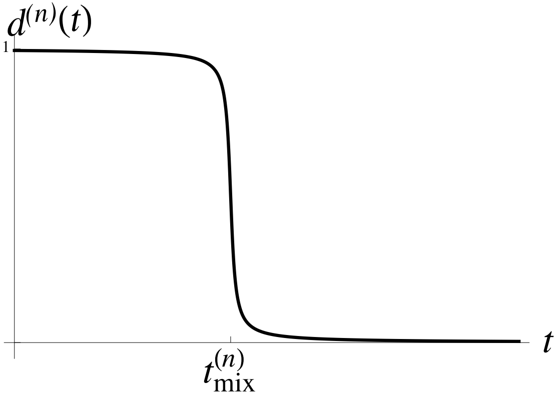
Cutoff for Markov Chains

David A. Levin

University of Oregon

LMS - EPSRC Durham Symposium 2017

Distance to equilibrium, a cartoon of cutoff:



These slides are at

`http://pages.uoregon.edu/dlevin/TALKS/durham.pdf`

In particular this document contains a bibliography.

The cutoff phenomenon for *families* of Markov chains was first identified in the groundbreaking works of Diaconis, Shahshahani and Aldous in the 1980's.

In 1996, P. Diaconis wrote:

> *At present writing, proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors.*

(From P. Diaconis (1996). "The cutoff phenomenon in finite Markov chains". In: *Proc. Nat. Acad. Sci. U.S.A.* 93.4, pp. 1659–1664. DOI: 10.1073/pnas.93.4.1659.)

Nonetheless, here, I will focus on examples where cutoff can be proved via probabilistic arguments, generally not requiring detailed analysis of eigenvalues and eigenvectors.

Early (and current) work focused on random walks on symmetric group (shuffling), and other groups.

See talks by Bernstein and Nestoridi for recent cutoff breakthroughs on such walks!

Also: see talk by Hermon for a different kind of example where the degree is bounded.

Plan

- Definitions, simple examples of cutoff and non-cutoff
- "Product condition" conjecture and counterexamples, product chains
- Strong stationary times
- Phase transition phenomenon for Glauber dynamics (some details for the mean-field case) and role of cutoff
- Bounded degree examples, path coupling and the biased exclusion process

Part I: Introduction and simple examples

- Definitions
- Coupling
- Lower bounds
- Examples

Most of these beginning examples were analyzed in by Diaconis, Graham, Shahshahani, Aldous.

The term “cutoff” appears first in work of Aldous and Diaconis (1986).
(Also “variation threshold” and “separation threshhold” in Aldous and Diaconis (1987).)

While exact calculation of eigenvalues and eigenvectors is possible in these examples and give much more precise information, the simple arguments I give here are useful because they are robust and often generalizable, unlike spectral methods. And they are good enough to yield cutoff.

Upper bounds via L^2 analysis only work when L^1 and L^2 mixing occur at the same time.

Set-up

- (X_t) is an ergodic Markov chain with transition matrix P on a finite state space $\mathcal{X}$.
- The unique stationary distribution is π satisfies $\pi = \pi P$.
- Let

$$d(t) = \max_{x\in\mathcal{X}} \|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}} = \max_{x\in\mathcal{X}} \frac{1}{2}\sum_z |P^t(x,z) - \pi(z)|$$
$$= \max_{x\in\mathcal{X}} \sup_{A\subset\mathcal{X}} |\mathbb{P}_x(X_t \in A) - \pi(A)|$$

  denote the **total variation** (L^1) distance between the law of X_t and π.
- Classical asymptotics: For a fixed chain, $d(t) \to 0$ as $t \to \infty$.

Other distances

- L^p distance, $1 < p \leq \infty$,

$$d_p(t) = \max_x \left\| 1 - \frac{P^t(x,y)}{\pi(y)} \right\|_{L^p(\pi)}$$

- Separation distance:

$$d_s(t) = 1 - \min_{x,y} \frac{P^t(x,y)}{\pi(y)} .$$

- Also useful:

$$\bar{d}(t) = \max_{x,y} \| P^t(x,\cdot) - P^t(y,\cdot) \|_{\mathrm{TV}} .$$

  It is $\bar{d}(t)$ which is submultiplicative: $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$.

- Note that $d(t) \leq d_s(t)$, also

$$d_s(2t) \leq 1 - (1 - \bar{d}(t))^2 \leq 2\bar{d}(t) \leq 4d(t) .$$

Families of chains

- Let $\{(X_t^{(n)})\}_{n=1}^{\infty}$ be a *sequence* of chains.
- The state spaces $\mathscr{X}^{(n)}$ and transition matrices depend on instance-size parameter n.
- For example, random walk on the n-cycle: $\mathscr{X}_n = \mathbb{Z}_n = \mathbb{Z} \bmod n\mathbb{Z}$.
- Let

$$t_{\mathrm{mix}}^{(n)}(\epsilon) = \min\{t : d^{(n)}(t) < \epsilon\}$$
$$t_{\mathrm{mix}}^{(n)} = t_{\mathrm{mix}}^{(n)}(1/4)\,.$$

- Our point of view: How does $t_{\mathrm{mix}}^{(n)}(\epsilon)$ scale as $n \to \infty$?

General references on mixing: Aldous and Fill (2002), Saloff-Coste (1997), L. Peres, and Wilmer (2009, 2017).

- For example, we will show via elementary arguments that there are constants c_1 and c_2 so that for random walk on the n-cycle,

$$c_1 n^2 \leq t_{\text{mix}}^{(n)} \leq c_2 n^2 .$$

- Is there a sharp constant $c_\star$ (independent of ϵ) such that $t_{\text{mix}}^{(n)}(\epsilon) \sim c_\star n^2$? (Cutoff.)
- Or is it that:

$$d^{(n)}(cn^2) \sim \phi(c)$$

where ϕ smoothly interpolates between $\phi(0) = 1$ and $\phi(\infty) = 0$.

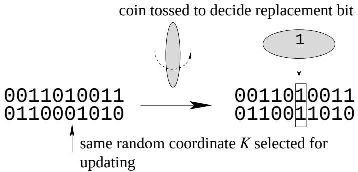
Tool: coupling

- Given pair of starting states, x, y, let (X_t, Y_t) be a Markov chain such that
  - (X_t) is a MC with transition matrix P started at x,
  - (Y_t) is a MC with transition matrix P started at y.
- Let $\tau = \min\{t \geq 0 \,:\, X_t = Y_t\}$.
- Suppose $X_t = Y_t$ for $t \geq \tau$.
- Doeblin:

$$
\begin{aligned}
\|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}} &= \frac{1}{2}\sum_z |\mathbb{P}_x(X_t = z) - \mathbb{P}_y(Y_t = z)| \\
&= \frac{1}{2}\sum_z |\mathbb{P}_x(X_t = z, \tau > t) - \mathbb{P}_y(Y_t = z, \tau > t)| \\
&\leq \mathbb{P}(\tau > t)
\end{aligned}
$$

- If $\bar{d}(t) = \max_{x,y}\|P^t(x,\cdot) - P^t(y,\cdot)\|$, then $d(t) \leq \bar{d}(t) \leq 2d(t)$.
- Note that $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$.

Lazy random walk on Hypercube: $\{0,1\}^n$



- Pick a coordinate K uniformly at random, and *refresh* by replacing current bit at K with an independent random bit.
- Couple two copies of the chain (X_t) and (Y_t) by choosing the same coordinate in both chains and refreshing both chains with the same bit.
- The chains must have met when every coordinate has been selected at least once!

- Reduces to "coupon collector problem":

$$\mathbb{P}(\tau > t) \leq \sum_{k=1}^{n} \mathbb{P}(\text{coordinate } k \text{ not selected}) = n\left(1 - \frac{1}{n}\right)^t .$$

- Taking $t = n\log n + cn$ yields $d(t) \leq e^{-c}$, whence

$$t_{\text{mix}}(\epsilon) \leq n\log n + n\log(1/\epsilon) .$$

- This is off by a factor of two; however, we will see later a modification which gives a sharp bound.

A better coupling for Random Walk on the Hypercube

- Need to only couple until within distance $\sqrt{n}$
- Then use diffusive behavior of Hamming distance.

Reduce to one-dimensional chain

Ehrenfest diffusion:

- Let W_t be the hamming weight, $W_t := \sum_{i=1}^{n} X^{(i)}(t)$.
- $\|\mathbb{P}_{\mathbf{1}}(X_t \in \cdot) - \pi\| = \|\mathbb{P}_n(W_t \in \cdot) - \pi_W\|$
- $$\mathbb{P}(W_{t+1} - W_t = x \mid \mathscr{F}_t) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{W_t}{2n} & x = -1 \\ \frac{n-W_t}{2n} & x = +1 \end{cases}.$$

- Since (X_t) is transitive, $d(t) = \|\mathbb{P}_{\mathbf{1}}(X_t \in \cdot) - \pi\|$.
- Couple lazy chains: toss a coin to decide which chain moves. Once the chains meet, they move together.
- Let (D_t) be the difference between the two Hamming weights.

$$\mathbb{E}_{z,w}[D_{t+1} - D_t \mid \mathscr{F}_t] = -\frac{D_t}{n}$$

- $$\mathbb{E}_{z,w}[D_t] \le (1 - n^{-1})^t D_0 \le n e^{-t/n}$$
- The stationary distributions cannot distinguish order $\sqrt{n}$ sites which are not mixed. (CLT.)
- Drive the expected distance down to $\sqrt{n}$ in $(1/2) n \log n$ steps.
- By comparison to simple random walk, only need $O(n)$ additional steps to hit zero.
- Thus $\mathbb{P}(D_t \neq 0)$ is small if $t = (1/2) n \log n + \alpha n$.

Lower bounds via a statistic

Proposition 1.

For $f : \mathscr{X} \to \mathbb{R}$, *define* $\sigma_\star^2 := \max\{\mathrm{Var}_\mu(f), \mathrm{Var}_\nu(f)\}$. *If*

$$|E_\nu(f) - E_\mu(f)| \geq r\sigma_\star$$

then

$$\left\|\mu - \nu\right\|_{\mathrm{TV}} \geq 1 - \frac{8}{r^2}.$$

In particular, if for a Markov chain (X_t) *with transition matrix* P *the function* f *satisfies*

$$|\mathbb{E}_x[f(X_t)] - E_\pi(f)| \geq r\sigma_\star,$$

then

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} \geq 1 - \frac{8}{r^2}.$$

Let $W(\boldsymbol{x}) = \sum_i x^{(i)}$ be the Hamming weight, and R_t be the number of coordinates not updated by time t.

$$\mathbb{E}_1(W(\boldsymbol{X}_t) \mid R_t) = R_t + \frac{(n - R_t)}{2} = \frac{1}{2}(R_t + n)\,,$$

so

$$\mathbb{E}_1(W(\boldsymbol{X}_t)) = \frac{n}{2}\left[1 + \left(1 - \frac{1}{n}\right)^t\right]$$

Also

$$\mathrm{Var}(W(\boldsymbol{X}_t)) \leq \frac{n}{4}$$

Thus

$$|E_\pi(W) - \mathbb{E}_1(W(\boldsymbol{X}_t)| = \sigma\sqrt{n}\left(1 - \frac{1}{n}\right)^t.$$

Thus if $t = \frac{1}{2}n\log n - n\alpha$, then

$$d(t) \geq 1 - 8e^{2-2\alpha}.$$

- There is a window of size n around $(1/2)n\log n$ where mixing occurs:

$$\lim_{\alpha\to-\infty}\liminf_{n} d(t_n+\alpha n)=1, \qquad \lim_{\alpha\to\infty}\limsup_{n} d(t_n+\alpha n)=0\,.$$

- This is an example of *cutoff*!

