CS 294 - Partition Functions

Joon Kim

Introduction

Def] Partition Function: polynomial $Z(\lambda) = \sum_{k=0}^{n} a_k \lambda^k$ where $a_k \geq 0$.

Goal: Compute $Z(\lambda)$ for various λ.

↳ Why? Intrinsic interest in combinatorics, coeff. $\{a_k\}$ are interesting, connection to Gibbs distribution ($\pi_\lambda(\sigma) = \frac{1}{Z(\lambda)} \lambda^{E(\sigma)}$ ↳ "energy" of config σ), derivatives of Z carry additional info, notion of phase transitions by singularities in $\log Z$

Course Scope: Complexity/Exact Computation, Approximation Algorithms, Approximating $Z(\lambda) \Leftrightarrow$ Sampling from π_λ, MCMC & Mixing Times, Correlation Decay (Spatial Properties), Geometry of Polynomials & Zeros, Connections with Phase Transitions

Counting Problems

Decision Problems $\longrightarrow$ Counting Problems (how many sat. assignments?)

ex) For bipartite graph G, $\exists$ perfect matching? If so, how many?

Let $Z(x) := \sum_k a_k x^k$ be the generating function for integer seq $\{a_k\}$.

ex) $a_k :=$ # of assignments that satisfy exactly k clauses of φ. Then, $a_m =$ # of satisfying assgns ($m =$ # clauses).

ex) $a_k :=$ # of matching of k edges. Then, $a_n =$ # of perfect matchings $\left(n = \frac{|V|}{2}\right)$.

Spin Systems

Def) Spin System: graph G, set of spins $\{1, \dots, q\}$. Configurations are assgns $\sigma : V \to [q]$. Hamiltonian $H(\sigma) := \sum_{(u,v)\in E} f(\sigma(u), \sigma(v)) + \sum_{v \in V} g(\sigma(v))$. ← symmetric

$\hookrightarrow \Pr[\text{system is in } \sigma] = \frac{1}{Z(\beta)} \exp(-\beta H(\sigma))$ where $\beta > 0$ is a parameter. (inverse temp.)

$Z(\beta) := \sum_\sigma \exp(-\beta H(\sigma))$. Let $\lambda = \exp(-\beta)$. Then, we can rewrite to $Z(\beta) = \sum_k a_k \lambda^k$ where $a_k =$ # of configs where $H(\sigma) = k$.

Examples of Spin Systems:

1) Ferromagnetic Ising Model: spin $\{\uparrow,\downarrow\}/\{+,-\}$.

+	−	+	+	+	−
−	−	+	−	−	+
+	+	−	+	−	+

$f(+,+)=f(-,-)=0,\ f(+,-)=f(-,+)=1.$

$g(+)=-h,\ g(-)=0$ where $h\geq 0$ is external.

$H(\sigma)=\#\text{Disagreements}(\sigma)-h\#\text{Pluses}(\sigma).$

$Z_G(\beta,h)=\sum_\sigma \exp(-\beta H(\sigma)) \longrightarrow Z_G(\lambda,\mu)=\sum_\sigma \lambda^{\#\text{Dis}(\sigma)}\mu^{\#\text{Plus}(\sigma)}$ where

$\lambda\leftarrow \overset{\leq 1}{\exp(-\beta)},\ \mu\leftarrow\exp(\beta h)\Rightarrow \sum_{j,k} a_{jk}\lambda^j\mu^k$ where $a_{jk}:=\#$ of configs with j disagreements and k pluses, enumerating all σs!

$\to$ If we set $\mu=1$ ($h=0$, no ext. field), $Z_G(\lambda)=\sum_j b_j\lambda^j$ where $b_j:=$ # of cuts in G with j edges $\Rightarrow$ Cut generating polynomial!

2) Anti-Ferromagnetic Ising: $f(+,+)=f(-,-)=1,\ f(+,-)=f(-,+)=0.$

$\hookrightarrow H(\sigma)=\#\text{Agreements}(\sigma)-h\#\text{Pluses}(\sigma)=|E|-\#\text{Dis}(\sigma)-h\#\text{Pluses}(\sigma)$

$\to Z_G(\lambda,\mu)=\lambda^{-|E|}\sum_{j,k}a_{jk}\lambda^j\mu^k$ where $\lambda\leftarrow\overset{\geq 1}{\underline{\exp(\beta)}},\ \mu\leftarrow\exp(\beta h).$

3) Potts Model: Ising, but an arbitrary number $q\geq 2$ of spins.

$f(i,j)=\begin{cases}1 & \text{if } i\neq j\\ 0 & \text{if } i=j\end{cases}$ for Ferro. $\to Z(\lambda)=\sum_{S_1\dots S_q}\lambda^{|E(S_1\dots S_q)|}$, $\lambda\in(0,1]$ ferro., $\lambda\geq 1$ antif.

4) Independent Sets/Hard-Core Model: $\sigma=\{\text{ind. sets in } G\}$.

$[q]=\{0,1\}$ where 1 is occupied, 0 unoccupied.

$f(1,1) = \infty$, $f(i,j) = 0$ otherwise. $g(k) = \begin{cases} \ell & \text{if } k=1 \\ 0 & \text{if } k=0. \end{cases}$

$\rightarrow Z(\lambda) = \sum_I \lambda^{|I|}$ where $\lambda = \exp(-\beta \ell) \rightarrow Z(\lambda) = \sum_k a_k \lambda^k$ where

$a_k :=$ # of ind. sets in G with exactly k vertices, $\lambda \in (0, \infty)$

* Spin Glass Model has (random) mixture of ferro. and antif. interactions.

Markov Random Fields

a.k.a. undirected graphical models. For a graph $G(V,E)$, RVs $\{X_v \mid v \in V\}$ (assume discrete) have the markov property s.t. if a set S seperates A and B, $X_A \perp\!\!\!\perp X_B$ given X_S.

Obs) Spin systems are MRFs.

Thm) Hammersly-Clifford: Let μ be an MRF on $G(V,E)$ with discrete r.v.s $\{X_v\}$ s.t. μ assigns nonzero prob. to every config. $\{X_v \mid v \in V\}$. Then, μ is equivalent to a spin system on G with Hamiltonian $H(x) := \sum_C \psi_C(X_C)$ where C ranges only over cliques in G.

Proof [Bessay]: WLOG, assume 0 is a valid value for all r.v.s and that $\pi(\vec{0}) = \pi(00..0) \neq 0$. Define $g(x) := \log \frac{\pi(x)}{\pi(\vec{0})}$. We can uniquely write $g(x) = \sum_{U \subseteq V} \varphi_U(X_U)$ for real valued function φ_U that only depend on X_U and satisfy $\varphi_U(x_U) = 0$ unless $x_U > \vec{0}$. E.g., sps all r.v.s are 0-1 valued.

$\rightarrow \varphi_{\{v\}}(x_v) = \log \frac{\pi(\vec{1}_v)}{\pi(\vec{0})}$ $\forall v$ where $\vec{1}_v$ assigns $x_v = 1$, $x_u = 0$ $\forall u \neq v$.

Now, we want $g(\vec{1}_{uv}) = \varphi_{uv}(\vec{1}_{uv}) + \varphi_u(\vec{1}_u) + \varphi_v(\vec{1}_v)$ by definition.

Thus, $\varphi_{uv}(x_{\{u,v\}}) = g(\vec{1}_{uv}) - \varphi_u(x_u) - \varphi_v(x_v)$, and build up iteratively. //

$\rightarrow \pi(x) = \pi(\vec{0}) \exp(\sum_{U \subseteq V} \varphi_U(x_U))$. We shall show that $\varphi_U \equiv 0$ $\forall U$ that is not a clique.

Consider u, v s.t. $(u,v) \notin E$. Let $u, v \in U$. For any valuation x, set $x' := x$ except $x'_u = 0$. Then $g(x) - g(x') = \log \frac{\pi(x)}{\pi(x')} = \log \frac{\pi(x_u | x_{V \setminus \{u\}})}{\pi(0 | x_{V \setminus \{u\}})}$ is independent of x_v by the Markov property. Also consider that $g(x) - g(x') = \sum_{U \ni u} \varphi_U(x_U)$, which should also be indep. of x_v.

Set $x_w = 0$ $\forall w \notin \{u,v\}$ and set $x_u \neq 0$. Then, $\varphi_u(x_u) + \varphi_{uv}(x_{uv})$ is indep. of x_v. $\Rightarrow \varphi_{uv} \equiv 0$. //

#P-Completeness

[Cook]

For NP-Completeness, we have 3-SAT as the canonical problem.

Def) #P: a function $f: \Sigma^* \rightarrow \mathbb{N}$ is in # if $\exists$ a poly. time NDTM M s.t. # of accepting comps. of M on input x is exactly $f(x)$.

Def) #P-Complete: $f \in \#P$ and all $g \in \#P$ reduces to it in poly. time.

A "parsimonious" reduction preserves the # of solutions.

Fact) $\exists$ many natural decision problems $\in P$ whose counting version is $\in$ #P-Complete.

ex) DNF-SAT: $(x_1 \land x_2 \land \bar{x}_3) \lor (\cdots\cdots) \rightarrow$ ORs of ANDs

Let φ be any boolean formula in 3-CNF. Then we can reduce $\#SAT(\varphi) = 2^n - \#DNFSAT(\neg\varphi) \Rightarrow$ DNF counting is #P-C. //

ex) Permanent: Let A be an $n \times n$ matrix. $per(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}$.

Thm) [Valiant] Computing $per(A)$ for a 0-1 matrix A is #P-Complete.

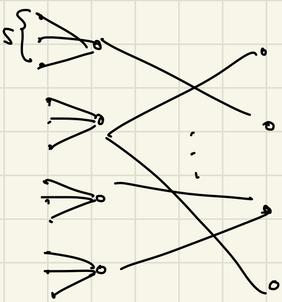
Fact) $per(A) =$ # of perfect matchings in G_A, the bipartite graph of A.

Proof of Thm: Observe that $per(A) =$ # of weighted cycle covers of G'_A, the directed graph of A. The goal is to take an arbitrary 3-CNF φ and construct a directed graph G_φ s.t. # of cycle covers of $G_\varphi =$ $\alpha(n,m) \cdot \#SAT(\varphi)$ where $n :=$ # of variables in φ, $m :=$ # of clauses in φ, and α is some poly. time pre-computable function. [Gadgets in slides/notes]

With appropriate gadget construction, $\alpha(m,n) := 4^{3m}$. Lastly, make all integer edge weights into $\{\pm 1\}$ by splitting, and since $\text{per}(A) \in [-N!, N!]$ where $N :=$ # of vertices in G_φ, we can actually compute $\text{per}(A) \bmod 2^{N^2}+1$ as $N! < 2^{N^2+1}$. Then, we can repace -1 by 2^{N^2}, and make them a series of N^2 edges of weight 2, and split. //

Thm) #MATCH(G) is #P-Complete. [For simplicity, assume G is bipartite]

Proof: Consider $Z_G(\lambda) = \sum_k m_k \lambda^k$ where $m_k :=$ # of k-edge matchings.



Consider an augmented graph G_s. Then, #MATCH(G_s) $= \sum_k m_k (s+1)^{n-k}$. If we have, say, $2n$ vertices, we can evaluate this for $(n+1)$ distinct s and compute all m_k by Lagrange Interpolation!

Thm') Evaluating $Z_G(\lambda) \in$ #P-Complete $\forall \lambda > 0$.

Proof: weight each original edge by λ. Then we get $Z_{G_s}(\lambda) = (1+s\lambda)^n Z_G\left(\frac{\lambda}{1+s\lambda}\right)$. This is the same hardness as above. //

Exact Counting

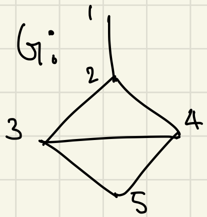
1) # Spanning Trees (Matrix-Tree Theorem)

2) FKT Algorithm for planar perfect matchings

"Holographic Algorithms" by Valiant

Thm) (Kirchoff's) Matrix-Tree Theorem: $\#ST(G) = (1,1)$ minor of $\underset{\text{Laplacian}}{\mathcal{L}(G)}$

$*\ \mathcal{L}(G) = \mathrm{diag}(d_i) - A$, (1,1) minor is determinant with first row and column deleted. (all (i,i) minors of $\mathcal{L}(G)$ are equal [Exercise])

ex) G: 

$$\begin{bmatrix} \cancel{1} & \cancel{-1} & \cancel{0} & \cancel{0} & \cancel{0} \\ \cancel{-1} & 3 & -1 & 0 & 0 \\ \cancel{0} & -1 & 3 & -1 & -1 \\ \cancel{0} & -1 & -1 & 3 & -1 \\ \cancel{0} & 0 & -1 & -1 & 2 \end{bmatrix}$$

$\#ST(G) = 8 = \det([\ ])$.

Proof: Let $B :=$ edge/vertex incidence matrix, e.g., for the above G,

	{1,2}	{2,3}	{2,4}	{3,4}	{3,5}	{4,5}
1	+1					
2	-1	+1	+1			
3		-1				
4			-1			
5						

$\mathcal{L} = BB^T$ [Exercise]. Then, $\det(\mathcal{L}) = \sum_S \det(B_S)\det(B_S^T)$ where B_S are square submatrices of B.

$\rightarrow \det(\mathcal{L}) = \sum_S \det(B_S)^2$. Similarly, $\det(\mathcal{L}_{(1,1)}) = \sum_S \det(B'_S)^2$ where $B' :=$ B with first row removed. To prove the theorem, we need to show that $\det(B'_S)^2 = \begin{cases} 1 & \text{if } S \text{ is a spanning tree} \\ 0 & \text{if } S \text{ is not a spanning tree.} \end{cases}$ Observe that

choosing some nonzero diagonal. If it is acyclic, it corresponds to a directed spanning tree rooted at vertex 1 and has $\det(B_S') = \pm 1$.
On the other hand, if the subset S contains a cycle, we can construct a linear combination of the cycling edge columns s.t. they sum to $\emptyset$, hence $\det(B_S') = \emptyset$. // just greedily set all row sums to $\emptyset$

Corollary) We can uniformly sample spanning trees by fixing a random edge and finding the decision tree prob. of including it in the ST.

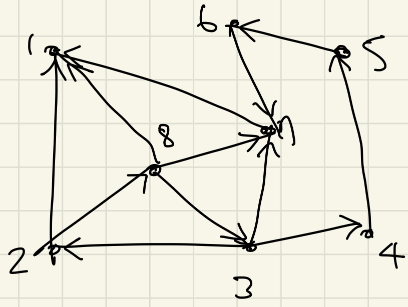
Alg) FKT Planar Perfect Matchings: Planar $G(V,E)$ where $|V| \overset{=2k}{}$ is even $\rightarrow$ # perf. match
(Inspired by dimer coverings of monomer-dimer system of lattice)
Let $\pi := \{\{i_1, j_1\}, \cdots, \{i_k, j_k\}\}$, a perfect matching. Let $A :=$ skew-symmetric matrix of G, $a_{ij} := \begin{cases} +1 & \text{if } \{i,j\} \in E \text{ and } i<j \\ -1 & \text{''} \quad i>j \\ 0 & \text{if } \{i,j\} \notin E \end{cases}$. The Pfaffian of A, $Pf(A) := \sum_{\pi} \text{sgn}(\pi) \prod_{l=1}^{k} a_{i_l, j_l}$ where $\text{sgn}(\pi) :=$ sign of $\begin{bmatrix} 1 & 2 & \cdots & (2k-1) & 2k \\ i_1 & j_1 & \cdots & i_k & j_k \end{bmatrix}$. Then $Pf(A) = \sum_{M} s(M)$ where M are perfect matchings and $s(M) := \text{sgn}(\pi_M) \prod_{l=1}^{k} a_{i_l, j_l}$.
Lemma: For any $2k \times 2k$ skew-symmetric matrix, $Pf(A)^2 = \det(A)$ [Optional Ex].
↳ $Pf(A)$ can be calculated in polytime.
Lemma 2: A Pfaffian orientation of G is an orientation of its edges s.t. $s(M)$ has the same sign for all perfect matchings M. Any planar graph G

has a Pf orientation. $\quad$ $*$:= # of edges of each orientation is odd.

Proof: Sps. $\exists$ orientation in which every even cycle C in G for which $G \setminus V(C)$ contains a perfect matching is oddly oriented*. Then, this is a Pf orientation. Let M_1, M_2 be any 2 pf. matchings. It is s.t.p. $s(M_1)\cdot s(M_2) = +1$. Take any cycle C in $M_1 \sqcup M_2$. By the condition, C must be oddly oriented. Analyzing, $s(M_1)s(M_2) = +1$. Now, sps. a planar drawing of G has an orientation in which the # of clockwise edges around (not necessarily even) face (except possibly the outer face) is odd. Then, this orientation is Pfaffian. This is a reduction to the previous more general statement. [In lecture notes, omitted here]

↳ Every planar graph has a Pf. orientation & can be found in polytime. (Consider a ST and its dual ST of faces. Greedily orienting suffices.)

ex) Graph from Note 3:



$$
\begin{array}{c}
\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array}
\begin{bmatrix}
0 & \boxed{-1} & & & & & \boxed{-1} & \boxed{-1} \\
\boxed{1} & 0 & \boxed{-1} & & & & & \boxed{1} \\
 & \boxed{1} & 0 & \boxed{1} & & & \boxed{1} & \boxed{-1} \\
 & & \boxed{-1} & 0 & \boxed{1} & & & \\
 & & & \boxed{-1} & 0 & \boxed{1} & & \\
 & & & & \boxed{-1} & 0 & \boxed{1} & \\
\boxed{1} & & \boxed{-1} & & & \boxed{-1} & 0 & \boxed{-1} \\
\boxed{1} & \boxed{-1} & \boxed{1} & & & & \boxed{1} & 0
\end{bmatrix}
\end{array}
$$

$\det(A) = 16$, # perfect matchings $= 4 = \sqrt{16}$

Approximations

$f: \Sigma^* \to \mathbb{R}^+$. What does it mean to approximate this function?

Def) Fully Polynomial (Randomized) Approximation Scheme (FPTAS/FPRAS): algorithm A s.t. on inputs (x, ε) where $x \in \Sigma^*$, $\varepsilon \in (0,1]$, outputs a (random) values $A(x,\varepsilon)$ s.t. $\Pr[(1+\varepsilon)^{-1} f(x) \leq A(x,\varepsilon) \leq (1+\varepsilon) f(x)] \geq 3/4$ and runs in time $\text{poly}(|x|, \varepsilon^{-1})$.

↳ we can boost $3/4$ with repeated trials and median finding [CS 174].

$\Rightarrow$ it suffices to run $t = O(\log \delta^{-1})$ trials to boost $3/4 \to 1-\delta$!

ex) Let $f(G)$ be # of proper colorings of G with q colors. Sps. $\exists$ an algo. that approximates $f(G)$ within constant C. Consider $G^{(m)}$, the m-copied version of G. Then $f(G^{(m)}) = (f(G))^m$. If we set $C^{1/m} = 1+\varepsilon$, $m = \frac{\log C}{\log(1+\varepsilon)} = O(\varepsilon^{-1})$. $\Rightarrow$ arbitrary precision with polynomial blowup

Approximate Counting $\xleftrightarrow{\text{equiv.}}$ Random Sampling $\to$ polytime alg. that outputs config σ from (close to) Gibbs distrib. ($\pi_\lambda(\sigma) = \frac{\lambda^k}{Z(\lambda)}$ where $H(\sigma) = k$)

For $Z(\lambda)$, we allow distribution $\hat{\pi}$ s.t. $\|\pi - \hat{\pi}\|_{TV} \leq \delta$.

Sampling $\to$ Counting Methods:

(1) ex) q-colorings, uniform. Define $\Omega :=$ set of proper q-coloring of G.

$|\Omega| = \frac{|\Omega_0|}{|\Omega_1|} \times \frac{|\Omega_1|}{|\Omega_2|} \times \cdots \times \frac{|\Omega_{n-1}|}{|\Omega_n|} \times \cancel{|\Omega_n|}^{1} =: \prod_{i=1}^{n} \frac{1}{Z_i} \times \cancel{|\Omega_n|}^{1}$ where $\Omega_i :=$ set of proper q-colorings of G with ordered vertices $1, \dots, i$ pinned to some values and $Z_i := \frac{|\Omega_i|}{|\Omega_{i-1}|} \leq 1$. Then, each Z_i can be estimated by sampling u.a.r. from Ω_{i-1} and observing proportion $\hat{Z}_i$ of colorings in which vertex i has coloring $c_i \leftarrow$ maximal proportion, to reduce variance.

Thm) Unbiased Estimator: Let $X_1, \dots, X_n$ be iid nonneg. r.v.s with $\mu = E[X_i]$, $\sigma^2 = \mathrm{Var}(X_i)$. Let $y = \frac{1}{t}\sum_{i=1}^{t} X_i$. Then $\Pr[|y-\mu| \geq \varepsilon\mu] \leq 1/4$ provided $t \geq \frac{4}{\varepsilon^2}\frac{\sigma^2}{\mu^2}$. Proof: Chebyshev.

$\rightarrow$ sample size of Z_i: $O(\frac{1}{\varepsilon^2} \times \frac{1}{Z_i}) \rightarrow O(\frac{1}{\varepsilon^2} \times q)$ since $\frac{1}{Z_i} \leq q$ w.h.p.

* the sampling blackbox algo. should be able to sample with prior pinnings!

↳ this is not a big problem as long as the problem is "self-reducible".

