Mixing Points on an Interval

Dana Randall *   Peter Winkler †

Abstract

We determine the mixing time (up to a constant factor) of the Markov chain whose state space consists of n "dots" on the unit interval, wherein a dot is selected uniformly at random and moved to a uniformly random point between its two neighbors. The method involves a two-step coupling for the upper bound, and an unusual probabilistic second-moment argument for the lower.

1 Introduction

Determination of mixing time for discrete Markov chains has proved to be a vital step in proving the efficiency of randomized approximation algorithms, with a bewildering variety of applications. One of the best sources for sampling problems is statistical physics where the state space of a Markov chain represents states of a physical system and sampling lends insight into the thermodynamic properties of the model. Although these models typically are discretized for the purposes of analysis, in reality physicists are often interested in continuous analogues defined in real Euclidean spaces.

For example, the so-called "microcanonical hard-core lattice gas model" consists of all configurations of n balls of radius r in a finite region R of $\mathbf{R}^n$, where balls are required to lie in non-overlapping positions. Although the centers of the balls can occupy any real point in $\mathbf{R}^n$, the model is typically discretized so that they lie on lattice points in $\Lambda_n \subseteq \mathbf{Z}^n$. In this setting, if r is chosen to be between $1/2$ and $\sqrt{2}/2$, then configurations correspond to independent sets on a lattice graph. While much analysis has focused on rigorous means of efficiently sampling discrete models such as independent sets in Λ_n, very little is known about sampling in the continuous setting.

The most powerful tools to date for sampling points in a continuous state space are walks used to estimate the volume of a convex body in $\mathbf{R}^n$ (see, e.g., [2], [3], [5], [6]). Notice that these walks provide a way to sample from the one-dimensional hard-core model where the balls lie in $R \subset \mathbf{R}$ because each configuration can be mapped to a point in the unit simplex. An example is the "hit-and-run" chain which preforms a random walk on points in the convex body as follows. If currently at a point v, we choose a random line through v and move to a new point chosen uniformly from the part of the line that lies inside the convex body. Lovász showed that this random walk converges in time $O^*(n^3)$, where the O^* notation suppresses log factors, and he conjectures that the true running time is much faster [5]. Unfortunately these random walks on the simplex are not very natural in the context of the hard-core model since they typically move all the balls in a single step. Our interest in this paper is studying a Markov chain that moves one ball at a time since most chains used in practice are local.

Alternative chains for sampling sets of points were considered in [4]. The first variant is a single particle, global chain that connects all pairs of states that differ by the position of one point. It is shown that the chain is rapidly mixing in $\mathbf{R}^n$ if the density of points is sufficiently low. The second variant examined is a single particle, local chain where steps involve moving one point to a new position in its current neighborhood. Although it is believed that this chain mixes rapidly in many settings, it was only shown that a discretized version of the chain mixes rapidly in one dimension. In this chain, the unit interval is subdivided into m discrete points and at each step one of the n particles is chosen and moved to an adjacent point, if that point is currently unoccupied. The mixing time for this chain was shown to be $O(n^3m^2(1-\rho)^2)$, where ρ is the density of particles.

In fact, most methods used to analyze discrete chains seem to break down in the continuous setting, explaining the scarcity of results. This seems to be more an artifact of the appearance of the "minimum stationary probability" parameter which often arises in the discrete cases, and not anything inherent in continuous chains which often seem to be efficient.

Here, we tackle a simple and natural chain whose stationary state is uniform on the unit n-simplex, that is, the set of ascending sequences of n points ("dots") in the unit interval. The chain keeps the dots in fixed relative order; moves consist of moving a random dot

*College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280

†Department of Mathematics, Dartmouth, Hanover, NH 03755-3551

to a point chosen uniformly at random between its neighbors. Notice that this can be viewed as variant on the "hit-and-run" chain on the simplex where the random line in which a move is made must be parallel to the coordinate axes. (This variant has been called the "rook's move" chain, as opposed to the "queen's move" in which any line through the current point is available.)

We formalize the moves of the "n dots on an interval" chain in the next section. Our main result is as follows.

THEOREM 1.1. *The mixing time of the "n dots on an interval" chain is $\Theta(n^3 \log n)$.*

Our methods extend typical discrete methods in two ways: by using two phases to couple, and by controlling a random process whose variance sometimes (but sufficiently rarely) gets out of hand.