Lazy random walk on the cycle

- Start two particles on $\mathbb{Z}_n$ at x and y.
- Flip a coin to decide which particle to move.
- The clockwise distance between the two particles is itself a simple random walk on $\{0, 1, \ldots, n\}$.
- Classical gambler's ruin bounds the expected time it takes to hit 0 or n.
- $$\mathbb{E}_{x,y}(\tau) \leq \frac{n^2}{4}$$
- Thus $t_{\text{mix}}(1/4) \leq 2n^2$.

Lower bound for cycle

- Consider the set $A = [n/4, 3n/4]$.
- Since $\pi(A) = 1/2$,
$$d(t) \geq \frac{1}{2} - P^t(0, A).$$
- By Chebyshev, if $t \leq n^2/32$, then $P^t(0, A) \leq 1/4$.
- Thus $t_{\text{mix}} \geq n^2/32$.

Cutoff

A family of chains has **cutoff** at mixing times $\{t_n\}$ with window w_n if $w_n = o(t_n)$ and

$$\lim_{\alpha\to\infty} \liminf_{n\to\infty} d(t_n - \alpha w_n) = 1$$

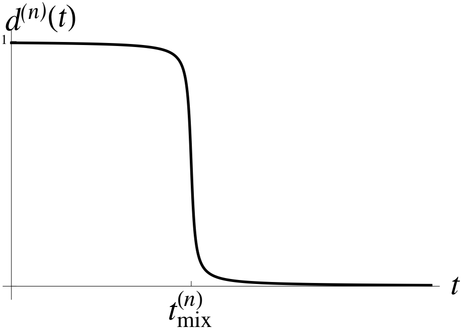
$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} d(t_n + \alpha w_n) = 0\,.$$

There is a cutoff iff for all ϵ

$$\lim_n \frac{t_{\mathrm{mix}}(\epsilon)}{t_{\mathrm{mix}}(1-\epsilon)} = 1\,.$$

When rescaling by t_{mix}, the chain has a cutoff if d approaches a step function:

$$\lim_n d^{(n)}(ct_{\mathrm{mix}}) = \begin{cases} 1 & c < 1 \\ 0 & c > 1 \end{cases}$$



The family has **pre-cutoff** if it satisfies

$$\sup_{0<\epsilon<1/2} \limsup_{n\to\infty} \frac{t_{\mathrm{mix}}^{(n)}(\epsilon)}{t_{\mathrm{mix}}^{(n)}(1-\epsilon)} < \infty.$$

Thus there are c_0 and c_1 such that

$$\liminf d^{(n)}(ct_{\mathrm{mix}}) = 1 \qquad [c < c_0]$$

$$\limsup d^{(n)}(ct_{\mathrm{mix}}) = 0 \qquad [c > c_1]$$

- Cutoff is often said to be a **high dimensional** phenomenon, with the hypercube being a key example.
- More generally, a stationary measure corresponding to a large number of independent or weakly independent variables give rise to the cutoff phenomenon. Examples: hypercube, or more generally product chains as dimension tends to infinity. Also high-temperature Ising model.
- Early examples all correspond to **high-degree** chains: walks on symmetric groups, hypercube all have degrees which are unbounded.
- Recently, cutoff explored in bounded degree graphs. (See the talk by Hermon on Friday for an example!)

Two arms of research

- Given specific family, find the mixing time, prove cutoff, and identify the window.
- Provide criteria for the existence of cutoff for classes of chains. ["Product condition", hitting time characterizations, etc.]

Part II

- Necessary condition for cutoff.
- A conjecture on cutoff.
- Counterexamples.

Spectral Analysis for Reversible Chains

- P is *reversible* if $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all x, y.
- If P is reversible then P has n real eigenvalues $1 = \lambda_1 > \lambda_2 > \cdots \lambda_2 > -1$.
- Let f_j be the eigenvector with eigenvalue λ_j of P. Then

$$4\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{j=2}^{n} f_j(x)^2 \lambda_j^{2t}.$$

  "This bound is both the key to our present understanding and a main method of proof for cutoff phenomena." (Diaconis 1996).
- A chain is *transitive* if for all x, y, there exists a bijection ϕ such that $\phi(x) = y$ and $P(\phi(z), \phi(w)) = P(z,w)$ for all z, w.
- Transitive chains satisfy:

$$4\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{j=2}^{n} \lambda_j^{2t}.$$

Relaxation time

- The *relaxation time* $t_{\text{rel}} = 1/(1-\lambda_\star)$, where $\lambda_\star = \max_{2 \le i \le n} |\lambda_i|$.
- The relaxation time is the time required for observations to be approximately uncorrelated.
- Time to mix from typical starting points.
- $$(t_{\text{rel}} - 1)\log(1/2\epsilon) \le t_{\text{mix}}(\epsilon) \le t_{\text{rel}} \log\left(\frac{1}{\epsilon \pi_{\min}}\right).$$

Hypercube

- Lazy random walk on $\{0,1\}^n$.
- Eigenvalues are $1 - \frac{j}{n}$ with multiplicity $\binom{n}{j}$.
-

$$\left\| \frac{P^t(x,\cdot)}{\pi(\cdot)} - 1 \right\|_2^2 = \sum_{k=1}^{n} \left(1 - \frac{k}{n}\right)^{2t} \binom{n}{k} \leq \sum_{k=1}^{n} e^{-2tk/n} \binom{n}{k} = (1 + e^{-2t/n})^n - 1$$

- The RHS is bound by $e^{e^{-2c}} - 1$ when

$$t = \frac{1}{2} n \log n + cn$$

- Note L^2 and TV mixing occur at the same time for hypercube.
- High multiplicity of the second eigenvalue yields cutoff.

Necessary gap condition

Proposition 2 (Cf. Aldous and Diaconis (1987) Proposition 7.8(b)).

If there is a pre-cutoff, then $t_{\mathrm{mix}}/(t_{\mathrm{rel}}-1) \to \infty$.

Proof.

Suppose that the ratio does not tend to infinity. There is an infinite set of integers J and $c_1 > 0$ such that

$$\frac{t_{\mathrm{rel}}-1}{t_{\mathrm{mix}}} \geq c_1 \quad n \in J\,.$$

Since

$$t_{\mathrm{mix}}(\epsilon) \geq (t_{\mathrm{rel}}-1)\log(1/2\epsilon)$$

Thus

$$\frac{t_{\mathrm{mix}}(\epsilon)}{t_{\mathrm{mix}}} \geq \frac{t_{\mathrm{rel}}-1}{t_{\mathrm{mix}}}\log(1/2\epsilon) \geq c_1\log(1/2\epsilon)$$

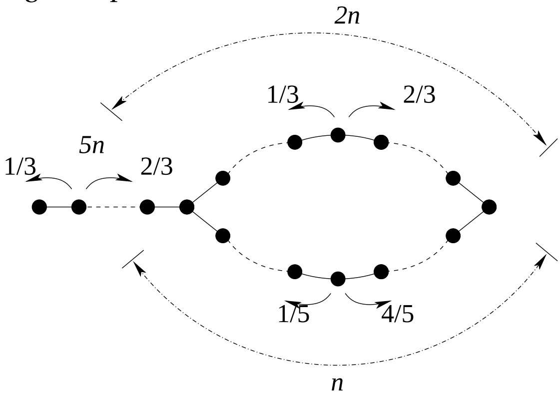
Let $\epsilon \to 0$. □

The cycle and the hypercube, revisited, and a conjecture

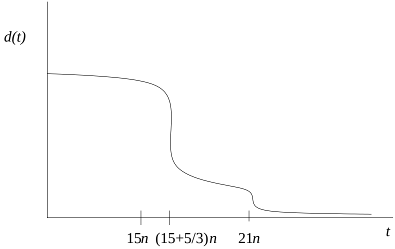
- For the cycle, $t_{\mathrm{mix}} \asymp n^2$ and $t_{\mathrm{rel}} \asymp n^2$, so there is no cutoff.
- For the hypercube $t_{\mathrm{mix}}/t_{\mathrm{rel}} \asymp \log n \to \infty$ and there is a cutoff.
- Many such examples led Y. Peres in 2004 to conjecture that the condition $t_{\mathrm{mix}}/t_{\mathrm{rel}} \to \infty$ is usually sufficient for cutoff.
- Note that for L^p distance, $1 < p \leq \infty$, the conjecture was proven by Chen and Saloff-Coste (2008).
- While there are counterexamples, it remains to find wide classes where it is true. It has been verified in many specific contexts.

Pre-cutoff, but no cutoff

The following example is due to D. Aldous:



The total variation distance looks like



Assume the right-most state has a loop.

- Since the stationary distribution grows geometrically from left-to-right, the chain mixes once it reaches near the right-most point.
- It takes about $15n$ steps for a particle started at the left-most endpoint to reach the fork. With probability about $3/4$, it first reaches the right endpoint via the bottom path. (This can be calculated using effective resistances
- When the walker takes the bottom path, it takes about $(5/3)n$ additional steps to reach the right. In fact, the time will be within order $\sqrt{n}$ of $(5/3)n$ with high probability.

- In the event that the walker takes the top path, it takes about $6n$ steps (again $\pm O(\sqrt{n})$) to reach the right endpoint.
- Thus the total variation distance will drop by $3/4$ at time $[15+(5/3)]n$, and it will drop by the remaining $1/4$ at around time $(15+6)n$.
  windows of order $\sqrt{n}$.
- Thus, the ratio $t_{\text{mix}}(\epsilon)/t_{\text{mix}}(1-\epsilon)$ will stay bounded as $n \to \infty$, but it does not tend to 1.

- Since there is a pre-cutoff, $t_{\mathrm{mix}}/t_{\mathrm{rel}} \to \infty$.
- Thus $t_{\mathrm{mix}}/t_{\mathrm{rel}} \to \infty$ not sufficient for cutoff.
- Other counterexamples due to I. Pak, H. Lacoin.