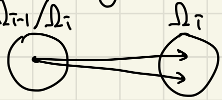
finally, we actually need $\varepsilon \mapsto \varepsilon/n$ bound for each trial for $(1+\frac{\varepsilon}{n})^n \sim 1+\varepsilon$, so each Z_i actually needs $O(\frac{n^2}{\varepsilon^2} q)$ trials, and since we have n Z_is, we need $O(\frac{n^3}{\varepsilon^2} q)$ samples (or by boosting, $O(\frac{n^2 \log n}{\varepsilon^2})$).

(2) q-coloring again, hard constraints. $|\Omega| = \frac{|\Omega_m|}{|\Omega_{m-1}|} \times \frac{|\Omega_{m-1}|}{|\Omega_{m-2}|} \cdots \times \frac{|\Omega_1|}{|\Omega_0|} \times \underbrace{|\Omega_0|}_{= q^n}$

$=: \prod_{i=1}^{n} \frac{1}{y_i} \times q^n$ where $\Omega_i :=$ set of q-colorings with only i edges present.

We can estimate $y_i = \frac{|\Omega_i|}{|\Omega_{i-1}|}$ by random sampling from Ω_{i-1} and observing proportion of them that is still consistent with i-th edge ($\sigma(u) \neq \sigma(v)$).

Claim: Assuming $q \geq \Delta + 1$ ($\Delta :=$ max deg. of G), $\frac{|\Omega_i|}{|\Omega_{i-1}|} \geq \frac{1}{2}$.

Proof: 

$|\Omega_i| \geq |\Omega_{i-1}| - |\Omega_i| \Rightarrow \frac{|\Omega_i|}{|\Omega_{i-1}|} \geq \frac{1}{2}$. //

↳ the rest of analysis is equivalent to (1).

(3) $Z(\beta) = \sum_\sigma \exp(-\beta H(\sigma))$. WLOG, assume that $H: \sigma \to \mathbb{N}^+$ and that $|H(\sigma)| \leq n^c$ for constant c. Goal is estimating $Z(\beta)$.

Let $Z(\beta) = \frac{Z(\beta_t)}{Z(\beta_{t-1})} \times \frac{Z(\beta_{t-1})}{Z(\beta_{t-2})} \times \cdots \times \frac{Z(\beta_1)}{Z(\beta_0)} \times Z(\beta_0)$ where $0 = \beta_0 < \beta_1 < \cdots < \beta_t = \beta$.

Choose $\beta_i = \beta_{i-1} + \frac{1}{n^c}$. $\to t \sim n^{2+c}$ since WLOG, $\beta \leq n^2$ (later justified).

$$\frac{Z(\beta_i)}{Z(\beta_{i-1})} = \frac{1}{Z(\beta_{i-1})} \sum_\sigma \exp(-\beta_i H(\sigma)) = \frac{1}{Z(\beta_{i-1})} \sum_\sigma \exp(-\beta_{i-1} H(\sigma)) \exp((\beta_{i-1} - \beta_i) H(\sigma))$$

$= \mathbb{E}_{\beta_{i-1}}[\exp((\beta_{i-1} - \beta_i) H(\sigma))]$, and this is just an observable.

$|\beta_{i-1} - \beta_i| = \frac{1}{n^c}$, $|H(\sigma)| \leq n^c$, so the r.v. $\exp(\sim)$ lies in const. range $[e^{-1}, e]$.

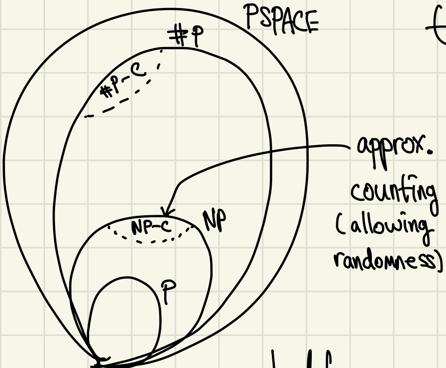
$\to \beta \leq n^2$ is a reasonable assumption since $\beta \geq n^2$ is a trivial ground state.

Counting $\longrightarrow$ Sampling: Consider the decision tree 

where each layer picks a color for each vertex. Every leaf corresponds to a proper coloring. As long as the counting error is $\leq 1 + \frac{1}{n}$, error is $(1 + \frac{1}{n})^n \leq e$, a constant. Let's say we sampled σ. $P_\sigma :=$ $\Pr[\text{reach leaf } \sigma]$ is known. We output σ w.p. $[2e\hat{N}P_\sigma]^{-1}$. Otherwise, reject & try again. $\hat{N} :=$ approx (upto factor 2) # of total colorings.

Claim: $1 \le 2e\hat{N}P_e \le 4e^2$. Proof: use bounds above.

Thm) $\forall f \in \#P$, $\exists$ a <u>polytime</u> randomized algorithm with an NP-oracle that satisfies the FPRAS requirements.



Algo: $\forall x \in \Sigma^*$, let $\overset{\subseteq\{0,1\}^m}{\Omega(x)}$ be the set of accepting computations of TM M on x. Repeatedly intersect $\Omega(x)$ with indep. uniform half-spaces in $\{0,1\}^m$. Define $S_t := \{z \in \Omega \mid z \cdot y_i = 0 \pmod 2$ for $1 \le i \le t\}$. Let $t^* := \min_t \{|S_t| = \emptyset\}$. Output 2^{t^*} as an estimate.

Analysis: Assume $|\Omega| = N$ lies in $2^{k-1} < N \le 2^k$. Observe that $|S_t|$ is a sum of N 0-1 r.v.s, each with prob 2^{-t}, pairwise indept.

$\rightarrow E[|S_t|] = N2^{-t}$. $Var(|S_t|) = N2^{-t}(1-2^{-t}) \le N2^{-t} = E[|S_t|]$.

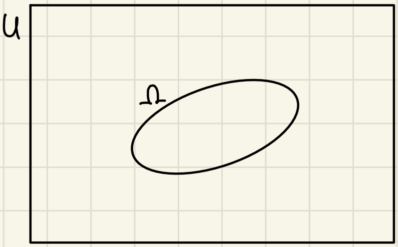
Claim: (i) if $t \ge k+3$, $Pr[|S_t| > 0] \le 1/8$. $\rightarrow E[|S_t|] = N2^{-t} \le 2^k 2^{-(k+3)} = 1/8$. By Markov's Inequality, $Pr[|S_t| > 0] \le 1/8$.

(ii) if $t \le k-4$, $Pr[|S_t| = 0] \le 1/8$. $\rightarrow E[|S_t|] = N2^{-t} \ge 2^{(k-1)} 2^{-(k-4)} = 8$.

By Chebyshev's Inequality, $Pr[|S_t| = 0] \le \frac{Var(|S_t|)}{E[|S_t|]^2} \le \frac{1}{E[|S_t|]} \le 1/8$. //

$\Rightarrow 2^{t^*}$ lies within a factor of 16 of N w.p. $\ge 3/4$. Replicate r TMs and take the r-th root to find a suitable $(1+\varepsilon)$ factor. //

Basic Monte Carlo



$U := \{0,1\}^m$, $\Omega := \{z \in \{0,1\}^m \mid z \text{ is a "solution"}\}$.

Repeat t times: pick $z \in U$ u.a.r.

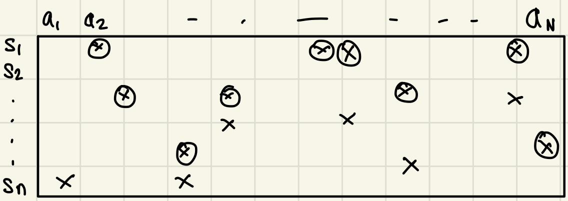
$X_i = \mathbb{1}\{z \in \Omega\}$. Output $\underbrace{|U| \cdot \frac{\sum^t X_i}{t}}$.

Fact: # samples to get a $(1+\varepsilon)$-approx. w.p. $\geq 3/4 \rightarrow t \geq \frac{4}{\varepsilon^2} \cdot \underbrace{\frac{1}{p}}_{\hookrightarrow \text{can be exp. large!}}$ where $p = \frac{|\Omega|}{|U|}$.

#DNF: φ is OR of ANDs. DNF $\in$ P, but #DNF $\in$ #P-Complete.

Let $S_i :=$ set of assignments that satisfy the i-th term. Then, $|\Omega| = |\bigcup_{i=1}^{m} S_i|$.

$|S_i| = 2^{n-r_i}$ where $r_i :=$ # of variables in i-th term. We use a new universe



$U = \{(a,i) \mid a \text{ satisfies the } i\text{-th term}\}$.

$\rightarrow$ pick $(a,i) \in U$ u.a.r. Set $X_j :=$ $\mathbb{1}\{S_i$ is the lowest numbered term satisfied by $a\}$. $\Omega := \{(a,i) \mid S_i \text{ is the lowest term satisfied by } i\}$. Repeat t times, and output $\underbrace{|U| \cdot \frac{\sum^t X_j}{t}}$.

Analysis: (i) $\frac{|\Omega|}{|U|} = \frac{\#\text{ of } \otimes \text{ in table}}{\#\text{ of } \times \text{ in table}} \geq 1/m$ since $\exists \otimes$ for every column.

$\rightarrow$ sample size is $t = O(m/\varepsilon^2)$.

(ii) implementation is efficient: picking $(a,i) \in U$ is poly time, checking $(a,i) \in \Omega$ is also easy. $\rightarrow O(nm)$ per sample.

$\Rightarrow$ total runtime $O(nm^2/\varepsilon^2)$, can improve to $(\frac{nm}{\varepsilon^2})$ [Karp/Luby]. //

* this extends to probabilistic DNF, $\Pr[\varphi$ is sat.$]$ when each $\Pr[X_i = T] = p_i$.

Network Reliability: Connected $G(V,E)$, fail pr. p_e for each edge. What is $P_{fail} := \Pr[G$ becomes disconnected under random edge failures$]$?

$\rightarrow P_{fail} := \Pr[\bigcup S_i]$ where $S_i :=$ i-th cut fails. However, # of cuts is large.

* Karger's MinCut Algorithm: while $|V| \geq 2$, pick an edge $(u,v) \in E$ u.a.r., merge u & v. output the remaining cut. If C is a min-cut, $\Pr[C$ is output$] \geq \frac{1}{\binom{n}{2}}$. [Analysis omitted, can be found in CS174 notes]

$\Pr[C_\alpha$ is output$] \geq 1/\binom{n}{2\alpha}$

$\hookrightarrow$ Corollary: # of min-cuts $\leq \binom{n}{2}$. Also, # of α-min-cuts (size $\leq \alpha c$) $\leq \binom{n}{2\alpha}$.

Ingredients for Algo: 1) probabilistic DNF Karp-Luby $\rightarrow$ FPRAS

2) Karger's Min-Cut $\rightarrow$ # of α-min-cuts $\leq \binom{n}{2\alpha} \leq n^{2\alpha} \rightarrow$ can be enumerated in poly time w.h.p. due to coupon collecting $(n^{2\alpha}\log(n^{2\alpha}))$

Algo) [Assume min-cut size is c, and $p_e = p\ \forall e$] $(G(V,E), p) \rightarrow P_{fail}$

- If $p^c \geq \frac{1}{n^4}$, then use basic Monte Carlo. (runtime $O(\varepsilon^{-2} \cdot n^4)$)
- Else, set $\alpha := 2 + \frac{1}{2}\log_n(2/\varepsilon)$.

Claim*: $\Pr[$some cut size $\geq \alpha c$ fails$] \leq \varepsilon p^c \leq \varepsilon P_{fail}$.

is "absorbed" into $(1+\varepsilon)P_{fail}$ error!

- Left with only α-min-cuts. Then, $\Pr[\text{some } \alpha\text{-min-cut fails}]$
$= \Pr\left[\bigvee_{i=1}^{t} (X_{e_{i_1}} \wedge X_{e_{i_2}} \wedge \cdots \wedge X_{e_{i_r}})\right]$ where $X_{e_{ij}}$ is 1 w.p. p. and
$t \leq \#$ of α-min-cuts $\leq n^{2\alpha} = \frac{2n^4}{\varepsilon}$.

$\hookrightarrow$ Use Karp-Luby to compute $(1+\varepsilon)$ approx. in poly$(n, 1/\varepsilon)$ time.

Proof of Claim*: Enumerate the cuts of size $\geq \alpha c$ in increasing order of size
$\alpha c \leq c_1 \leq c_2 \cdots$.

(i) Consider the first $n^{2\alpha}$ of these cuts. $\forall$ such cuts, $\Pr[\text{cut fails}] \leq p^{\alpha c}$.
$\rightarrow \Pr[\text{any of such cuts fails}] \leq n^{2\alpha} \cdot p^{\alpha c}$. By assumption, $p^c = n^{-(4+\delta)}$.
$\rightarrow \leq n^{2\alpha} \cdot n^{-(4+\delta)\alpha} = n^{-(2+\delta)\alpha}$.

(ii) $\forall \beta > 0$, $\#$ of β-min-cuts $\leq n^{2\beta}$. Hence, $c_{n^{2\beta}} \geq \beta c \Rightarrow c_k \geq \frac{c}{2}\log_n k$ ($c_i = |i\text{-th cut}|$)
$\rightarrow p^{c_k} \leq p^{c/2 \log_n k} = n^{-(4+\delta)(\log_n k)/2} = k^{-(2+\delta/2)}$.
$\rightarrow \Pr[\text{any of remaining cuts fail}] \leq \sum_{k > 2^\alpha} k^{-(2+\delta/2)} \leq \int_{n^{2\alpha}}^{\infty} x^{-(2+\delta/2)}\,dx \leq n^{-(2+\delta)\alpha}$.

$\Rightarrow \Pr[\text{any of cuts size} \geq \alpha c \text{ fails}] \leq 2n^{-(2+\delta)\alpha}$. Substituting $\alpha = 2 + \frac{1}{2}\log_n(2/\varepsilon)$,
$\leq \varepsilon n^{-(4+\delta)} = \varepsilon p^c \leq \varepsilon P_{fail}$.//

* Calculating $P_{\text{not-fail}} = 1 - P_{fail}$ is actually not reducible to P_{fail}!
[Guo, Jerrum] came up with a different algorithm for this.

Markov Chain Monte Carlo

Input: implicit description of a very large set Ω, weight function $w: \Omega \to \mathbb{R}^+$.

Goal: sample $\sigma \in \Omega$ from distribution $\pi(\sigma) := \frac{w(\sigma)}{\sum_{\sigma'} w(\sigma')}$.

$\to$ Equivalently, compute $\sum_{\sigma' \in \Omega} w(\sigma')$, the partition function.

Def) Markov Chain: sequence of r.v.s s.t. $\Pr[X_{t+1} = y \mid X_0, \dots, X_t] = \Pr[X_{t+1} = y \mid X_t]$.

Then we can let $P_{ij} := \Pr[X_{t+1} = j \mid X_t = i]$. P is the <u>transition matrix</u>.

$\to P$ is <u>stochastic</u>, i.e., $P_{xy} \geq 0$, $\sum_y P_{xy} = 1 \ \forall x$. [Denote $P_{xy} \equiv P(x,y)$]

Notation: $p_x^{(t)} :=$ distribution of X_t given that $X_0 = x$.

$\to p_x^{(t+1)} = p_x^{(t)} P$, i.e., $p_x^{(t+1)}(u) = \sum_z p_x^{(t)}(z) P(z,y)$.

Def) Reversibility: $\exists$ distribution π s.t. $\pi(x) P(x,y) = \pi(y) P(x,y) \ \forall x,y$.

↳ think of the "flow" of $x \to y$ and $y \to x$ being same! $x \rightleftarrows y$

↳ then we can represent P as an <u>undirected graph</u>, $Q(x,y) = \pi(x) P(x,y)$

Thm) Fundamental Theorem of MC: If P is <u>irreducible</u> and <u>aperiodic</u>, (and <u>finite</u>) then it has a unique stationary distribution $\pi > 0$; this is the unique left eigenvector of P with eigenvalue 1. Also, $p^{(t)}(x,y) \xrightarrow[t \to \infty]{} \pi(y) \ \forall x,y$.

Def) Irreducibility: $\forall x,y,\ \exists t$ s.t. $P^{(t)}(x,y) > 0$.

Def) Aperiodic: $\forall x,y,\ \gcd\{t \mid P^{(t)}(x,y) > 0\} = 1$.

Fact: If $\overset{\swarrow \text{rev, aper.}}{P}$ is reversible w.r.t. π, then $\pi > 0$ is the unique stationary dist.

Proof: $(\pi P)(y) = \sum_x \pi(x) P(x,y) = \sum_x \pi(y) P(y,x) = \pi(y) \cancelto{1}{\sum_x P(y,x)}$. //

Fact: If P is reversible, then P is <u>similar</u> to a symmetric matrix. Hence, all eigenpairs will be real and P is diagonalizable.

Proof: Let $D := \mathrm{diag}(\sqrt{\pi(x)})$. Then $DPD^{-1}(x,y) = \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}} P(x,y)$
$= \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}} P(y,x) = DPD^{-1}(y,x)$. //

Thm) [Perron-Frobenius]: Any irreducible, aperiodic stochastic matrix P has an eigenvalue $\lambda_1 = 1$ with unique left eigenvector $e_1 > 0$ and all other eigenvalues λ_i satisfy $|\lambda_i| < 1$.

↳ This implies the Fund. Thm. of MC since we can represent $p^{(0)} = \sum \alpha_i \overset{\text{eigenvector}}{\overset{\downarrow}{e_i}}$,
$p^{(t)} = p^{(0)} P^t = \sum \alpha_i \lambda_i^t e_i \xrightarrow[t\to\infty]{} \alpha_1 e_1 = e_1$ ($\alpha_1 = 1$ b/c conservation of "mass").

* λ_2 dictates how fast $p^{(t)}$ converges to e_1, and the quantity $(1-\lambda_2)$ is called the <u>spectral gap</u>.

ex1) Random walk on undirected graph $G(V,E)$, $\Omega = V$.

at state $u \in V$, pick v u.a.r. from neighbor(u) and transition to u

[Exercise!]

- irreducible iff G is connected, aperiodic iff G is not bipartite.
- reversible w.r.t. $\pi(x) = \frac{\deg(x)}{2|E|}$.

ex2) Card Shuffling, Ω = set of permutations of n cards.

$\pi(\sigma) = 1$ (uniform dist.)

(i) Random Transposition: pick two positions i, j. switch them. ($\to i=j$ is allowed)

- irreducible & aperiodic. reversible w.r.t. π as actions are symmetric.

(ii) Top-in-at-Random: Take top card, move to a uniform random position.

- irreducible & aperiodic. reversible? $\to$ no, it's not symmetric. However, we can observe that P is doubly stochastic. $\to$ uniform π must be stationary.

(iii) Gilbert-Shannon-Reeds Riffle: split deck into L & $R \sim Bi(n, \frac{1}{2})$.

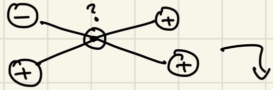
drop cards from L/R w.p. $\frac{|L|}{|L|+|R|} \Big/ \frac{|R|}{|L|+|R|}$ (equiv. to uniform interleaving).

- irreducible? think of L being only the top card. reduces to (ii). ✓
- aperiodic? think of L being empty. ✓
- not reverible, but doubly stochastic (consider the "inverse riffle").

ex3) Glauber Dynamics, spin system on $G(V,E)$, spins $= \{1, \dots, q\}$.

Hamiltonian $H(\sigma)$, Gibbs $\pi(\sigma) \propto \exp(-\beta H(\sigma))$. $\Omega = \{\sigma : V \to [q]\}$.

in config σ, pick vertex $v \in V$ u.a.r., resample $\sigma(v)$ from correct conditional dist. $\pi(\cdot \mid \sigma_{V \setminus v})$.



Consider Ising, $\pi(\sigma) \propto \lambda^{\#\mathrm{Dis}(\sigma)}$, $\lambda = \exp(-2\beta)$. $\to \sigma(v) \propto \begin{cases} \lambda^3 & (-) \\ \lambda & (+) \end{cases}$.

- irreducible? trivial since we can set any v to anything. (no hard constraints)

↳ (in general, this is not true. consider q-coloring on q-clique.)

- aperiodic? trivial.
- reversible? $\pi(\sigma)P(\sigma,\tau) = \pi(\tau)P(\tau,\sigma) \to \frac{\pi(\sigma)}{\pi(\tau)} = \frac{P(\cdot,\sigma)}{P(\cdot,\tau)}$. ✓

ex4) Metropolis Process. Ω, $w : \Omega \to \mathbb{R}^+$, sample $\pi(\sigma) \propto w(\sigma)$.

construct an undirected connected graph on Ω ("neighborhood structure")

define a proposal dist. $K(x,\cdot)$ on $N(x)$, neighbors of x s.t. $K(x,y) > 0$ $\forall y \in N(x)$. at state x, pick $y \in N(x)$ w.p. $K(x,y)$.

go to y w.p. $\min\left\{1, \frac{w(y)K(y,x)}{w(x)K(x,y)}\right\}$. (trivially irred. & aper.)

- reversible w.r.t. π? take neighbors x,y, $w(y)K(y,x) \leq w(x)K(x,y)$.

$w(x)P(x,y) = \cancel{w(x) \cdot K(x,y)} \cdot \frac{w(y)K(y,x)}{\cancel{w(x)K(x,y)}} = w(y)K(y,x) = w(y)P(y,x)$. //

ex5) Swendsen-Wang Algorithm. q-state Potts model, $[q]$.