2 n Dots on an Interval

To avoid confusion with other points on the interval, the n designated points defining the chain will be called *dots* and the chain itself will be called "n dots on an interval." The position of the kth dot at time t will be denoted by $x_k(t)$ or, if the time is understood, by x_k. The state space S of the Markov chain then consists of the simplex of points $\bar{x} = (x_1, \ldots, x_n) \in [0,1]^n$ for which $x_1 \le x_2 \le \cdots \le x_n$.

In the related hard-core model from statistical physics, the state space is actually n non-overlapping balls of radius r, centered at $y_1, \ldots, y_n$. Notice that there is a bijection between the sets of balls and the sets of dots defined by mapping the ith ball to $x_i = (y_i + (2i-1)r)/s$, where $s = 1 + 2nr$. In the remainder of the paper we refer to the configuration of dots, but the theorems apply to the hard-core balls as well.

A step of the chain is determined by choosing a uniformly random integer k between 1 and n and an independent real λ chosen from the Lebesgue distribution on $[0,1]$. If the previous state of the chain was $\bar{x}$, the new state will be

$$\bar{x}' = (x_1, \ldots, x_{k-1}, (1-\lambda)x_{k-1} + \lambda x_{k+1}, x_{k+1}, \ldots, x_n)$$

where we interpret $x_0 = 0$ and $x_{n+1} = 1$; in other words, if $k = 1$, $x'_1 = \lambda x_2$ and if $k = n$, $x'_n = (1-\lambda)x_{n-1} + \lambda$ with the remaining coordinates unchanged.

THEOREM 2.1. *The stationary state of the "n dots on an interval" chain is uniform on the simplex $\{\bar{x} : x_1 \le x_2 \le \cdots \le x_n\}$.*

Proof. This follows from detail balance (or simply noticing that the position of x_k in a uniform point from the simplex is indeed uniformly random between x_{k-1} and x_{k+1}.)

It is convenient here to define the *mixing time* of a Markov chain $\mathbf{X}$ to be the least number t of steps such that, beginning at any state, the total variation distance $\frac{1}{2}\|\mathbf{X}(t) - \sigma\|_1$ between the state of $\mathbf{X}$ at time t and the stationary distribution σ of $\mathbf{X}$ is less than $\frac{1}{4}$.

3 The Upper Bound

We will prove Theorem 1.1 in two stages, starting here with the upper bound.

LEMMA 3.1. *The mixing time of the "n dots on an interval" chain is $O(n^3 \log n)$.*

Proof. It suffices to exhibit a coupling which will achieve coalescence of any chain $\mathbf{X}$ with a stationary chain in time $\mathrm{O}(n^3 \log n)$, with probability at least $\frac{3}{4}$. Our coupling proceeds in two phases, first bringing the dots of one chain close to the corresponding dots of the other, then aligning them exactly.

The natural coordinate-wise partial order imposes a lattice structure on S, where, for example, the join of $\bar{x}$ and $\bar{y}$ is $(\max(x_1, y_1), \ldots, \max(x_n, y_n))$. If chains $\mathbf{X}$ and $\mathbf{Y}$ are run in parallel and coupled so as to employ the same k and λ at each step, then it is immediate that $\bar{x} \ge \bar{y}$ implies $\bar{x}' \ge \bar{y}'$; i.e., this "linear" coupling respects the partial order. We will use the linear coupling for the first phase and a second coupling, which also respects the partial order, for phase two.

Instead of coupling an arbitrary chain with a stationary chain, we simultaneously couple chains beginning at *every* possible state, keeping track of chains $\mathbf{X}$ and $\mathbf{Y}$ which begin, respectively, at the extreme states $\bar{1}$ and $\bar{0}$. (This is sometimes referred to as a "complete coupling.") When $\mathbf{X}$ and $\mathbf{Y}$ reach the same state, all other pairs of chains are unified as well since they are sandwiched between these two. The advantage of having to worry only about $\mathbf{X}$ and $\mathbf{Y}$ is that they are comparable in the partial order; thus we will always have each dot of $\mathbf{X}$ to the right of the corresponding dot of $\mathbf{Y}$.

