Advanced Algorithms

Spectral methods and algorithms

尹一通　栗师　**刘景铖**

Recap

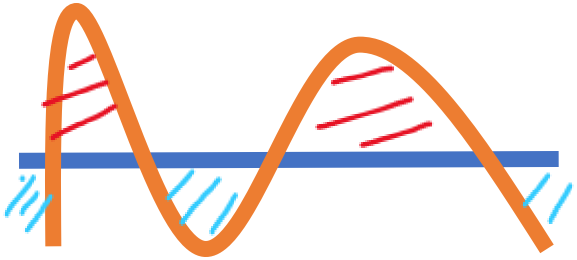
Previous lecture:
Random walks on undirected graphs

- Fundamental theorem of Markov chains
- Spectral analysis
- Mixing time
- Random sampling

What next?
Random walks on undirected graphs

- From sampling to counting and MCMC
- Expander graphs and random walks

Coupling of two distributions



Given distributions p and q over $[n]$, a **coupling** between them is a joint distribution μ over $[n] \times [n]$ such that the marginals are p and q, respectively:

$$\sum_{j \in [n]} \mu(i,j) = p(i)$$

$$\sum_{i \in [n]} \mu(i,j) = q(j)$$

Independently joining p and q is obviously a coupling. More interesting are when they are not independent.

Theorem

For any distributions p and q, and any coupling μ between them, $d_{TV}(p,q) \le \Pr_{(X,Y)\sim\mu}[X \neq Y]$

Furthermore, there is a coupling μ such that $d_{TV}(p,q) = \Pr_{(X,Y)\sim\mu}[X \neq Y]$

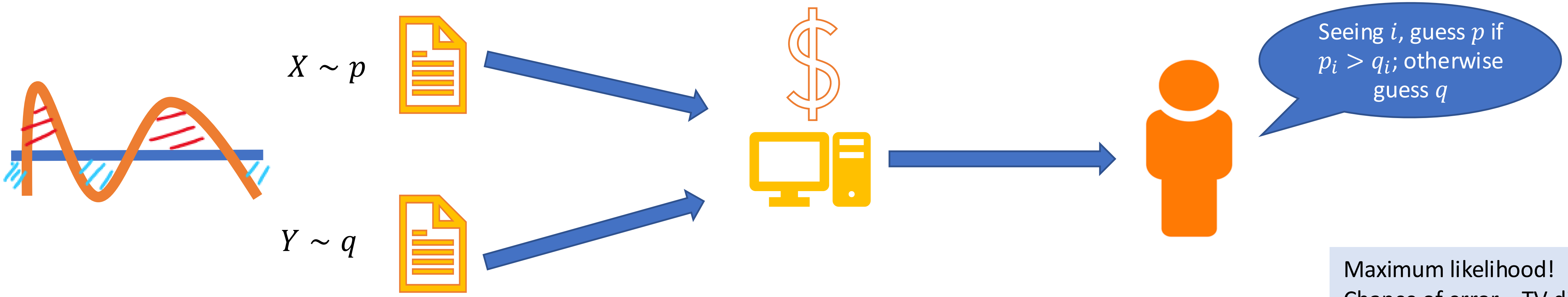
Intuitively, the best we can do is to make the random variables equal in the overlapping regions, that is, $\min\{p_i, q_i\}$; then with the remaining probability, they must be unequal.

Note that the region in red, and the region in light blue have the same area.

Coupling vs Indistinguishing game

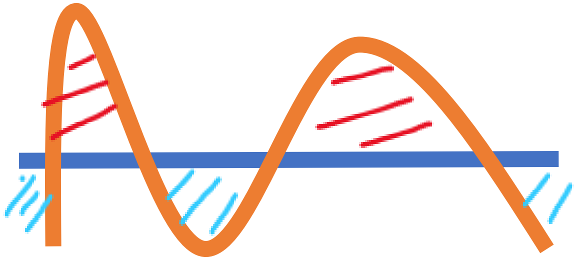
TV distance is also known as statistical distance

- A game to distinguish two distributions p and q over $[n]$
- Player A draw a sample $X \sim p$ and a sample $Y \sim q$
- Player A flips a fair coin to decide which sample to send to Player B
- Player B now needs to guess which distribution does it came from



Maximum likelihood!
Chance of error = TV distance!

Coupling of two random walks



Let (X_t) and (Y_t) be two copies of a Markov chain over $[n]$. A **coupling** between them is a joint **process** (X_t, Y_t) over $[n] \times [n]$ such that

1. Marginally, viewed in isolation, (X_t) and (Y_t) are both copies of the original chain
2. $X_t = Y_t \Rightarrow X_{t+1} = Y_{t+1}$

Basically, one can think of two random walkers on the same graph G

In isolation, they each behave faithfully as a random walk on G

But their moves could be dependent

The coupling technique is to design a joint moving process, such that

- The two random walkers meet quickly
- Once they meet, they make identical moves thereafter

Then by the coupling theorem, we know that the time they meet will roughly be an upperbound of mixing time

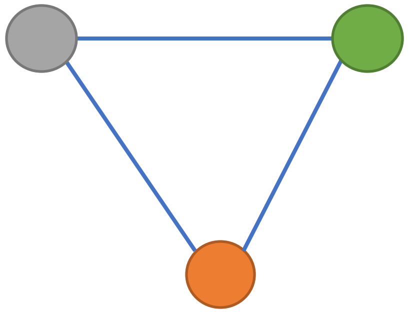
Random walk on the hypercube

- Start with $\sigma \in \{0,1\}^n$
- Pick a coordinate $i \in [n]$ u.a.r., and $b \in \{0, 1\}$ u.a.r.
- Update $\sigma_i = b$

To analyze its mixing time, we consider the following coupling
Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)
At each step, we let them choose the same coordinate i and same b

Then, the time that they perfectly couple together is exactly the coupon collecting time!
Note that the probability of not collecting the coupon i after r rounds is at most $\left(1 - \frac{1}{n}\right)^r$
By a union bound, the probability of not collecting all the coupons after $n \ln \frac{n}{\epsilon}$ rounds is at most ϵ
So, the ϵ-mixing time for a random walk on the hypercube is $n \ln \frac{n}{\epsilon}$