$\pi(\sigma) \propto \lambda^{\#\mathrm{Dis}(\sigma)}$, $\lambda = \exp(-2\beta)$. in config σ, delete all disagreeing (bichromatic) edges. in the remaining subgraph, retain each edge w.p $p=(1-\lambda)$, otherwise delete it. assign uniformly random spins to each remaining connected components. go to the resulting configuration.

[Exercise] Check that SW dynamic is reversible w.r.t. π.

↳ consider all intermediate configurations after the $p=(1-\lambda)$ thinning.

Coupling

Def) Total Variance Distance: For two prob. dist. μ, υ on Ω,

$$\|\mu-\upsilon\|_{(TV)} := \frac{1}{2}\sum_{\sigma\in\Omega}|\mu(\sigma)-\upsilon(\sigma)| = \max_{A\subseteq\Omega}|\mu(A)-\upsilon(A)|.$$

Def) Coupling: For any two prob. dist. μ, υ on Ω, ω on $(\Omega\times\Omega)$ s.t.

$$\mu(x)=\sum_y \omega(x,y)\ \forall x,\ \upsilon(y)=\sum_x \omega(x,y)\ \forall y.$$

Thm) Coupling Lemma: Let X, Y be RVs with dist. μ, υ on Ω. Then,

(i) $Pr[X\neq Y]\geq\|\mu-\upsilon\|_{TV}$, (ii) $\exists$ coupling of (μ,υ) s.t. $Pr[X\neq Y]=\|\mu-\upsilon\|$.

Proof: Consider the coupling $Pr[X=Y=Z=\sigma]=\min\{\mu(\sigma),\upsilon(\sigma)\}$.

Setup of Fundamental Thm of MC: P is an irred & aperiodic MC, and assume P has a stationary distribution π. Let $\Delta_x(t)=\|P_x^{(t)}-\pi\|_{TV}$,

$\Delta(t) = \max_x \Delta_x^{(t)}$, $D_{xy}(t) = \|P_x^{(t)} - P_y^{(t)}\|$, $D(t) = \max_{x,y} D_{xy}^{(t)}$. Then,

$\Delta(t) \le D(t) \le 2\Delta(t)$ by lower bound and TI.

Claim 1: $\Delta_x(t)$ is non-increasing with t.

Proof: Let $X_0 = x$, $Y_0 \sim \pi$. Fix t and couple X_t, Y_t s.t. $\Pr[X_t \ne Y_t] = \|p_x^{(t)} - \pi\|$.

Couple (X_{t+1}, Y_{t+1}) as follows: if $X_t = Y_t$, $X_{t+1} = Y_{t+1}$. Else, evolve independently.

$\Delta_x^{(t+1)} = \|p_x^{(t+1)} - \pi\| \le \Pr[X_{t+1} \ne Y_{t+1}] \le \Pr[X_t \ne Y_t] = \Delta_x^{(t)}$.

Claim 2: $\forall s,t \in \mathbb{N}$, $D(s,t) \le D(s)D(t)$.

Proof: Let $X_0 = x$, $Y_0 = y$. Couple $P_x^{(t)}, P_y^{(t)}$ s.t. $\Pr[X_t \ne Y_t] = \|p_x^{(t)} - p_y^{(t)}\|_{TV}$.

Then, if $X_t = Y_t$, set $X_{t+i} = Y_{t+i}$ for $i = 1, \dots, s$. Else, let $x' = X_t$, $y' = Y_t$,

where $x' \ne y'$. Use coupling lemma to couple distributions X_{t+s}, Y_{t+s}

conditioned on $X_t = x'$, $Y_t = x'$ s.t. $\Pr[X_{t+s} \ne Y_{t+s} \mid X_t = x', Y_t = y'] = \|p_{x'}^{(s)} - p_{y'}^{(s)}\|$

$= D_{x'y'}(S) \le D(S)$. Then, $D_{xy}(S+t) = \|p_x^{(S+t)} - p_y^{(S+t)}\| \le \Pr[X_{t+s} \ne Y_{t+s}]$

$\le D(s) D_{xy}(t) \le D(s) D(t)$. //

Claim 3: $D(t) < 1$ for some finite t.

Proof: Irreducibility & Aperiodicity $\Rightarrow P^t(x,y) > 0$ $\forall x,y \in \Omega$ for some finite t.

Proof of Fund. Thm. of MC: $\|p_x^{(t)} - \pi\|_{TV} \xrightarrow{t\to\infty} 0$ $\forall x$ suffices. Let t be s.t.

$D(t) = 1 - \delta$ for some $\delta > 0$. $\forall k > 0$, $\Delta(kt) \le D(kt) \le D(t)^k \le (1-\delta)^k$.

$\Rightarrow \Delta(t) \to 0$ as $t \to \infty$, so $\Delta_x(t) \to 0$. //

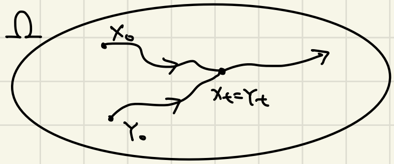
Let $T_x(\varepsilon) := \min\{t \mid \Delta_x(t) = \varepsilon\}$, $T(\varepsilon) := \max_x T_x(\varepsilon)$, $T_{mix} := T(\frac{1}{2e})$.

Claim: $\Delta(t) \leq \exp(-\lfloor \frac{t}{T_{mix}} \rfloor)$.

Proof: $\Delta(kT_{mix}) \leq D(kT_{mix}) \leq D(T_{mix})^k \leq (2\Delta(T_{mix}))^k \leq e^{-k}$. //

Corollary: $T(\varepsilon) \leq T_{mix} \cdot \lceil \ln(\varepsilon^{-1}) \rceil$.

Coupling for Mixing Times



Def) Coupling for MC: a pair process (X_t, Y_t) s.t. (i) if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$, (ii) $\forall t, a, b$,

$\Pr[X_{t+1} = b \mid X_t = a] = \Pr[Y_{t+1} = b \mid Y_t = a] = P(a,b)$.

Def) $T_{xy} := \min\{t \mid X_t = Y_t, X_0 = x, Y_0 = y\}$ under a given coupling.

Claim: $\Delta(t) \leq \max_{x,y}\{\Pr[T_{xy} > t]\}$.

Proof: $\Delta(t) \leq D(t) = \max_{x,y} \| p_x^{(t)} - p_y^{(t)} \|_{TV} \leq \max_{x,y} \Pr[X_t \neq Y_t \mid X_0 = x, Y_0 = y]$

$= \max_{xy} \Pr[T_{xy} > t]$. //

Corollary: for any coupling, $T_{mix} \leq \max_{x,y} \Pr[T_{xy} > \frac{1}{2e}] \leq \max_{x,y} 2e E[T_{xy}]$.

ex1) Random Walk on cube $\{0,1\}^n$. $|\Omega| = 2^n$. Make the graph lazy.
(self-loop w.p. $\frac{1}{2}$)

If moving, pick $i \in [n]$ and flip the i-th bit. Equivalently, pick $i \in [n]$ and $b \in \{0,1\}$ u.a.r., and set $X_i \leftarrow b$.

Coupling: X_t & Y_t both use the same i & b. $\rightarrow d(X_t, Y_t)$ is nonincreasing.

$E[\text{time until } d(X_t, Y_t) = 0] \leq$ coupon-collecting n coupons $= n \ln n + \dots$.

$\Rightarrow T_{mix} \leq n \ln n + O(n)$, $T_{mix} = \frac{1}{2} n \ln n$.

ex2) Top-in-at-Random Shuffle: Symmetric, so inverse path is isomorphic to our real path. Analyzing the inverse is easier.

Coupling: (X_t, Y_t), choose a random card (not position!). Then, the top k cards that are ordered cannot be disturbed.

$\Rightarrow$ again coupon-collecting $\rightarrow T_{mix} = n \ln n + \dots$ //

ex3) Random Transposition Shuffle: Pick position i & card c. Exchange c with card at position i.

Coupling: (X_t, Y_t) chooses same i & c. $\overset{d_t =}{d(X_t, Y_t)} :=$ # of disagreements.

(i) if c is in same position in X_t & Y_t, $d_{t+1} = d_t$.

(ii) else, if card at position i is same, $d_{t+1} = d_t$. else, $d_{t+1} \leq d_t - 1$.

$\Rightarrow d_t$ is non-increasing. $\Pr[d_{t+1} \leq d_t - 1] = \left(\frac{d_t}{n}\right)^2 \rightarrow T_{mix} = O\left(\sum_{t=n}^{1} \left(\frac{n}{d_t}\right)^2\right)$

(* tight T_{mix} is $\frac{1}{2} n \ln n + \dots$) $\quad = O(n^2)$.

Graph Coloring

Connected Undirected $G(V,E)$, # of colors q, assume max degree Δ.

↳ Sample from uniform distribution over proper q-colorings of G.

Conjecture: Standard Glauber dynamics on q-colorings has mixing time poly(n) (actually $O(n\log n)$) provided that $q \geq \Delta+2$.

Glauber Dynamics: pick $u \in V$, $c \in Q$ u.a.r. recolor if possible.

Claim: G.D. is connected (irreducible) if $q \geq \Delta+2$.

Proof Sketch: we can always greedily recolor every vertex.

Claim: When $q \geq 4\Delta+1$, mixing time for G.D. is $O(n\log n)$.

Proof: Coupling. Both X_t & Y_t choose the same v and c.

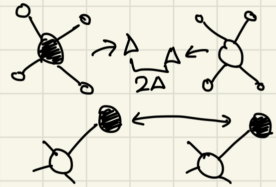
Distance $d(X_t, Y_t) := $ # of disagreeing vertices.

"Good" move $(d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1)$ when v is a disagreeing vertex & color c is not in $N(v)$ for both X_t & Y_t.

"Bad" move $(d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1)$ when v is an agreeing vertex & color c is in $N(v)$ of either X_t or Y_t but not both.

Other moves are neutral (distance stays the same).

At some state, # good moves $\ge d_t(q-2\Delta)$.



bad moves $\le \Delta d_t \cdot 2$. $\Rightarrow q \ge 4\Delta$ ensures that

of good moves $\ge d_t(q-2\Delta) \ge d_t(2\Delta) \ge$ # bad moves!

$\Rightarrow E[d_{t+1} | X_t, Y_t] \le d_t - \frac{d_t(q-2\Delta)}{qn} + \frac{2d_t\Delta}{2} = d_t\left(1 - \frac{q-4\Delta}{qn}\right)$.

$\to E[d_t | X_0, Y_0] \le \left(1 - \frac{q-4\Delta}{qn}\right)^t \cdot \overset{\text{upper bound on } d_0}{n} = \left(1 - \frac{1}{\alpha n}\right)^t n$ where $\alpha := \frac{q}{q-4\Delta}$.

$\Pr[d_t > 0 | X_0, Y_0] = \Pr[d_t \ge 1 | X_0, Y_0] \le E[d_t | X_0, Y_0] \le \left(1 - \frac{1}{\alpha n}\right)^t n$

$\le \varepsilon$ if we take $t = C_\varepsilon\left(\frac{q}{q-4\Delta}\right) n \ln n$. $\Rightarrow T_{mix} = O(n \ln n)$. //

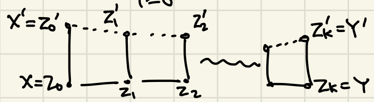
Path Coupling: a <u>pre-metric</u> on Ω is a connected graph with the edge weights s.t. every edge is a shortest path. Say x, y are adjacent if $\exists$ edge (x,y).

Thm) [Bubley, Dyer]: Sps. $\exists$ a coupling $(X,Y) \to (X', Y')$ only defined on pairs (X,Y) that are adjacent in pre-metric, and s.t.

$^*E[d(X',Y') | X,Y] \le (1-\alpha) d(X,Y)$ for some $\alpha \in [0,1]$ where d is the metric defined by the pre-metric. Then, this coupling can be extended to a coupling on all pairs (X,Y) also satisfying the above condition $*$.

Proof: For arbitrary (X,Y), let $X = Z_0 \to \cdots \to Z_k = Y$ be a

shortest path in the pre-metric. $E[d(X',Y')|X,Y] \le$
$E[\sum_{i=0}^{k-1} d(Z_i', Z_{i+1}')|X,Y] = \sum_{i=0}^{k-1} E[d(Z_i', Z_{i+1}')|Z_i, Z_{i+1}] \le (1-\alpha)\sum_{i=0}^{k-1} d(Z_i, Z_{i+1})$
$= (1-\alpha)\, d(X,Y)$. //



For colorings, we now extend Ω to allow invalid colorings. However, there are no transitions into invalid colorings.

Claim: we can get down to $q \ge 2\Delta+1$ with $O(n\log n)$ mixing time.
Proof: we are now considering X, Y that differ in exactly only v_0.
Consider the coupling s.t. if $v \in N(v_0)$, then couple $\text{color}(v_0^X), \text{color}(v_0^Y)$.
"Good" moves when $v = v_0$, $c \notin \text{Color}(N(v_0)) \to$ # good moves $\ge 1 \times (q-\Delta)$
"Bad" moves when $v \in N(v_0)$, $c = (\text{color}(v_0^Y), \text{color}(v_0^X)) \to$ # bad moves $\le \Delta$.
$\Rightarrow T_{mix} = O(n\log n)$ provided that $q \ge 2\Delta+1$ by path coupling. //

Thm) [Vigoda]: $q \ge \frac{11}{6}q + 1$ is sufficient.
Thm) [Hayes, Vigoda, et Al]: $q \ge 1.76\Delta+1$ is suff. if $\Delta = \Omega(\log n)$ and G is triangle-free.
Proof Sketch: # available colors $\ge q-\Delta$ is too pessimistic. Let $A(X,v)$:= # available colors at v in coloring X. Suff. to have $A(X,v) \ge \Delta$.

Consider a uniformly random coloring & assume $N(v)$ are independent.
$\to E[A(X,v)] = q(1-\frac{1}{q})^{\Delta} \propto q e^{-\Delta/q} > \Delta$ provided that $q \geq \alpha\Delta$ where α is unique solution to $x = e^{1/x} \to \alpha \simeq 1.76$.

"$q \geq \Delta + 3$ is suff. provided that girth is a large constant dependent on Δ."
For $q < \Delta$, it is NP-Hard to even approximate # of q-colorings.
For $q \leq \Delta/2$, it is " decide if $\exists$ a q-coloring.

Dobrushin Conditions

Def) Influence Matrix: for a spin system on G with Gibbs dist. π, $R := (\rho_{ij})$, an $n \times n$ matrix, s.t. $\rho_{ij} := \max_{\sigma,\tau \in S_j} \{ \| \pi_i(\sigma, \cdot) - \pi_i(\tau, \cdot) \|_{TV} \}$
and where $S_j := \{ (\sigma, \tau) \mid \sigma, \tau \text{ agree everywhere except } j \}$

Def) Spectral (L2 Operator) Norm: $\|A\| := \sup_{v \neq 0} \frac{\|Av\|_2}{\|v\|_2}$

Thm) [Hayes]: If $\|R\| \leq (1-\delta)$ for some $\delta > 0$, then the mixing time of Glauber dynamics is $\leq \frac{c}{\delta} n \ln n$ for some constant c.
Proof: $(X_t), (Y_t)$. Coupling - at every step, X_t & Y_t choose the same vertex i & use a maximal coupling for the update, i.e.,

$\Pr[X_{t+1}(i) \neq Y_{t+1}(i) \mid X_t, Y_t] = \| \pi_i(X_t, \cdot) - \pi_i(Y_t, \cdot) \|$.

$\to$ Write $p_t(i) := \Pr[X_t(i) \neq Y_t(i) \mid X_0, Y_0]$. coupling argument!

Then $p_{t+1}(i) \leq (1 - \frac{1}{n}) p_t(i) + \frac{1}{n} \sum_{j \in N(i)} \rho_{ij} p_t(j)$. [Exercise]

Let $A := \frac{n-1}{n} I + \frac{1}{n} R$. Then $p_{t+1} \leq A p_t \Rightarrow p_t \leq A^t p_0$.

$\|A\| \leq \frac{n-1}{n} \|I\| + \frac{1}{n} \|R\| \leq (1 - \delta/n)$. Now, $\Pr[X_t \neq Y_t] \leq \|p_t\|_1$ union bound

$\leq \sqrt{n} \|p_t\|_2 \leq \sqrt{n} \|A^t p_0\|_2 \leq \sqrt{n} \|A^t\| \|p_0\|_2 \leq \sqrt{n} (1 - \frac{\delta}{n})^t \cdot \sqrt{n}$

$= n (1 - \frac{\delta}{n})^t \Rightarrow T_{mix} = O(\frac{n}{\delta} \ln n)$. //

ex1) Hard-Core Model: G, spins $:= \{0, 1\}$. Ω = ind. sets of G.

$\pi(\sigma) \propto \lambda^{|\sigma|}$. If some $j \in N(i)$ is occupied, $\Pr[i \text{ occ.}] = 0$. Else,

$\Pr[i \text{ occ.}] = \frac{\lambda}{1+\lambda}$. $\to \rho_{ij}$, worst case is when all $j' \in N(i) - j$ are unocc.,

so $\rho_{ij} = \frac{\lambda}{1+\lambda}$. $\to R = \frac{\lambda}{1+\lambda} A_G$. adjacency matrix of G $\to \|R\| = \frac{\lambda}{1+\lambda} \|A_G\| = \frac{\lambda}{1+\lambda} \cdot \lambda_0(G)$. max eigenv.

We know that $\lambda_0(G) \leq \Delta$. $\to \|R\| \leq \frac{\lambda}{1+\lambda} \Delta$. $\Rightarrow$ Setting $\underline{\lambda \leq \frac{1-\delta}{\Delta}}$,

we get $\|R\| \leq (1 - \delta)$. //

ex2) Ferromag. Ising Model: Spins $\{+, -\}$, $\pi(\sigma) \propto \lambda^{\#\mathrm{Dis}(\sigma)}$.

$\Pr[\sigma(i) = +] = \frac{\lambda^{d-}}{\lambda^{d+} + \lambda^{d-}}$ where $d\pm :=$ # of neighbors with sign $\pm$.

Worst case influence is when $\deg(i)$ is odd & $|d^+ - d^-| = 1$.

$\to$ This gives a bound on $\rho_{ij} \le \frac{\lambda^{-1/2} - \lambda^{1/2}}{\lambda^{-1/2} + \lambda^{1/2}} = \tanh(\beta/2)$ where $\lambda = \exp(-\beta)$.

So $R \le \tanh(\beta/2) A_G \Rightarrow \|R\| \le \tanh(\beta/2) \cdot \lambda_0(G) \le \tanh(\beta/2) \cdot \Delta$

$\Rightarrow$ Rapid mixing provided $\underline{\tanh(\beta/2) \le \frac{1-\delta}{\Delta}}$. //

ex3) Linear Extensions: Partial order $\precsim$ on n elements. Sample uniformly/ count # of linear extensions of $\precsim$ to a total order. (#P-Complete)

Dynamics: pick a position $p \in [n-1]$. Exchange $(p, p+1)$ if legal. (w.p. $1/2$, do nothing)

[Exercise: Check irreducibility (diameter $\binom{n}{2}$)]

Pre-metric: σ, τ are adjacent iff they differ at exactly i & j, $i<j$.

$\underline{d(\sigma,\tau) := j-i}$.

Path Coupling: Case 1) $j \ne i+1$. both σ, τ pick the same p & attempt.

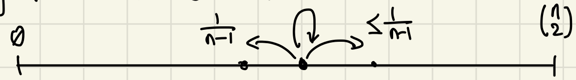
Case 2) $j = i+1$. both σ, τ pick the same p. if $p \ne i$, make the same move. if $p = i$, couple doing nothing in σ & making a move in τ, vice versa.

Analysis: Case I) $p \notin \{i-1, i, j-1, j\}$. moves in σ, τ are identical. $\hookrightarrow$ no change.

Case II) $p = i-1$ or $p = j$. Symmetrical, so WLOG assume $p = i-1$. Worst outcome is d increasing by 1 w.p. $\le 2 \cdot \frac{1}{2(n-1)}$.

Case III) $p = i$ or $p = j-1$. if $j \ne i+1$, then d decreases by 1 since every such attepted move <u>must be legal</u>. $\to \Pr[\text{decrease}] = 2 \cdot \frac{1}{2(n-1)}$.

if $j=i+1$, then d decreases surely from 1 to $\emptyset$.

→ we are doing a random walk 

⇒ $O(n^5 \log n) \simeq O(n^5)$ for high prob. to hit $\emptyset$.

* Tight bound is $O(n^3 \log n)$ [Wilson] (For any partial ordering!)

↳ If we use p sampling w.p. $\propto p(n-p)$, we get this tight bound.

Functional Analysis of MCMC

Motivation: view the MC as an <u>operator</u> on its state space.

Let P be the transition matrix of an ergodic finite MC,
state space Ω, and its stationary distribution π.

Def) Expectation & Variance: $\forall$ real-valued function $\varphi : \Omega \to \mathbb{R}$,

$$E_\pi[\varphi] := \sum_{x\in\Omega} \pi(x)\varphi(x), \quad Var_\pi[\varphi] := \sum_{x\in\Omega} \pi(x)\cdot\left(\varphi(x) - E_\pi[\varphi]\right)^2$$

The transition matrix P is an operator on φ, i.e. $P\varphi(x) := \sum_{y\in\Omega} P(x,y)\varphi(y)$.