For the first phase, we make use of a "gap function" of the type employed by David Wilson [8, 9]; later we will use a similar function to bound the mixing time from below. For convenience we define $\bar{z} := \bar{x} - \bar{y}$ where $\bar{x}$ and $\bar{y}$ are the states of $\mathbf{X}$ and $\mathbf{Y}$; then always $\bar{z} \ge \bar{0}$. We let $\bar{z}' := \bar{x}' - \bar{y}'$ represent the updated vector after a step of the coupled chains. For notational convenience, we define $z_0 = z_{n+1} = 0$.

Note that over all choice of k and λ, the expected value $\mathrm{E}[x'_k]$ of the kth coordinate of $\mathbf{X}$ will be $\frac{n-1}{n}x_k +$

$\frac{1}{n}(x_{k-1}+x_{k+1})/2$. Define

$$f(\bar z)=\sum_{k=1}^{n}\sin\big(\frac{k\pi}{n+1}\big)z_k$$

and observe that after a coupled step of the chains,

$$\begin{aligned}
\mathrm{E}[f(\bar z')] &= \sum_{k=1}^{n}\sin\big(\frac{k\pi}{n+1}\big)\Big(\frac{n-1}{n}z_k+\frac{1}{n}\cdot\frac{z_{k-1}+z_{k+1}}{2}\Big)\\
&= \sum_{k=1}^{n}\Big(\frac{n-1}{n}\sin\big(\frac{k\pi}{n+1}\big)+\frac{1}{2n}\sin\big(\frac{(k-1)\pi}{n+1}\big)\\
&\qquad +\sin\big(\frac{(k+1)\pi}{n+1}\big)\Big)z_k\\
&= \sum_{k=1}^{n}\Big(\frac{n-1}{n}\sin\big(\frac{k\pi}{n+1}\big)+\frac{1}{n}\cos\big(\frac{\pi}{n+1}\big)\sin\big(\frac{k\pi}{n+1}\big)\Big)z_k\\
&= \Big(\frac{n-1}{n}+\frac{\cos(\frac{\pi}{n+1})}{n}\Big)\sum_{k=1}^{n}\sin\big(\frac{k\pi}{n+1}\big)z_k\\
&= \Big(1-\frac{1}{n}\Big(1-\cos\big(\frac{\pi}{n+1}\big)\Big)\Big)f(\bar z)\ .
\end{aligned}$$

Thus the expected value of f has dropped by a factor of $1-\gamma_n$ where

$$\gamma_n=\frac{1}{n}\Big(1-\cos\big(\frac{\pi}{n+1}\big)\Big)>\frac{\pi^2}{2}\frac{1}{(n+1)^3}.$$

If we set $\bar x(0)=\bar 1$ and $\bar y(0)=\bar 0$, we have $\bar z(0)=\bar 1$ and, by concavity of the sine function between 0 and π,

$$f(\bar z(0))<(n+1)\int_0^1\sin\pi u\,du=\frac{2(n+1)}{\pi}.$$

If we run the coupled chains for $t_1=12((n+1)^3/\pi^2)\log n$ steps (where all logs are base e here), we have

$$\begin{aligned}
\mathrm{E}[f(\bar z(t_1))] &< (1-\gamma_n)^{t_1}\frac{2(n+1)}{\pi}\\
&< \Big(1-\frac{\pi^2}{2}\frac{1}{(n+1)^3}\Big)^{t_1}\cdot\frac{2n}{\pi}\\
&< e^{-6\log n}\frac{2n}{\pi}\\
&= 2n^{-5}/\pi.
\end{aligned}$$

Since $f(\bar z)$ can never be negative, we see using Markov's inequality that with probability at least .8, $f(\bar z(t_1))<10n^{-5}/\pi$.

The purpose of obtaining this bound is to shrink the distance $z_k=x_k-y_k$ for each k. Since all z_k's are positive and the smallest coefficient in f is $\sin(\frac{\pi}{n+1})>\frac{\pi}{n+2}$, we have (with probability at least .8)

$$z_k<\frac{10n^{-5}}{\pi}\cdot\frac{n+2}{\pi}<1.1\,n^{-4}$$

for sufficiently large n. For convenience we set $\delta:=1.1n^{-4}$ so that we now have $0\le x_k-y_k\le\delta$ for all k—again with probability at least .8.

