Review

- Designing Chains: $\mu \in \mathbb{R}^{\Omega}_{\geq 0}$ $\quad N \in \mathbb{R}^{\Omega\times\Omega'}_{\geq 0}$

noise

$$\pi(x,y) := \mu(x) N(x,y) \qquad N^{\circ}(y,x) = \pi(x|y)$$

Time-reversible Chain: NN° with stationary μ

Examples: Glauber dynamics, block dynamics, delete/add or add/delete in S.T., hit and run, restricted Gaussian, Langevin dynamics

- Transport/Wasserstein:

$$W(\nu,\nu') = \min\{\mathbb{E}[d(X,Y)] \mid X\sim\nu,\ X'\sim\nu'\}$$

- Contraction of $W \implies$ mixing time bound

Strategy: Couple two $X_0 \to X_1$, $X_0' \to X_1'$ for all deterministic pairs (X_0, X_0') of Stat.

$$\mathbb{E}[d(X_1, X_1')] < (1-c)\, d(X_0, X_0')$$

- Coloring: If $q \geq 4\Delta+1$ we got

$$W(\nu P, \nu' P) \leq \left(1 - \frac{q-4\Delta}{qn}\right) W(\nu,\nu')$$

Metropolis Chain

- Path Coupling: If d is a shortest-path metric, enough to look at adjacent starts.

$$x = x_0 \sim x_1 \sim x_2 \sim \cdots \sim x_\ell = x'$$

$$W(1_x P, 1_{x'} P) \underset{\text{HW}}{\leq} \sum_i W(1_{x_i} P, 1_{x_{i+1}} P) \leq (1-c)\sum_i d(x_i, x_{i+1})$$

$$= (1-c)\, d(x, x')$$

Plan for Today

- Coloring with $q \geq 2\Delta+1$
- Dobrushin's conditions
  - * Hardcore model
  - * Ising model
- Intro to Spectral Analysis
- Fourier analysis

Path Coupling for Colorings [Jerrum]

Take adjacent colorings X_0, X_0' ← even improper

$X_0(w) = a \qquad X_0'(w) = b \qquad (a \neq b)$

w: differing vertex.

Coupling:

- Pick same vertex v.
- If $v = w$, or $v \not\sim w$, pick same c.
- If $v \sim w$:

good   bad

$c: 1, 2, \ldots, a, \ldots, b, \ldots, q$

$c': 1, 2, \ldots, a, \ldots, b, \ldots, q$

$$\mathbb{E}[d(X_1, X_1')] \leq 1 - \underbrace{\frac{1}{n}}_{\text{picking } w} \cdot \frac{q-\Delta}{q} + \frac{\Delta}{n} \cdot \frac{1}{q}$$

$$= 1 - \frac{q - 2\Delta}{nq}.$$

As long as $q \geq 2\Delta + 1 \Rightarrow t_{mix} = O\left(\frac{nq}{q-2\Delta} \cdot \lg n\right)$

Metropolis mixes fast.

What about Glauber dynamics?

- Use comparison ← future lectures
- Use Dobrushin's conditions

Product Spaces

Suppose μ is on $\Omega = \Omega_1 \times \cdots \times \Omega_n$

$\{1, \ldots, q\}$ in coloring.

Glauber dynamic:

$$X = (X(1), \ldots, X(n)) \xrightarrow{\text{pick } i \text{ u.a.r.}} X' = (X(1), \ldots, \underset{i\text{-th index}}{Z}, \ldots, X(n))$$

Z is sampled from i-th marginal of μ

Conditioned on $X(1), \ldots, X(i-1), X(i+1), \ldots, X(n)$.

Def: Call this $\mu_i(\cdot \mid X(-i))$.

Goal: Show contraction of W based on

Hamming dist: $d(X, X') = \#\{i : X(i) \neq X'(i)\}$

Dobrushin's Influence Matrix

Adjacent X, X': unique j where $X(j) \neq X'(j)$.