↳ $P\varphi(x)$ is the "one-step average" of φ starting at x.

Similarly, $P^t\varphi(x) := \sum_{y\in\Omega} P^t(x,y)\varphi(y)$ is the "t-step average".

Obs) π is stationary, so $E_\pi[P^t\varphi] = E_\pi[\varphi]$ $\forall t$, and as $t\to\infty$,
$P^t\varphi$ converges to $E_\pi[\varphi]$. ⇒ $Var_\pi[P^t\varphi] \to \emptyset$ as $t\to\infty$!

The convergence of MC can be linked to convergence of $\mathrm{Var}_\pi[P^t\varphi]$!

Def) Dirichlet Form: for functions $\varphi, \psi : \Omega \to \mathbb{R}$, $\mathcal{E}_P(\varphi,\psi) := \langle \varphi, \mathcal{L}\psi \rangle_\pi$

where Laplacian $\mathcal{L} := I - P$ and inner product $\langle f, g \rangle_\pi := \sum_x \pi(x) f(x) g(x)$

(also $E_\pi[fg]$!)

The symmetric Dirichlet form $\mathcal{E}_P(\varphi,\varphi) := \langle \varphi, \mathcal{L}\varphi \rangle_\pi$

$\mathbb{1}\{x=y\}$

$$= \sum_x \pi(x)\varphi(x)\left[(I-P)\varphi(x)\right] = \sum_x \pi(x)\varphi(x) \sum_y \left(I(x,y) - P(x,y)\right)\varphi(y)$$

$$= \sum_{xy} \pi(x)\varphi(x)\left(I(x,y) - P(x,y)\right)\varphi(y) = \sum_x \pi(x)\varphi(x)^2 - \sum_{xy} \pi(x)\varphi(x)P(x,y)\varphi(y)$$

$$= \frac{1}{2}\sum_{xy} \pi(x)P(x,y)\left(\varphi(x)^2 + \varphi(y)^2\right) - \sum_{xy} \pi(x)\varphi(x)P(x,y)\varphi(y)$$

$$= \frac{1}{2}\sum_{xy} \pi(x)P(x,y)\left(\varphi(x) - \varphi(y)\right)^2 \,//$$

Comparing to the variance $\mathrm{Var}_\pi[\varphi] := \sum_x \pi(x)\left(\varphi(x) - E_\pi(x)\right)^2$

$$= \sum_x \pi(x)\varphi(x)^2 - E_\pi[\varphi]^2 \quad \text{(use variance formula } \mathrm{Var}(x) = E[x^2] - E[x]^2\text{)}$$

$$= \frac{1}{2}\sum_{xy} \pi(x)\pi(y)\left(\varphi(x)^2 + \varphi(y)^2\right) - \sum_x \pi(x)\varphi(x)\sum_y \pi(y)\varphi(y)$$

$$= \frac{1}{2}\sum_{xy} \pi(x)\pi(y)\left(\varphi(x) - \varphi(y)\right)^2 .$$

Obs) $\mathcal{E}_P(\varphi,\varphi)$ and $\mathrm{Var}_\pi[\varphi]$ are similar, except that $\mathcal{E}_P(\varphi,\varphi)$ is "local".

Def) Poincare Constant: $\alpha := \inf_{\varphi \text{ non-const}} \dfrac{\mathcal{E}_P(\varphi,\varphi)}{\mathrm{Var}_\pi[\varphi]}$.

$\geq 1/2$ self-loop $\forall \sigma$

Thm) For any lazy, ergodic P and any initial state $x \in \Omega$,

$$T_x(\varepsilon) \leq \frac{1}{\alpha}\left(2\ln(\varepsilon)^{-1} + \ln(4\pi(x))^{-1}\right).$$

Proof: The proof of the theorem relies on the following lemma:

Lemma*: $\forall \varphi:\Omega\to\mathbb{R}$, $\mathrm{Var}_\pi[P\varphi] \le \mathrm{Var}[\varphi] - \mathcal{E}_P(\varphi,\varphi)$. [proof deferred]

↳ This immediately gives a $(1-\alpha)$ contraction at every step.

Corollary: $\forall$ non-constant $\varphi:\Omega\to\mathbb{R}$, $\mathrm{Var}_\pi[P^t\varphi] \le (1-\alpha)^t\,\mathrm{Var}_\pi[\varphi]$.

Relating $\mathrm{Var}_\pi[P^t\varphi]$ to variation distance, we get by Cauchy-Schwarz

$$\underbrace{\|p_x^{(t)}-\pi\|_{TV}^2}=\left(\frac{1}{2}\sum_\sigma |p_x^{(t)}(\sigma)-\pi(\sigma)|\right)^2=\frac{1}{4}\left(\sum_\sigma \pi(\sigma)\left|\frac{p_x^{(t)}(\sigma)}{\pi(\sigma)}-1\right|\right)^2$$

$$\le \frac{1}{4}\left(\sum_\sigma \pi(\sigma)\left(\frac{p_x^{(t)}(\sigma)}{\pi(\sigma)}-1\right)^2\right)\cancel{\left(\sum_\sigma \pi(\sigma)(1)^2\right)}^{\to 1}$$

$$=\frac{1}{4}\left(\sum_\sigma \pi(\sigma)\left(\frac{p_x^{(t)}(\sigma)}{\pi(\sigma)}-E_\pi\left[\frac{p_x^{(t)}}{\pi}\right]\right)\right)\quad \left[E_\pi\left[\frac{p_x^{(t)}}{\pi}\right]=\sum_\sigma \cancel{\pi(\sigma)}\frac{p_x^{(t)}(\sigma)}{\cancel{\pi(\sigma)}}=1\right]$$

$$=\underbrace{\frac{1}{4}\mathrm{Var}\left[\frac{p_x^{(t)}}{\pi}\right]}.$$

Now define $\varphi := \frac{p_x^{(0)}}{\pi}$. Note that $P^t\varphi \ne \frac{p_x^{(t)}}{\pi}$.

Def) Time Reversal: $P^*(x,y) := \frac{\pi(y)}{\pi(x)}P(y,x)$. Some properties:

(i) P^* is ergodic and has the same π.

$$(\pi P^*)(x)=\sum_y \pi(y)P^*(y,x)=\sum_y \cancel{\pi(y)}\frac{\pi(x)}{\cancel{\pi(y)}}P(x,y)=\sum_y P(x,y)\pi(x)=\pi(x).$$

(ii) $(P^*)^*=P$. Let $Q:=P^*$. $Q^*(x,y)=\frac{\pi(x)}{\pi(y)}P^*(y,x)=\frac{\cancel{\pi(x)}}{\cancel{\pi(y)}}\frac{\cancel{\pi(y)}}{\cancel{\pi(x)}}P(x,y).$

(iii) $P^*=P$ iff P is reversible. $P(x,y)=\frac{P(y)}{P(x)}P(y,x)=P^*(x,y)\ \forall x,y.$

(iv) $\mathcal{E}_{P^*}(\varphi,\varphi)=\mathcal{E}_P(\varphi,\varphi)\ \forall\varphi$. $\mathcal{E}_{P^*}(\varphi,\varphi)=\frac{1}{2}\sum_{xy}\pi(x)P^*(x,y)(\varphi(x)-\varphi(y))^2$

$$=\frac{1}{2}\sum_{xy}\cancel{\pi(x)}\frac{\pi(y)}{\cancel{\pi(x)}}P(y,x)(\varphi(x)-\varphi(y))^2=\frac{1}{2}\sum_{yx}\pi(y)P(y,x)(\varphi(y)-\varphi(x))^2$$

$$=\frac{1}{2}\sum_{xy}\pi(x)P(x,y)(\varphi(x)-\varphi(y))^2.$$

$\Rightarrow$ The Poincare constant for P and P^* are the same!

Obs) $P^{*t}\varphi = P^{*t}\frac{p_x^{(0)}}{\pi} = \frac{p_x^{(0)}P^t}{\pi} = \frac{p_x^{(t)}}{\pi}$

Proof: $\left(P^*\frac{p_x^{(0)}}{\pi}\right)(y) = \sum_{z\in\Omega} P^*(y,z)\frac{p_x^{(0)}(z)}{\pi(z)} = \sum_z \frac{\cancel{\pi(z)}}{\pi(y)}P(z,y)\frac{p_x^{(0)}(z)}{\cancel{\pi(z)}}$

$= \sum_z P(z,y)\frac{p_x^{(0)}(z)}{\pi(y)} = \frac{p_x^{(1)}(y)}{\pi(y)}.$ Induct on t to get $P^{*t}\varphi = \frac{p_x^{(t)}}{\pi}$. //

$\Rightarrow 4\|p_x^{(t)}-\pi\|_{TV}^2 \le \mathrm{Var}_\pi\left[\frac{p_x^{(t)}}{\pi}\right] = \mathrm{Var}_\pi[P^{*t}\varphi] \le (1-\alpha)^t\mathrm{Var}[\varphi].$

Since $\mathrm{Var}_\pi[\varphi] = \frac{1}{\pi(x)} - 1 \le \frac{1}{\pi(x)}$, setting $t = \frac{1}{\alpha}(2\ln\varepsilon^{-1} + \ln(4\pi(x))^{-1})$

ensures $(1-\alpha)^t\mathrm{Var}_\pi[\varphi] \le 4\varepsilon^2 \Rightarrow \|p_x^{(t)}-\pi\|_{TV} \le \varepsilon$. //

Proof of Lemma*: P is lazy, so $P := \frac{1}{2}(I+\hat{P})$ where $\hat{P}$ is stochastic & has same π. Then, $[P\varphi](x) = \sum_y P(x,y)\varphi(y)$

$= \frac{1}{2}\varphi(x) + \frac{1}{2}\sum_y \hat{P}(x,y)\varphi(y) = \frac{1}{2}\sum_y \hat{P}(x,y)(\varphi(x)+\varphi(y)).$

WLOG, assuming that $E_\pi[\varphi] = 0$ (shifting $\varphi(x)$ doesn't affect $\mathrm{Var}_\pi[X]$),

$\mathrm{Var}_\pi[\varphi] = \sum_x \pi(x)([P\varphi](x))^2 = \frac{1}{4}\sum_x \pi(x)\left(\sum_y \hat{P}(x,y)(\varphi(x)+\varphi(y))\right)^2$

$\le \frac{1}{4}\sum_{xy}\pi(x)\hat{P}(x,y)(\varphi(x)+\varphi(y))^2$ (due to Cauchy-Schwarz).

Also, $\mathrm{Var}_\pi[\varphi] = \frac{1}{2}\sum_x \pi(x)\varphi(x)^2 + \frac{1}{2}\sum_y \pi(y)\varphi(y)^2$

$= \frac{1}{2}\sum_{xy}\pi(x)\varphi(x)^2\hat{P}(x,y) + \frac{1}{2}\sum_{xy}\pi(x)\hat{P}(x,y)\varphi(y)^2$

$= \frac{1}{2}\sum_{xy}\pi(x)\hat{P}(x,y)(\varphi(x)^2+\varphi(y)^2).$

Taking differences, $\mathrm{Var}_\pi[\varphi] - \mathrm{Var}_\pi[P\varphi] \ge \frac{1}{4}\sum_{xy}\pi(x)\hat{P}(x,y)(\varphi(x)-\varphi(y))^2.$

Observe that entries in $\hat{P}$ are twice as large as P except for diagonal entries. However, the diagonals disappear in this sum, so the RHS $= \frac{1}{2}\sum_{xy} \pi(x)\hat{P}(x,y)(\varphi(x)-\varphi(y))^2 = \mathcal{E}_P(\varphi,\varphi)$.

$\Rightarrow \mathrm{Var}_\pi[\varphi] - \mathrm{Var}_\pi[P\varphi] \geq \mathcal{E}_P(\varphi,\varphi)$. //

If P is reversible ($\pi(x)P(x,y) = \pi(y)P(y,x)$), then an alternative proof is given by spectral graph theory.

→ Recall that if P is reversible, $S := DPD^{-1}$ where $D := \mathrm{diag}(\sqrt{\pi(x)})$ is non-negative & symmetric $\Rightarrow$ ev of P is same as S & all are real.

By Perron-Frobenius, spectrum of P is $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n > -1$.

Then, the initial distribution $p^{(0)} = \sum_i \alpha_i \underset{}{\overset{\swarrow \text{evecs}}{e_i}}$ where $e_1 = \pi$, $\alpha_1 = 1$.

$\rightarrow p^{(0)}P^t = \sum_i \alpha_i \lambda_i^t e_i$, and since $|\lambda_i| < 1 \ \forall i > 1$, $p^{(0)}P^t \xrightarrow[t\to\infty]{} \pi$, and the slowest decay is given by $\max_{i>1} |1-\lambda_i|$. If P is lazy, $\lambda_i \geq 0 \ \forall i$, so the rate of convergence is given by <u>spectral gap $(1-\lambda_2)$</u>.

Claim: $\forall$ ergodic, reversible, lazy MC P and $\forall x \in \Omega$, [Exercise!]

$\Delta_x(t) \leq \frac{\lambda_2^t}{2\sqrt{\pi(x)}}$, hence $\tau_x(\varepsilon) \leq \frac{1}{1-\lambda_2}\left(\ln \varepsilon^{-1} + \frac{1}{2}\ln(4\pi(x))^{-1}\right)$.

Claim: for an ergodic & reversible P, $1-\lambda_2 = \inf_{\varphi \text{ non-const}} \frac{\mathcal{E}_P(\varphi,\varphi)}{\mathrm{Var}_\pi[\varphi]}$

Proof: $\mathcal{L} := I - P$ has evals $\mu_i = (1-\lambda_i)$, so $0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_n < 2$

with the same evecs. So the spectral gap of P is just μ_2.

Suppose P is symmetric, then μ_1 minimizes the Raleigh Quotient, $\mu_1 = \inf_{\varphi \neq 0} \frac{\langle \varphi, L\varphi \rangle}{\langle \varphi, \varphi \rangle}$ where $\langle \varphi, \psi \rangle := \sum_x \varphi(x)\psi(x)$. This is given by the principal evec $\vec{1}$. Now, $\mu_2 = \inf_{\varphi \perp \vec{1}} \frac{\langle \varphi, L\varphi \rangle}{\langle \varphi, \varphi \rangle} = \inf_{\varphi \not\parallel \vec{1}} \frac{\langle \varphi, L\varphi \rangle}{\langle \varphi, \varphi \rangle - \langle \varphi, 1 \rangle^2}$.

Extending to reversible P, apply this to $(I - DPD^{-1})$ to obtain $\mu_2 = \inf_{\varphi \perp \vec{1}} \frac{\langle \varphi, L\varphi \rangle_\pi}{\langle \varphi, \varphi \rangle_\pi} = \inf_{\varphi \not\parallel \vec{1}} \frac{\langle \varphi, L\varphi \rangle}{\langle \varphi, \varphi \rangle - \langle \varphi, \vec{1} \rangle^2}$ [Exercise!] where $\langle \varphi, \psi \rangle_\pi$ is defined the same as before. Then, this is just $\frac{\mathcal{E}_P(\varphi, \varphi)}{\mathrm{Var}_\pi[\varphi]}$. //

Multicommodity Flow

To estimate the Poincare constant α, we incorporate flow into the MC.

Here are some basic definitions for an ergodic MC P with s.d. π:

Capacity/Ergodic Flow: $C(e) := \pi(z)P(z, z')$ $\forall$ directed $e = (z, z')$.

↳ represents the flow of probability mass on e when stationary

Demand: $D(x,y) := \pi(x)\pi(y)$ $\forall (x,y) \in (\Omega \times \Omega)$.

Flow: $f : \mathcal{P} \to \mathbb{R}^+ \cup \{0\}$ where $\mathcal{P} := \bigcup_{xy} \mathcal{P}_{xy}$ and $\mathcal{P}_{xy}$ is the set of all simple paths from x to y s.t. $\sum_{p \in \mathcal{P}_{xy}} f(p) = D(x,y)$ $\forall (x,y) \in (\Omega \times \Omega)$.

Cost of flow: $\rho(f) := \max_e f(e)/C(e)$ where $f(e) := \sum_{p \ni e} f(p)$.

Length of flow: $\ell(f) := \max_{p : f(p) > 0} |p|$, length of longest flow-carrying path.

Thm) $\forall$ ergodic P, $\forall$ flow f for P, $\alpha \geq \frac{1}{\rho(f)\ell(f)}$.

Proof: start from symmetrized variance equation and using $D(x,y)$,

$$2\,\mathrm{Var}_\pi[\varphi] = \sum_{xy} \pi(x)\pi(y)(\varphi(x)-\varphi(y))^2 = \sum_{xy}\sum_{p\in P_{xy}} f(p)(\varphi(x)-\varphi(y))^2.$$

Use the following telescoping sum over paths: $\varphi(x)-\varphi(y) = \sum_{(u,v)\in p}(\varphi(u)-\varphi(v))$.

$$\Rightarrow 2\,\mathrm{Var}_\pi[\varphi] = \sum_{xy}\sum_{p\in P_{xy}} f(p)\Big(\sum_{(u,v)\in p}(\varphi(u)-\varphi(v))\Big)^2 \text{ [by Cauchy-Schwarz]}$$

$$\leq \sum_{xy}\sum_{p\in P_{xy}} f(p)\cdot|p|\cdot\sum_{(u,v)\in p}(\varphi(u)-\varphi(v))^2 \text{ [switch order of summations]}$$

$$= \sum_{e=(u,v)}(\varphi(u)-\varphi(v))^2\sum_{p\ni e} f(p)\cdot|p| \leq \ell(f)\sum_{e=(u,v)}(\varphi(u)-\varphi(v))^2\sum_{p\ni e} f(p).$$

[Recall that $\sum_{p\ni e} f(p) = f(e)$ and $\rho(f) = \max_e f(e)/C(e)$]

$$\leq \ell(f)\rho(f)\sum_{e=(u,v)}(\varphi(u)-\varphi(v))^2 C(e) \text{ [substitute } C(e)=\pi(u)P(u,v)]$$

$$= \ell(f)\rho(f)\sum_{e=(u,v)}(\varphi(u)-\varphi(v))^2\pi(u)P(u,v) = 2\,\ell(f)\rho(f)\,\mathcal{E}_P(\varphi,\varphi).$$

$$\Rightarrow \mathrm{Var}_\pi[\varphi] \leq \ell(f)\rho(f)\,\mathcal{E}_P(\varphi,\varphi) \Rightarrow \alpha = \inf_{\varphi \text{ non-const}} \frac{\mathcal{E}_P(\varphi,\varphi)}{\mathrm{Var}_\pi[\varphi]} \geq \frac{1}{\ell(f)\rho(f)}. \quad //$$

Corollary: $\forall$ ergodic, lazy MC P and $\forall$ flow f, the mixing time

$$T_{mix}(\varepsilon) \leq \ell(f)\rho(f)\big(2\ln\varepsilon^{-1} + \ln(4\pi(x))^{-1}\big) \Rightarrow T_{mix} = O(\ell(f)\rho(f)\ln(\underbrace{\pi_{min}}_{\min_{x\in\Omega}\pi(x)})^{-1}).$$

↳ T_{mix} is essentially bounded by $\rho(f)$ since $\underbrace{\ell(f)}_{\approx \text{diameter of MC}}$ & $\underbrace{\ln(\pi_{min})^{-1}}_{\pi \text{ uniform, then } O(n)}$ are usually small.

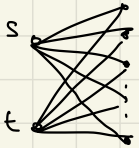
ex1) Hypercube: lazy random walk on $\{0,1\}^n$. Let $N=2^n$ be # vertices.

$\pi(x) = \frac{1}{N}$. $C(u,v) = \pi(u)P(u,v) = \frac{1}{N}\cdot\frac{1}{2n} = \frac{1}{2Nn}$ $\forall (u,v)\in E$. $D(x,y) = \frac{1}{N^2}$.

Intuitively, flow maximizing $\frac{1}{\rho(f)\,\ell(f)}$ spreads across all shortest paths of (x,y).
Using the symmetry of the hypercube, $f(e)=f(e')$ $\forall e,e'\in E$, so
$f(e)=\frac{\sum_{e\in E} f(e)}{|E|}=\frac{\frac{1}{N^2}\sum_{x,y\in V}\mathrm{dist}(x,y)}{Nn}=\frac{n/2}{Nn}=\frac{1}{2N}$ where we use the fact that the average distance $\frac{1}{N^2}\sum_{x,y}\mathrm{dist}(x,y)$ is $n/2$ for hypercubes.
$\Rightarrow \rho(f)=\max_{e\in E} f(e)/C(e)=\frac{1}{2N}\cdot 2Nn=n$. All flows take shortest paths, so $\ell(f)=n$. Thus, $\tau_{mix}=O(\rho(f)\,\ell(f)\ln \pi_{min}^{-1})=\underline{O(n^3)}$.
↳ This is not tight, since we know that $\tau_{mix}\sim\frac{1}{2}n\ln n$. The slack is induced by the heavy-duty approximations used in this method.
* for hyper cubes, α is actually $\sim\frac{1}{n}$, so it is still not tight ($O(n^2)$).

ex2) Random Walk on a Line: MC for lazy walk on line $\{1,2,\dots,N\}$, self-loop w.p. ½ except endpoints w.p. ¾. π is uniform $\rightarrow \pi(x)=\frac{1}{N}$.
$C(e)=\pi(x)P(x,y)=\frac{1}{4N}$ $\forall$ non self-loop edges, $D(x,y)=\frac{1}{N^2}$ $\forall x,y$.
There is a unique path for each pair, $f((i,i+1))=i(N-i)\frac{1}{N^2}\le 1/4$ (max in middle) is unique.
Then the cost is $\rho(f)=\max_e \frac{f(e)}{C(e)}=\frac{1/4}{1/(4N)}=N$, and length is $\ell(f)=N$.
Thus, $\alpha\ge\frac{1}{\rho(f)\,\ell(f)}\ge\frac{1}{N^2}\Rightarrow \tau_{mix}=O(N^2\log N)$, tight to $\log N$ factor.

ex3) Random Walk on $K_{2,N}$: lazy random walk w.p. ½, small side is s.t.