For the second phase of the coupling, we choose at each step a dot k and *two* independent, uniform reals λ and μ from the unit interval. If it happens that $x_{k-1}>y_{k+1}$ (this will be a rare event), we ignore μ and execute a step of the linear coupling employed in phase one.

Otherwise, consider a chain $\mathbf{W}$ caught between $\mathbf{X}$ and $\mathbf{Y}$, currently in state $\bar w=(w_1,\dots,w_n)$. We divide the interval between w_{k-1} and w_{k+1} into three parts: $I:=[w_{k-1},x_{k-1}]$; $J:=[x_{k-1},y_{k+1}]$; and $K:=[y_{k+1},w_{k+1}]$. Here, as before, we adhere to the convention that $x_0=y_0=0$ and $x_{n+1}=y_{n+1}=1$ so that we don't have to handle the cases $k=1$ and $k=n$ separately. We will use μ to decide which subinterval to place w'_k into, then λ to determine the particular spot.

Accordingly, let $p_{\mathbf{W}}:=|I|/(|I|+|J|+|K|)$ and $q_{\mathbf{W}}:=(|I|+|J|)/(|I|+|J|+|K|)$. If $\mu<p_{\mathbf{W}}$ we put $w'_k=(1-\lambda)w_{k-1}+\lambda x_{k-1}$; if $p_{\mathbf{W}}\le\mu\le q_{\mathbf{W}}$ we put $w'_k=(1-\lambda)x_{k-1}+\lambda y_{k+1}$; and if $\mu>q_{\mathbf{W}}$ we put $w'_k=(1-\lambda)y_{k+1}+\lambda w_{k+1}$. It is easy to verify that this coupling is faithful to the individual chains and preserves the partial order.

Among all possible $\mathbf{W}$, the one maximizing $p_{\mathbf{W}}$ will be $\mathbf{Y}$, with $p_{\mathbf{Y}}=(y_{k+1}-x_{k-1})/(y_{k+1}-y_{k-1})$; and the one minimizing $q_{\mathbf{W}}$ will be $\mathbf{X}$, with $q_{\mathbf{X}}=(y_{k+1}-x_{k-1})/(x_{k+1}-x_{k-1})$. It follows that with probability $q_{\mathbf{X}}-p_{\mathbf{Y}}$, namely when $p_{\mathbf{Y}}\le\mu\le q_{\mathbf{X}}$, w'_k will be the same for every chain; in this case we say we have achieved (or possibly only preserved) a "match" at k.

Sandwiched between chains $\mathbf{X}$ and $\mathbf{Y}$ is a stationary chain $\mathbf{U}$, whose state at any time is distributed as a uniformly random point $(u_1,\dots,u_n)$ chosen from the unit hypercube $[0,1]^n$, with coordinates reordered to be ascending. Let $\varepsilon=(n/\log n)^{-3}$; we claim that with probability tending to one as n grows, at no time between $t_1=10((n+1)^3/\pi^2)\log n$ and $t_2:=t_1+n(\log n)^2$ (the era of phase two of the coupling) will there ever be a k such that u_{k-1} and u_{k+1} fall within distance ε of each other.

To see this, note that such an accident (at any particular time), for $k\in[2,n-1]$, implies that for some $i\in\{1,2,\dots,\lfloor 1/\varepsilon\rfloor\}$ three of our random points from the hypercube landed in the interval $[(i-1)\varepsilon,(i+1)\varepsilon]$.

The probability of such an event is less than

$$\binom{n}{3}(2\varepsilon)^3(1/\varepsilon) < \tfrac{4}{3}n^3\varepsilon^2 = \tfrac{4}{3}n^{-3}(\log n)^6.$$

The cases $k = 1$ and $k = n$ require two points falling in $[0, \varepsilon]$ or in $[1-\varepsilon, 1]$, respectively, adding only the ignorable amount

$$2\binom{n}{2}\varepsilon^2 < n^2\varepsilon^2 = n^{-4}(\log n)^6.$$

The probability that such an accident occurs anywhere in $n(\log n)^2$ steps of $\mathbf{U}$ is thus bounded by

$$\tfrac{4}{3}n^{-3}(\log n)^6 \cdot n(\log n)^2 = \tfrac{4}{3}n^{-2}(\log n)^8,$$

which tends to zero as claimed.