Recap: Graph coloring



Given an undirected graph with max. degree Δ and k colors
Goal: generate a k-coloring uniformly at random

This is presumably harder than deciding if there is a k-coloring
Nevertheless, the following random walk has a stationary distribution uniform over all k-colorings:

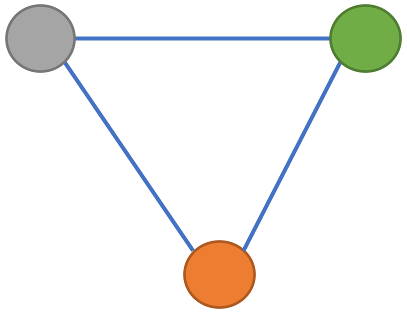
- Start with any k-coloring σ
- Pick a vertex v and a color c uniformly at random, recolor v with c if it is legal; otherwise do nothing;

This Markov chain is irreducible provided that $k \geq \Delta + 2$, and aperiodic

We prove rapid mixing assuming $k \geq 4\Delta + 1$, based on a coupling argument, and explain ideas for $k \geq 2\Delta + 1$
State of the art: $k \geq (\frac{11}{6} - \epsilon)\Delta$ for a small ϵ, or $k \geq \Delta + 3$ for sufficiently large girth graphs

This is known as the Metropolis chain
Other chains: Glauber dynamics, Wang–Swendsen–Kotecký chain, …

Coupling for Graph Coloring



- Start with any k-coloring σ
- Pick a vertex v and a color c uniformly at random, recolor v with c if it is legal; otherwise do nothing

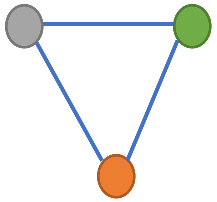
Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)

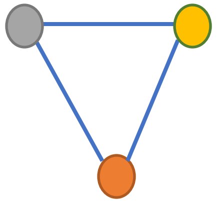
At each step, we let them choose the same vertex v and same color c

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Unlike the previous example, d_t can increase now

We need to consider Good Moves that decrease d_t, and balance them with Bad Moves that increase d_t





Start with any k-coloring σ
Pick a vertex v and a color c u.a.r.,
recolor v with c if legal

Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)

At each step, we let them choose the same vertex v and same color c

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Good Moves that decrease d_t:

If we chose a disagreeing vertex v, and color c does not appear in the neighborhood of v in X_t or Y_t, this is a good move

Because we can safely recolor a disagreeing vertex v with color c, and they agree from then on

Let g_t be the number of good moves (among all possible kn choices)

There are d_t vertices to choose from, and each disagreeing vertex has a neighborhood of at most Δ colors in either process, so each disagreeing vertex has $k - 2\Delta$ "safe colors"

$$g_t \geq d_t(k - 2\Delta)$$

Start with any k-coloring σ
Pick a vertex v and a color c u.a.r., recolor v with c if legal

Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)
At each step, we let them choose the same vertex v and same color c
Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Bad Moves that increase d_t: a legal move in one process but not the other
This happens when (and only when) the chosen color c is already the color of some neighbor of v in one process but not the other

In other words, v must be a neighbor of some disagreeing vertex u, and c must be the color of u in either X_t or Y_t

Let b_t be the number of bad moves (among all possible kn choices)
There are d_t choices of disagreeing vertex u, then Δ choices for v, then 2 for X_t or Y_t

$$b_t \leq 2\Delta d_t$$

Start with any k-coloring σ
Pick a vertex v and a color c u.a.r.,
recolor v with c if legal

Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, (X_t) and (Y_t)
At each step, we let them choose the same vertex v and same color c
Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Combined: $\mathbb{E}[d_{t+1}|d_t] = d_t + \frac{b_t - g_t}{kn} \le d_t + d_t\frac{4\Delta - k}{kn} \le d_t\left(1 - \frac{1}{kn}\right)$

Since $d_0 \le n$, we have $\mathbb{E}[d_t|d_0] \le 1/e$ for $t = 2k\, n \ln n$. Thus,

$$d_{TV}(p_t, \pi) \le \Pr_{(X_t, Y_t)\sim\mu}[X_t \ne Y_t] \le \Pr[d_t > 0|X_0, Y_0] = \Pr[d_t \ge 1|X_0, Y_0] \le \mathbb{E}[d_t|d_0] \le 1/e$$

This concludes that the ϵ-mixing time is $O\left(nk \log\frac{n}{\epsilon}\right)$

To improve this to $k \ge 2\Delta + 1$, one tries to pair bad moves in (X_t) but blocked in (Y_t), with bad moves in (Y_t) but blocked in (X_t)

From sampling to counting

Now that we have a Markov chain that outputs a k-coloring σ almost uniformly at random from all proper colorings after $O\left(nk \log \frac{n}{\epsilon}\right)$ steps

Can we estimate the total number of proper colorings?

This task is known as ***approximate counting***

For many natural concrete problems

$$\textbf{\textit{ApproxCount}} \equiv \textbf{\textit{ApproxSample}} \equiv \textbf{\textit{UniformSample}} \subset \textbf{\textit{ExactCount}}$$

The first three are in $\mathrm{BPP}^{\mathrm{NP}}$, while ***ExactCount*** is #P

From sampling to counting

Denote the number of proper colorings of a graph G by Z_G
We start by finding an arbitrary proper coloring σ in G
Then, we reveal the colors in σ one by one
We count how many proper colorings are consistent with the revealed colors

Let Z_i be the number of proper colorings τ such that
in the first i coordinates, τ agrees with σ

Notice that $Z_0 = Z_G, Z_n = 1$, and

$$Z_G = \frac{Z_0}{Z_1} \cdot \frac{Z_1}{Z_2} \cdots \frac{Z_{n-1}}{Z_n}$$

From sampling to counting

Let Z_i be the number of proper colorings τ such that

in the first i coordinates, τ agrees with σ

Notice that $Z_0 = Z_G, Z_n = 1$, and

$$Z_G = \frac{Z_0}{Z_1} \cdot \frac{Z_1}{Z_2} \cdots \frac{Z_{n-1}}{Z_n}$$

Suppose we estimate each ratio within $\left(1 \pm \frac{\epsilon}{2n}\right) \cdot \frac{Z_{i+1}}{Z_i}$ except with prob. $\frac{\delta}{n}$