Stationary distribution is $\pi(s)=\pi(t)=1/4$, $\pi(x)=\frac{1}{2N}$ $\forall x\neq s,t$.

$C(e)=\frac{1}{8N}$ $\forall e$, $D(s,t)=\frac{1}{16}$, $D(s,x)=D(t,x)=\frac{1}{8N}$, $D(x,y)=\frac{1}{(4N)^2}$.

If we send flow through one shortest path, then $s\to t$ & $t\to s$ has $1/16$ units, so $\rho(f)\geq\frac{1/16}{1/(8N)}=\frac{N}{2}$, and $l(f)=2\Rightarrow T_{mix}=O(N)$, which is a bad bound since true mixing time is $\Theta(1)$ due to trivial coupling.

If we distribute flow across all shortest paths, then

$$\max_e f(e)\leq\frac{1}{8N}+\frac{1}{16N}+\frac{1}{4N^2}\cdot\frac{N-1}{2}\leq C\times\frac{1}{N} \text{ for constant } C.$$

$\Rightarrow \rho(f)=\frac{C/N}{1/8N}=8C$, so mixing time from s or t is $O(1)$, tight.

Flow Encoding: how to count paths & calculate flows in general MCs

Let $|\Omega|=N$ be the size of ergodic, lazy MC, where N is exponential in n, the natural measure of problem size. Assume π is uniform ($\pi(x)=\frac{1}{N}$), $P(x,y)\geq\frac{1}{poly(n)}$ $\forall$ non-zero transitions, i.e. degree is not too large. Then, $C(u,v)=\pi(u)\cdot P(u,v)\geq\frac{1}{N poly(n)}$. To get a polytime T_{mix}, we need f s.t. $l(f)\leq poly(n)$ and $\frac{f(e)}{C(e)}\leq poly(n)$ $\forall e$. Hence, we need $f(e)\leq\frac{poly(n)}{N}$.

On the other hand, # edges $\leq N\times poly(n)$, and total flow is $\sum_{xy}\frac{1}{N^2}\simeq 1$.

Thus, some edge must carry at least $\frac{1}{N\times poly(n)}$ flow, $f(e)\geq\frac{1}{N\times poly(n)}$.

$\Rightarrow$ any good flow is optimal upto a polynomial factor.

Suppose flow $x \to y$ goes along a single path γ_{xy}. Let paths(e) be the set of paths through e under f. Then $f(e) = |\text{paths}(e)| \times \frac{1}{N^2}$.

Since we want $f(e) \lesssim \text{poly}(n)/N$, we need $\underbrace{|\text{paths}(e)| \lesssim \text{poly}(n) \cdot |\Omega| \ \forall e.}$

Multiple paths from x to y is the same for averaged values, and π being non-uniform is same except everything is weighted accordingly.

$\Rightarrow$ we need to compare $|\text{paths}(e)|$ with $N = |\Omega|$. $\to$ injective mapping!

Def) Encoding: for a flow f, a set of functions $\eta_e : \text{paths}(e) \to \Omega$ s.t.

(i) η_e is injective, (ii) $\forall \beta \lesssim \text{poly}(n)$, $\pi(x)\pi(y) \le \beta \pi(z) \pi(\eta_e(x,y))$ $\forall x, y \in \text{paths}(e)$, $e = (z,z')$

$\hookrightarrow$ (i) says η_e is an injection, and (ii) says η_e is in addition weight-preserving upto poly factor β. If π is uniform, $\beta = 1$ satisfies trivially.

Claim: If $\exists$ encoding for f as above, then $\rho(f) \le \beta \max\limits_{P(z,z')>0} \frac{1}{P(z,z')}$.

Proof: Let $e = (z,z')$ be an arbitrary edge. Then $f(e) = \sum\limits_{(x,y) \in \text{paths}(e)} \pi(x)\pi(y)$ $\le \beta \sum\limits_{(x,y) \in \text{paths}(e)} \pi(z)\pi(\eta_e(x,y)) \le \beta \pi(z)$, using properties (i) and (ii). Finally, $C(e) = \pi(z) P(z,z')$, so $\frac{f(e)}{C(e)} \le \beta \cdot \frac{1}{P(z,z')}$. //

ex) Hypercube revisited: use encodings, not symmetry of hypercube.

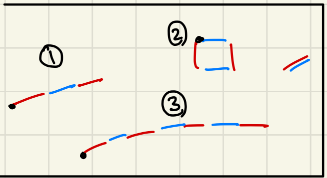
$C(e) = \frac{1}{2nN}$, $D(x,y) = \frac{1}{N^2}$. Consider f that sends $x \to y$ flow through a "left-right bit fixing path" γ_{xy}. Then, $\ell(f) = n$ trivially.

Consider an arbitrary edge $e=(z,z')$ where they differ in bit i.
Consider any pair $(x,y)\in paths(e)$. Observe that y agrees with z in the first $(i-1)$ bits and x agrees with z in the last $(n-i)$ bits.
We define $\eta_e(x,y) := x_1x_2\dots x_i y_{i+1} y_{i+2}\dots y_n$ s.t. $\eta_e(x,y)$ is the string that agrees with x on first i bits and with y on the rest. (missing information!)
$\rightarrow$ We can recover a unique pair (x,y) from $\eta_e(x,y)$ and e, so η_e is an injection. Since π is uniform, it is trivially weight-preserving.
$\Rightarrow \eta_e$ is a valid encoding! So $\rho(f) \leq \max_{z,z'} \frac{1}{P(z,z')} \leq 2n$.
Upto a constant, we recover the same $T_{mix} = O(n^3)$ bound.

Matchings

$G(V,E)$, $\Omega :=$ set of matchings in G. $\pi_\lambda(M) := \frac{\lambda^{|M|}}{Z(\lambda)}$, $Z(\lambda) := \sum_k m_k \lambda^k$
where $m_k :=$ # matchings of size k. Assume $\lambda \geq 1$.
MC: in a matching M, w.p. ½ do nothing. Else, pick $e=(u,v)$ u.a.r.
If u & v are both unmatched, match them ($M \mapsto M+e$). If $e \in M$, $M \mapsto M-e$ w.p. $1/\lambda$. If exactly one of u/v is matched, go to $M+e-e'$ where e' is the edge touching u/v. Else, do nothing.

How to get from $x \to y$, $x, y \in M$? The union $x+y$ is alternating paths. Order the alt. paths with some ordering. "Unwind" each path in order.



If we consider some transition in $x \to y$, we can infer (x,y) from z and some "fill-in" $\eta_e(x,y)$.

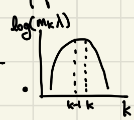
This is injective, and $\pi(x)\pi(y) \le \overbrace{\lambda^2}^{\beta}\pi(z)\pi(\eta_e(x,y))$. Also, $\min P(z,z') = \frac{1}{2|E|\lambda}$. $\to T_{mix} = O(\overbrace{\lambda^2 \cdot |E|\lambda}^{\rho} \cdot \overbrace{|V|}^{\ell} \cdot \log(\pi(x_0)))$. But we can always from a maximal matching x_0, and $\pi(x_0) \ge \frac{1}{\#\text{matchings}} \ge \frac{1}{2^{|E|}}$.

$\Rightarrow \underline{T_{mix} = O(\lambda^3 |V||E|^2)}$. //

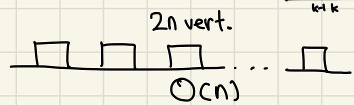
Now, we have a black-box $\xrightarrow{G,\lambda} \boxed{A} \longrightarrow M$ and A runs in $\text{poly}(G,\lambda)$.

1. Approximating $Z(\lambda)$: $Z(\lambda) = \frac{Z(\lambda_r)}{Z(\lambda_{r-1})} \cdots \frac{Z(\lambda_1)}{Z(\lambda_0)} \cdot Z(\lambda_0)$ where $\lambda_0 = \emptyset < \lambda_1 < \cdots < \lambda_r = \lambda$ and $\lambda_i = \lambda_{i-1}(1+\frac{1}{n})$. $\frac{Z(\lambda_i)}{Z(\lambda_{i-1})} = E_{\lambda_{i-1}}\left[\left(\frac{\lambda_i}{\lambda_{i-1}}\right)^{|M|}\right] \le (1+\frac{1}{n})^n \le e$. $\to$ Low variance, good approximation.

2. Approximating coeff.s $\{m_k\}$: Use the fact that $\{m_k\}$ is log-concave, i.e. $m_{k-1}m_{k+1} \le m_k^2 \; \forall k$ [Exercise, use injection?]. Then $\log(m_k \lambda^k)$ is concave, so we can adjust its peak. $m_k = \frac{m_k}{m_{k-1}} \times \cdots \times \frac{m_1}{m_0} \times m_0$, each term can be approximated by finding λ s.t. distribution peaks at m_{k-1} & m_k $\to \lambda = \frac{m_{k-1}}{m_k}$.

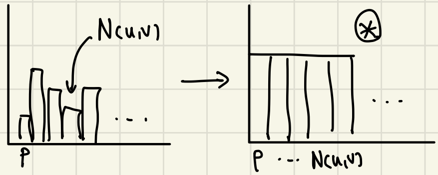


*Problem in graphs where $\frac{m_{n-1}}{m_n} \gg \text{poly}(n)$. e.g.,



where $M_n = 1$ but $m_{n-1} = \exp(cn)$ due to each square being indep!

Idea: consider a modified MC with $\Omega = M_{n-1} \cup M_n$ and uniform π.



"Hole weight" $w(u,v) := \frac{|P|}{N(u,v)}$

Weight of matching $w(M) := \begin{cases} w(u,v) & \text{if } M \in N(u,v) \\ 1 & \text{if } M \in P \end{cases}$

Using Metropolis, we can get the distribution ⊛ as the new stationary.

$\to T_{mix} = O(\mathrm{poly}(n))$, π puts weight at $\Omega(1/n^2)$ on P.

... except for the fact that we don't know $\frac{|P|}{N(u,v)}$ values!

Introduce edge weights $\{\lambda_e\}_{e \in [n] \times [n]}$. Initialize $\lambda_e \leftarrow 1 \ \forall e$.

Reduce λ_e for $e \notin E$ down to $\ll \frac{1}{n!}$.

Each iteration, we reduce $\lambda_e \to \lambda_e/2$ for one $e \notin E$. $\to$ #iters is poly(n).

Then $\lambda(M) = \prod_{e \in M} \lambda_e$ will be negligible $\forall M \in \Omega$ ($\exists e \in M$ s.t. $e \notin E$).

Let $\lambda(P) = \sum_{M \in P} \lambda(M)$, $\lambda(N(u,v)) = \sum_{M \in N(u,v)} \lambda(M)$.

Hole weights $w_\lambda(u,v) := \frac{\lambda(P)}{\lambda(N(u,v))}$. $w(M) = \begin{cases} w_\lambda(u,v) & \text{if } M \in N(u,v) \\ 1 & \text{if } M \in P \end{cases}$.

For current values $\{\lambda_e\}$, now cut $\lambda_e \to \lambda_e/2$.

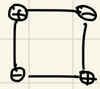
Claim: By observing the output of the MC with new edge weights $\{\lambda_e\}$, we can learn the required hole weights $\overset{w_\lambda(u,v)}{=} \frac{\lambda(P)}{\lambda(N(u,v))}$ w/ arbitrary accuracy.

Proof Sketch: Current MC has π s.t. satisfying $\pi(N(u,v)) = \frac{\lambda(N(u,v))\, w'(u,v)}{Z}$,
↳ can calculate using flows
old hole wts

$\pi(P) = \frac{\lambda(P)}{Z}. \to \omega'(u,v) = \frac{\pi(N(u,v))}{\lambda(N(u,v))} \times \frac{\lambda(P)}{\pi(P)}.$

Recall that $\omega_\lambda(u,v) = \frac{\lambda(P)}{\lambda(N(u,v))} = \omega'(u,v) \times \frac{\pi(P)}{\pi(N(u,v))}$. Induction complete. //

Ferromagnetic Ising Model

 $G(V,E)$

$\sigma = \{\pm 1\}^V$, $\omega(\sigma) = \exp\left(\beta \sum_{(u,v) \in E} \sigma(u)\sigma(v) + \beta h \sum_{u \in V} \sigma(u)\right)$.

$\pi(\sigma) = \omega(\sigma) / Z(\beta,h)$ where $Z(\beta,h) = \sum_\sigma \omega(\sigma)$.

Glauber dynamics: pick $v \in V$ u.a.r., reassign $\sigma(v)$ according to conditional distribution given its neighbors.

Thm) For a $\sqrt{n} \times \sqrt{n}$ box in $\mathbb{Z}^2$, when $h = 0$, T_{mix} of Gl. dyn. is:

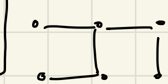
$O(n \log n)$ if $\beta < \beta_c$, $\exp(\Omega(\sqrt{n}))$ if $\beta > \beta_c$, $O(\text{poly}(n))$ if $\beta = \beta_c$.

↳ how do we deal with the supercritical ($\beta > \beta_c$) case?

Idea: Transform Ising spin configs $\Omega \to$ "subgraphs-world" configs Ω'

where $Z(\beta,h) = C_{\beta,h} Z'(\eta, \xi)$ where $\eta = \tanh(\beta)$, $\xi = \tanh(\beta h)$.

Subgraphs world: $\Omega' = \{A \mid A \subseteq E\}$. $\boxed{\omega'(A) = \eta^{|A|} \xi^{|\#\text{odd deg.}(A)|}}$



$Z'(\beta,h) = \sum_A \omega'(A)$, $Z(\beta,h) = \sum_\sigma \omega(\sigma)$.

$\omega'(A) = \eta^6 \xi^6 \rightarrow$ 

Claim: $Z(\beta,h) = C_{\beta,h} Z'(\eta,\xi)$ where $C_{\beta,h} = (\cosh\beta)^{|E|} \cdot (2\cosh(\beta h))^{|V|}$.

Thm) "Glauber dynamics" of subgraphs word has $T_{mix} = O(\text{poly}(n, \xi^{-1}))$.

($\forall$ graphs, $\forall \beta$).

$\to$ if $|h| > \varepsilon > 0$, then $T_{mix} = O(poly(n))\ \forall \beta$. ($Z(\beta,h) \simeq Z(\beta,0)$ for $\beta \ll 1/n$).

Proof of Claim: Use $\exp(x) = \cosh(x)(1+\tanh(x))$.

$$Z(\beta,h) = (\cosh\beta)^{|E|}(\cosh\beta h)^{|V|} \sum_{\sigma} \prod_{(u,v)} (1+\tanh(\beta\sigma(u)\sigma(v))) \prod_{u} (1+\tanh(\beta h\sigma(u)))$$

$$= 2^{-|V|} C_{\beta,h} \sum_{\sigma} \prod_{(u,v)} (1+\sigma(u)\sigma(v)\overbrace{\tanh(\beta)}^{\eta}) \prod_{u} (1+\sigma(u)\overbrace{\tanh(\beta h)}^{\xi})$$

$$= 2^{-|V|} C_{\beta,h} \sum_{\sigma} \Big(\sum_{A\subseteq E} \prod_{(u,v)\in A} \eta\cdot\sigma(u)\cdot\sigma(v)\Big)\Big(\sum_{U\subseteq V} \sum_{u\in U} \xi\cdot\sigma(u)\Big)$$

$$= 2^{-|V|} C_{\beta,h} \sum_{A\subseteq E} \sum_{U\subseteq V} \sum_{\sigma} W(A,U,\sigma) \text{ where } W(A,U,\sigma) := \prod_{(u,v)\in A} \eta\sigma(u)\sigma(v) \prod_{u\in U} \xi\sigma(u).$$

Observe that $\sum_{\sigma} W(A,U,\sigma) = \begin{cases} 0 & \text{if } U \neq odd(A) \\ 2^{|V|}\eta^{|A|}\xi^{|odd(A)|} & \text{if } U = odd(A). \end{cases}$ //

Entropy & Log-Sobolev Constant

Motivation: Stronger T_{mix} bounds than Poincare constant! (log vs log log)

CTMC: P irreducible, heat kernel $H_t(x,y) = \Pr[X_t = y \mid X_0 = x]$, computed as

$$H_t(x,y) := \sum_{k=0}^{\infty} \frac{e^{-t}t^k}{k!} P^k(x,y) = \exp(-tL) \text{ where } L = I-P \text{ is Laplacian.}$$

$*\ \exp(A) = I + A + \frac{1}{2!}A^2 + \cdots$ is a Taylor expansion on matrices.

not needed in cont. time!

Lemma: Let P be irreducible (not necessarily aperiodic) MC with st.dist. π.

$\tilde{P} = \frac{1}{2}(I+P)$ is the lazy version of P. $\forall$ initial state x,

(i) for suff. large k, if $\|\tilde{P}_x^{(k)} - \pi\|_{TV} \leq \varepsilon$, then $\|h_x^{(k)} - \pi\|_{TV} \leq 2\varepsilon$.

(ii) for suff. large t, if $\|h_x^{(t)} - \pi\|_{TV} \leq \varepsilon$, then $\|\tilde{P}_x^{(\lceil 4t \rceil)} - \pi\|_{TV} \leq 2\varepsilon$.

Proof Sketch: $\|\tilde{P}_x^{(k)} - \pi\|_{TV} \leq \varepsilon \to \tilde{H}$ after $2k$ steps, $\Pr$[fewer than k transitions happened] $= \Pr[\mathrm{Po}(2k) < k] \to \emptyset$ as $k \to \infty$. $\tilde{H}_{2t} = \tilde{H}$.

$\|h_x^{(t)} - \pi\|_{TV} \leq \varepsilon \to$ same holds after extra t P transitions. # of transitions in P is $\mathrm{Po}(t) + t$. $\to 4t$ transitions in $\tilde{P} \to$ # trans. in P is $\mathrm{Bin}(4t, \frac{1}{2})$.

$\|(\mathrm{Po}(t) + t) - \mathrm{Bin}(4t, \frac{1}{2})\|_{TV} \to \emptyset$ as $t \to \infty$.

Previously, $\mathrm{Var}_\pi[P\varphi] - \mathrm{Var}_\pi[\varphi] \leq -\mathcal{E}_p(\varphi, \varphi)$. (difference form)

Thm) For an irreducible P & heat kernel H_t, $\forall \varphi : \Omega \to \mathbb{R}$,

$\frac{d}{dt} \mathrm{Var}_\pi[H_t\varphi] = -2\mathcal{E}_p(H_t\varphi, H_t\varphi)$. (derivative form)

Proof: Let $\varphi_t := H_t\varphi$. Since $H_t = \exp(-tL)$, $\frac{d}{dt}\varphi_t = -L\varphi_t$.

$\frac{d}{dt}\mathrm{Var}_\pi[\varphi_t] = \frac{d}{dt}\left[E_\pi[\varphi_t]^2 - \cancelto{0}{E_\pi[\varphi_t^2]}\right] = \frac{d}{dt}\left[\sum_x \pi(x) \cdot \varphi_t(x)^2\right]$

$= 2\sum_x \pi(x)\varphi_t(x)\frac{d}{dt}\varphi_t(x) = -2\sum_x \pi(x)\varphi_t(x)[L\varphi_t(x)] = -2\mathcal{E}_p(\varphi_t, \varphi_t)$. //

Corollary: $\forall \varphi$, $\frac{d}{dt}\mathrm{Var}_\pi[H_t\varphi] \leq -2\alpha \mathrm{Var}_\pi[H_t\varphi]$ where $\alpha := \inf_{\text{non-const}} \frac{\mathcal{E}_p(\varphi_t, \varphi_t)}{\mathrm{Var}_\pi[\varphi_t]}$.

Corollary: $T_x(\varepsilon) \leq \frac{1}{2\alpha}(2\ln\varepsilon^{-1} + \ln(4\pi(x))^{-1}) = O(\frac{1}{\alpha}\ln\pi_{min}^{-1})$.

Def) Entropy: $\mathrm{Ent}_\pi[\varphi] := \sum_x \pi(x)\varphi(x)\log\varphi(x) - E_\pi[\varphi]\log E_\pi[\varphi]$

$= \sum_x \pi(x)\varphi(x)\log\frac{\varphi(x)}{E_\pi[\varphi]}$.

Set $\varphi = \frac{\mu}{\pi}$ where μ is a prob. dist. over Ω. $\to E_\pi[\varphi] = 1$, and

$\mathrm{Ent}_\pi[\varphi] = \sum_x \mu(x) \log \frac{\mu(x)}{\pi(x)} = D_{KL}(\mu \| \pi)$, the KL-Divergence.

Goal: bound rate of decay of $D(p_x^{(t)} \| \pi) = \mathrm{Ent}_\pi\left[\frac{p_x^{(t)}}{\pi}\right]$.

Def) (Modified) Log-Sobolev Constant; $\rho := \inf_{\substack{\varphi \geq 0 \\ \mathrm{Ent}_\pi[\varphi] \neq 0}} \frac{\mathcal{E}_P(\varphi, \log \varphi)}{\mathrm{Ent}_\pi[\varphi]}$.