Recap

A family of chains has **cutoff** at mixing times $\{t_n\}$ with window w_n if $w_n = o(t_n)$ and

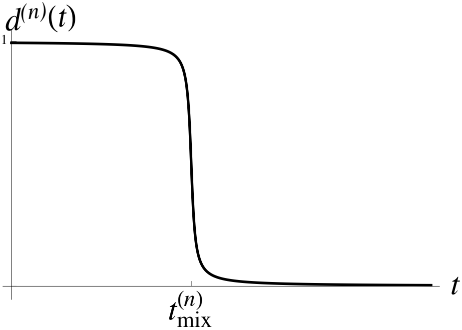
$$\lim_{\alpha \to \infty} \liminf_{n \to \infty} d(t_n - \alpha w_n) = 1$$
$$\lim_{\alpha \to \infty} \limsup_{n \to \infty} d(t_n + \alpha w_n) = 0 \,.$$

There is a cutoff iff for all ϵ

$$\lim_n \frac{t_{\mathrm{mix}}(\epsilon)}{t_{\mathrm{mix}}(1-\epsilon)} = 1 \,.$$

When rescaling by t_{mix}, the chain has a cutoff if d approaches a step function:

$$\lim_n d^{(n)}(ct_{\text{mix}}) = \begin{cases} 1 & c < 1 \\ 0 & c > 1 \end{cases}$$



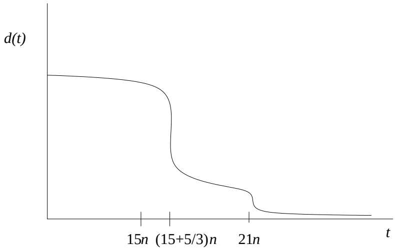
The family has **pre-cutoff** if it satisfies

$$\sup_{0<\epsilon<1/2} \limsup_{n\to\infty} \frac{t_{\mathrm{mix}}^{(n)}(\epsilon)}{t_{\mathrm{mix}}^{(n)}(1-\epsilon)} < \infty .$$

Thus there are c_0 and c_1 such that

$$\liminf d^{(n)}(ct_{\mathrm{mix}}) = 1 \qquad [c < c_0]$$

$$\limsup d^{(n)}(ct_{\mathrm{mix}}) = 0 \qquad [c > c_1]$$



Summary of our two simple examples: random walk on hypercube and cycle

- Lazy SRW on the hypercube has

$$\frac{1}{2}n\log n - c_1 n\log(1/\epsilon) \le t_{\mathrm{mix}}(\epsilon) \le \frac{1}{2}n\log n + c_2\log(1/\epsilon)$$

- The mixing time is the time to randomize all but $O(\sqrt{n})$ sites. This time is concentrated around $(1/2)n\log n$.
- Alternatively, diagonalizing shows that the L^2 distance satisfies

$$d_2(t) = [1+o(1)]\sqrt{n}e^{-t/n},$$

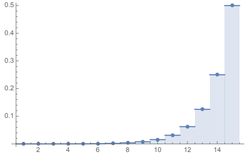
since the second eigenvalue $1-1/n$ has multiplicity n.
- L^2 mixing and L^1 mixing coincide in this case.
- Lazy SRW on the cycle does not have a cutoff, as $\mathrm{gap}\cdot t_{\mathrm{mix}} = t_{\mathrm{mix}}/t_{\mathrm{rel}}$ remains bounded.
- Hypercube suggested products should have cutoff.

Biased Random Walk on n-Path

- Suppose (X_t) is nearest-neighbor random walk on n-path:

$$\mathbb{P}(X_{t+1} - X_t = +1 \mid \mathscr{F}_t) = p$$
$$\mathbb{P}(X_{t+1} - X_t = -1 \mid \mathscr{F}_t) = 1 - p$$

  Let $\beta = 2p - 1 > 0$ be the bias.
- Since π is geometric; $1 - o(1)$ of the mass is within $O(1)$ of n.



- The chain mixes once it is in a neighborhood of n.
- By the Central Limit Theorem,

$$X_t \sim \beta t + c\sqrt{t} Z,$$

  where Z is a standard Normal random variable.
- Thus need $t = \frac{n}{\beta}$ to mix, and
- There is window of $O(\sqrt{n})$.

Trees

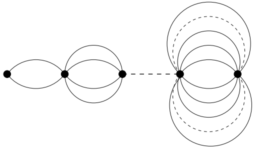
- Despite the fact that the distance to the origin on an d-ary tree behaves like a biased random walk, the random walk on the tree does not have cutoff. (But now the worst starting location is on the boundary.)
- Consider the binary tree of depth k. Couple lazy random walks (X_t) and (Y_t) by selecting one at random to move, until they are at the same level; Once at the same level, move both towards root or away from root together.
- They must have met by the time it takes for X_t to hit leaves and then root.

Commute time via effective resistance

- Commute time identity:

$$2|E|\mathscr{R}(a \leftrightarrow b) = \mathbb{E}_a(\tau_b) + \mathbb{E}_b(\tau_a)\,.$$

- For binary tree, to find effective resistance from root to boundary, glue together all vertices at the same level:



- The commute time from leaves to root for non-lazy walk is

$$2|E|\mathscr{R}(\rho \leftrightarrow \partial T) = 2(n-1)\sum_{j=1}^{k} 2^{-j} \leq 2n$$

- Thus $t_{\text{mix}} \leq 16n$

Lower bound via Cheeger constant

Let

$$Q(A, A^c) = \sum_{x \in A, y \in A^c} \pi(x) P(x, y).$$

If

$$\Phi_A = \frac{Q(A, A^c)}{\pi(A)},$$

then for $\pi(A) \leq 1/2$,

$$t_{\text{mix}} \geq \frac{1}{4\Phi_A}.$$

(Sinclair and Jerrum 1989).

- Lower bound for binary tree uses Cheeger constant.
- Take S to be the "right" tree below the root.
- $\Phi(S) = 1/[2(n-2)]$; thus $t_{\mathrm{mix}} \geq \frac{n-2}{2}$.
- We also have $t_{\mathrm{rel}} \leq t_{\mathrm{mix}} \leq c_1 n$.
- Sinclair and Jerrum (1989) implies that $t_{\mathrm{rel}} \geq 1/2\Phi_\star$, whence $t_{\mathrm{rel}} \geq c_2 n$.
- Conclude that both $t_{\mathrm{mix}} \asymp n$ and $t_{\mathrm{rel}} \asymp n$, whence there is no cutoff.

More on trees

- There are examples of trees with cutoff (Peres and Sousi 2015b). This exploits the relation between hitting time and mixing as developed in Peres and Sousi (2015a) (also Oliveira (2012).)
- In fact, the condition $t_{\mathrm{mix}}/t_{\mathrm{rel}} \to \infty$ is sufficient for cutoff on trees (Basu, Hermon, and Peres 2017).
- Basu, Hermon, and Peres (2017) establish for the parameter

$$\mathrm{hit}_{1/2}(\epsilon) = \min\{t : \left[\max_x \max_{A:\pi(A)\geq 1/2} \mathbb{P}_x(\tau_A > t)\right] \leq \epsilon\}$$

that there is a cut-off for a family of lazy reversible chain if and only if

$$\mathrm{hit}_{1/2}(\epsilon) - \mathrm{hit}_{1/2}(1-\epsilon) = o(\mathrm{hit}_{1/2}(1/4))$$

Product chains

Suppose that P_i is a transition matrix on $\mathscr{X}_i$ for $i = 1, 2, \ldots, n$. Define for $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{X} = \prod_{i=1}^{n} \mathscr{X}_i$,

$$\tilde{P}_i(\boldsymbol{x}, \boldsymbol{y}) := \begin{cases} P_i(x^{(i)}, y^{(i)}) & \text{if } x^{(j)} = y^{(j)} \text{ for } j \neq i, \\ 0 & \text{otherwise.} \end{cases},$$

Let

$$P = \frac{1}{n} \sum_{i=1}^{n} \tilde{P}_i,$$

so P corresponds to choosing a coordinate at random and making a move according to P_i in that coordinate.

Product chains studied in Diaconis and Saloff-Coste (1996).

Cf. Barrera, Lachaud, and Ycart (2006) and Aldous and Diaconis (1987)

Theorem 1.

Suppose, for $i = 1, \ldots, n$*, the spectral gap* γ_i *for the chain with reversible transition matrix* P_i *is bounded below by* γ *and the stationary distribution* $\pi^{(i)}$ *satisfies* $\sqrt{\pi^{(i)}_{\min}} \geq c_0$*, for some constant* $c_0 > 0$*. If* P *is the matrix above, then the Markov chain with matrix* P *satisfies*

$$t_{\text{mix}}^{\text{cont}}(\epsilon) \leq \frac{1}{2\gamma} n \log n + \frac{1}{\gamma} n \log(1/[c_0\epsilon]). \tag{1}$$

If the spectral gap $\gamma_i = \gamma$ *for all i, then*

$$t_{\text{mix}}^{\text{cont}}(\epsilon) \geq \frac{n}{2\gamma}\left\{\log n - \log\left[8 \log\left(1/(1-\epsilon)\right)\right]\right\}. \tag{2}$$

- Lazy chains have a cutoff if and only if the corresponding continuous time chains have a cutoff. (Chen and Saloff-Coste 2013).

Hellinger distance

- The Hellinger distance

$$d_H(\mu,\nu) = \sqrt{\sum_x \left(\sqrt{\mu(x)} - \sqrt{\nu(x)}\right)^2}$$

satisfies for $\mu = \prod \mu^{(i)}$ and $\nu = \prod \nu^{(i)}$

$$d_H^2(\mu,\nu) \leq \sum d_H^2(\mu^{(i)},\nu^{(i)})\,.$$

- Also

$$\|\mu-\nu\|_{\mathrm{TV}} \leq d_H(\mu,\nu)\,,$$

and if $\mu \ll \nu$, then

$$d_H(\mu,\nu) \leq \|\frac{d\mu}{d\nu} - 1\|_{L^2(\nu)}$$

As for discrete-time chain, the spectral decomposition of P gives for reversible chains that

$$2\|\mathbb{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \le \frac{e^{-\gamma t}}{\pi_{\min}}$$

The continuous time chain $\boldsymbol{X}_t$ satisfies

$$\mathbb{P}_x(\boldsymbol{X}_t = y) = \prod_{i=1}^{n} \mathbb{P}_x(X^{(i)}_{t/n} = y^{(i)})\,.$$

Thus

$$d_H^2(P^t(\boldsymbol{x},\cdot),\boldsymbol{\pi}) \le \sum d_H^2(\mathbb{P}_x(X^{(i)}_{t/n} \in \cdot),\pi_i) \le \sum \| \frac{\mathbb{P}_x(X^{(i)}_{t/n} \in \cdot)}{\pi_i} - 1\|_2^2 \le \frac{ne^{-2\gamma t}}{c_0^2}\,.$$

- $Y_n = (X_1^{(n)}(t), \ldots, X_n^{(n)}(t))$, where $\{X_i^{(n)}\}_{i=1}^n$ are iid.
- Lacoin (2015): For any sequence (X_n)

$$\limsup \frac{t_{\text{mix}}(1-\epsilon)}{t_{\text{mix}}(\epsilon)} \leq 2\,.$$

- Easiest to see that it also holds for separation; if D is for the product

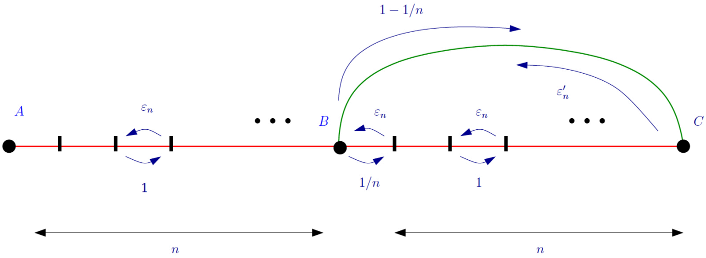
$$D_s^{(n)}(t) = 1 - (1 - d_s^{(n)}(t))^n$$

-

$$t_s(n^{-2/3}) \leq T_s^n(1-\epsilon) \leq T^n(\epsilon) \leq t_s^n(n^{-4/3}) \leq 2\,t_s^n(n^{-2/3})\,.$$

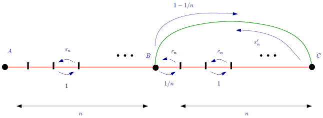
- Holds for TV distance via Hellinger distance.

Lacoin Example



From Lacoin (2015)

$$\epsilon_n = 2^{-n^2}.$$



- The stationary mass is at C except $o(1)$; the chain mixes once it reaches C.
- With high probability, before the chain reaches C, it makes no backtrack, and it takes the shortcut.
- Thus a single copy of the chain has a cutoff at n.
- The product chain reaches C^n in about time $2n$ if one of the coordinates takes the "long way", which occurs with non-zero probability.
- Thus the hitting time of C^n has some mass concentrated near n and some concentrated at $2n$.