Then multiplying them all together gives $(1 \pm \epsilon) \cdot Z_G$ except with prob. δ

From sampling to counting

Let Z_i be the number of proper colorings τ such that: in the first i coordinates, τ agrees with σ

Let π_i be the uniform distribution of proper colorings τ such that: in the first i coordinates, τ agrees with σ

Recall that $Z_0 = Z_G$, $Z_n = 1$, and

$$Z_G = \frac{Z_0}{Z_1} \cdot \frac{Z_1}{Z_2} \cdots \frac{Z_{n-1}}{Z_n}$$

How do we estimate each ratio $\frac{Z_{i+1}}{Z_i}$?

We run a Markov chain that samples from π_i, and use Monte Carlo method to estimate how many are counted in Z_{i+1}

Markov chain: in the first i coordinates, we fix the colorings as in σ, and only update the remaining $n - i$ coordinates

Monte Carlo: given a sample τ, we check if $\tau_{i+1} = \sigma_{i+1}$

Sampling from π_i is an unbiased estimator for the ratio:

$$E_{\tau\sim\pi_i}\,[[\tau_{i+1} = \sigma_{i+1}]] = \frac{Z_{i+1}}{Z_i}$$

Sampling from a rapidly mixing Markov chain p_t only introduces a small bias (recall the def. of TV distance):

$$\left|E_{\tau\sim p_t}\,[[\tau_{i+1} = \sigma_{i+1}]] - E_{\tau\sim\pi_i}\,[[\tau_{i+1} = \sigma_{i+1}]]\right| \le d_{TV}(p_t, \pi_i)$$

$$d_{TV}(p_t, \pi) = \max_{S\subseteq[n]} |p_t(S) - \pi(S)|$$

From sampling to counting

We run a Markov chain that samples from π_i, and use Monte Carlo method to estimate how many are counted in Z_{i+1}

Markov chain: in the first i coordinates, we fix the colorings as in σ, and only update the remaining

Monte Carlo: given a sample τ, we check if $\tau_{i+1} = \sigma_{i+1}$

Sampling from π_i is an unbiased estimator for the ratio:

$$E_{\tau\sim\pi_i}\left[[\tau_{i+1} = \sigma_{i+1}]\right] = \frac{Z_{i+1}}{Z_i}$$

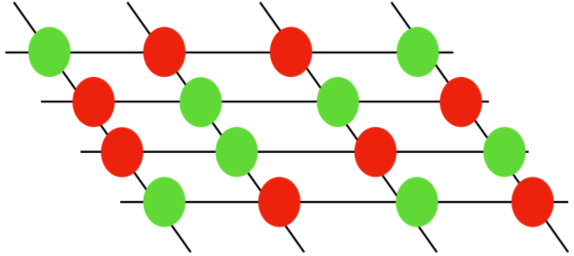
Variance can also be bounded because $\frac{Z_{i+1}}{Z_i}$ is strictly between (0,1): $\left(\frac{k-\Delta-1}{k-\Delta}\right)^{\Delta} \cdot \frac{1}{k} \leq \frac{Z_{i+1}}{Z_i} \leq \frac{1}{k-\Delta}$

It suffices to take the average over poly$\left(n, \frac{1}{\epsilon}, \frac{1}{\delta}\right)$ samples

Then apply Chebyshev's inequality

See Chapter 14.4 of LPW book

Upperbound for the ratio follows from having many colors available
lowerbound from bounding the prob. that any neighbors take the same color

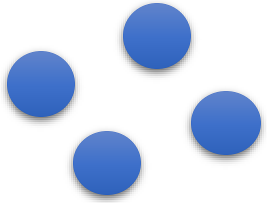


Integration, sampling, and inference

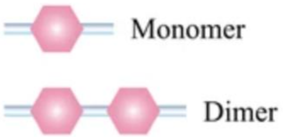
Statistical physics model

- Ferromagnets

- Hardcore lattice gas



- Monomer dimer



- Spin systems

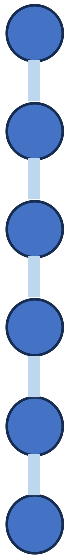
Combinatorial interpretation

- Cuts generating polynomial
- Independence polynomial

$$\forall S \in \mathrm{Ind}(G), \quad \mu(S) := \frac{\lambda^{|S|}}{Z(G)}$$

- Matching polynomial
- Constraint satisfaction problem
  - Example: 3-SAT
  - Variables: $x_1, x_2, \dots, x_n \in \{T, F\}$
  - Constraints: $(x_1 \; OR \; x_2 \; OR \; \overline{x_3}) \; AND \; \dots$

1D Ising model



We consider 1D Ising model (Lenz 1920, Ising 1925)
Ising (1925) exactly solved this model in his thesis to show that there is no *phase transition* in 1D

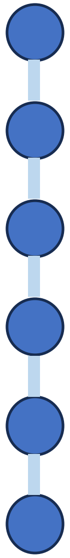
We will see a ``coupling proof’’ of a related phenomenon:
  the Glauber dynamics mixes rapidly in 1D even with mixed interactions

Configuration: Each vertex v of a path graph gets a spin $\sigma_v \in \{\pm 1\}$
Interaction: For every adjacent pair of vertices, there is an *interaction strength* $J_{x,x+1}$
The **energy** of a configuration is given by the *Hamiltonian* function

$$H(\sigma) \coloneqq \sum_x J_{x,x+1} \cdot \sigma_x \cdot \sigma_{x+1}$$

The Boltzmann/Gibbs distribution is given by $\Pr[\sigma] \propto \exp\big(H(\sigma)\big)$
The normalizing “constant” is known as the ***partition function*** $Z(J) \coloneqq \sum_\sigma \exp\big(H(\sigma)\big)$
Combinatorially, partition functions can be seen as generating polynomials