Thm°) $T_x(\varepsilon) \leq \frac{1}{\rho}(2 \ln \varepsilon^{-1} + \ln \ln \pi_x^{-1}) \Rightarrow T_{mix} = O\left(\frac{1}{\rho} \log \log \pi_{min}^{-1}\right)$.

Thm*) $\forall \varphi : \Omega \to \mathbb{R}^+$, $\frac{d}{dt} \mathrm{Ent}_\pi\left[\frac{\varphi_t}{\pi}\right] = -\mathcal{E}_P(\varphi_t, \log \varphi_t)$ where $\varphi_t := \frac{\varphi H_t}{\pi}$.

Proof*: $\varphi_t = \frac{\varphi H_t}{\pi} = H_t^*\left(\frac{\varphi}{\pi}\right)$ where H_t^* is the heat kernel of reversalization of P,

$p^*(x,y) = \frac{\pi(y) P(x,y)}{\pi(x)}$ [Exercise]. $\to \frac{d}{dt} \varphi_t = \frac{d}{dt} H_t^*\left(\frac{\varphi}{\pi}\right) = -L^* \varphi_t$.

$\frac{d}{dt} \mathrm{Ent}_\pi[\varphi_t] = \frac{d}{dt}\left[E_\pi[\varphi_t \log \varphi_t] - E_\pi[\varphi_t] \log E_\pi[\varphi_t]\right]$ ($\sum_x \pi(x) \frac{d}{dt} \varphi_t(x) = 0$ due to conserv. of mass)

$= \frac{d}{dt} \sum_x (\pi(x) \varphi_t(x) \log \varphi_t(x)) = \sum_x \pi(x)(1 + \log \varphi_t(x)) \frac{d}{dt} \varphi_t(x)$

[Exercise]

$= -\sum_x \pi(x) \log \varphi_t(x) [L^* \varphi_t](x)$. L^* is adjoint on $\langle \cdot, \cdot \rangle_\pi$, $\langle \eta, L^* \psi \rangle_\pi = \langle \psi, L\eta \rangle_\pi$.

$\to -\sum_x \pi(x) \varphi_t(x) [L \log \varphi_t](x) = -\mathcal{E}_P(\varphi_t, \log \varphi_t)$. //

Corollary: $\forall \varphi : \Omega \to \mathbb{R}^+$, $\mathrm{Ent}_\pi[\varphi] \neq 0$, $\mathrm{Ent}_\pi\left[\frac{\varphi H_t}{\pi}\right] \leq \exp(-\rho t) \mathrm{Ent}_\pi\left[\frac{\varphi}{\pi}\right]$.

Proof°: Set $\varphi = P_x^{(0)}$ (initial dist. concentrated at x).

Recall $\mathrm{Ent}_\pi\left[\frac{P_x^{(t)}}{\pi}\right] = D_{KL}(p_x^{(t)} \| \pi)$. $\to D_{KL}(p_x^{(t)} \| \pi) \leq \exp(-\rho t) D_{KL}(p_x^{(0)} \| \pi)$

$= \exp(-\rho t) \log \pi(x)^{-1}$. Pinsker's Inequality: $2\|p - \pi\|_{TV}^2 \leq D_{KL}(p \| \pi)$.

$\Rightarrow \|p_x^{(t)} - \pi\|_{TV} \leq \varepsilon$ $\forall t \geq \frac{1}{\rho}\left(\ln \frac{1}{2\varepsilon^2} + \ln \ln \pi(x)^{-1}\right) \Rightarrow \underline{T_{mix} = O\left(\frac{1}{\rho} \ln \ln \pi_{min}^{-1}\right)}$

Fact[Exercise]: if P is reversible, $\mathcal{E}_P(\varphi,\psi)=\frac{1}{2}\sum_{xy}\pi(x)P(x,y)(\varphi(x)-\varphi(y))(\psi(x)-\psi(y))$

ex) Random Walk on Hypercube: $\Omega=\{0,1\}^n$. Decompose $\Omega=\Omega_1\times\cdots\times\Omega_n$. $\leftarrow\{0,1\}$

Base Case: $\overset{0}{\bullet}$—$\overset{1}{\bullet}$ $P=\begin{bmatrix}0&1\\1&0\end{bmatrix}$. Let φ be an arbitrary non-negative function s.t. $\alpha:=\varphi(0)$, $\beta:=\varphi(1)$. $\underline{\mathcal{E}_P(\varphi,\log\varphi)}=\frac{1}{2}\sum_{xy}\pi(x)P(x,y)(\varphi(x)-\varphi(y))\log\left(\frac{\varphi(x)}{\varphi(y)}\right)$

$=\frac{1}{2}(a-b)\log\frac{a}{b}$. $\underline{\mathrm{Ent}_\pi[\varphi]}=E_\pi[\varphi\log\varphi]-E_\pi[\varphi]\log E_\pi[\varphi]$

$=\frac{a\log a+b\log b}{2}-\left(\frac{a+b}{2}\right)\log\left(\frac{a+b}{2}\right)$. $\Rightarrow$ Find $\min_C\left\{\max_{a,b\geq 0}\frac{a\log a+b\log b-(a+b)\log\left(\frac{a+b}{2}\right)}{(a-b)\log\left(\frac{a}{b}\right)}\leq C\right\}$.

With some Hessian argument [omitted, in notes], $C=1/4\Rightarrow\underline{\rho\geq 4}$. //

If we have some operator P_i on Ω_i, $P=\sum_{i=1}^n q_i(I\otimes I\otimes\cdots P_i\otimes\cdots I)$ where $\sum_i^n q_i=1$.

Claim*: Let $(\Omega_1,P_1),(\Omega_2,P_2)$ be two reversible MCs on disjoint state spaces Ω_1 & Ω_2, stat. dist. π_1,π_2. Then the combined MC $q_1(P_1\otimes I)+q_2(I\otimes P_2)$ is reversible w.r.t. $(\pi_1\otimes\pi_2)$ [ex] and $\rho\geq\min\{\rho_1q_1,\rho_2q_2\}$ where ρ_1,ρ_2 are log-Sobolev const. for P_1,P_2.

Corollary: for $\{0,1\}^n$, $n=1\to\rho\geq 4$. $n>1\to\rho\geq 4/n$ by induction.

Proof: $\Omega=\Omega_1\times\{0,1\}^{n-1}$. $P=\frac{1}{n}(P_1\otimes I)+(1-\frac{1}{n})(I\otimes P_{n-1})$. By IH, $\rho_1\geq 4$, $\rho_{n-1}\geq\frac{4}{n-1}\to\rho\geq\min\left\{\frac{4}{n},\frac{n-1}{n}\cdot\frac{4}{n-1}\right\}=\frac{4}{n}$. // $\Rightarrow T_{mix}=O(n\ln n)$!!!
(tight for hypercubes)

To prove the Claim*, we first introduce two Lemmas.

Lemma 1: $\mathcal{E}_P(\varphi,\psi) = q_1 \sum_{x_2} \pi_2(x_2)\, \mathcal{E}_{P_1}(\varphi(\cdot,x_2),\psi(\cdot,x_2)) + q_2 \sum_{x_1} \pi_1(x_1) \cdot$
$\mathcal{E}_{P_2}(\varphi(x_1,\cdot),\psi(x_1,\cdot))$.

Lemma 2: $\mathrm{Ent}_\pi[\varphi] \le \sum_{x_1} \pi_1(x_1)\, \mathrm{Ent}_{\pi_2}[\varphi(x_1,\cdot)] + \sum_{x_2} \pi_2(x_2)\, \mathrm{Ent}_{\pi_1}[\varphi(\cdot,x_2)]$.

Proof of Claim*: $\mathcal{E}_P(\varphi, \log\varphi) = q_1 \sum_{x_2} \pi_2(x_2)\, \mathcal{E}_{P_1}(\sim) + q_2 \sum_{x_1} \pi_1(x_1)\, \mathcal{E}_{P_2}(\sim)$

$\ge q_1 \rho_1 \sum_{x_2} \pi_2(x_2)\, \mathrm{Ent}_{\pi_1}[\varphi(\cdot,x_2)] + q_2 \rho_2 \sum_{x_1} \pi_1(x_1)\, \mathrm{Ent}_{\pi_2}[\varphi(x_1,\cdot)]$

$\ge \min\{q_1\rho_1, q_2\rho_2\} \cdot \left[\sum_{x_2} \pi_2(x_2)\, \mathrm{Ent}_{\pi_1}[\varphi(\cdot,x_2)] + \sum_{x_1} \pi_1(x_1)\, \mathrm{Ent}_{\pi_2}[\varphi(x_1,\cdot)]\right]$

$\ge \min\{q_1\rho_1, q_2\rho_2\} \cdot \mathrm{Ent}_\pi[\varphi]$. //

$x := (x_1,x_2),\ y := (y_1,y_2)$

Proof of Lemma 1: $\mathcal{E}_P(\varphi,\psi) = \frac{1}{2} \sum_{xy} \pi(x) P(x,y)(\varphi(x)-\varphi(y))(\psi(x)-\psi(y))$

$= \frac{q_1}{2} \sum_{x_1,x_2,y} \pi_1(x_1)\pi_2(x_2) P_1(x_1,y_1)(\varphi(x_1,x_2)-\varphi(y_1,x_2))(\psi(x_1,x_2)-\psi(y_1,x_2))$

$+ \frac{q_2}{2} \sum_{x,y_1,y_2} \pi_1(x_1)\pi_2(x_2) P_2(x_2,y_2)(\qquad\qquad)(\qquad\qquad)$

$= q_1 \sum_{x_2} \pi_2(x_2)\, \mathcal{E}_{P_1}(\varphi(\cdot,x_2),\psi(\cdot,x_2)) + q_2 \sum_{x_1} \pi_1(x_1)\, \mathcal{E}_{P_2}(\varphi(x_1,\cdot),\psi(x_1,\cdot))$. //

Proof of Lemma 2: $\mathrm{Ent}_\pi[\varphi] = \sum_x \pi(x)\varphi(x) \log\left(\frac{\varphi(x)}{E_\pi[\varphi]}\right)$

$= \sum_{x_1,x_2} \pi_1(x_1)\pi_2(x_2)\varphi(x_1,x_2) \log\left(\frac{\varphi(x_1,x_2)}{\sum_{x_1,x_2}\pi_1(x_1)\pi_2(x_2)\varphi(x_1,x_2)}\right)$

$= \underbrace{\sum_{x_1,x_2} \overbrace{\pi_1(x_1)\pi_2(x_2)\varphi(x_1,x_2)} \log\left(\frac{\varphi(x_1,x_2)}{\sum_{x_2}\pi_2(x_2)\varphi(x_1,x_2)}\right)} + \underbrace{\sum_{x_1,x_2} \overbrace{\qquad\qquad} \log\left(\frac{\sum_{x_2}\pi_2(x_2)\varphi(x_1,x_2)}{\sum_{x_1,x_2}\pi_1(x_1)\pi_2(x_2)\varphi(x_1,x_2)}\right)}$

$\hookrightarrow \sum_{x_1} \pi_1(x_1)\, \mathrm{Ent}_{\pi_2}[\varphi(x_1,\cdot)]$ / use log-Sobolev inequality: $\sum a_i \log\frac{\sum a_i}{\sum b_i} \le \sum a_i \log\left(\frac{a_i}{b_i}\right)\ \forall\, a \ge 0,\ b > 0$

$\to\ \le \sum_{x_1} \pi_1(x_1) \sum_{x_2} \pi_2(x_2)\varphi(x_1,x_2) \log\left(\frac{\cancel{\pi_2(x_2)}\varphi(x_1,x_2)}{\cancel{\pi_2(x_2)}\sum_{x_1}\pi_1(x_1)\varphi(x_1,x_2)}\right) = \sum_{x_2} \pi_2(x_2) \sum_{x_1} \pi_1(x_1)\varphi(x_1,x_2) \log(\sim)$

$= \sum_{x_2} \pi_2(x_2)\, \mathrm{Ent}_{\pi_1}[\varphi(\cdot, x_2)]$. //

Matroid Bases

Def) Matroid: $\mathcal{M}(\underset{\text{Ground set}}{E}, \underset{\text{Indep. set}}{\mathcal{I}})$, $\mathcal{I} \subseteq 2^E$ where (i) $I \in \mathcal{I}$ and $J \subseteq I \Rightarrow J \in \mathcal{I}$. → downward closure

(ii) $I, J \in \mathcal{I}$, $|I| > |J| \Rightarrow \exists e \in I \setminus J$ s.t. $J + e \in \mathcal{I}$. → augmentation property

Def) Basis: maximal indep. set. of a matroid $\mathcal{M}$.

Fact: all bases have same cardinality → call it "rank" of $\mathcal{M}$.

ex) Graphical Matroids: Connected $G(V,E)$, $E = E$. $\mathcal{I}$ = sets of forests.

→ Bases are spanning trees.

ex) Uniform Matroids: E, a finite set. $\mathcal{I}$ = subsets of E of size $\leq k$.

→ Bases are k-subsets of E.

ex) Linear/Representabe Matroids: $n \times m$ matrix over $\mathbb{F}$, $m \geq n$. E = columns.

$\mathcal{I}$ = linearly independent subsets of E. → bases are maximally indep. columns.

* We assume a membership oracle; given $J \subseteq E$, tells if $J \in \mathcal{I}$.

Algorithm; use a suitable MCMC to sample bases to $\mathcal{B}$, set of bases.

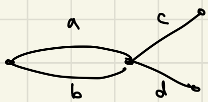
Natural algorithm: <u>Bases Exchange</u> MCMC. With current basis B, pick $e \in B$ u.a.r and pick f u.a.r. from $F=\{f \in E \mid B\setminus\{e\}\cup\{f\}=\mathcal{I}\}$.

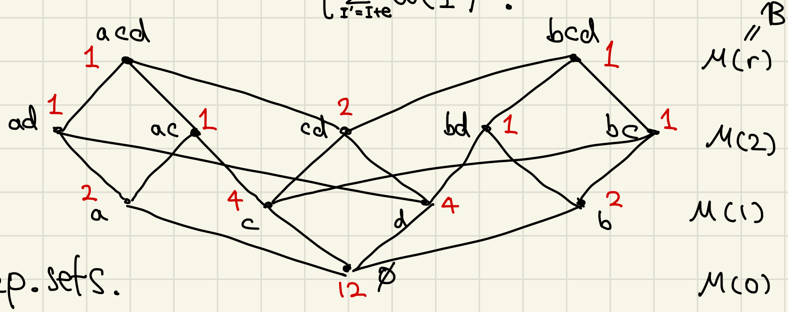
Irreducibility? $\forall B \to B'$, $\{e_1,\dots,e_{2k}\}=B\oplus B'$. By augmentation property, if we throw out e_1 from B, $\exists\, e_j$ s.t. $B\setminus e_1+e_j \in \mathcal{I}$. induct.//

Thm) T_{mix} of Basis Exchange MC is $O(n\log n)$ (optimal!) where $n=|E|$ (can be improved to $r\log r$)

Def) Weight Function: $w:\mathcal{I}\to\mathbb{Z}^+$. $w(I)=\begin{cases} 1 & \text{if } I\in\mathcal{I} \\ \sum_{I'=I+e} w(I') \end{cases}$.

ex)





Obs) $w(\emptyset)=|\mathcal{B}|\times(r!)$

$M(k)$:= k-th level indep. sets.

π_k on M_k: $\pi_k(I) := \dfrac{w(I)}{Z_k}$ where $Z_k=\sum_{I\in M_k} w(I)$.

Def) Contraction: $\forall I \in \mathcal{J}$, $M_I=(E\setminus I, \mathcal{I}_I)$ $\{J\subseteq E\setminus I \mid J\cup I=J\}$.

Write $\pi_{I,k}$ for conditional distribution on M_I, $\pi_{I,k}(J) := \begin{cases} \frac{k!\,w(J)}{w(I)} & \text{if } J\supseteq I \\ \emptyset & \text{o.w.} \end{cases}$

A Down-Up Walk $P_k^{\vee}$ on $M(k)$: remove $e\in I$ uar, let $S=\{J\in M(k) \mid J=I-e+e' \text{ for some } e'\}$. Move to $J\in S$ w.p. $\propto w(J)$.

Modified Thm) $\forall\, 2\le k\le r$, ML-S const. of $P_k^{\vee}$ satisfies $\rho_k \ge \frac{1}{k}$

$\Rightarrow T_{mix} = O\left(\frac{1}{\rho_r}\times \log\log \pi_{min}^{-1}\right) = O(r(\log r+\log\log n)) = O(r\log n)$!   ($|\mathcal{B}|\le\binom{n}{r}\le n^r$)

Let $A_{k-1}(I,J) := \mathbb{1}\{\exists e \text{ s.t. } J=I+e\}$. $\begin{array}{c} \\ M(k-1) \end{array}\begin{pmatrix} \cdots & \vdots \\ & 1 \end{pmatrix}$ $M(k)$ Also let

$P^{\downarrow}_{k-1} := \frac{1}{k} A^{T}_{k-1}$, $P^{\uparrow}_{k-1} := A_{k-1}$ with (I,J) replaced by $\frac{\omega(J)}{\omega(I)}$. Then $P^{\vee}_{k} = P^{\downarrow}_{k} P^{\uparrow}_{k-1}$.

Claim: [Exercise] $\forall I,J$, $\pi_k(I) P^{\downarrow}_{k}(I,J) = \pi_{k-1}(J) P^{\uparrow}_{k-1}(J,I)$.

$\forall f^{(k)} : \mathcal{M}(k) \to \mathbb{R}^{\geq 0}$ and any $i<k$, define $f^{(i)} := P^{\uparrow}_{i} P^{\uparrow}_{i+1} \cdots P^{\uparrow}_{k-1} f^{(k)}$.

Fact: Always assume wlog functions have mean 1, i.e. $E_{\pi_k}[f^{(k)}] = 1$.

Then this implies that $E_{\pi_i}[f^{(i)}] = 1$ $\forall i<k$ as well.

Main Lemma: $\forall k \geq 2$, $\forall f^{(k)} : \mathcal{M}_k \to \mathbb{R}^{\geq 0}$ with $E_{\pi_k}[f^{(k)}] = 1$, we have

$\mathrm{Ent}_{\pi_{k-1}}[f^{(k-1)}] \leq (1-\frac{1}{k}) \mathrm{Ent}_{\pi_k}[f^{(k)}]$. (Entropy Contraction)

Lemma $\to$ Thm: s.t.p. $\rho_k = \inf_{f^{(k)} \sim} \frac{\mathcal{E}_{P^{\vee}_{k}}(f^{(k)}, \log f^{(k)})}{\mathrm{Ent}_{\pi_k}[f^{(k)}]}$.

Claim: $\mathrm{Ent}_{\pi_{k-1}}[f^{(k-1)}] \geq \mathrm{Ent}_{\pi_k}[f^{(k)}] - \mathcal{E}_{P^{\vee}_{k}}(f^{(k)}, \log f^{(k)})$. If Claim holds,

$$\mathcal{E}_{P^{\vee}_{k}}(f^{(k)}, \log f^{(k)}) \geq \mathrm{Ent}_{\pi_k}[f^{(k)}] - \mathrm{Ent}_{\pi_{k-1}}[f^{(k-1)}]$$
$$\geq \frac{1}{k} \mathrm{Ent}_{\pi_k}[f^{(k)}] \text{ by the Main Lemma} \to \rho_k \geq \frac{1}{k}. //$$

Proof of Claim: $\mathrm{Ent}_{\pi_{k-1}}[f^{(k-1)}] = \sum_{I \in \mathcal{M}(k-1)} \pi_{k-1}(I) f^{(k-1)}(I) \log(f^{(k-1)}(I))$

$$= \sum_{I \in \mathcal{M}(k-1)} \pi_{k-1}(I) [P^{\uparrow}_{k-1} f^{(k)}](I) \log([P^{\uparrow}_{k-1} f^{(k)}](I))$$

$$= \sum_{I \in \mathcal{M}(k-1), J \in \mathcal{M}(k)} \pi_{k-1}(I) P^{\uparrow}_{k-1}(I,J) f^{(k)}(J) \log\Big(\sum_{L \in \mathcal{M}(k)} P^{\uparrow}_{k-1}(I,L) f^{(k)}(L)\Big)$$

$$= \sum_{I,J \in \mathcal{M}(k-1), \mathcal{M}(k)} \pi_k(J) f^{(k)}(J) P^{\vee}_{k}(J,I) \log(\ \text{''}\ ) \text{ [by "detailed balance"]}$$

$$\geq \sum_{\substack{I \in \mathcal{M}(k-1) \\ J,L \in \mathcal{M}(k)}} \pi_k(J) f^{(k)}(J) P^{\downarrow}_{k}(J,I) P^{\uparrow}_{k-1}(I,L) \log(f^{(k)}(L)) \text{ [by Jensen's Inequality]}$$

$$= \sum_{J,L \in \mathcal{M}(k)} \pi_k(J) f^{(k)}(J) P^{\vee}_{k}(J,L) \log(f^{(k)}(L))$$

$= \mathrm{Ent}_{\pi_k}[f^{(k)}] - \mathcal{E}_{P_k^\vee}(f^{(k)}, \log f^{(k)})$. //

Proof of Main Lemma: Induction step for $k>2$.

Assume $\mathrm{Ent}_{\pi_{k-1}}[f^{(k-1)}] \le (1-\frac{1}{k})\,\mathrm{Ent}_{\pi_k}[f^{(k)}]$.