- If τ is the hitting time of C, then

$$n\mathbb{P}_A(\tau > cn) = n\mathbb{P}_A(\tau > cn \,|\, \text{long})\frac{1}{n} + n\mathbb{P}_A(\tau > cn \,|\, \text{short})(1 - \frac{1}{n})$$

$$\to \begin{cases} \infty & c < 1 \\ 1 & c \in (1,2) \\ 0 & c > 2\,. \end{cases}$$

- If τ^{prod} is the hitting time of C^n,

$$\begin{aligned} d^{\text{prod}}(cn) &= \mathbb{P}(\tau^{\text{prod}} > cn) + o(1) \\ &= 1 - (1 - \mathbb{P}(\tau > cn))^n \\ &\to \begin{cases} 1 & c < 1 \\ 1 - e^{-1} & c \in (1,2) \\ 0 & c > 2 \end{cases} \end{aligned}$$

Note that each coordinate is **low entropy**: essentially determined by a highly biased coin flip.

Thus, there are caveats to "high dimensions have cutoff": need reasonable entropy in each dimension.

We see here that lack of cutoff is related to lack of concentration of hitting time.

Part III

- Strong Stationary Times
- The Ising Model
- Path coupling and the biased exclusion process

Strong Stationary Times

- A *strong stationary time* τ is a stopping time such that
  - $\mathbb{P}_x(X_\tau \in \cdot) = \pi$, and
  - τ and X_τ are independent.
- The separation distance

$$s_x(t) = \max_y \left[1 - \frac{P^t(x,y)}{\pi(y)} \right]$$

satisfies

$$\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq s_x(t)\,.$$

Lemma 2.

If τ is a SST then

$$s_x(t) \leq \mathbb{P}_x(\tau > t)\,.$$

Proof.

From the definition,

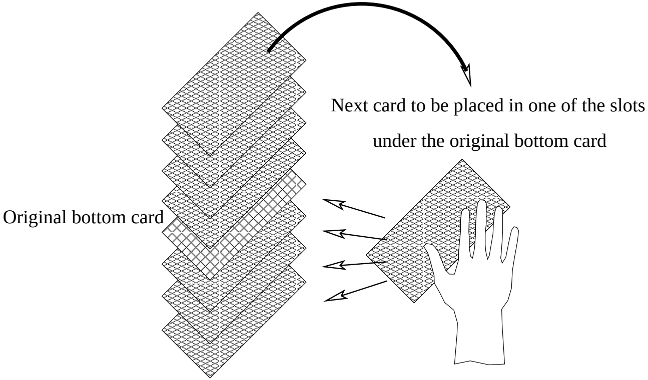
$$\mathbb{P}_x\{\tau \leq t, X_t = y\} = \mathbb{P}_x\{\tau \leq t\}\pi(y). \tag{3}$$

Fix $x \in \mathcal{X}$. Observe that for every $y \in \mathcal{X}$,

$$\begin{aligned} 1 - \frac{P^t(x,y)}{\pi(y)} &= 1 - \frac{\mathbb{P}_x\{X_t = y\}}{\pi(y)} \\ &\leq 1 - \frac{\mathbb{P}_x\{X_t = y, \tau \leq t\}}{\pi(y)} = \mathbb{P}\{\tau > t\}\,. \end{aligned}$$

□

Top-to-random card shuffle



Theorem 3.

If there exists a halting state for x, then τ is optimal:

$$s_x(t) = \mathbb{P}_x(\tau > t)\,.$$

Proof.

If y is a halting state for starting state x and the stopping time τ, then

$$\begin{aligned} 1 - \frac{P^t(x,y)}{\pi(y)} &= 1 - \frac{\mathbb{P}_x\{X_t = y\}}{\pi(y)} \\ &\leq 1 - \frac{\mathbb{P}_x\{X_t = y, \tau \leq t\}}{\pi(y)}. \end{aligned}$$

is an equality for every t. Therefore, if there exists a halting state for starting state x, then

$$s_x(t) = \mathbb{P}_x(\tau > t).$$

$\square$

Example: Top-to-random insertion. Let τ be one shuffle after the first time that the next-to-bottom card comes to the top.
Since this is a sum of geometrics (+1), the coupon collector analysis applies:

$$\mathbb{P}_x(\tau > n\log n + cn) \le e^{-\alpha}.$$

In fact Erdös and Rényi (1961) show that

$$\mathbb{P}_x(\tau < n\log n + cn) \sim e^{-e^{-c}}$$

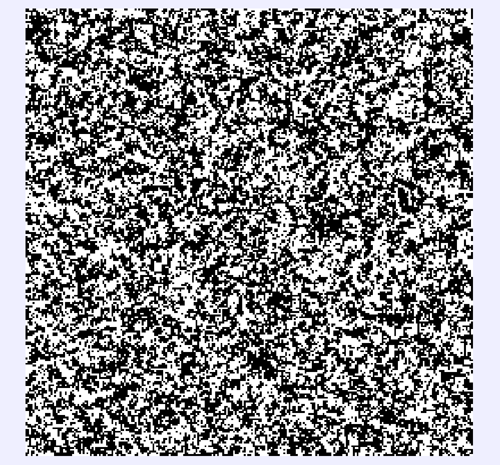
The optimality of τ shows that there is a separation cut-off at $n\log n$ with window of size n.

Separation vs Total variation cutoff

- Separation and total-vartiation cutoffs are not equivalent. (Hermon, Lacoin, and Peres 2016).
- They are for birth-and-death chains (Ding, Lubetzky, and Peres 2010).

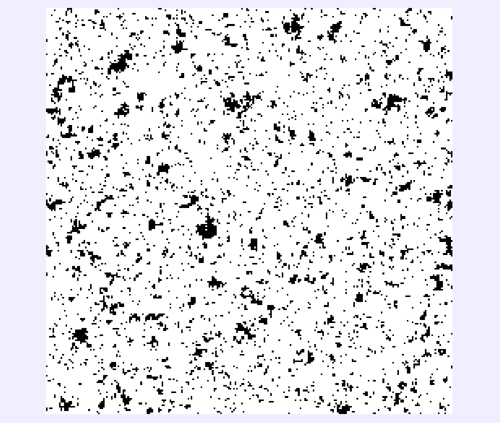
Three regimes for the Ising model

High temperature ($\beta < \beta_c$):



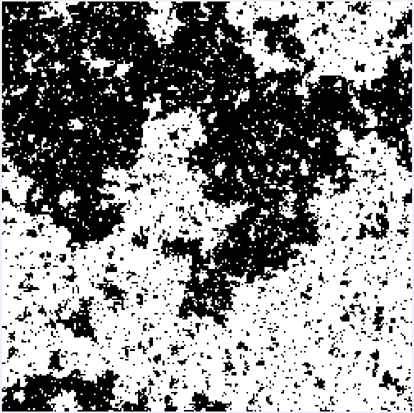
Three regimes for the Ising model

low temperature ($\beta > \beta_c$),



Three regimes for the Ising model

critical temperature ($\beta = \beta_c$),



Mixing behavior of Glauber dynamics for Ising

- Picture to confirm, precise definitions to follow. Glauber dynamics for the Ising model on graph with n vertives.
- At high temperature, mixing is as fast as possible $n \log n$, with cutoff.
- At critical temperature, the mixing is polynomial in n with no cut-off
- At low temperature, mixing is exponential in n.
- At low temperature, confined to one of the phases, there is fast mixing with a cutoff.
- Lattice cases beyond the scope of these lectures, but we will provide some details for the complete graph.
- Key idea is that careful couplings can give bounds sharp enough to prove a cutoff.

Introduction to Glauber dynamics for Ising model

Let $G_n = (V_n, E_n)$ be a graph with $N = |V_n| < \infty$ vertices.

The nearest-neighbor *Ising model* on G_n is the probability distribution on $\{-1,1\}^{V_n}$ given by

$$\mu(\sigma) = Z(\beta)^{-1} \exp\left(\beta \sum_{(u,v)\in E_n} \sigma(u)\sigma(v)\right),$$

where $\sigma \in \{-1,1\}^{V_n}$.

The interaction strength β is a parameter which has physical interpretation as $\frac{1}{\text{temperature}}$.

Glauber dynamics

The (single-site) *Glauber dynamics* for μ is a Markov chain (X_t) having μ as its stationary distribution.

Transitions are made from state σ as follows:

1. a vertex v is chosen uniformly at random from V_n.
2. The new state σ' agrees with σ everywhere except possibly at v, where $\sigma'(v) = 1$ with probability

$$\frac{e^{\beta S(\sigma,v)}}{e^{\beta S(\sigma,v)} + e^{-\beta S(\sigma,v)}}$$

where

$$S(\sigma, v) := \sum_{w: w \sim v} \sigma(w).$$

Note the probability above equals the μ-conditional probability of a positive spin at v, given that that all spins agree with σ at vertices different from v.

Glauber dynamics

The (single-site) *Glauber dynamics* for μ is a Markov chain (X_t) having μ as its stationary distribution.

Transitions are made from state σ as follows:

1. a vertex v is chosen uniformly at random from V_n.
2. The new state σ' agrees with σ everywhere except possibly at v, where $\sigma'(v) = 1$ with probability

$$\frac{e^{\beta S(\sigma,v)}}{e^{\beta S(\sigma,v)} + e^{-\beta S(\sigma,v)}}$$

where

$$S(\sigma, v) := \sum_{w: w \sim v} \sigma(w).$$

Note the probability above equals the μ-conditional probability of a positive spin at v, given that that all spins agree with σ at vertices different from v.

Glauber dynamics

The (single-site) *Glauber dynamics* for μ is a Markov chain (X_t) having μ as its stationary distribution.

Transitions are made from state σ as follows:

1. a vertex v is chosen uniformly at random from V_n.
2. The new state σ' agrees with σ everywhere except possibly at v, where $\sigma'(v) = 1$ with probability

$$\frac{e^{\beta S(\sigma,v)}}{e^{\beta S(\sigma,v)} + e^{-\beta S(\sigma,v)}}$$

where

$$S(\sigma, v) := \sum_{w : w \sim v} \sigma(w).$$

Note the probability above equals the μ-conditional probability of a positive spin at v, given that that all spins agree with σ at vertices different from v.

A conjecture

For the Glauber dynamics on graph sequences with bounded degree, $t^n_{\mathrm{mix}} = \Omega(n \log n)$. (Hayes and Sinclair 2007)

For Ising, lower bound of $\Omega(n \log n)$ from Ding and Peres (2009); simple proof from Nestoridi (2017).

If the Glauber dynamics for a sequence of transitive graphs satisfies $t^n_{\mathrm{mix}} = O(n \log n)$, is there is a cut-off? (Peres)

Mean field case

Take $G_n = K_n$, the complete graph on the n vertices: $V_n = \{1, \ldots, n\}$, and E_n contains all $\binom{n}{2}$ possible edges.

The total interaction strength should be $O(1)$, so replace β by β/n. The probability of updating to a $+1$ is then

$$\frac{e^{\beta(S-\sigma(v))/n}}{e^{\beta(S-\sigma(v))/n} + e^{-\beta(S-\sigma(v))/n}}$$

where S is the *total magnetization*

$$S = \sum_{i=1}^{n} \sigma(i).$$

The statistic S is almost sufficient for determining the updating probability.

The chain (S_t) will be key to analysis of the dynamics.

Mean field case

Take $G_n = K_n$, the complete graph on the n vertices: $V_n = \{1, \ldots, n\}$, and E_n contains all $\binom{n}{2}$ possible edges.

The total interaction strength should be $O(1)$, so replace β by β/n. The probability of updating to a +1 is then

$$\frac{e^{\beta(S-\sigma(v))/n}}{e^{\beta(S-\sigma(v))/n} + e^{-\beta(S-\sigma(v))/n}}$$

where S is the *total magnetization*

$$S = \sum_{i=1}^{n} \sigma(i).$$

The statistic S is almost sufficient for determining the updating probability.