1D Ising model



Configuration: Each vertex v of a path graph gets a spin $\sigma_v \in \{\pm 1\}$
Interaction: For every adjacent pair of vertices, there is an *interaction strength* $J_{x,x+1}$
The **energy** of a configuration is given by the *Hamiltonian* function

$$H(\sigma) \coloneqq \sum_x J_{x,x+1} \cdot \sigma_x \cdot \sigma_{x+1}$$

The Boltzmann/Gibbs distribution is given by $\Pr[\sigma] \propto \exp\big(H(\sigma)\big)$
The normalizing "constant" is known as the ***partition function*** $Z(J) \coloneqq \sum_\sigma \exp\big(H(\sigma)\big)$
Combinatorially, partition functions can be seen as generating polynomials

Suppose that for some $J > 0, \forall x,\ J_{x,x+1} = J$, smaller cuts are preferred, "ferro-magnets"
by setting $\beta = \exp(-J)$,

$$Z(J) = \sum_{S \subset V} \beta^{|E(S,S^C)| - (|E| - |E(S,S^C)|)} = \beta^{-|E|} \sum_{S \subset V} \beta^{2|E(S,S^C)|}$$

Suppose that for some $J < 0, \forall x,\ J_{x,x+1} = J$, larger cuts are preferred, "Antiferro-magnets"
Suppose that $\forall x,\ J_{x,x+1} = 0$, then there is no interaction in the system, a product distribution

Counting, sampling vs. inference

Computing $(1 \pm \epsilon)\, Z$

Equivalent to

- **(Approximate) sampling**

Sampling from the Gibbs distribution?

- **Approximate inference**

Given partial observation of the system, what can you infer about the rest?

- …

Given a Gibbs distribution π, how do you design a Markov chain with π as its stationary distribution?

Reversible Markov chains

Markov chains are random walks on directed graphs in general

The analog of “random walks on undirected graphs” are reversible MC

Definition

Let π be a distribution. A Markov chain P is *reversible* with respect to π if

$$\pi(x)P(x,y) = \pi(y)P(y,x), \qquad \forall x,y.$$

A symmetric transition matrix P is trivially reversible w.r.t the uniform distribution

This is also known as the *detailed balance* condition

Observation

If a Markov chain P is reversible w.r.t. π, then π is a stationary distribution for P.

Proof.

$$(\pi P)(y) = \sum_x \pi(x)P(x,y) = \sum_x \pi(y)P(y,x) = \pi(y)\sum_x P(y,x) = \pi(y).$$

Reversible Markov chains

Markov chains are random walks on directed graphs in general

The analog of “random walks on undirected graphs” are reversible MC

Definition

Let π be a distribution. A Markov chain P is *reversible* with respect to π if

$$\pi(x)P(x,y) = \pi(y)P(y,x), \qquad \forall x, y$$

A symmetric transition matrix P is trivially reversible w.r.t the uniform distribution

This is also known as the *detailed balance* condition

Observation

If a Markov chain P is reversible w.r.t. π, then π is a stationary distribution for P.

Observation

If a Markov chain P is reversible w.r.t. π, then P is similar to a symmetric matrix.

Proof. Let $T = diag\left(\sqrt{\pi(x)}\right)$. Then, reversibility means $\sqrt{\frac{\pi(x)}{\pi(y)}}P(x,y) = \sqrt{\frac{\pi(y)}{\pi(x)}}P(y,x)$, so TPT^{-1} is symmetric

Glauber Dynamics

A general way to construct Markov chains with stationarity

$$\pi(\sigma) = \frac{w(\sigma)}{Z}$$

Let σ be the current state:

- Choose a vertex v u.a.r
- Update the spin $\sigma_v \leftarrow \tau$ with probability $\propto \pi\left(\sigma_{\setminus v} \cup \{\sigma_v = \tau\}\right)$

Note that for Ising model, one only needs to know how many neighbors of v is assigned +/-

In other words, it suffices to know $\sigma_{N(v)}$

Aside: Phase transitions and Markov chains

How fast can a system out-of-equilibrium, return to

a unique thermal equilibrium (Gibbs measure)?

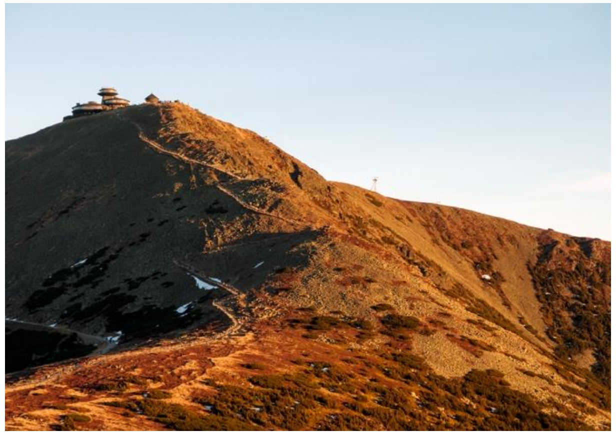
Glauber dynamics is a Markov chain (algorithm) that also attempts to model such a process

- **Rapid (thermally) mixing**

- **Torpid mixing**

Example: Mean-field model of Imitation

Binary vote: option A or option B



N individuals/voters, each time an individual i will cast a vote

$$\sigma_i(t) = \begin{cases} +1, & \text{option A} \\ -1, & \text{option B} \end{cases}$$

They are influenced by the global trend, i.e. by the "mean-field"

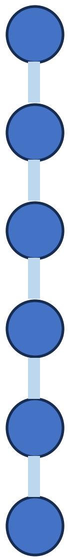
$$\Pr[\sigma_i(t+1) = +1] \propto \exp(\text{global trend(t)} \times J),$$

$$\text{global trend(t)} = \sum_{j=1}^{N} \sigma_j(t)$$

J = "inverse temperature" decides how much individuals rely on the dominating trend

Essentially, the Glauber dynamics for Ising model on a complete graph

1D Ising model with mixed interactions



Ising (1925) exactly solved this model in his thesis to show that there is
no *phase transition* in 1D