We now have the tools to show that $q_x - p_y$ is very close to 1—in fact, so close that with probability tending to 1, every step in the second phase of the coupling will result in creating or preserving a match. Any step that fails to perfectly align these points we call a mismatch.

Fix some time t between t_1 and t_2 and assume that no mismatch has occurred so far. In that case, all corresponding pairs of dots from $\mathbf{X}$ and $\mathbf{Y}$ are either at the same point or (if k hasn't yet been "called") exactly where they were at time t_1, which was within distance δ. Examining first the denominator of p_y,

$$\begin{aligned} y_{k+1} - y_{k-1} &> -(u_{k+1} - y_{k+1}) + (u_{k+1} - u_{k-1}) \\ &\qquad +(u_{k-1} - y_{k-1}) \\ &> 0 + \varepsilon - \delta \end{aligned}$$

and thus

$$\begin{aligned} p_y &< \frac{y_{k+1} - x_{k-1}}{\varepsilon - \delta} < \frac{\delta}{\varepsilon - \delta} \\ &< \frac{1.1\, n^{-4}}{(n/\log n)^{-3} - 1.1\, n^{-4}} < \frac{1.2}{n(\log n)^3} \end{aligned}$$

for large n. A similar argument gives the same bound for $1 - p_x$.

We thus have that with probability at least

$$1 - \Big(\frac{2.4}{n(\log n)^3}\Big)\big(n(\log n)^2\big) = 1 - \frac{2.4}{n}$$

every step of phase two is a match. Since $n(\log n)^2$ steps is enough to ensure that all n values of k are selected at least once (being one log factor greater than the expected number of steps for the "coupon collector's problem") we have perfectly aligned every pair of points with probability almost .8, proving the upper bound.

Remark. The same argument can be used to provide an upper bound on the mixing time of a discrete version of the "n dots on an interval" chain. In this formulation, we consider configurations consisting of n non-overlapping dots on the interval $\{1, \ldots, m\}$, with $m \geq n$. These configurations correspond precisely to r-combinations, representing the distinct ways of putting n indistinguishable objects into $r = m - n$ bins.

The discrete version of the "n dots on an interval" chain is a well-known one-dimensional particle process defined as follows. At each step we choose $k \in \{1, \ldots, n\}$ and move x_k to any of the unoccupied sites between x_{k-1} and x_{k+1} (where we fix $x_0 = 0$ and $x_{n+1} = m+1$). A local version of this chain in which dots are moved distance at most one in each step was studied previously by Kannan et al. and gives a bound of $O(n^3(m-n)^2)$ [4]. Their coupling argument can be modified for the "n dots on an interval" chain as well, giving the same upper bound of $O(n^3)$ when m–n is constant. In the case of general m, we can improve the upper bound on "n points on an interval" by using the same eigenfunction f defined in lemma 3.1. This gives an upper bound of $O(n^3 \log n)$ for the mixing time of this chain, removing the dependence on m altogether.

4 The Lower Bound

We conclude the proof of Theorem 1.1 by showing the matching lower bound.

Lemma 4.1. *The mixing time of the "n dots on an interval" chain is $\Omega(n^3 \log n)$.*

Proof. It is relatively easy to show that the mixing time is $\Omega(n^3)$; the extra log factor is more subtle. To get the weaker bound, we start from state $\bar{1}$ and show that after $t_3 = n^3$ steps, the event $A :=$ "$f(2\bar{x}(t_3) - \bar{1}) < 0$" has probability less than $\frac{1}{4}$. Since by symmetry A has probability $\frac{1}{2}$ for $\bar{x}$ drawn from the stationary distribution σ, the event A witnesses the fact that $\mathbf{X}(t_3)$ is at total variation distance greater than $\frac{1}{4}$ from σ.

Letting $\bar{z} := 2\bar{x} - 1$ instead of $\bar{x} - \bar{y}$, we still have $\mathrm{E}[f(\bar{z}')] = (1 - \gamma_n)f(\bar{z})$, this time needing an upper bound for γ_n:

$$\gamma_n = \frac{1}{n}\Big(1 - \cos\big(\frac{\pi}{n+1}\big)\Big) < \frac{\pi^2}{2}n^{-3}.$$

Then

$$\begin{aligned} \mathrm{E}[f(\bar{z}(t_3))] &> (1 - \gamma_n)^{t_1} f(\bar{z}(0)) \\ &> \Big(1 - \frac{\pi^2}{2}n^{-3}\Big)^{n^3} f(\bar{z}(0)) \\ &> e^{-2/9} f(\bar{z}(0)) \\ &> .75 f(\bar{z}(0)) \end{aligned}$$

for large n. Since $f(\bar{z}(t_3))$ cannot exceed $f(\bar{z}(0))$, the probability that it is below zero at time t_3 is less than $\frac{1}{4}$, as desired.