To prove: $\mathrm{Ent}_{\pi_k}[f^{(k)}] \le (1-\frac{1}{k+1})\,\mathrm{Ent}_{\pi_{k+1}}[f^{(k+1)}]$.

Decompose $\pi_k = \sum_e \pi_1(e)\,\pi_{e,k-1}$.

By Standrd Procedure [Ex] $\mathrm{Ent}_{\pi_k}[f^{(k)}] = \mathrm{Ent}_{\pi_1}[f^{(1)}] + \sum_{e\in E} \pi_1(e)\,\mathrm{Ent}_{\pi_{e,k-1}}[f^{(k)}]$.

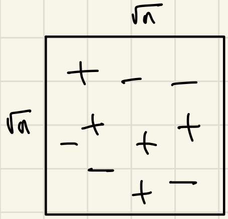
By IH applied to all of $f^{(k)}, f^{(k-1)}, \dots, f^{(2)}$, $\mathrm{Ent}_{\pi_k}[f^{(k)}] \ge k\,\mathrm{Ent}_{\pi_1}[f^{(1)}]$.

$\rightarrow \sum_{e\in E} \pi_1(e)\,\mathrm{Ent}_{\pi_{e,k-1}}[f^{(k)}] \ge (k-1)\,\mathrm{Ent}_{\pi_1}[f^{(1)}]$.

$$
\begin{aligned}
\mathrm{Ent}_{\pi_{k+1}}[f^{(k+1)}] &= \mathrm{Ent}_{\pi_1}[f^{(1)}] + \sum_e \pi_1(e)\,\mathrm{Ent}_{\pi_{e,k}}[f^{(k+1)}] \\
&\ge \mathrm{Ent}_{\pi_1}[f^{(1)}] + \sum_e \pi_1(e)\,\frac{k}{k-1}\,\mathrm{Ent}_{\pi_{e,k-1}}[f^{(k)}] \quad [\text{By IH on } M_e] \\
&= \mathrm{Ent}_{\pi_1}[f^{(1)}] + \frac{1}{k(k-1)}\sum_e\left(\pi_1(e)\,\mathrm{Ent}_{\pi_{e,k-1}}[f^{(k)}]\right) + \frac{k+1}{k}\sum_e(\,{}''\,) \\
&\ge \mathrm{Ent}_{\pi_1}[f^{(1)}] + \frac{k-1}{k(k-1)}\,\mathrm{Ent}_{\pi_1}[f^{(1)}] + \frac{k+1}{k}\sum_e \pi_1(e)\,\mathrm{Ent}_{\pi_{e,k-1}}[f^{(k)}] \\
&= \frac{k+1}{k}\left(\mathrm{Ent}_{\pi_1}[f^{(1)}] + \sum_e \pi_1(e)\,\mathrm{Ent}_{\pi_{e,k-1}}[f^{(k)}]\right) = \frac{k+1}{k}\,\mathrm{Ent}_{\pi_k}[f^{(k)}]. \;//
\end{aligned}
$$

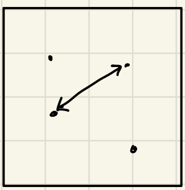
*Base Case of $k=2$ is deferred to Lecture Notes. It uses a completely orthogonal set of analyses from the induction. This content may be coved later if time permits.

Correlation Decay

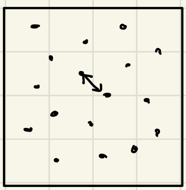


$\emptyset$ field, λ interaction. For some λ_e, $\lambda < \lambda_e \Rightarrow$ no long-range correlations as $n \to \infty$ & GD has $T_{mix} = O(n \log n)$.
OTOH, $\lambda > \lambda_e \Rightarrow$ long-range correlations exist as $n \to \infty$ and GD has $T_{mix} = \exp(\Omega(\sqrt{n}))$.

Hard-Core Model: $G(V,E)$, Ω = indep. sets in G, $\pi(I) = \frac{\lambda^{|I|}}{Z(\lambda)}$, $Z(\lambda) = \sum_k \underset{\substack{\uparrow \\ \#\text{IS of size } k}}{a_k} \lambda^k$

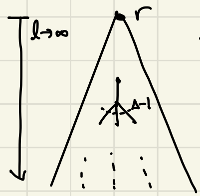


(λ small)



(λ large)

Tree view: infinite Δ-regular tree $T(\Delta)$



Fact: $\forall \Delta, \exists \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$ s.t. "phase transition" occurs at $\lambda = \lambda_c(\Delta)$.

Thm) For graphs of max degree $\Delta \geq 3$:

(i) [Weitz] If $\lambda < \lambda_c(\Delta)$ then $\exists$ FPTAS for $Z(\lambda)$ — deterministic!

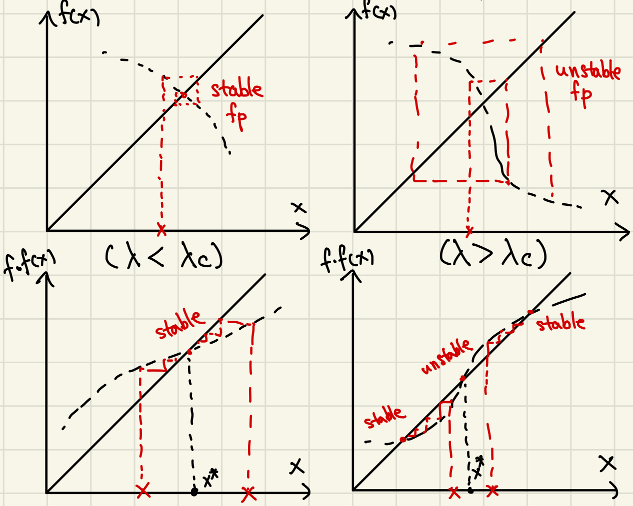
(ii) [Sly] If $\lambda > \lambda_c(\Delta)$ then $\nexists$ FPRAS for $Z(\lambda)$ (unless RP = NP)

Def) Weak Spatial Mixing (WSM): Let S be any subset of vertices in $T(\Delta)$ and τ be some arbitrary boundary configuration of S. Let P_r^{τ} be the probability that the root r is occupied. WSM holds if $|P_r^{\tau} - P_r^{\tau'}| \leq \exp(-c \cdot \text{dist}(r,S))$ for some $c > 0$ $\forall \tau, \tau'$ configs of S.

Proof of Fact: Let $p_v := \Pr_{T_v}[\sigma(v)=0]$, prob. that root v is unoccupied.

$$p_v = \frac{Z_{T_v}[\sigma(v)=0]}{\underbrace{Z_{T_v}[\sigma(v)=0]+Z_{T_v}[\sigma(v)=1]}_{Z_{T_v}}} = \frac{\prod_{i=1}^{d} Z_{T_{v_i}}}{\prod_{i=1}^{d} Z_{T_{v_i}} + \lambda\cdot\prod_{i=1}^{d} Z_{T_{v_i}}[\sigma(v_i)=0]} = \frac{1}{1+\lambda\cdot\prod_{i=1}^{d} p_{v_i}}.$$

Translation invariance $\to p_v = p \ \forall v. \Rightarrow p = \frac{1}{1+\lambda p^d}$, p is f.p. of $f(x)=\frac{1}{1+\lambda x^d}$.



stability is $f'(x^*) < 1$.

$(f\circ f)'(x) = f'(f(x))f'(x)$.

At x^*, $(f\circ f)'(x) = f'(x^*)^2$.

So $(f\circ f)'(x^*) \lessgtr 1 \Leftrightarrow f'(x^*) \gtrless -1$.

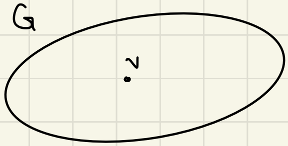
At $x=x^*$, $x^* = f(x^*) = \frac{1}{1+\lambda x^{*d}}$. (a)

$f'(x^*) = -1 = -\lambda\frac{x^{*(d-1)}}{(1+\lambda x^{*d})^2}$. (b)

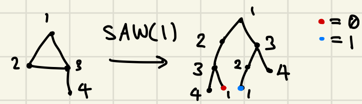
(a) $\to$ (b) shows $d(x^*-1)=1 \Rightarrow x^* = \frac{d-1}{d}$. plug into (a) $\to \lambda = \frac{d^d}{(d-1)^{(d+1)}}$. //

Sketch of (i) of Theorem: 1. SAW construction 2. Spatial mixing for truncation

Consider graph G with max degree Δ. Compute $\Pr[v \text{ occupied}]$? $\lambda < \lambda_c(\Delta)$.



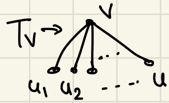
- Construct a Self-Avoiding Walk Tree $T_{SAW}(G,v)$. → can't go back immediately
- Use corr. decay on $T_{SAW}(G,v)$ to truncate the tree at level $O(\log n)$ and do explicit reccurence.



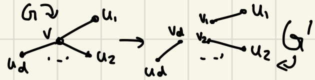
Claim 1: $\Pr_G[\sigma(v)=0] = \Pr_{T_{SAW}(G,v)}[\sigma(v)=0]$.

Proof: Let $R_v := \frac{\Pr[\sigma(v)=1]}{\Pr[\sigma(v)=0]}$. s.t.p. that $\underline{R_v^G = R_v^{T_v}}$.

$$R_v^{T_v} = \frac{Z_{T_v}[\sigma(v)=1]}{Z_{T_v}[\sigma(v)=0]} = \frac{\lambda \cdot \prod_{i=1}^{d} Z_{T_{u_i}}[\sigma(u_i)=0]}{\prod_{i=1}^{d}\left(Z_{T_{u_i}}[\sigma(u_i)=0]+Z_{T_{u_i}}[\sigma(u_i)=1]\right)} = \lambda \prod_{i=1}^{d} \frac{1}{1+R_{u_i}^{T_{u_i}}}.$$



$$R_v^{G} = \frac{\Pr_G[\sigma(v)=1]}{\Pr_G[\sigma(v)=0]} = \lambda^{-(d-1)} \frac{\Pr_{G'}[\sigma(v_1)=\cdots=\sigma(v_d)=1]}{\Pr_{G'}[\sigma(v_1)=\cdots=\sigma(v_d)=0]}$$



$$= \lambda^{-(d-1)} \frac{\Pr_{G'}[\sigma(v_1)=1 \mid \sigma(v_2)=\cdots=\sigma(v_d)=0]}{\Pr_{G'}[\sigma(v_1)=0 \mid \sigma(v_2)=\cdots=\sigma(v_d)=0]} \cdot \frac{\Pr[\sigma(v_2)=1 \mid \sigma(v_1)=1, \sigma(v_3)=\cdots=\sigma(v_d)=0]}{\Pr[\sigma(v_2)=0 \mid \cdots]} \cdots$$

(Using $\frac{\Pr[AB]}{\Pr[CD]} = \frac{\Pr[A|D]}{\Pr[C|D]} \cdot \frac{\Pr[B|A]}{\Pr[D|A]}$)

$$= \lambda^{-(d-1)} \prod_{i=1}^{d} R_{v_i}^{G',\tau_i} \text{ where } \tau_i(v_j) = \mathbb{1}\{j<i\}.$$

$$\to R_{v_i}^{G',\tau_i} = \frac{\lambda Z_{G'-v_i}^{\tau_i}[\sigma(u_i)=0]}{Z_{G'-v_i}^{\tau_i}[\sigma(u_i)=0]+Z_{G'-v_i}^{\tau_i}[\sigma(u_i)=1]} = \frac{\lambda}{1+R_{u_i}^{G'-v_i,\tau_i}}.$$

$$\to R_v^{G} = \lambda^{-(d-1)} \prod_{i=1}^{d} \frac{\lambda}{1+R_{u_i}^{G'-v_i,\tau_i}} = \lambda \prod_{i=1}^{d} \frac{1}{1+R_{u_i}^{G'-v_i,\tau_i}}.$$

$R_{u_i}^{G'-v_i,\tau_i} = R_{u_i}^{T_{u_i}}$ by induction on subtree. //

Def) Strong Spatial Mixing: $|P_r^{\tau} - P_r^{\tau'}| \le \exp(-c \cdot \mathrm{dist}(r, \hat{S}))$ where $\tau = \tau'$ for $S \setminus \hat{S}$.

Thm2) For hard-core model on T_Δ, WSM $\Rightarrow$ SSM.

Proof: Fix boundary conditions accordingly. Let $p_v = \Pr[\sigma(v)=0]$.

As before, $p_v = \frac{1}{1+\lambda \prod_{i=1}^{d} p_{v_i}} = f(p_{v_1}, \dots, p_{v_n})$.

Need to show that $|p_v - p_v'| \le (1-\gamma) \max_i |p_{v_i} - p_{v_i}'|$. This is false. But,

Def) Message/Potential: cont. diff. function $\varphi : [\frac{1}{1+\lambda}, 1] \to \mathbb{R}$ with positive derivative. $\to$ same holds for φ^{-1}.

$\to$ Use $\varphi(x) = \frac{1}{s} \log\left(\frac{x}{s-x}\right)$ where $s := \frac{d+1}{d}$.

Work with $\varphi(p_v)$ instead of p_v. $p_{v_i} \to p_i$. $m = \varphi(p_v)$, $m_i = \varphi(p_i)$.

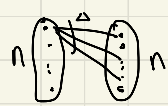
$\to m = \varphi(f(\varphi^{-1}(m_1), \varphi^{-1}(m_2), \ldots, \varphi^{-1}(m_n))) = F(m_1, \ldots, m_n)$.

Key Fact: $\forall \lambda < \lambda_c(\Delta)$, $\exists \gamma > 0$ s.t. $\forall \vec{m}, \vec{m}' \in \left(\varphi\left[\frac{1}{1+\lambda}, 1\right]\right)^d$,

$|F(\vec{m}) - F(\vec{m}')| \leq (1-\gamma)|\vec{m} - \vec{m}'|_\infty$. $\to$ Implies original goal. //

Sketch of (ii) of Theorem: We will actually show the following:

[Sly] A poly time algorithm for sampling ind. sets from Gibbs distribution $\pi(I) = \frac{1}{Z(\lambda)} \lambda^{|I|}$ would allow us to solve MAXCUT in poly time w.h.p. (for $\lambda > \lambda_c(\Delta)$).



Consider the bipartite graph G with independent, random Δ-deg. perfect matchings. An (α, β)-ind. set has αn vertices in L, βn in R.

Let $Z_G^{(\alpha,\beta)}(\lambda) := \sum_{(\alpha,\beta) \text{ IS } I} \lambda^{(\alpha+\beta)n}$.

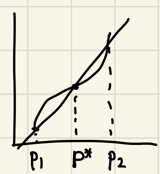
Claim: $E[Z_G^{\alpha,\beta}(\lambda)] = \exp\{\Phi_\lambda(\alpha,\beta) \cdot n \cdot (1+o(1))\}$ where

$\Phi_\lambda(\alpha,\beta) := (\alpha+\beta)\ln\lambda + H(\alpha) + H(\beta) + \Delta\left[(1-\beta)H\left(\frac{\alpha}{1-\beta}\right) - H(\alpha)\right]$

and $H(x) = -x\ln x - (1-x)\ln(1-x)$ (two-point entropy).

Proof: $E[Z_G^{\alpha,\beta}(\lambda)] = \lambda^{(\alpha+\beta)n} \sum_{|I_L| = \alpha n, |I_R| = \beta n} \Pr[I_L \cup I_R \text{ is ind. set}]$

$= \lambda^{(\alpha+\beta)n} \binom{n}{\alpha n}\binom{n}{\beta n} \Pr[I_L \cup I_R \text{ is ind. set}] = \lambda^{(\alpha+\beta)n}\binom{n}{\alpha n}\binom{n}{\beta n}\left[\frac{\binom{n-\beta n}{\alpha n}}{\binom{n}{\alpha n}}\right]^\Delta$.



Using $\ln\binom{n}{cn} \sim nH(c) + o(1)$, we get the result. //

Fact: $\Phi_\lambda(\alpha,\beta)$ satisfies: (i) max on line $\alpha = \beta$ is at $\alpha = \beta = p^*$.

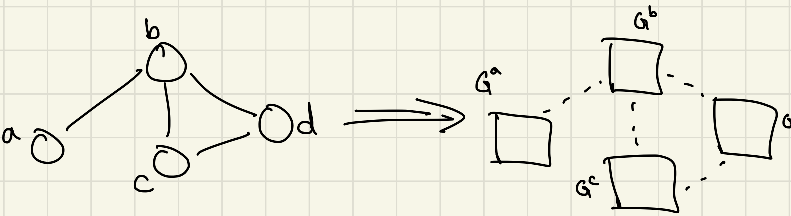
(ii) if $\lambda < \lambda_c(\Delta)$ then $\alpha = \beta = p^*$ is a global maximum.

(iii) if $\lambda > \lambda_c(\Delta)$ then (p^*, p^*) is a saddle point and (p_1, p_2) and (p_2, p_1) are two global maxima. (ind. sets are likely to be lop-sided)

Fact*: w.h.p. over the random bipartite graph G of degree Δ,
$Z_G^{\alpha,\beta}(\lambda) > \frac{1}{n} E[Z_G^{\alpha,\beta}(\lambda)]$ $\forall \alpha, \beta$ not close to $0, 1$.

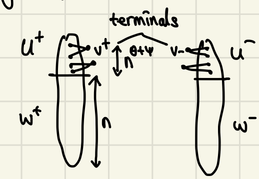
Thm) Glauber Dynamics for ind. sets with $\lambda > \lambda_c(\Delta)$ has mixing time $\exp(n)$ w.h.p. over G. (Intuition: conductance btwn $p_1 \leftrightarrow p_2$ is small)

Thm) Sly's Reduction: H (input for Max Cut) $\to H^G$ (input for sampling ind. sets)



(max degree Δ)

consider the modified gadget:



we have n^θ trees with n^ψ vertices each in terminals.

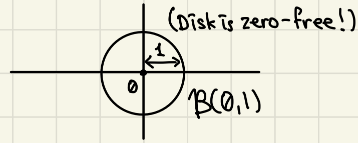
also, $0 < \theta + \psi < 1$.

Fact: For any Δ and $\lambda > \lambda_c(\Delta)$, $\exists$ constants θ, ψ with $\theta + \psi < 1$ s.t.

(i) $\Pr_\sigma[S(\sigma) = +] \geq 1/n$ and $\Pr_\sigma[S(\sigma) = -] \geq 1/n$, ($S$ is the $\pm$ phase)

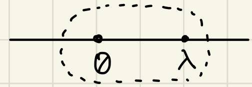
(ii) Conditioned on phase S, marginal of Gibbs dist. on terminals is very close to a product dist. with parameters p_1, p_2. i.e.,

$\max_{\sigma_v}\left\{\frac{\pi(\sigma_v|S=t)}{Q_v^+[\sigma_v]} - 1\right\} \le n^{-2\theta}$. [deferred to notes due to time]



Lee-Yang Theorem

$Z(\lambda)=\sum_k a_k\lambda^k$ where $a_k\ge 0$. What if $\lambda\in\mathbb{C}$?



Ferromagnetic Ising Model $Z_G(\lambda,\mu)=\sum_{S\subseteq V}\lambda^{|E(S,\bar{S})|}\cdot\mu^{|S|}$ where $\lambda\in[0,1]$, $\mu\ge 0$.

↳ $Z(\mu)=\sum a_k\mu^k$ where $a_k=\sum_{|S|=k}\lambda^{|E(S,\bar{S})|}$. Let $f(\mu):=\log Z(\mu)$.

Claim: $\forall \varepsilon\le 1/4,\ \mu$, if $|\tilde{f}-f(\mu)|\le\varepsilon$, then $|\exp(\tilde{f})-Z(\mu)|\le 4\varepsilon Z(\mu)$.

Proof: $|\exp(\tilde{f})-Z(\mu)|=|\exp(\tilde{f})-\exp(f(\mu))|=\exp(f(\mu))|\exp(\overbrace{\tilde{f}-f(\mu)}^{\varepsilon})-1|$

$\le \exp(f(\mu))\cdot 4\varepsilon\le 4\varepsilon Z(\mu)$ (Geometric fact!)

Taylor expansion of $f(\mu)=\log Z(\mu)$ around $\mu=0$: $f(\mu)=\sum_{i=0}^{\infty}\frac{f^{(i)}(0)\mu^i}{i!}$.

Truncate it after m terms, $f_m(\mu)=\sum_{i=0}^{m}\frac{f^{(i)}(0)\mu^i}{i!}$. $|f_m(\mu)-f(\mu)|\le\varepsilon$?

Write $Z(\mu)=\prod_{i=1}^{n}\left(1-\frac{\mu}{r_i}\right)$ where $\{r_i\}$ are complex roots of Z s.t. $|r_i|\ge \ell$.

↳ $f(\mu)=\sum_{i=1}^{n}\log\left(1-\frac{\mu}{r_i}\right)=-\sum_{i=1}^{n}\sum_{j=0}^{\infty}\frac{1}{j}\left(\frac{\mu}{r_i}\right)^j$.

$|f_m(\mu)-f(\mu)|\le n\cdot\sum_{j=m+1}^{\infty}\frac{(\mu/\ell)^j}{j}\le n\frac{(\mu/\ell)^{m+1}}{m+1}\sum_{j=0}^{\infty}(\mu/\ell)^j=n\frac{(\mu/\ell)^{m+1}}{(m+1)(1-|\mu/\ell|)}$.