We see a ``coupling proof'' of a related phenomenon:
the Glauber dynamics mixes rapidly in 1D even with mixed interactions

We assume that all interactions have the same absolute magnitude,
but may not have the same sign

$$|J_{x,x+1}| = |J_{y,y+1}|, \forall x, y$$

Glauber dynamics:

- Choose a vertex v u.a.r
- The neighbors of v are $v - 1$ and $v + 1$
- Update $\sigma_v \leftarrow \begin{cases} +1, \text{with probability} \propto \exp(J_{v-1,v} \cdot \sigma_{v-1} + J_{v,v+1} \cdot \sigma_{v+1}) \\ -1, \text{with probability} \propto \exp(-J_{v-1,v} \cdot \sigma_{v-1} - J_{v,v+1} \cdot \sigma_{v+1}) \end{cases}$

Simplifying assumption: $|J_{x,x+1}| = |J_{y,y+1}|, \forall x, y$

Ising model on a path graph

Glauber dynamics:

- Choose a vertex v u.a.r
- The neighbors of v are $v-1$ and $v+1$
- Update $\sigma_v \leftarrow \begin{cases} +1, \text{with probability} \propto \exp(J_{v-1,v} \cdot \sigma_{v-1} + J_{v,v+1} \cdot \sigma_{v+1}) \\ -1, \text{with probability} \propto \exp(-J_{v-1,v} \cdot \sigma_{v-1} - J_{v,v+1} \cdot \sigma_{v+1}) \end{cases}$

GD1

Let $\theta := \frac{2\exp(-|J_{v-1,v}|-|J_{v,v+1}|)}{\exp(-|J_{v-1,v}|-|J_{v,v+1}|)+\exp(|J_{v-1,v}|+|J_{v,v+1}|)}$

An equivalent description of the dynamics:

- Choose a vertex v u.a.r
- With probability θ, we update $\sigma_v \leftarrow \text{Bernoulli}\left(\frac{1}{2}\right)$
- With probability $1-\theta$, we look at the value $K := J_{v-1,v} \cdot \sigma_{v-1} + J_{v,v+1} \cdot \sigma_{v+1}$
  - If $K \in \{\pm(|J_{v-1,v}| + |J_{v,v+1}|)\}$, update $\sigma_v \leftarrow \text{sign}(K)$
  - If $K = 0$, update $\sigma_v \leftarrow \text{Bernoulli}\left(\frac{1}{2}\right)$

GD2

Simplifying assumption: $|J_{x,x+1}| = |J_{y,y+1}|, \forall x, y$

Ising model on a path graph

Let $\theta := \frac{2\exp(-|J_{v-1,v}|-|J_{v,v+1}|)}{\exp(-|J_{v-1,v}|-|J_{v,v+1}|)+\exp(|J_{v-1,v}|+|J_{v,v+1}|)}$

An equivalent description of the dynamics:

- Choose a vertex v u.a.r
- With probability θ, we update $\sigma_v \leftarrow \mathrm{Bernoulli}\left(\frac{1}{2}\right)$
- With probability $1-\theta$, we look at the value $K := J_{v-1,v} \cdot \sigma_{v-1} + J_{v,v+1} \cdot \sigma_{v+1}$
    - If $K \in \{\pm(|J_{v-1,v}| + |J_{v,v+1}|)\}$, update $\sigma_v \leftarrow \mathrm{sign}(K)$
    - If $K = 0$, update $\sigma_v \leftarrow \mathrm{Bernoulli}\left(\frac{1}{2}\right)$

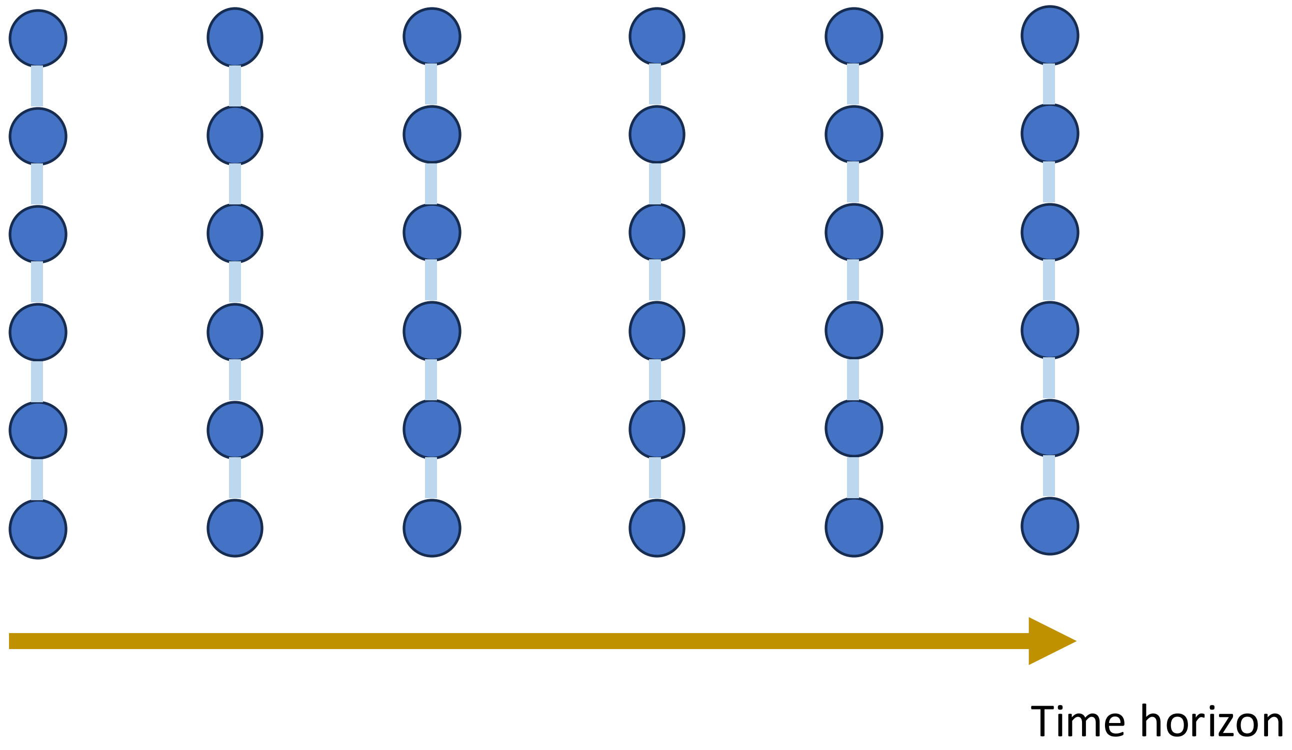
GD2

Furthermore, this is equivalent to the following:

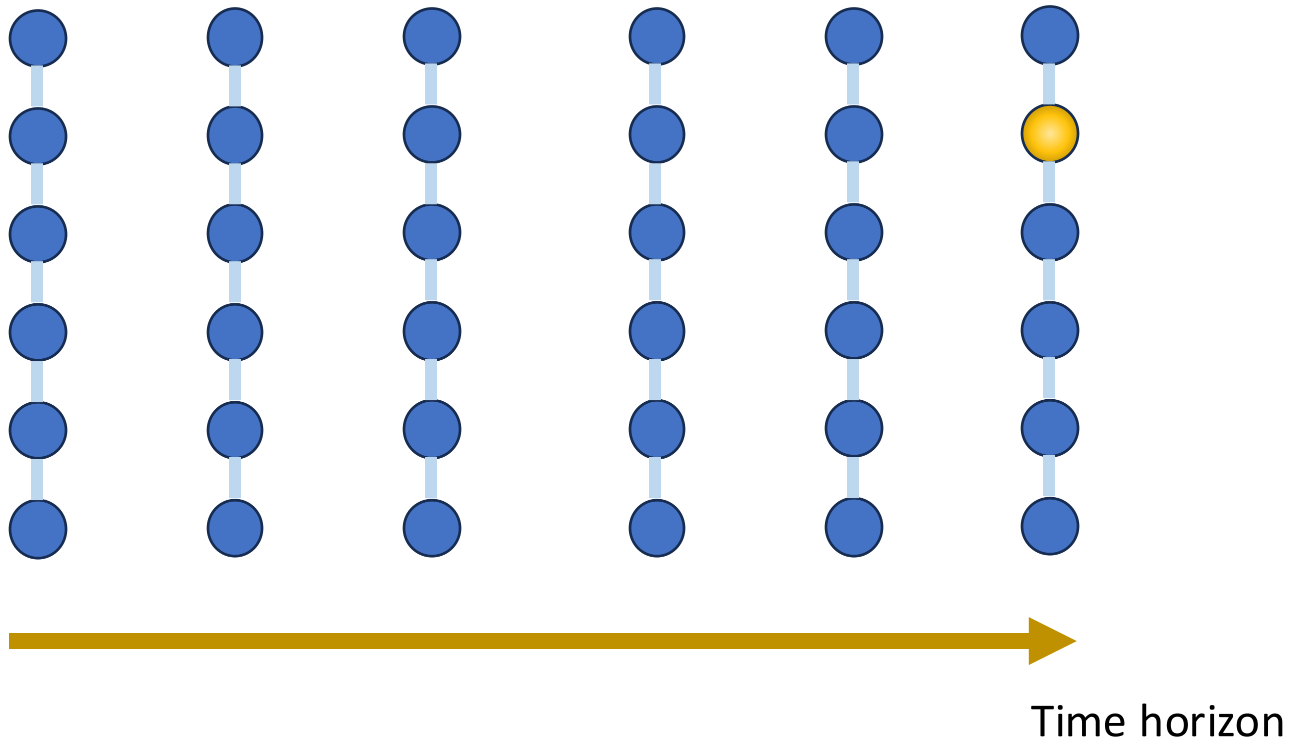
- Choose a vertex v u.a.r
- With probability θ, we update $\sigma_v \leftarrow \mathrm{Bernoulli}\left(\frac{1}{2}\right)$
- With probability $1-\theta$, we choose a random neighbor $u \in \{v-1, v+1\}$ and update $\sigma_v \leftarrow \mathrm{sign}(J_{u,v}\sigma_u)$

GD3

Ising model on a path graph



Ising model on a path graph



This is known as "information percolation"