The chain (S_t) will be key to analysis of the dynamics.

Mean field case

Take $G_n = K_n$, the complete graph on the n vertices: $V_n = \{1, \dots, n\}$, and E_n contains all $\binom{n}{2}$ possible edges.

The total interaction strength should be $O(1)$, so replace β by β/n. The probability of updating to a $+1$ is then

$$\frac{e^{\beta(S-\sigma(v))/n}}{e^{\beta(S-\sigma(v))/n} + e^{-\beta(S-\sigma(v))/n}}$$

where S is the *total magnetization*

$$S = \sum_{i=1}^{n} \sigma(i).$$

The statistic S is almost sufficient for determining the updating probability.

The chain (S_t) will be key to analysis of the dynamics.

Mean field has $t_{\mathrm{mix}} = O(n \log n)$

A consequence (that can be obtained, e.g., from Aizenman and Holley (1987)) of the Dobrushin-Shlosman uniqueness criterion: For the Glauber dynamics on K_n, if $\beta < 1$, then

$$t_{\mathrm{mix}} = O(n \log n).$$

(See also Bubley and Dyer (1998).)

High temperature mean-field

Theorem 4 (L.-Luczak-Peres 2010).

Let (X_t^n) be the Glauber dynamics for the Ising model on K_n. If $\beta < 1$, then $t_{\mathrm{mix}}(\epsilon) = (1 + o(1))\frac{n \log n}{2(1-\beta)}$ and there is a cut-off.

In fact, we show that there is *window* of size $O(n)$ centered about

$$t_n = \frac{1}{2(1-\beta)} n \log n.$$

That is,

$$\limsup_n d_n(t_n + \gamma n) \to 0 \quad \text{as} \quad \gamma \to \infty.$$

and

$$\liminf_n d_n(t_n + \gamma n) \to 1 \quad \text{as} \quad \gamma \to -\infty.$$

Critical temperature mean-field

Theorem 5 (L.-Luczak-Peres 2010).

Let (X_t^n) be the Glauber dynamics for the Ising model on K_n. If $\beta = 1$, then there are constants c_1 and c_2 so that

$$c_1 n^{3/2} \leq t_{\mathrm{mix}} \leq c_2 n^{3/2}.$$

Low temperature mean-field

If $\beta > 1$, then

$$t_{\text{mix}}^{n} > c_1 e^{c_2 n}.$$

This can be established using Cheeger constant – there is a bottleneck going between states with positive magnetization and states with negative magnetization.

Arguments for exponentially slow mixing in the low temperature regime go back at least to Griffiths, Weng and Langer (1966)

Our results show that once this barrier to mixing is removed, the mixing time is reduced to $n \log n$.

Low temperature mean-field

If $\beta > 1$, then

$$t^n_{\text{mix}} > c_1 e^{c_2 n}.$$

This can be established using Cheeger constant – there is a bottleneck going between states with positive magnetization and states with negative magnetization.

Arguments for exponentially slow mixing in the low temperature regime go back at least to Griffiths, Weng and Langer (1966)

Our results show that once this barrier to mixing is removed, the mixing time is reduced to $n \log n$.

Low temperature mean-field

If $\beta > 1$, then

$$t_{\text{mix}}^{n} > c_1 e^{c_2 n}.$$

This can be established using Cheeger constant – there is a bottleneck going between states with positive magnetization and states with negative magnetization.

Arguments for exponentially slow mixing in the low temperature regime go back at least to Griffiths, Weng and Langer (1966)

Our results show that once this barrier to mixing is removed, the mixing time is reduced to $n \log n$.

Low temperature mean-field

If $\beta > 1$, then

$$t_{\text{mix}}^{n} > c_1 e^{c_2 n}.$$

This can be established using Cheeger constant – there is a bottleneck going between states with positive magnetization and states with negative magnetization.

Arguments for exponentially slow mixing in the low temperature regime go back at least to Griffiths, Weng and Langer (1966)

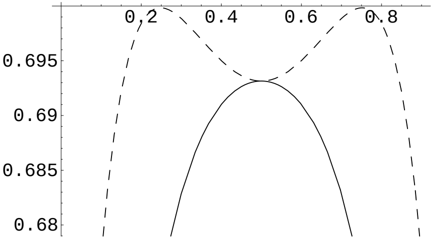
Our results show that once this barrier to mixing is removed, the mixing time is reduced to $n \log n$.

The bottleneck

Let A_k be those configurations with magnetization k. Then

$$\mu(A_{\lfloor \alpha n \rfloor}) = \frac{1}{Z(\beta)} e^{-n[\phi(\alpha)+o(1)]}.$$

The function ϕ changes shape at $\beta = 1$:



The bottleneck

If Ω^+ are configurations with strictly positive magnetization,

$$\frac{Q(\Omega^+, (\Omega^+)^c)}{\pi(A_{\lfloor n/2 \rfloor})} \leq \frac{\exp n[\phi(1/2) + o(1)]}{\exp n[\phi(\alpha_0) + o(1)]}.$$

If $\beta > 1$, there is α_0 so that $\phi(\alpha_0) \geq \phi(1/2)$ and then

$$\phi_S \leq c_1 e^{-c_2 n}.$$

Truncated dynamics for low temperature mean-field

If the bottleneck at zero magnetization is removed by truncating the dynamics at zero magnetization, then the chain converges fast:

Theorem 6 (L.-Luczak-Peres).

Let $\beta > 1$. Let (X_t) be the Glauber dynamics on K_n, restricted to the set of configurations with non-negative magnetization. Then $t_{\text{mix}}^n = O(n \log n)$.

Ding, Lubetzky, and Peres (2009a) show that in fact there is a cutoff for the censored dynamics.

Proof idea for low temperature

Use coupling: Show that for arbitrary starting states, can run together two copies of the chain so that the chains meet with high probability in $O(n \log n)$ steps.

- First show that the magnetizations will agree after $O(n \log n)$ steps, when chains are run independently. (Hard part – involves hitting time calculations.)
- After magnetizations agree, couple the chains as below.

Proof idea for low temperature

Use coupling: Show that for arbitrary starting states, can run together two copies of the chain so that the chains meet with high probability in $O(n \log n)$ steps.

- First show that the magnetizations will agree after $O(n \log n)$ steps, when chains are run independently. (Hard part – involves hitting time calculations.)
- After magnetizations agree, couple the chains as below.

Proof idea for low temperature

Use coupling: Show that for arbitrary starting states, can run together two copies of the chain so that the chains meet with high probability in $O(n \log n)$ steps.

- First show that the magnetizations will agree after $O(n \log n)$ steps, when chains are run independently. (Hard part – involves hitting time calculations.)
- After magnetizations agree, couple the chains as below.

A coupling (any temperature)

Write (X_t) and $(\tilde{X}_t)$ for the two chains. We assume that $S(X_t) = S(\tilde{X}_t)$.

Let J be the vertex selected for updating in X_t, and let $s \in \{-1, 1\}$ be the spin used to update $X_t(J)$.

The $\tilde{X}$-chain will also be updated with the spin s at a site $\tilde{J}$ which has $\tilde{X}_t(\tilde{J}) = X_t(J)$, although it will not always be that $J = \tilde{J}$.

If $X_t(J) = \tilde{X}_t(J)$, then update both chains at J.

If $X_t(J) \neq \tilde{X}_t(J)$, then pick $\tilde{J}$ uniformly at random from

$$\{i : \tilde{X}_t(i) \neq X_t(i) \text{ and } \tilde{X}_t(i) = X_t(J)\}.$$

A coupling (any temperature)

Write (X_t) and $(\tilde{X}_t)$ for the two chains. We assume that $S(X_t) = S(\tilde{X}_t)$.

Let J be the vertex selected for updating in X_t, and let $s \in \{-1, 1\}$ be the spin used to update $X_t(J)$.

The $\tilde{X}$-chain will also be updated with the spin s at a site $\tilde{J}$ which has $\tilde{X}_t(\tilde{J}) = X_t(J)$, although it will not always be that $J = \tilde{J}$.

If $X_t(J) = \tilde{X}_t(J)$, then update both chains at J.

If $X_t(J) \neq \tilde{X}_t(J)$, then pick $\tilde{J}$ uniformly at random from

$$\{i : \tilde{X}_t(i) \neq X_t(i) \text{ and } \tilde{X}_t(i) = X_t(J)\}.$$

A coupling (any temperature)

Write (X_t) and $(\tilde{X}_t)$ for the two chains. We assume that $S(X_t) = S(\tilde{X}_t)$.

Let J be the vertex selected for updating in X_t, and let $s \in \{-1, 1\}$ be the spin used to update $X_t(J)$.

The $\tilde{X}$-chain will also be updated with the spin s at a site $\tilde{J}$ which has $\tilde{X}_t(\tilde{J}) = X_t(J)$, although it will not always be that $J = \tilde{J}$.

If $X_t(J) = \tilde{X}_t(J)$, then update both chains at J.

If $X_t(J) \neq \tilde{X}_t(J)$, then pick $\tilde{J}$ uniformly at random from

$$\{i : \tilde{X}_t(i) \neq X_t(i) \text{ and } \tilde{X}_t(i) = X_t(J)\}.$$

A coupling (any temperature)

Write (X_t) and $(\tilde{X}_t)$ for the two chains. We assume that $S(X_t) = S(\tilde{X}_t)$.

Let J be the vertex selected for updating in X_t, and let $s \in \{-1, 1\}$ be the spin used to update $X_t(J)$.

The $\tilde{X}$-chain will also be updated with the spin s at a site $\tilde{J}$ which has $\tilde{X}_t(\tilde{J}) = X_t(J)$, although it will not always be that $J = \tilde{J}$.

If $X_t(J) = \tilde{X}_t(J)$, then update both chains at J.

If $X_t(J) \neq \tilde{X}_t(J)$, then pick $\tilde{J}$ uniformly at random from

$$\{i : \tilde{X}_t(i) \neq X_t(i) \text{ and } \tilde{X}_t(i) = X_t(J)\}.$$

A coupling (any temperature)

Write (X_t) and $(\tilde{X}_t)$ for the two chains. We assume that $S(X_t) = S(\tilde{X}_t)$.

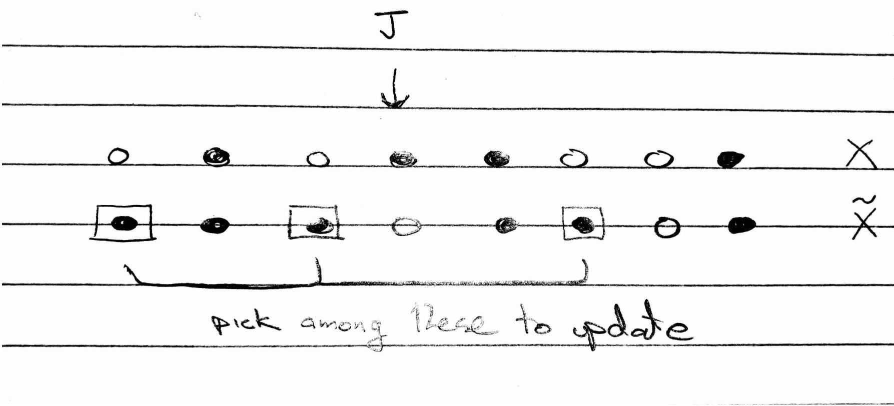
Let J be the vertex selected for updating in X_t, and let $s \in \{-1, 1\}$ be the spin used to update $X_t(J)$.

The $\tilde{X}$-chain will also be updated with the spin s at a site $\tilde{J}$ which has $\tilde{X}_t(\tilde{J}) = X_t(J)$, although it will not always be that $J = \tilde{J}$.