To get the extra log factor into t to establish tight upper and lower bounds, we must control the variance of $f(\bar{z}(t))$ to conclude that the function stays fairly close to its mean. Let $R(t) := \mathrm{E}\big[(f(t{+}1) - f(t))^2\big]$, with R an upper bound for $R(t)$. Lemma 5 of Wilson [9] gives a handy quantitative formulation of this second-moment method, which has been used in the past by Diaconis and Shashahani [1] and Lee and Yau [7] to get lower bounds for mixing times.

For our purposes Wilson's lemma says the following.

Lemma [9]: *If a function Φ on the state space of a Markov chain satisfies $\mathrm{E}[\Phi(\bar{x}(t+1))|\bar{x}(t)] = (1-\gamma)\Phi(\bar{x}(t))$, and $\mathrm{E}[(\Phi(\bar{x}(t+1)) - \Phi(\bar{x}(t)))^2] \le R$, then after fewer than*

$$\frac{\log \Phi_{\max} + \frac{1}{2}\log\frac{\gamma\varepsilon}{4R}}{-\log(1-\gamma)}$$

the variation distance from stationarity exceeds $1-\varepsilon$.

Substituting $f(\bar{z}) = f(2\bar{x} - \bar{1})$ for $\Phi(\bar{x})$, $\frac{\pi^2}{2}n^{-3}$ for γ, $f(\bar{1}) \sim 2n/\pi$ for $\Phi_{\max}$, and $2/3$ for ε, we see that the desired log factor will appear in t provided $R = \mathrm{O}(n^{-1-\delta})$ for some $\delta > 0$.

Alas, it appears that our $R(t)$ can be as large as $\Theta(1/n)$. If dot k is chosen for adjustment at time t, it will move to the right with probability $\frac{x_{k+1}-x_k}{x_{k+1}-x_{k-1}}$, causing an expected square increment of $c_k^2(z_{k+1} - z_k)^2/3$, where c^k is the kth coefficient of f; moving left gives $c_k^2(z_k - z_{k-1})^2/3$. Hence, altogether, the expected squared change in $f(\bar{z})$ is

$$\frac{1}{n}\sum_{k=1}^{n}\frac{c_k^2}{3}g(k),$$

where

$$g(k) = \Big(\frac{(z_{k+1}-z_k)^3 + (z_k - z_{k-1})^3}{z_{k+1}-z_{k-1}}\Big)$$

$$= \Big((z_{k+1}-z_k)^2 - (z_{k+1}-z_k)(z_k-z_{k-1}) + (z_k-z_{k-1})^2\Big).$$

In the worst case imaginable, $x_1, \ldots, x_m$ are near 0 and the rest of the dots near 1, for m around $n/2$; then c_m and c_{m+1} are close to 1 and $R(t) \sim \frac{1}{n}(\frac{1}{3}2^2 + \frac{1}{3}2^2) = 8/3n$.

But is this sort of unpleasantness likely? Early on, when dots are just beginning to peel away from 1, they are spread out causing $x_k - x_{k-1}$ to have size of order 1, but these are small k's with correspondingly small c_k's. In fact, we will show that $R(t) = \mathrm{o}(1/n)$ at all times with high probability. We still won't be able to use Wilson's lemma directly, but we can instead adapt the proof.