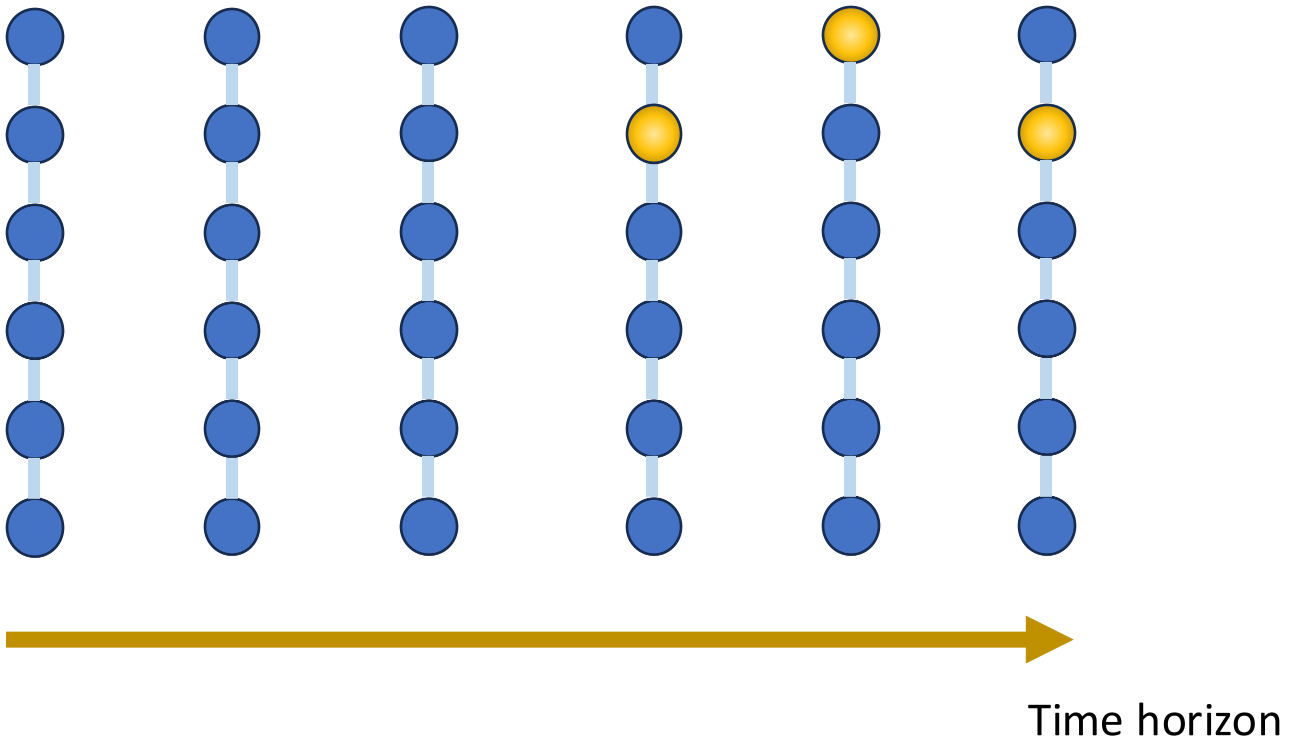
Ising model on a path graph

A random walk on $\mathbb{Z}_n$ that dies with probability θ

After $\frac{1}{\theta}\ln n$ updates per vertex, the probability that there is any surviving branch is small

Balls into bins tells us that this can be achieved after $\frac{C}{\theta} n \ln n$ steps of Glauber dynamics for large C

Chernoff + union bound



Expander Graphs

- Combinatorial: graphs with good expansion
- Probabilistic: graphs in which random walks mix rapidly
- Algebraic: graphs with large spectral gap

Let G be a d-regular graph, and let $d = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq -d$ be the spectrum of its adjacency matrix.

We will be interested in the spectral radius, given by

$$\alpha := \max\{\alpha_2, |\alpha_n|\}$$

If α is much smaller than d, we have good **spectral expansion**.

There are many nice properties associated with expander graphs

Among others, say if we want more than one sample in MCMC, do we have to resample entirely?

Cauchy-Schwarz inequality:
$\langle u, v\rangle \leq \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$

Expander Mixing lemma

Intuitively, an expander can be seen as an approximation to the complete graph, because edges are distributed evenly

Induced edges: $E(S,T) \coloneqq \{(u,v): u \in S, v \in T, uv \in E\}$
We also allow non-disjoint S,T, in which case an edge can be counted twice.

Expander Mixing lemma

Let G be a d-regular graph with n vertices. If the spectral radius of G is α, then for every $S \subseteq [n], T \subseteq [n]$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \alpha\sqrt{|S||T|}.$$

Proof: Note that $E(S,T) = \chi_S^T A \chi_T$. Let $\chi_S = \sum_i a_i v_i, \chi_T = \sum_i b_i v_i$, where $\{v_i\}$ is an orthonormal basis for A, with eigenvalues $\{\alpha_i\}$.

$$E(S,T) = \frac{d|S||T|}{n} + \sum_{i\geq 2} \alpha_i a_i b_i .$$

By Cauchy-Schwarz,

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \alpha \|a\|_2 \|b\|_2 = \alpha \|\chi_S\|_2 \|\chi_T\|_2 = \alpha\sqrt{|S||T|}$$

Expander Mixing lemma

Intuition: Expander mixing lemma tells us that a spectral expander looks like a random graph.

Exercise: Let G be a d-regular graph with spectral radius α. Show that the size of the maximum independent set of G is at most $\frac{\alpha n}{d}$.

Use this result to conclude that the chromatic number is at least $\frac{d}{\alpha}$.

Converse to Expander Mixing lemma

(By Bilu and Linial)
Suppose that for every $S \subseteq [n], T \subseteq [n]$ with $S \cap T = \emptyset$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha\sqrt{|S||T|}.$$

Then all but the largest eigenvalue of A in absolute value is at most

$$O\left(\alpha\left(1 + \log\frac{d}{a}\right)\right)$$

- Proof is based on LP duality
- Would be nice to see an analog of Trevisan's Cheeger's rounding proof

Existence of expanders

- Complete graphs are obviously the best expanders in terms of “expansion” (in all three notions of “expansion”)
- What’s interesting is the existence of sparse expanders: e.g. d-regular expanders for constant d

- A random d-regular graph is a (combinatorial) expander with high probability
- However, deterministic and explicit construction of expanders seems to be much harder to come up with

Alon-Boppana Bound

- For d-regular graphs, how small can the spectral radius be?
- Ramanujan graphs: graphs whose spectral radius are at most $2\sqrt{d-1}$

Alon-Boppana Bound

Let G be a d-regular graph with n vertices, and α_2 be the second largest eigenvalue of its adjacency matrix. Then

$$\alpha_2 \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{\lfloor \mathrm{diam}(\mathrm{G})/2 \rfloor}$$

Alon-Boppana Bound

<u>An easy lower bound on spectral radius</u>

Let G be a d-regular graph with n vertices, and α be its spectral radius. Then $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

Proof: Consider $\mathrm{Tr}(A^2)$. Counting length-2 walks we have

$$\mathrm{Tr}(A^2) \geq nd$$

On the other hand, $\mathrm{Tr}(A^2) = \sum_i \alpha_i^2 \leq d^2 + (n-1)\alpha^2$.

Combined, we have $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

For the Alon-Boppana bound, one may consider $\mathrm{Tr}\left(A^{2k}\right)$.

Trace method/trick

Random walks in expanders

- We knew that it mixes rapidly, in time $O\left(\frac{\log n}{1-\epsilon}\right)$ for $\alpha = \epsilon d$.
- Perhaps surprisingly, not just the final vertex is close to the uniform distribution, but the entire sequence of walks looks like a sequence of independent samples for many applications.
- In fact, expander random walks can fool many test functions:

  ***Expander random walks: a Fourier-analytic approach*, by Cohen, Peri and Ta-Shma**

Probability amplification

Say you have a randomized algorithm that fails with probability β

To boost success probability, we can run it multiple times until it succeed

Run independently for t rounds, the failure probability becomes β^t

Q: Can we save randomness while still achieving the same probability amplification?