If $X_t(J) = \tilde{X}_t(J)$, then update both chains at J.

If $X_t(J) \neq \tilde{X}_t(J)$, then pick $\tilde{J}$ uniformly at random from

$$\{i : \tilde{X}_t(i) \neq X_t(i) \text{ and } \tilde{X}_t(i) = X_t(J)\}.$$



J
X
X̃
pick among these to update

A coupling, continued

If D_t is the number of sites where X_t and $\tilde{X}_t$ disagree, then when $D_t \geq 0$,

$$\mathbb{E}[D_{t+1} \mid \mathscr{F}_t] \leq \left[1 - \frac{c_1}{n}\right] D_t.$$

It takes $O(n \log n)$ steps to drive this expectation down to ϵ.

A coupling, continued

If D_t is the number of sites where X_t and $\tilde{X}_t$ disagree, then when $D_t \geq 0$,

$$\mathbb{E}[D_{t+1} \mid \mathscr{F}_t] \leq \left[1 - \frac{c_1}{n}\right] D_t.$$

It takes $O(n \log n)$ steps to drive this expectation down to ϵ.

Magnetization chain: key equation

If $S_t = \sum_{i=1}^{n} X_t(i)$, then for $S_t \geq 0$,

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\left[\frac{S_t}{n} - \tanh(\beta S_t / n)\right].$$

$\beta < 1$

When $\beta < 1$, using the inequality $\tanh(x) \le x$ for $x \ge 0$ shows that for $S_t \ge 0$,

$$\mathbb{E}[S_{t+1} \mid \mathscr{F}_t] \le S_t \left(1 - \frac{1-\beta}{n}\right)$$

Need $[2(1-\beta)]^{-1} n \log n$ steps to drive $\mathbb{E}[S_t]$ to $\sqrt{n}$.

Additional $O(n)$ steps needed for magnetization to hit zero. (Compare with simple random walk.)

Can couple two versions of the chain so that the magnetizations agree by the time the magnetization of the top chain hits zero.

Once magnetizations agree, use a two-dimensional process to bound time until full configurations agree. Takes an additional $O(n)$ steps.

$\beta < 1$

When $\beta < 1$, using the inequality $\tanh(x) \leq x$ for $x \geq 0$ shows that for $S_t \geq 0$,

$$\mathbb{E}[S_{t+1} \mid \mathscr{F}_t] \leq S_t \left(1 - \frac{1-\beta}{n}\right)$$

Need $[2(1-\beta)]^{-1} n \log n$ steps to drive $\mathbb{E}[S_t]$ to $\sqrt{n}$.

Additional $O(n)$ steps needed for magnetization to hit zero. (Compare with simple random walk.)

Can couple two versions of the chain so that the magnetizations agree by the time the magnetization of the top chain hits zero.

Once magnetizations agree, use a two-dimensional process to bound time until full configurations agree. Takes an additional $O(n)$ steps.

$\beta < 1$

When $\beta < 1$, using the inequality $\tanh(x) \le x$ for $x \ge 0$ shows that for $S_t \ge 0$,

$$\mathbb{E}[S_{t+1} \mid \mathscr{F}_t] \le S_t \left(1 - \frac{1-\beta}{n}\right)$$

Need $[2(1-\beta)]^{-1} n \log n$ steps to drive $\mathbb{E}[S_t]$ to $\sqrt{n}$.

Additional $O(n)$ steps needed for magnetization to hit zero. (Compare with simple random walk.)

Can couple two versions of the chain so that the magnetizations agree by the time the magnetization of the top chain hits zero.

Once magnetizations agree, use a two-dimensional process to bound time until full configurations agree. Takes an additional $O(n)$ steps.

$\beta < 1$

When $\beta < 1$, using the inequality $\tanh(x) \le x$ for $x \ge 0$ shows that for $S_t \ge 0$,

$$\mathbb{E}[S_{t+1} \mid \mathscr{F}_t] \le S_t \left(1 - \frac{1-\beta}{n}\right)$$

Need $[2(1-\beta)]^{-1} n \log n$ steps to drive $\mathbb{E}[S_t]$ to $\sqrt{n}$.

Additional $O(n)$ steps needed for magnetization to hit zero. (Compare with simple random walk.)

Can couple two versions of the chain so that the magnetizations agree by the time the magnetization of the top chain hits zero.

Once magnetizations agree, use a two-dimensional process to bound time until full configurations agree. Takes an additional $O(n)$ steps.

$\beta < 1$

When $\beta < 1$, using the inequality $\tanh(x) \le x$ for $x \ge 0$ shows that for $S_t \ge 0$,

$$\mathbb{E}[S_{t+1} \mid \mathscr{F}_t] \le S_t \left(1 - \frac{1-\beta}{n}\right)$$

Need $[2(1-\beta)]^{-1} n \log n$ steps to drive $\mathbb{E}[S_t]$ to $\sqrt{n}$.

Additional $O(n)$ steps needed for magnetization to hit zero. (Compare with simple random walk.)

Can couple two versions of the chain so that the magnetizations agree by the time the magnetization of the top chain hits zero.

Once magnetizations agree, use a two-dimensional process to bound time until full configurations agree. Takes an additional $O(n)$ steps.

Two-dimensional chain

	1	2	3					·········					n
σ_0	+	+	+	+	+	+	-	-	-	-	-	-	-
				u_0						v_0			
X_t	+	+	+	-	-	-	+	+	+	+	-	-	-
		$A(X_t)$			$B(X_t)$				$C(X_t)$			$D(X_t)$	

$$U_t = |\{i : X_t(i) = \sigma_0(i) = +1\}|$$
$$V_t = |\{i : X_t(i) = \sigma_0(i) = -1\}|.$$

We have

$$\|\mathbb{P}_{\sigma_0}\{X_t \in \cdot\} - \mu\|_{\text{TV}} = \|\mathbb{P}_{\sigma_0}\{(U_t, V_t) \in \cdot\} - \mu_2\|_{\text{TV}}.$$

Two-dimensional chain

	1	2	3				$\cdots\cdots\cdots$					n
σ_0	+	+	+	+	+	+	-	-	-	-	-	-
	u_0						v_0					
X_t	+	+	+	-	-	-	+	+	+	+	-	-
	$A(X_t)$			$B(X_t)$			$C(X_t)$				$D(X_t)$	

$$
\begin{aligned}
U_t &= |\{i : X_t(i) = \sigma_0(i) = +1\}| \\
V_t &= |\{i : X_t(i) = \sigma_0(i) = -1\}|.
\end{aligned}
$$

We have

$$
\|\mathbb{P}_{\sigma_0}\{X_t \in \cdot\} - \mu\|_{\mathrm{TV}} = \|\mathbb{P}_{\sigma_0}\{(U_t, V_t) \in \cdot\} - \mu_2\|_{\mathrm{TV}}.
$$

Two-dimensional chain (continued)

	1	2	3					⋯⋯⋯					n
σ_0	+	+	+	+	+	+	-	-	-	-	-	-	-
	u_0						v_0						
X_t	+	+	+	-	-	-	+	+	+	+	-	-	-
	$A(X_t)$			$B(X_t)$			$C(X_t)$				$D(X_t)$		
$\tilde{X}_t$	+	+	+	+	-	-	+	+	+	-	-	-	-
	$A(\tilde{X}_t)$				$B(\tilde{X}_t)$		$C(\tilde{X}_t)$			$D(\tilde{X}_t)$			

If $R_t = U_t - \tilde{U}_t$, then

$$\mathbb{E}[R_{t+1} - R_t \mid X_t] \leq 0.$$

If $R_{t+1} - R_t > 0$ with probability bounded away from zero, can compare to simple random walk.

Holds if U_t/n, V_t/n not near 0 or 1, which is true after initial phase, provided σ_0 is not too unbalanced.

After the initial phase, $R_t = O(\sqrt{n})$.

Two-dimensional chain (continued)

	1	2	3					·········					n
σ_0	+	+	+	+	+	+	-	-	-	-	-	-	-
				u_0						v_0			
X_t	+	+	+	-	-	-	+	+	+	+	-	-	-
		$A(X_t)$			$B(X_t)$				$C(X_t)$			$D(X_t)$	
$\tilde{X}_t$	+	+	+	+	-	-	+	+	+	-	-	-	-
			$A(\tilde{X}_t)$			$B(\tilde{X}_t)$		$C(\tilde{X}_t)$				$D(\tilde{X}_t)$	

If $R_t = U_t - \tilde{U}_t$, then

$$\mathbb{E}[R_{t+1} - R_t \mid X_t] \le 0.$$

If $R_{t+1} - R_t > 0$ with probability bounded away from zero, can compare to simple random walk.

Holds if $U_t/n, V_t/n$ not near 0 or 1, which is true after initial phase, provided σ_0 is not too unbalanced.

After the initial phase, $R_t = O(\sqrt{n})$.

Two-dimensional chain (continued)

	1	2	3					....					n
σ_0	+	+	+	+	+	+	-	-	-	-	-	-	-
				u_0						v_0			
X_t	+	+	+	-	-	-	+	+	+	+	-	-	-
		$A(X_t)$			$B(X_t)$			$C(X_t)$				$D(X_t)$	
$\tilde{X}_t$	+	+	+	+	-	-	+	+	+	-	-	-	-
			$A(\tilde{X}_t)$			$B(\tilde{X}_t)$		$C(\tilde{X}_t)$				$D(\tilde{X}_t)$	

If $R_t = U_t - \tilde{U}_t$, then

$$\mathbb{E}[R_{t+1} - R_t \mid X_t] \leq 0.$$

If $R_{t+1} - R_t > 0$ with probability bounded away from zero, can compare to simple random walk.

Holds if $U_t/n, V_t/n$ not near 0 or 1, which is true after initial phase, provided σ_0 is not too unbalanced.

After the initial phase, $R_t = O(\sqrt{n})$.

Two-dimensional chain (continued)

	1	2	3				....					n
σ_0	+	+	+	+	+	+	-	-	-	-	-	-	-
				u_0						v_0			
X_t	+	+	+	-	-	-	+	+	+	+	-	-	-
		$A(X_t)$			$B(X_t)$				$C(X_t)$			$D(X_t)$	
$\tilde{X}_t$	+	+	+	+	-	-	+	+	+	-	-	-	-
			$A(\tilde{X}_t)$			$B(\tilde{X}_t)$		$C(\tilde{X}_t)$				$D(\tilde{X}_t)$	

If $R_t = U_t - \tilde{U}_t$, then

$$\mathbb{E}[R_{t+1} - R_t \mid X_t] \leq 0.$$

If $R_{t+1} - R_t > 0$ with probability bounded away from zero, can compare to simple random walk.

Holds if $U_t/n, V_t/n$ not near 0 or 1, which is true after initial phase, provided σ_0 is not too unbalanced.

After the initial phase, $R_t = O(\sqrt{n})$.

$\beta = 1$. Why $n^{3/2}$?

Expanding $\tanh(x) = x - x^3/3 + \cdots$ in the key equation yields

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\frac{1}{3}\left(\frac{S_t}{n}\right)^3 .$$

Need $t = \Theta(n^{3-2\alpha})$ steps for $\mathbb{E}[S_t] = n^{\alpha}$.

By comparison with nearest-neighbor random walk, need additional $n^{2\alpha}$ steps to hit zero.

Total expected time to hit zero is

$$O(n^{3-2\alpha}) + O(n^{2\alpha})$$

The choice $\alpha = 3/4$ makes the powers equal, and gives hitting time with expectation $n^{3/2}$.

Once the magnetizations agree, need additional $O(n \log n)$ to make the configurations agree.