It turns out to be useful to reconsider the linear coupling, this time between $\mathbf{X}$ and a stationary chain $\mathbf{U}$. Let us note first that at all times, x_i/u_i is decreasing (not necessarily strictly) in k for $k = 1, \ldots, n+1$. Certainly it starts that way; now observe that for $k > 1$,

$$\frac{x_{k-1}}{u_{k-1}} \ge \frac{x_{k+1}}{u_{k+1}}$$

implies

$$\frac{x_{k-1}}{u_{k-1}} \ge \frac{(1-\lambda)x_{k-1} + \lambda x_{k+1}}{(1-\lambda)u_{k-1} + \lambda u_{k+1}} \ge \frac{x_{k+1}}{u_{k+1}}.$$

When the first dot is chosen, i.e., $k = 1$, we get equality since $x_1'/u_1' = \lambda x_2/\lambda u_2 = x_2/u_2$. We conclude that the sequence x_i/u_i remains decreasing, and in particular, throughout the coupling, $x_{k-1}/x_k \ge u_{k-1}/u_k$ for all k. Thus $\frac{x_k - x_{k-1}}{x_k} \le \frac{u_k - u_{k-1}}{u_k}$ and

$$x_k - x_{k-1} \le x_k\frac{u_k - u_{k-1}}{u_k} \le \frac{u_k - u_{k-1}}{u_k}.$$

Now, given the value of u_k, the dots $u_1, \ldots, u_{k-1}$ are uniformly distributed between 0 and u_k; thus if $k = n^\alpha$,

$$\begin{aligned}
&\Pr\Big(x_k - x_{k-1} > n^{-\alpha}(\log n)^5\Big)\\
&< \Pr\Big(\frac{u_k - u_{k-1}}{u_k} > n^{-\alpha}(\log n)^5\Big)\\
&= \big(1 - n^{-\alpha}(\log n)^5\big)^{n^\alpha - 1}\\
&\sim \exp(-(\log n)^5) = n^{-5}.
\end{aligned}$$

Over the course of $t_4 = \frac{1}{10}n^3\log n$ steps, there are fewer than $n^4 \log n$ opportunities for such a "big gap" to appear so with probability at least $1 - \log n/n$, we will always have $x_k - x_{k-1} < n^{-\alpha}(\log n)^5$ where $\alpha = \log_n k$.

Recalling that the coefficient $c_k = \sin\frac{k\pi}{n+1} < \frac{k\pi}{n+1}$, we have that when $k = n^\alpha$, with probability as large as we wish, the kth term

$$\frac{c_k^2}{12}\Big((z_{k+1}-z_k)^2 - (z_{k+1}-z_k)(z_k-z_{k-1}) + (z_k-z_{k-1})^2\Big)$$

of $R(t)$ is bounded by

$$\big(\frac{\pi n^\alpha}{n+1}\big)^2\big(2n^{-\alpha}(\log n)^5\big)^2 < 4\pi^2(\log n)^{10}n^{-2}.$$

Since $R(t)$ is the mean of these terms, it is itself bounded by $R = 4\pi^2(\log n)^{10}n^{-2}$, which is small enough for our purposes with a factor of $n^{1-\delta}$ to spare.

We now define a sequence of random variables $\Phi(t)$ by $\Phi(t) := f(\bar{x}(t))$ for $t \le \mathbf{T}$, where $\mathbf{T}$ is the first time a bad gap appears. For $t > \mathbf{T}$, we put $\Phi(t) := (1-\gamma)^{t-\mathbf{T}}\Phi(\mathbf{T})$.

Now we have $\mathrm{E}[(\Phi(t+1)-\Phi(t))^2] \le R$ at all times, since after time $\mathbf{T}$ it is just $\gamma^2\Phi(t)^2 \le \gamma^2 f(\bar{1})$. We still cannot apply Wilson's lemma, since Φ is not a function of the state of $\mathbf{X}$, but we mimic part of its proof as follows.

We still have that $\mathrm{E}[\Phi(t)] = (1-\gamma)^t\Phi(0)$ for all t, and letting $\Delta\Phi$ denote $\Phi(t+1)-\Phi(t)$, we have

$$\Phi(t+1)^2 = \Phi(t)^2 + 2\Phi(t)\Delta\Phi + (\Delta\Phi)^2,$$

and hence

$$\begin{aligned}\mathrm{E}[\Phi(t+1)^2|\Phi(t)] &= (1-2\gamma)\Phi(t)^2 + \mathrm{E}[(\Delta\Phi)^2|\Phi(t)]\\ &\le (1-2\gamma)\Phi(t)^2 + R,\end{aligned}$$

so by induction,

$$\mathrm{E}[\Phi(t)^2] \le (1-2\gamma)^t\Phi(0)^2 + \frac{R}{2\gamma}.$$

Therefore

$$\begin{aligned}\mathrm{Var}[\Phi(t)] &= \mathrm{E}[\Phi(t)^2] - \mathrm{E}[\Phi(t)]^2\\ &\le \big((1-2\gamma)^t - (1-2\gamma)^{2t}\big)\Phi(0)^2 + \frac{R}{2\gamma}\\ &\le \frac{R}{2\gamma}.\end{aligned}$$