Imagine a random walk on the $N = 2^n$ random bits

There is a set B of size βN that we try to escape from (or avoid)

We want that the escape probability close to β^t

Q: Can we use a sparse expander instead of a complete graph for the random walk?

Hitting property of expander walks

Let G be a d-regular graph with n vertices, $\alpha = \epsilon d$ be its spectral radius and B be a set of size at most βn.

Then, starting from a uniformly random vertex, the probability that a t-step random walk has never escaped from B, denoted by $P(B, t)$, is at most $(\beta + \epsilon)^t$.

$\Pr[X_0 \in B, X_1 \in B, X_2 \in B, \ldots, X_t \in B]$

Remarks before a proof:

- Compare this to a sequence of independent samples.
- Expander mixing lemma is like $t = 2$: Note that $\varphi(S) = \Pr(X_2 \notin S \mid X_1 \sim \pi_S)$
- Bound can be strengthened → see Chapter 4 of ***Pseudorandomness*, by Vadhan**
- Applications to error reduction for randomized algorithms
  - Instead of using kt bits of randomness, only need $k + O(t \log d)$
  - for one-sided error, escaping the bad set of "random bits"
  - for two-sided error, a Chernoff type bound can also be shown → then take the majority of the answers

Hitting property of expander walks

$$\Pi_B = \begin{matrix} & B & V\backslash B \\ B & I_B & 0 \\ V\backslash B & 0 & 0 \end{matrix}$$

$\Pi_B \Pi_B = \Pi_B$

$W = \frac{1}{d} A$

$u = \frac{1}{n} \vec{1}$

Proof. Observe that $P(B,t) = \|(\Pi_B W)^t \Pi_B u\|_1$

To see this, notice that $\Pr[X_0 \in B] = \|\Pi_B u\|_1$

$$\Pr[X_0 \in B, X_1 \in B] = \|\Pi_B W \Pi_B u\|_1$$

And so on and so forth.

Suppose that we can show $\forall f: f$ is a probability distribution, we have

$$\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$$

Then,

$$\begin{aligned} \|(\Pi_B W)^t \Pi_B u\|_1 &\leq \sqrt{n}\|(\Pi_B W)^t \Pi_B u\|_2 \\ &= \sqrt{n}\|(\Pi_B W \Pi_B)^t u\|_2 \\ &\leq \sqrt{n}(\beta + \epsilon)^t \|u\|_2 \\ &= (\beta + \epsilon)^t \end{aligned}$$

Cauchy-Schwarz inequality:

$\langle u, v\rangle \leq \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$

$$\Pi_B = \begin{array}{c} \\ B \\ V\backslash B \end{array} \begin{array}{cc} B & V\backslash B \\ \begin{bmatrix} I_B & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$W = \frac{1}{d}A$ has $\lambda_2(W^\top W) = \epsilon^2$

Hitting property of expander walks

Proof (cont'd):It remains to show $\forall f$: f is a probability distribution,

$$\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$$

Without loss of generality, we can assume f is supported only on B.

$$\|\Pi_B W \Pi_B f\|_2 = \|\Pi_B W f\|_2 = \|\Pi_B W(u+v)\|_2 \le \|\Pi_B u\|_2 + \|\Pi_B W v\|_2$$

$u = \frac{1}{n}\vec{1}$, so $\frac{\langle f,u\rangle}{\langle u,u\rangle}u = u$, then $v \perp \vec{1}$

Next, $\|\Pi_B W v\|_2 \le \|Wv\|_2 \le \epsilon\|v\|_2 \le \epsilon\|f\|_2$.

On the other hand, $\|\Pi_B u\|_2 = \sqrt{\frac{\beta}{n}} \le \beta\|f\|_2$,

where last inequality follows from Cauchy-Schwarz:

$$1 = \|f\|_1 = \langle 1_B, f\rangle \le \sqrt{\beta n}\|f\|_2$$

Combined together, we have $\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$ as desired.

Cauchy-Schwarz inequality:

$$\langle u, v\rangle \le \sqrt{\langle u,u\rangle} \cdot \sqrt{\langle v,v\rangle}$$

Hitting property of expander

To get a tail bound, consider
$P(S,t) = \left\|\Pi_{Z_t} W \Pi_{Z_{t-1}} W \ldots \Pi_{Z_1} u\right\|_1$
where $S = (Z_t, Z_{t-1}, \ldots, Z_1)$
indicates whether $Z_i \in \{B, \bar{B}\}$

Proof. Observe that $P(B,t) = \|(\Pi_B W)^t \Pi_B u\|_1$

Suppose that we can show $\forall f: f$ is a probability distribution, we have $\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$. Then,

$$\|(\Pi_B W)^t \Pi_B u\|_1 \leq \sqrt{n}\|(\Pi_B W)^t \Pi_B u\|_2 = \sqrt{n}\|(\Pi_B W \Pi_B)^t u\|_2 \leq \sqrt{n}(\beta + \epsilon)^t \|u\|_2 = (\beta + \epsilon)^t$$

It remains to show $\forall f: f$ is a probability distribution,

$$\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$$

Without loss of generality, we can assume f is supported only on B.

$$\|\Pi_B W \Pi_B f\|_2 = \|\Pi_B W f\|_2 = \|\Pi_B W(u + v)\|_2 \leq \|\Pi_B u\|_2 + \|\Pi_B W v\|_2$$

Next, $\|\Pi_B W v\|_2 \leq \|W v\|_2 \leq \epsilon\|v\|_2 \leq \epsilon\|f\|_2$.

On the other hand, $\|\Pi_B u\|_2 = \sqrt{\frac{\beta}{n}} \leq \beta\|f\|_2$,

The last inequality follows from Cauchy-Schwarz:

$$1 = \|f\|_1 = \langle 1_B, f\rangle \leq \sqrt{\beta n}\|f\|_2$$

Combined together, we have $\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$ as desired.