To achieve $|f_m(\mu)-f(\mu)|\le\varepsilon/4$, $m\ge\frac{1}{\log|\mu/\ell|^{-1}}\left(\log\frac{4n}{\varepsilon}+\log\frac{1}{1+|\mu/\ell|}\right)=O(\log(n/\varepsilon))$.

Algorithm: compute $f_m(\mu)$ for $m=C\cdot(\log n/\varepsilon)$. Output $\exp(f(\mu))$.

Computing $f_m(\mu)$: $f'(\mu)=\frac{1}{Z(\mu)}\frac{dZ(\mu)}{d\mu} \to \frac{dZ(\mu)}{d\mu} = Z(\mu)f'(\mu)$.

$Z^{(m)}(0)=\sum_{j=0}^{m-1}\binom{m-1}{j}Z^{(j)}(0)f^{(m-j)}(0)$ [exercise; by induction]

↳ Non-singular triangular system $\Rightarrow$ can compute all $f^{(i)}(0)$ w.r.t. $\{Z^{(j)}(0)\}_{j=0}^{m}$

Finding $Z^{(j)}(0)$ is exhaustive enumeration of $\binom{n}{j}=O(n^{\log(n/\varepsilon)})$ subsets in G

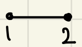
$\Rightarrow$ quasi-polynomial algorithm (if Δ is bounded, polynomial)

Reminder: $Z_G(\lambda,\mu)=\sum_{S\subseteq V}\lambda^{|E(S,\bar{S})|}\cdot\mu^{|S|}=\sum_{|S|=k}a_k\mu^k$.

Thm) Lee-Yang: For any finite, connected graph G and fixed $\lambda<1$, all the zeros of the Ising model partition function on $\mathbb{C}$ lie on $B(0,1)$.

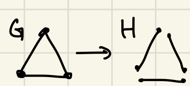
Corollary: $\exists$FPTAS for $Z_G(\lambda,\mu)$ $\forall G, \forall\lambda, \mu\neq 1$. (b/c $Z(\mu)=\mu^n Z(\mu^{-1})$).

Proof: [Asano] write $Z_G(\mu_1,\dots,\mu_n):=\sum_{S\subseteq V}\lambda^{|E(S,\bar{S})|}\cdot\prod_{i\in S}\mu_i$. Say that Z_G is "Lee-Yang" if $\forall i, |\mu_i|>1 \Rightarrow Z_G(\mu_1,\dots,\mu_n)\neq 0$. S.t.p. Z_G is L-Y.

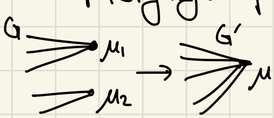
Structural Induction on G: (i) 

$Z(\mu_1,\mu_2)=\mu_1\mu_2+\lambda(\mu_1+\mu_2)+1$.

Suppose $Z(\mu_1,\mu_2)=0$. $\to |\mu_2|=\left|\frac{1+\lambda\mu_1}{\lambda+\mu_1}\right| \to$ if $|\mu_1|>1, |\mu_2|\leq 1$. //

(ii) For vertex-disjoint G, H, if Z_G & Z_H are both L-Y, $Z_{G\uplus H}$ is also L-Y since roots of $Z_{G\uplus H}$ are union of roots of Z_G & Z_H. 

(iii) Merging: sps. by induction that Z_G is L-Y. Fix all $\mu_3\dots\mu_n$ s.t. $|\mu_i|>1$



$\forall 3\leq i\leq n$. $Z_G(\mu_1,\mu_2)=A\mu_1\mu_2+B\mu_1+C\mu_2+D$.

We know that Z_G is L-Y by induction. Thus the above cannot be $\emptyset$ if $|\mu_1|, |\mu_2| > 1$. Hence roots of $A\mu^2 + (B+C)\mu + D$ have magnitude ≤ 1.

Hence magnitude of product of these roots is $|\frac{D}{A}| \leq 1$.

(Technicality: show $A \neq \emptyset$).

$Z_{G'}(\mu) = A\mu + D \rightarrow$ root of this is $\mu = -D/A$, so $|\mu| < 1$. //

* this does not work with $\mu \xrightarrow{\lim} 1$ (cannot approximate zero-field)

// End of official lectures, below are Student Presentations

Langevin Algorithm for Log-Concave [Altschuler & Talwar, COLT 23']

Continuous energy function $f: \mathbb{R}^d \rightarrow \mathbb{R}$, sample $x \in \mathbb{R}^d$ w.p. $\propto e^{-f(x)} \rightarrow \pi$

Assumptions on f: f is convex and M-smooth. $\rightarrow \lambda_{max}(\nabla^2 f) \leq M$ $\rightarrow \forall x,y,\ \|\nabla f(x) - \nabla f(y)\| \leq M\|x-y\|$

Motivation: Sampling from / Approximating volume of convex body

Langevin Diffusion: Start from $x_0 \in \mathbb{R}^d$. Consider the following diff eq,

$dX_t = -\nabla f(X_t) \cdot dt + \sqrt{2}\, dB_t$. $X_t \rightarrow \pi$ as $t \rightarrow \infty$.

Langevin Algorithm (Discrete): step size $\eta > 0$, convex subset $K \subseteq \mathbb{R}^d$.

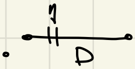
$X_{t+1} = \underbrace{\Pi_K}_{\text{projection to } K}(X_t + \eta \cdot \nabla f(X_t) + Z_t)$ where $Z_t \sim N(\emptyset, 2\eta \cdot I_d)$

Question: How fast is T_{mix} for Langevin Algorithm?

$$\begin{tikzcd} LA_\eta \arrow[r, "t\to\infty"] \arrow[d, "\eta\to 0"] & \pi_\eta \arrow[d, "\eta\to 0"] \\ LD \arrow[r, "t\to\infty"] & \pi \end{tikzcd}$$

Main Thm) Let $K \subseteq \mathbb{R}^d$ with diameter D. Let $f: \mathbb{R}^d \to \mathbb{R}$ be a convex & M-smooth potential. Assume step size $\eta < 2/M$. Consider the of above LA_η. $\underbrace{T_{\text{mix}(\substack{TV\\KL})}(1/4) \le O\left(\frac{D^2}{\eta}\right)}$.

Remarks: 1) $O\left(\frac{D^2}{\eta}\right)$ is also a lower bound! Consider a 1-D line.



2) Stochastic update $f = \frac{1}{n}\sum_i f_i$ also works WLOG.

3)

	LA	GD
Task:	sample x w.p. $\propto e^{-f(x)}$	find $\min_x\{f(x)\}$
Iteration:	$X_t - \eta\cdot\nabla f(X_t) + N(0, 2\eta I_d)$	$X_t - \eta\cdot\nabla f(X_t)$
Error bound:	$e^{-\frac{D^2}{\eta t}}$	$\frac{\eta t}{D^2}$

Def) α-Contractivity: $\phi: \mathbb{R}^d \to \mathbb{R}$ s.t. $\forall x, y \in \mathbb{R}^d$, $\|\phi(x) - \phi(y)\| \le \alpha \cdot \|x - y\|$

Obs) GD step and projection step are 1-contractive! (uses smoothness)

Consider $X_0, X_0' \in K$. Run $X_0 \to \cdots \to X_t$, $X_0' \to \cdots \to X_t'$ with LA_η.

Claim: $TV \le \sqrt{\frac{1}{2}\cdot KL}$ for this coupling.

1) Data Processing Inequality: $X \sim \mu$, $y \sim \nu$. Then $D_{KL}(\phi(x) \| \phi(y)) \le D_{KL}(x \| y)$.

2) DPI $\Rightarrow D_{KL}(X+Z \| y+Z) \le D_{KL}((X,Z) \| (y,Z)) = D_{KL}(X \| y) + \cancel{D_{KL}(Z \| Z)}^{\nearrow 0}$.

3) Shifted KL $D_{KL}^{(z)}(X\|y) := \inf_{(X,X')\sim\Gamma} D_{KL}(X'\|y)$ where $W_\infty(X,X') \leq z$. $\overset{\|X-X'\|_2}{}$

↳ monotonically decreases with z.

Claims: 1') for 1-contractive ϕ, $D_{KL}^{(z)}(\phi(x)\|\phi(y)) \leq D_{KL}^{(z)}(x\|y)$.

2') $\forall a \geq 0$, $D_{KL}(X+Z\|y+Z) \leq D_{KL}^{(z+a)}(x\|y) + \frac{a^2}{2\sigma}$ where $Z\sim N(0,\sigma^2 I)$.

Proof: 1') $(x,x')\sim(\mu,\mu')$. $\|x-x'\| \leq z \Rightarrow \|\phi(x)-\phi(x')\| \leq z$.

$\Rightarrow D_{KL}^{(z)}(\phi(x)\|\phi(y)) \leq D_{KL}(\phi(x')\|\phi(y)) \leq D_{KL}(X'\|y) = D_{KL}^{(z)}(X\|y)$. //

2') $D_{KL}(X+Z\|y+Z) = D_{KL}(\overbrace{X-W}^{X'}+W+Z\|y+Z) \leq D_{KL}((X', W+Z)\|(y,Z))$

$= D_{KL}(X'\|y) + \mathbb{E}_{X',X,y}[D_{KL}(W+Z\|Z) \mid X',X,y] = D_{KL}(X'\|y) + \frac{E[\|W\|_2^2]}{2\sigma^2}$ $\rightarrow a^2$. //

Consider the coupling (y_t, y_t') with ϕ. $\begin{cases} y_{t+1} = \phi_t(y_t) + Z_t \\ y_{t+1}' = \phi_t(y_t') + Z_t. \end{cases}$

Prop: $KL(y_T\|y_T') \leq \frac{\|y_0-y_0'\|_2^2}{2\sigma^2 T} \leq \frac{D^2}{2\sigma^2 T}$.

MCMC for Computing Volumes [Cousins, Vempala 18']

Problem: Compute volume of convex body $K \in \mathbb{R}^n$ given a membership oracle

↳ requires $(1-\varepsilon)\mathrm{vol}(K) \leq V \leq (1+\varepsilon)\mathrm{vol}(K)$ w.p. $1-p$.

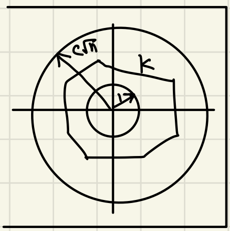
Complexity measured by # of oracle calls

DFK : $\tilde{O}(\mathrm{poly}(n, 1/\varepsilon)) = \tilde{O}\left(\frac{n^3}{\varepsilon^2} + n^4\right)$ $\qquad \mathbb{E}_{x \sim K}[\|x\|^2] = O(n)$

Assume that $B_n \subseteq K \subseteq C\sqrt{n}\, B_n$, that is, K is "well-rounded".

takes $\tilde{O}(n^4)$

Ideally, K should be isotropic, $\mathbb{E}_{x \sim K}[xx^T] = I_n$. Then, the problem is easy due to concentration bound decaying exponentially out of $C\sqrt{n}\, B_n$.



Define $F(\sigma^2) := \int_{\mathbb{R}^n} f(\sigma^2; x)\, dx$ where

$f(\sigma^2, x) = \mathbb{1}\{x \in K\} \cdot \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$. Then $F(\infty) = \int_K dx = \mathrm{vol}(K)$.

Strategy: incrementally approximate $\frac{F(\sigma_{i+1}^2)}{F(\sigma_i^2)}$.

Algorithm Sketch: consider some σ^2 schedule of $O(\frac{1}{n}) \to O(n) \to \infty$.

sample $x_j \sim \mu_i$ where $\mu_i(x) := \frac{f(\sigma_i^2, x)}{F(\sigma_i^2)}$. estimate $\frac{F(\sigma_{i+1}^2)}{F(\sigma_i^2)}$ with

Monte Carlo: $\frac{1}{k} \sum_{j=1}^{k} \frac{f(\sigma_{i+1}^2, x_j)}{f(\sigma_i^2, x_j)}$. Then, $V = F(\infty) \simeq \prod \frac{F(\sigma_{i+1}^2)}{F(\sigma_i^2)} \cdot F(\sigma_0^2)$.

To sample $x_j \sim \mu_i$ efficiently, we use a Speedy Walk that has conductance $\phi = \Omega\left(\frac{\delta}{\sigma\sqrt{n}}\right)$ where $\delta \leq \frac{\sigma}{8\sqrt{n}}$. [Proof omitted]

Gaussian Cooling: σ^2 scheduled in two phases, $O(\frac{1}{n}) \to 1$, $1 \to O(n)$.

phase 1 $\to \sigma_{i+1}^2 \propto \sigma_i^2(1+\frac{1}{n})$, phase 2 $\to \sigma_{i+1}^2 \propto \sigma_i^2(1+\frac{\sigma_i^2}{n})$

Analysis: Let $Y_j := \frac{f(\sigma_{i+1}^2, X_j)}{f(\sigma_i^2, X_j)}$, $X_j \sim \mu_i$, $\hat{Y} = \frac{1}{k}\sum_{j=1}^{k} Y_j$. $E[Y] = \frac{F(\sigma_{i+1}^2)}{F(\sigma_i^2)}$. (unbiased)

Y is concentrated around $E[Y]$ with Chebyshev, and $O(\frac{1}{\varepsilon^2})$ samples suffice.

Each sampling takes mixing $T \sim \max\{1, \sigma^2\} \cdot n^2$. First phase requires $\tilde{O}(n)$ updates x $\tilde{O}(\frac{1}{\varepsilon^2})$ samples x $\tilde{O}(n^2)$ sampling time $= \tilde{O}(\frac{n^3}{\varepsilon^2})$.

Second phase has $\tilde{O}(\frac{n}{\sigma^2})$ updates x $\tilde{O}(\frac{1}{\varepsilon^2})$ samples x $\tilde{O}(\sigma^2 \cdot n^2)$ sample time $= \tilde{O}(\frac{n^3}{\varepsilon^2})$. $\Rightarrow$ Total runtime $\tilde{O}(n^4 + \frac{n^3}{\varepsilon^2})$ (isotropic transformation & SW)

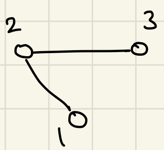
Spectral Independence [ALG]

Problem: $\mu: 2^{[n]} \to [0,1]$, hard-core model, Glauber dynamics.

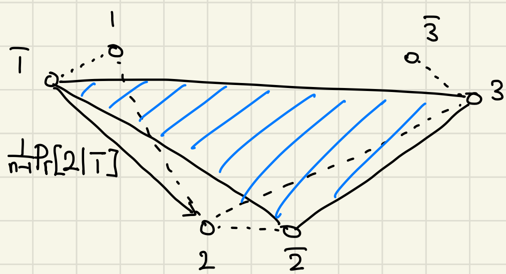
pick i u.a.r., move to $S \cup \{i\}$ w.p. $\frac{\mu(S\cup\{i\})}{\mu(S\cup\{i\}) + \mu(S\setminus\{i\})}$.

Simplicial Complex: X^μ, make two nodes $i, \bar{i}$ for each i, representing inclusion/exclusion of i in the IS. Then draw valid edges w.p. $\frac{1}{n-1}\Pr[j|i]$.

ex)



$\Longrightarrow$



Then, each face is a valid IS.

Generally, conditioning on m states has transitions $\frac{1}{n-m-1}\Pr[j \mid i_1,\dots,i_m]$.

We view transitions as down-up $P_d^{\vee}(\sigma,\sigma')$.

(pure) d-dimensional SC

Local Expander: X^{μ} is $(\alpha_0,\dots,\alpha_{d-2})$ LE if conditioned on i items,

conditionals $(2(n-i)\times 2(n-i))$

$\lambda_2(P_{n-i}) \le \alpha_i$. By [Alev & Lau 20'], $1-\lambda_2(P_d^{\vee}) \le \frac{1}{d}\prod_{k=0}^{d-2}(1-\alpha_k)$.

$(n\times n)$

Influence Matrix: $\Psi_\mu(i,j) := \Pr[j \mid i] - \Pr[j \mid \bar{i}]$.

η-Spectral Independence: $\lambda_{max}(\Psi_\mu) \le \eta$.

$(\eta_0,\dots,\eta_{n-2})$-SI: $\lambda_{max}(\Psi_{\mu(\cdot|S)}) \le \eta_i$ where $|S|=i$.

↳ This allows $1-\lambda_2(P_d^{\vee}) \le \frac{1}{n}\prod_{k=0}^{n-2}\left(1-\frac{\eta_k}{n-k-1}\right)$.

Quantum Metastability [BCV 25']

Background: consider spin system on $G(V,E)$, $|V|=n$.

	Classical	Quantum
configurations	$v\in V$ to $\{\pm 1\}$	$v\in V$ to $\mathbb{C}^2$; $tr(\rho)=1$, PSD
state	$\sigma: V\to\{\pm1\}$	quantum state $\rho\in\mathbb{C}^{2^n\times 2^n}$
Hamiltonian	$H(\sigma)$	H is an operator.
Gibbs?	distribution $\pi(\sigma)\propto e^{-\beta H(\sigma)}$	state $\rho\propto e^{-\beta H}$
Markov Generator	heat kernel $H_t(x,y)=e^{tL}$; $L = P-I$	Lindbladian $\mathcal{L}$

Quantum Gibbs sampling? design $\mathcal{L}$ s.t. $\mathcal{L}[\rho]=0$? $\simeq 0$? (metastable)

↳ Are metastable states useful? Do they satisfy properties of ρ?

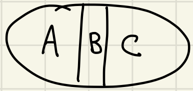
Def) ε-metastability: $\|\mathcal{L}[\sigma]\|_1 \le \varepsilon$.

Prop: Fix $T>0$, any initial state σ_0. Then the time-averaged state $\bar{\sigma}_T := \frac{1}{T}\int_0^T e^{t\mathcal{L}}[\sigma_0]dt$ is ε-metastable with $\varepsilon < \frac{2}{T}$.

Proof: equation of motion is $\frac{d}{dt}\sigma_t = \mathcal{L}[\sigma_t]$. $\to \sigma_T - \sigma_0 = \int_0^T \mathcal{L}[\sigma_t]dt$

$= \mathcal{L}\left[\int_0^T \sigma_t dt\right] = T\cdot\mathcal{L}[\bar{\sigma}_T]$. $\to T\cdot\|\mathcal{L}[\bar{\sigma}_T]\|_1 \le \|\sigma_T\| + \|\sigma_0\| \le 2$. //

$I(A;C|B)=0$.

Classical (Global) Markov Property:  $\sigma \sim \pi$. $\Rightarrow \sigma_A \perp \sigma_C | \sigma_B$

Area Law: $S(x) := \sum_x p(x)\log\left(\frac{1}{p(x)}\right)$, $I(x;y) := S(x)+S(y)-S(x,y) \le S(x)$

$I(x;y,z) := I(x;y) - I(x,z|y)$. $\Rightarrow I(A;A_c)_\pi \le |\partial A|$ (boundary of A)

Quantum Area Law: $S(\rho) := \mathrm{Tr}[\rho\log\rho]$. $I(A;C|B)_\rho = 0$? not in local.

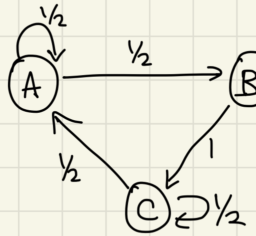
Thm1) $\forall\varepsilon$-metastable state σ, $\forall A\subseteq[n]$, $I(A;\bar{A})_\sigma \le 2\beta\|2H\|_\infty + n^2\varepsilon^\lambda e^{\mu|A|}$.

Local Approx. Markov Property: $I(A;C|B)_\rho$ exponentially decays, A or C is "small." something like $I(A;C|B) \lesssim e^{\min(|A|,|C|)}\cdot e^{-\mathrm{dist}(A,C)}$.

Thm2) $\forall\varepsilon$-MS σ, $\forall A\subseteq[n]$, $\exists \underset{\text{recovery map}}{R}$ s.t. $\forall$ noise $\mathcal{N}_A$, $\|\sigma - R[\mathcal{N}_A[\sigma]]\|_1 \le n e^{\mu|A|}\varepsilon^\lambda$.

Bottom Line w/o proof: Metastability $\Rightarrow \cdots \overset{\text{(Thm2)}}{\Rightarrow}$ Locally Markov $\overset{\text{(Thm 1)}}{\Rightarrow}$ Area Law

Coupling From the Past [Propp & Wilson 96']



Naïve MC: for $i \in [N]$, sample random seed $r_0, \dots, r_t$.

start from $M=2$ initial states, evolve with r_i.

first time they meet → coupled, return that state.

Observe: if we collapse at state B at time A, then at time $(t-1)$, we must have collapsed at state A. ⇒ Naïve forward MC never outputs state B!

Let $f : \Omega \times [0,1] \to \Omega$ be a "global" coupling, i.e. $\forall x \in \Omega$, $r \sim U([0,1])$, $\Pr_r[f(x,r)=y] = P(x,y)$. Fix a time t, then $f_t(x) := f(x, r_t)$.

From t_1 to t_2: $F_{t_1}^{t_2}(x) := (f_{t_2} \circ \cdots \circ f_{t_1})(x)$ (composition of f_t)

Prop: $\Pr[F_{t_1}^{t_2}(x) = y] = P^{t_2 - t_1}(x,y)$. Proof: Induction.

Let T be the first time when $\forall x \in \Omega$, $|F_0^T(x)| = 1$.

↳ Is $F_0^T(x) = \pi(x)$? No, but motivates coupling from past, $F_{-M}^0(x)$!

Thm) F_{-M}^0 has the same distribution as π.