We then have by Chebyshev's inequality that

$$\Pr[|\Phi(t) - \mathrm{E}[\Phi(t)]| \ge \sqrt{5R/2\gamma}] \le \frac{1}{5}.$$

For $t_4 = \frac{1}{10}n^3\log n$ we have

$$\begin{aligned}\mathrm{E}[\Phi(t_4)] &\ge (1-\gamma)^{t_4}\frac{2n}{\pi}\\ &\ge \big(1-\frac{\pi^2}{2}n^{-3}\big)^{n^3\log n/10}\frac{2n}{\pi}\\ &\ge \exp(-2/\pi^2)\frac{2}{\pi}n^{1-\pi^2/20}\end{aligned}$$

which is order $n^{0.5065\ldots}$, while

$$\sqrt{R/10\gamma} \le \sqrt{\frac{20\pi^2(\log n)^{10}n^{-2}}{\pi^2 n^{-3}}} = \sqrt{20}(\log n)^5 n^{1/2}.$$

Thus, in particular, the probability that $\Phi(t_4)$ is positive is at least $4/5$ for n large.

We claim now that the event "$f(\bar{x}(t_4)) > 0$" witnesses the fact that the total variation between $\mathbf{X}(t_4)$ and the stationary distribution is $1/4$ or more. We only need to make $\Pr[\mathbf{T} < t_4] \le 1/20$ to ensure that

$$\Pr[f(\bar{x}(t_4)) > 0] \ge \Pr[\Phi(t_4) > 0] - \frac{1}{20} \ge \frac{3}{4}$$

exceeding $\Pr[f(\bar{u}(t_4)) > 0] = 1/2$, where $\mathbf{U}$ is stationary, by at least $1/4$.

5 Remarks

We have actually shown that the mixing time of "n dots on an interval" lies between $0.1n^3\log n$ and $1.2158542n^3\log n$, asymptotically, with room for improvement in both constants. Experience with other Markov chains suggests that there is probably some constant c such that the time to reach total variation ε is $(c-\mathrm{o}(1))n^3\log n$ for any small ε.

There are some variations of "n dots on an interval" that are natural to consider: for example, "n dots on a circle" (labelled or not) and similar chains where a point is first selected uniformly at random, and then the dot to its right or left is moved to that point. These chains are also under study by the authors, but appear to require different methods from the chain considered here; results will appear in subsequent papers.

References

[1] P. Diaconis and M. Shashahani, Time to reach stationarity in the Bernoulli-Laplace diffusion model, *SIAM J. Math. Anal.* **81** #1 (1987), 208–218.

[2] M. Dyer, A. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex sets, *J. for the Association for Computing Machinery* **38** (1991), 1–17.

[3] R. Kannan, L. Lovász, and M. Simonovits, Random walks and an $O^*(n^5)$ volume algorithm, *Random Structures and Algorithms* **11** (1997), 1–50.

[4] R. Kannan, M.W. Mahoney, and R. Montenegro, Rapid mixing of several Markov chains for a hard-core model, *Proc. 14th Annual ISAAC* (2003), 663–675.

[5] L. Lovász, Hit-and-run mixes fast, *Math. Prog.* **86** (1998), 443–461.

[6] L. Lovász and S. Vempala, Logconcave Functions: Geometry and Efficient Sampling Algorithms, *Proc. 44th Annual FOCS,* (2003), 640–649.

[7] T.-Y. Lee and H.-T. Yau, Logarithmic Sobolev inequalities for some models of random walks, Ann. Prob. **26** #4 (1998), 1855–1873.

[8] D.B. Wilson, Mixing times of lozenge tiling and card shuffling Markov chains, *Annals of Appl. Probab.* **14** (2004), 274–325.

[9] D.B. Wilson, Mixing time of the Rudvalis shuffle, *Elect. Comm. in Probab.* **8** (2003), 77–85.