and simplifying yields $\beta = \frac{1}{2}$. By symmetry this also represents the expected change in distance for particle $j+1$.

Putting these pieces together to determine the overall expected change, we find that $E(\Delta\delta(\sigma_1, \sigma_2)) \le \frac{1}{k}(-1 + \frac{1}{2} + \frac{1}{2}) \le 0$. (An inequality, as it is possible, say, that $j = 2$ so there is no bad move for the fixed particle 1.) □

Whenever $\delta(\sigma_1, \sigma_2) > 0$, the probability that the distance changes in one step is at least $1/k$. Using the path coupling theorem, we have a bound on the mixing time for the Markov chain $\tilde{P}_k$ on $\mathcal{P}_k$:

$$\begin{aligned}\tau_{\tilde{P}_k}(\varepsilon) &\le \lceil \frac{e(k(n-k-1))^2}{\frac{1}{k}} \log(\frac{1}{\varepsilon}) \rceil\\ &= \mathrm{O}(n^5 \log(\frac{1}{\varepsilon})).\end{aligned}$$

Recall that P_k, the true projection arising from partitioning $\mathcal{S}_k$, is also a Markov chain on $\mathcal{P}_k$ but with single-site dynamics using Metropolis transitions. Our analysis of the heat bath algorithm allows us to deduce that this Markov chain is also rapidly mixing.

Lemma 6.6 *The Markov chain P_k on $\mathcal{P}_k$ with single-site dynamics is mixing in polynomial time.*

This theorem follows from the Diaconis/Saloff-Coste comparison method (see [2, 9]); we leave these details for the full paper.

We now have shown in lemmas 6.2 and 6.6 that the restrictions defined by the decomposition $\mathcal{S}_k = \dot{\cup}_T \mathcal{S}_{k,T}$, as well as the projection, are all mixing in polynomial time. Appealing to corollary 4.4 with $\alpha = \frac{1}{8n}$ and $\beta = \frac{1}{4}$, we find

Lemma 6.7 *The Markov chain $\widehat{\mathcal{M}}$ on $\mathcal{S}_k$ is rapidly mixing.*

6.3. Mixing for $\mathcal{S}$: The final word

A polynomial bound on the mixing time for $\widehat{\mathcal{M}}$ now follows from all of our previous work. By lemma 6.7 the restrictions to each set $\mathcal{S}_k$ are all rapidly mixing, and the mixing time of the projection followed from the log-concavity of the sets $\mathcal{S}_k$ (lemma 5.4). Using corollary 4.4, we can bound the mixing time of the chain $\widehat{\mathcal{M}}$ on all of $\mathcal{S}$. Note that in this case we have $\alpha = \frac{1}{4n(\lambda+1)}$ and $\beta = \frac{1}{\lambda n}$.

Theorem 6.8 *The Markov chain $\widehat{\mathcal{M}}$ on $\mathcal{S}$ is rapidly mixing.*

Finally, a simple application of the comparison theorem establishes a polynomial bound for the mixing time of the original simpler Markov chain $\mathcal{M}$ on $\mathcal{S}$. The full analysis of all of these steps will be included in the final version of this paper.

Acknowledgements

The authors are very grateful to Alan Sokal for observing that theorem 4.1 could be used to greatly simplify the statement and proof of theorem 4.2. We also thank Claire and Rick Kenyon for useful discussions.

References

[1] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. *Proc. 38th Annual IEEE Symposium on Foundations of Computer Science* 223–231, 1997.

[2] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability*, **3**:696–730, 1993.

[3] M. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures and Algorithms*, **13**:285–317, 1998.

[4] M.R. Jerrum and A.J. Sinclair. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, **82**:93–133, 1989.

[5] M. Luby, D. Randall, and A.J. Sinclair. Markov chain algorithms for planar lattice structures. *Proc. 36th IEEE Symposium on Foundations of Computer Science*, 150–159, 1995.

[6] N. Madras and D. Randall. Factoring graphs to bound mixing rates. *Proc. 37th Annual IEEE Symposium on Foundations of Computer Science*, 194–203, 1996.

[7] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, (to appear).

[8] L. McShine and P. Tetali. On the mixing time of the triangulation walk and other Catalan structures. *Randomization methods in algorithm design*, 147–160, *DIMACS Series on Discrete Mathematics and Theoretical Computer Science*, **43**, 1999.

[9] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics*, **41**:1598–1615, 2000.

[10] B. van Rensburg. Adsorbing staircase walks and staircase polygons. *Annals of Combinatorics*, **3**:451–473, 1999.

[11] E.J.J. van Rensburg. Collapsing and adsorbing polygons. *Journal of Physics A: Mathematical and General*, **31**:8295–8306, 1998.

[12] A.J. Sinclair. *Algorithms for random generation & counting: a Markov chain approach.* Birkhäuser, Boston, 1993.

[13] R. Stanley. *Enumerative Combinatorics*, Vol. II, Cambridge University Press, 1999.

[14] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. Preprint, 1997.