$\beta = 1$. Why $n^{3/2}$?

Expanding $\tanh(x) = x - x^3/3 + \cdots$ in the key equation yields

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\frac{1}{3}\left(\frac{S_t}{n}\right)^3 .$$

Need $t = \Theta(n^{3-2\alpha})$ steps for $\mathbb{E}[S_t] = n^\alpha$.

By comparison with nearest-neighbor random walk, need additional $n^{2\alpha}$ steps to hit zero.

Total expected time to hit zero is

$$O(n^{3-2\alpha}) + O(n^{2\alpha})$$

The choice $\alpha = 3/4$ makes the powers equal, and gives hitting time with expectation $n^{3/2}$.

Once the magnetizations agree, need additional $O(n \log n)$ to make the configurations agree.

$\beta = 1$. Why $n^{3/2}$?

Expanding $\tanh(x) = x - x^3/3 + \cdots$ in the key equation yields

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\frac{1}{3}\left(\frac{S_t}{n}\right)^3 .$$

Need $t = \Theta(n^{3-2\alpha})$ steps for $\mathbb{E}[S_t] = n^\alpha$.

By comparison with nearest-neighbor random walk, need additional $n^{2\alpha}$ steps to hit zero.

Total expected time to hit zero is

$$O(n^{3-2\alpha}) + O(n^{2\alpha})$$

The choice $\alpha = 3/4$ makes the powers equal, and gives hitting time with expectation $n^{3/2}$.

Once the magnetizations agree, need additional $O(n \log n)$ to make the configurations agree.

$\beta = 1$. Why $n^{3/2}$?

Expanding $\tanh(x) = x - x^3/3 + \cdots$ in the key equation yields

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\frac{1}{3}\left(\frac{S_t}{n}\right)^3 .$$

Need $t = \Theta(n^{3-2\alpha})$ steps for $\mathbb{E}[S_t] = n^\alpha$.

By comparison with nearest-neighbor random walk, need additional $n^{2\alpha}$ steps to hit zero.

Total expected time to hit zero is

$$O(n^{3-2\alpha}) + O(n^{2\alpha})$$

The choice $\alpha = 3/4$ makes the powers equal, and gives hitting time with expectation $n^{3/2}$.

Once the magnetizations agree, need additional $O(n \log n)$ to make the configurations agree.

$\beta = 1$. Why $n^{3/2}$?

Expanding $\tanh(x) = x - x^3/3 + \cdots$ in the key equation yields

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\frac{1}{3}\left(\frac{S_t}{n}\right)^3 .$$

Need $t = \Theta(n^{3-2\alpha})$ steps for $\mathbb{E}[S_t] = n^\alpha$.

By comparison with nearest-neighbor random walk, need additional $n^{2\alpha}$ steps to hit zero.

Total expected time to hit zero is

$$O(n^{3-2\alpha}) + O(n^{2\alpha})$$

The choice $\alpha = 3/4$ makes the powers equal, and gives hitting time with expectation $n^{3/2}$.

Once the magnetizations agree, need additional $O(n \log n)$ to make the configurations agree.

$\beta = 1$. Why $n^{3/2}$?

Expanding $\tanh(x) = x - x^3/3 + \cdots$ in the key equation yields

$$\mathbb{E}[S_{t+1} - S_t \mid \mathscr{F}_t] \approx -\frac{1}{3}\left(\frac{S_t}{n}\right)^3 .$$

Need $t = \Theta(n^{3-2\alpha})$ steps for $\mathbb{E}[S_t] = n^\alpha$.

By comparison with nearest-neighbor random walk, need additional $n^{2\alpha}$ steps to hit zero.

Total expected time to hit zero is

$$O(n^{3-2\alpha}) + O(n^{2\alpha})$$

The choice $\alpha = 3/4$ makes the powers equal, and gives hitting time with expectation $n^{3/2}$.

Once the magnetizations agree, need additional $O(n \log n)$ to make the configurations agree.

More detailed picture

Subsequently, Ding, Lubetzky, and Peres (2009b) showed that if $\beta = 1 - \delta$ with $\delta n^2 \to \infty$, the dynamics has a cutoff at time $\frac{n}{2\delta}\log(\delta^2 n)$ with window size n/δ.

If $\beta = 1 \pm \delta$ with $\delta n^2 = O(1)$, the mixing time is $\Theta(n^{3/2})$ with no cutoff.

Note that at low temperature, the dimension is effectively reduced.

The chain is more analogous to the barbell:

Despite the apparent high dimensionality, there is not cutoff.

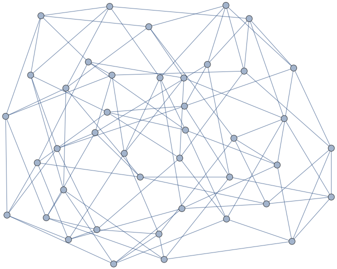
Recent breakthroughs on Ising

- Lubetzky and Sly (2013) and Lubetzky and Sly (2016) show that there is cutoff on $\mathbb{Z}_n^d$ for $\beta < \beta_c$.
- Lubetzky and Sly (2012) show that on $\mathbb{Z}_n^2$, at β_c, the mixing is polynomial in n.

For Potts:

- Lubetzky and Sly (2014) show cutoff for Glauber dynamics for Potts on $\mathbb{Z}_n^d$ for β small.
- Cuff, Ding, Louidor, Lubetzky, Peres, and Sly (2012) give the complete picture on the complete graph for Potts.

Random walk on d-regular graphs



- Many early examples of cutoff were for chain sequences with unbounded degree (such as the hypercubes).
- Simple random walk on random d-regular graphs on n vertices is an example where the degree is bounded.
- Lubetzky and Sly: For random d-regular graphs, cutoff at $\frac{d}{d-2}\log_{d-1}(n)$ with window $O(\sqrt{\log(n)})$
- This established a conjecture of Durrett.

- Lubetzky and Sly use that random d-regular graphs are locally tree-like.
- Thus can compare to the walk on the tree.
- More delicate analysis of the geometry yield results for the *non-backtracking random walk*: With high probability,

$$\lceil \log_{d-1}(dn) \rceil \leq t_{\text{mix}}(\epsilon) \leq \lceil \log_{d-1}(dn) \rceil + 1$$

- Eliminating the noise produced by backtracking moves reduces the window to constant size.
- In fact, it was observed by Peres that cutoff for SRW can be reduced to cutoff for non-backtracking RW (NBRW).
- This is used in Lubetzky and Peres to show cutoff for SRW on Ramanujan graphs (expanders with optimal gap)

Proof idea in L-S for upper bound

- A *cover tree* at u is a map $\phi : T_d \to G$ so that $\phi(\rho) = u$ and ϕ maps the neighbors of w to the neighbors of $\phi(w)$.
- If (X_t) is SRW on T_d, then $\phi(X_t)$ is SRW on G.
- Key estimate: if w and u are "nice" points separated by distance around $\log_{d-1}(\log(n))$, and j is near $\log_{d-1}(n)$, then

$$\mathbb{P}(\phi(X_t) = v \mid |X_t| = j) \gtrsim \frac{1}{n}.$$

- Thus

$$\mathbb{P}(W_t = v) \gtrsim \mathbb{P}(|X_t| \text{ near } \log_{d-1}(n)) \frac{1}{n}$$

- The CLT for X_t guarantees that if

$$t = \frac{d}{d-2}\log_{d-1} n + \alpha\sqrt{\log_{d-1}(n)}$$

then the above is

$$(1 + o(1))\frac{1}{n}(1 - \Phi(-c_1\alpha)).$$

Some further results

- Berestycki, Lubetzky, Peres, and Sly (2015) prove that *from random starting point*, SRW on supercritical random graphs has cutoff.
- Worst case mixing is slower, due to dangling paths. (Fountoulakis and Reed, and Benjamini, Kozma and Wormald)

Path coupling

Theorem 7 (Bubley and Dyer).

Suppose the state space $\mathcal{X}$ of a Markov chain is the vertex set of a graph with length function ℓ defined on edges. Let ρ be the corresponding path metric. Suppose that for each edge $\{x, y\}$ there exists a coupling (X_1, Y_1) of the distributions $P(x, \cdot)$ and $P(y, \cdot)$ such that

$$\mathbb{E}_{x,y}\left(\rho(X_1, Y_1)\right) \leq \rho(x, y)e^{-\alpha} \tag{4}$$

Then

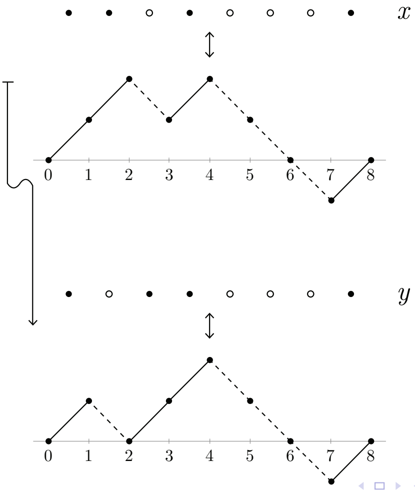
$$d(t) \leq e^{-\alpha t}\mathrm{diam}(\mathcal{X}),$$

and consequently

$$t_{\mathrm{mix}}(\epsilon) \leq \lceil \frac{-\log(\epsilon) + \log(\mathrm{diam}(\mathcal{X}))}{\alpha} \rceil.$$

Path coupling: Biased Exclusion Process

Descriptions either as k particles on $\{1, 2, \ldots, n\}$ or nearest-neighbor paths:



- With probability p, place a local min, with probability $1-p$ place a local max. The bias is $\beta = 2p-1$.
- Proper choice of a metric in path coupling can be powerful tool.

Pre-cutoff

Theorem 8 (L.-Peres 2016).

Consider the β-biased exclusion process on $\{1,2,\ldots,n\}$ with k particles. We assume that $k/n \to \rho$ for $0 < \rho \leq 1/2$.

1. *If $n\beta \leq 1$, then*
$$t_{\mathrm{mix}}^{(n)} \asymp n^3 \log n. \tag{5}$$
2. *If $1 \leq n\beta \leq \log n$, then*
$$t_{\mathrm{mix}}^{(n)} \asymp \frac{n \log n}{\beta^2}. \tag{6}$$
3. *If $n\beta > \log n$ and $\beta < \text{const.} < 1$, then*
$$t_{\mathrm{mix}}^{(n)} \asymp \frac{n^2}{\beta}. \tag{7}$$

Moreover, in all regimes, the process has a pre-cutoff.

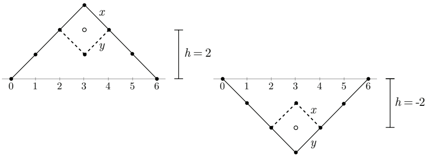
Related results

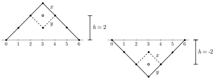
- Lacoin 2016 showed cut-off at $\pi^{-2} n^3 \log n$ for the unbiased case.
- Labbé and Lacoin 2016 recently showed cutoff for fixed bias.

Upper bound via path coupling

- The following path coupling argument is due to Greenberg, Pascoe, and Randall 2009:
- Let $\alpha = \sqrt{p/q}$; if x and y differ at a single neighbor, let

$$\ell(x,y) = \alpha^{n-k+h}.$$





- Let x be the upper configuration, and y the lower. Here the edge between $v-2$ and $v-1$ is "up", while the edge between $v+1$ and $v+2$ is "down", in both x and y.
- If v is selected, the distance decreases by α^{n-k+h}.
- If either $v-1$ or $v+1$ is selected, and a local minimum is selected, then the lower configuration y is changed, while the upper configuration x remains unchanged. Thus the distance increases by $\alpha^{n-k+h-1}$ in that case. We conclude that

$$\begin{aligned}\mathbb{E}_{x,y}[d(X_1,Y_1)]-d(x,y) &= -\frac{1}{n-1}\alpha^{h+n-k}+\frac{2}{n-1}p\alpha^{h+n-k-1}\\ &= \frac{\alpha^{h+n-k}}{n-1}\left(\frac{2p}{\alpha}-1\right)=\frac{\alpha^{h+n-k}}{n-1}\left(2\sqrt{pq}-1\right).\end{aligned}$$

- In all cases, if $\delta = 1 - 2\sqrt{p(1-p)} > 0$, then

$$\mathbb{E}_{x,y}[d(X_1, Y_1)] = d(x,y)\left(1 - \frac{\delta}{n-1}\right) \leq d(x,y)e^{-\frac{\delta}{n-1}} .$$

- By the path coupling technique of Bubley and Dyer, it is enough to check that distance contracts for neighboring states: As $\delta > \beta^2/2$

$$t_{\mathrm{mix}}(\epsilon) \leq \frac{2n}{\beta^2}\left[\log(1/\epsilon) + \log(\mathrm{diam})\right]$$

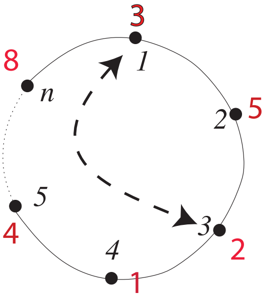
  If $\beta \to 0$, then

$$t_{\mathrm{mix}}(\epsilon) \leq \frac{2n}{\beta^2}\left[\log(\epsilon^{-1}) + n[\beta + O(\beta^2)] - 2\log\beta + O(\beta)\right] .$$

- If $\beta = 1/n$ then $t_{\mathrm{mix}}(\epsilon) = O(n^3 \log n)$, as in unbiased case.
- Need different method for $\beta < 1/n$.

Selected Open Problems

- Cyclic-to-random transpositions.



- Used in cryptographic algorithm RC4
- Lower bound of $cn \log n$ in Peres, Sinclair, and Mossel (2004) and upper bound of $n \log n$ Saloff-Coste and Zúñiga (2007).

- Glauber dynamics for Ising at high temperature on any transitive graph.
- Potts model on lattice down to critical temperature. Energy for Potts is

$$H(\sigma) = -\sum_{i \sim j} \mathbf{1}(\sigma(i) = \sigma(j))$$

Aldous, D. and P. Diaconis (1986). "Shuffling cards and stopping times". In: *Amer. Math. Monthly* 93.5, pp. 333–348. ISSN: 0002-9890. DOI: 10.2307/2323590. URL: http://dx.doi.org/10.2307/2323590.

– (1987). "Strong uniform times and finite random walks". In: *Adv. in Appl. Math.* 8.1, pp. 69–97. ISSN: 0196-8858. DOI: 10.1016/0196-8858(87)90006-6. URL: http://dx.doi.org/10.1016/0196-8858(87)90006-6.

Aldous, D. and J. A. Fill (2002). *Reversible Markov Chains and Random Walks on Graphs*. Unfinished monograph, recompiled 2014, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

Barrera, J., B. Lachaud, and B. Ycart (2006). "Cut-off for *n*-tuples of exponentially converging processes". In: *Stochastic Process. Appl.* 116.10, pp. 1433–1446. ISSN: 0304-4149. DOI: 10.1016/j.spa.2006.03.003. URL: http://dx.doi.org/10.1016/j.spa.2006.03.003.

Basu, R., J. Hermon, and Y. Peres (2017). "Characterization of cutoff for reversible Markov chains". In: *The Annals of Probability* 45.3, pp. 1448–1487. DOI: 10.1214/16-AOP1090.

Berestycki, N., E. Lubetzky, Y. Peres, and A. Sly (2015). "Random walks on the random graph". In: arXiv: 1504.01999v3 [math.PR].

Chen, G.-Y. and L. Saloff-Coste (2008). "The Cutoff Phenomenon for Ergodic Markov Processes". In: *Electronic Journal of Probability* 13, pp. 26–78. DOI: 10.1214/EJP.v13-474.

– (2013). "Comparison of cutoffs between lazy walks and Markovian semigroups". In: *J. Appl. Probab.* 50.4, pp. 943–959. ISSN: 0021-9002. DOI: 10.1239/jap/1389370092. URL: http://dx.doi.org/10.1239/jap/1389370092.

Cuff, P., J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly (2012). "Glauber dynamics for the mean-field Potts model". In: *J. Stat. Phys.* 149.3, pp. 432–477. DOI: 10.1007/s10955-012-0599-2.

Diaconis, P. and L. Saloff-Coste (1996). "Logarithmic Sobolev inequalities for finite Markov chains". In: *Ann. Appl. Probab.* 6.3, pp. 695–750. ISSN: 1050-5164. DOI: 10.1214/aoap/1034968224. URL: http://dx.doi.org/10.1214/aoap/1034968224.

Diaconis, P. (1996). "The cutoff phenomenon in finite Markov chains". In: *Proc. Nat. Acad. Sci. U.S.A.* 93.4, pp. 1659–1664. DOI: 10.1073/pnas.93.4.1659.

Ding, J. and Y. Peres (2009). "Mixing time for the Ising model: a uniform lower bound for all graphs". In: arXiv: 0909.5162 [math.PR].

Ding, J., E. Lubetzky, and Y. Peres (2009a). "Censored Glauber dynamics for the mean field Ising model". In: *J. Stat. Phys.* 137.3, pp. 407–458. ISSN: 0022-4715. DOI: 10.1007/s10955-009-9859-1. URL: http://dx.doi.org/10.1007/s10955-009-9859-1.

– (2009b). "The mixing time evolution of Glauber dynamics for the mean-field Ising model". In: *Comm. Math. Phys.* 289.2, pp. 725–764. ISSN: 0010-3616. DOI: 10.1007/s00220-009-0781-9. URL: http://dx.doi.org/10.1007/s00220-009-0781-9.

Ding, J., E. Lubetzky, and Y. Peres (2010). "Total variation cutoff in birth-and-death chains". In: *Probab. Theory Related Fields* 146.1-2, pp. 61–85. ISSN: 0178-8051. DOI: 10.1007/s00440-008-0185-3. URL: http://dx.doi.org/10.1007/s00440-008-0185-3.

Greenberg, S., A. Pascoe, and D. Randall (2009). "Sampling biased lattice configurations using exponential metrics". In: *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, Philadelphia, PA, pp. 76–85.

Hayes, T. P. and A. Sinclair (2007). "A general lower bound for mixing of single-site dynamics on graphs". In: *Ann. Appl. Probab.* 17.3, pp. 931–952. ISSN: 1050-5164. DOI: 10.1214/105051607000000104. URL: http://dx.doi.org/10.1214/105051607000000104.

Hermon, J., H. Lacoin, and Y. Peres (2016). "Total variation and separation cutoffs are not equivalent and neither one implies the other". In: *Electron. J. Probab.* 21, Paper No. 44, 36. ISSN:

1083-6489. DOI: 10.1214/16-EJP4687. URL: http://dx.doi.org/10.1214/16-EJP4687.

Labbé, C. and H. Lacoin (2016). “Cutoff phenomenon for the asymmetric simple exclusion process and the biased card shuffling”. In: arXiv: 1610.07383v1 [math.PR].

Lacoin, H. (2015). “A product chain without cutoff”. In: *Electronic Communications in Probability* 20, pp. 1–10. DOI: 10.1214/ECP.v20-3765.

– (2016). “Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion”. In: *Ann. Probab.* 44.2, pp. 1426–1487. ISSN: 0091-1798. DOI: 10.1214/15-AOP1004.

Levin, D. A., M. J. Luczak, and Y. Peres (2010). “Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability”. In: *Probab. Theory Related Fields* 146.1-2, pp. 223–265. ISSN: 0178-8051. DOI: 10.1007/s00440-008-0189-z.

Levin, D. A. and Y. Peres (2016). "Mixing of the Exclusion Process with Small Bias". In: *J. Stat. Phys.* 165.6, pp. 1036–1050. DOI: 10.1007/s10955-016-1664-z.

– (2017). *Markov chains and mixing times, second edition*. With contributions by Elizabeth L. Wilmer and with a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI. URL: http://pages.uoregon.edu/dlevin/MARKOV.

Levin, D. A., Y. Peres, and E. L. Wilmer (2009). *Markov chains and mixing times*. With a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI, pp. xviii+371. ISBN: 978-0-8218-4739-8.

Lubetzky, E. and A. Sly (2012). "Critical Ising on the square lattice mixes in polynomial time". In: *Comm. Math. Phys.* 313.3, pp. 815–836. DOI: 10.1007/s00220-012-1460-9.

– (2013). "Cutoff for the Ising model on the lattice". In: *Invent. Math.* 191.3, pp. 719–755. DOI: 10.1007/s00222-012-0404-5.

Lubetzky, E. and A. Sly (2014). “Cutoff for general spin systems with arbitrary boundary conditions”. In: *Comm. Pure Appl. Math.* 67.6, pp. 982–1027. ISSN: 0010-3640. DOI: 10.1002/cpa.21489.

– (2016). “Information percolation and cutoff for the stochastic Ising model”. In: *J. Amer. Math. Soc.* 29.3, pp. 729–774. DOI: 10.1090/jams/841.

Nestoridi, E. (2017). “Optimal strong stationary times for random walks on the chambers of a hyperplane arrangement”. In: arXiv: 1706.00310 [math.PR].

Oliveira, R. I. (2012). “Mixing and hitting times for finite Markov chains”. In: *Electron. J. Probab.* 17, no. 70, 12. ISSN: 1083-6489. DOI: 10.1214/EJP.v17-2274. URL: http://dx.doi.org/10.1214/EJP.v17-2274.

Peres, Y., A. Sinclair, and E. Mossel (2004). “Shuffling by Semi-Random Transpositions”. In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* 00, pp. 572–581. ISSN: 0272-5428. DOI: doi.ieeecomputersociety.org/10.1109/FOCS.2004.60.

Peres, Y. and P. Sousi (2015a). “Mixing times are hitting times of large sets”. In: *J. Theoret. Probab.* 28.2, pp. 488–519. ISSN: 0894-9840. DOI: 10.1007/s10959-013-0497-9. URL: http://dx.doi.org/10.1007/s10959-013-0497-9.

– (2015b). “Total variation cutoff in a tree”. In: *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6* 24.4, pp. 763–779. DOI: 10.5802/afst.1463.

Saloff-Coste, L. and J. Zúñiga (2007). “Convergence of some time inhomogeneous Markov chains via spectral techniques”. In: *Stochastic Process. Appl.* 117.8, pp. 961–979. ISSN: 0304-4149. DOI: 10.1016/j.spa.2006.11.004. URL: http://dx.doi.org/10.1016/j.spa.2006.11.004.

Saloff-Coste, L. (1997). “Lectures on finite Markov chains”. In: *Lectures on probability theory and statistics (Saint-Flour, 1996).* Vol. 1665. Lecture Notes in Math. Springer, Berlin, pp. 301–413. DOI: 10.1007/BFb0092621. URL: http://dx.doi.org/10.1007/BFb0092621.

Sinclair, A. and M. Jerrum (1989). "Approximate counting, uniform generation and rapidly mixing Markov chains". In: *Inform. and Comput.* 82.1, pp. 93–133. ISSN: 0890-5401. DOI: 10.1016/0890-5401(89)90067-9. URL: http://dx.doi.org/10.1016/0890-5401(89)90067-9.