We write $X \sim_j X'$.

Influence: How much do marginals differ in Glauber when resampling coord i.

$$I[j \to i] = \max\{ d_{TV}(\mu_i(\cdot \mid X(-i)), \mu_i(\cdot \mid X'(-i))) \mid X \sim_j X' \}.$$

Note: The value of $X(i), X'(i)$ do not matter, so think of X, X' as in $\Omega_1 \times \cdots \times \widehat{\Omega_i} \times \cdots \times \Omega_n$

Note: We have $I[i \to i] = 0$.

Influence Matrix:

$$\begin{array}{cc} & \downarrow \text{col } j \\ \text{row } i \to & \begin{bmatrix} & I[j \to i] & \end{bmatrix} \end{array}.$$

Informal Thm: If I is "small" $\Rightarrow$ fast mixing.

Example. μ is uniform over $\{0,1\}^n$.

$$I[j \to i] = 0 \Rightarrow I = 0$$

Example. Coloring with a palette of q colors.

$\mu_i(\cdot \mid X(-i)) :=$ uniform over $\{1, \ldots, q\} - \{X(v) \mid v \text{ neighbor of } i\}$.

What is $I[j \to i]$:

$$\begin{cases} * \; j \not\sim i : & 0 \\ * \; j \sim i : & \dfrac{1}{q - \Delta} \end{cases}$$

Worst-case happens when neighbors of i have diff colors in both X, X'.

$I \leqslant \frac{1}{q-\Delta} \cdot A$

entry-wise

adjacency matrix

Thm: If columns of I sum to $\leq 1-\delta$

$$\Rightarrow W(\underbrace{\nu P, \nu' P}_{\text{Glauber}}) \leq \left(1-\frac{\delta}{n}\right) W(\nu, \nu')$$

Proof: Use path coupling: $X_0 \sim_j X_0'$.

$$\left.\begin{array}{l} X_0 \longrightarrow X_1 \\ X_0' \longrightarrow X_1' \end{array}\right\} \text{Couple}$$

- Pick same coord i
- Maximally couple replacement $\rightarrow$ the one defining $d_{TV}(\cdot,\cdot)$

$$d(X_0, X_0') = 1$$

$$\mathbb{E}[d(X_1, X_1')] \leq \underbrace{\frac{1}{n}}_{\text{pick } i=j} \times 0 + \sum_{i \neq j} \frac{1}{n} \times (1 + I[j \to i])$$

$$= \frac{n-1}{n} + \frac{\sum_i I[j \to i]}{n} \leq 1 - \frac{\delta}{n}.$$

Corollary: For coloring $I \leq \frac{1}{q-\Delta} A$, so

column sums $\leq \frac{\Delta}{q-\Delta} = 1 - \frac{q-2\Delta}{q-\Delta}$.

For $q \geq 2\Delta+1$ we have $1 - \frac{q-2\Delta}{n(q-\Delta)}$

contraction of $W \Longrightarrow$

$$t_{mix} = O\left(\frac{q-\Delta}{q-2\Delta} \cdot n \lg n\right)$$

Dobrushin matrix is very useful for spin systems / graphical models / etc.

$$\mu(X(1), \ldots, X(n)) \propto \phi_1(X(1), X(2)) \cdot \phi_2(X(3)) \cdots$$

If i and j conditioned on everything else independent $\Longrightarrow I[j \to i] = 0$.

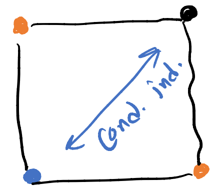
$\uparrow$ this happens when i, j do not appear in the same factor

Example. (Coloring)

Define $\Phi : [q]^2 \to \{0,1\}$ as $\Phi(a,b) = \mathbb{1}[a \neq b]$

$$\mu \propto \prod_{u \sim v} \Phi(X(u), X(v))$$

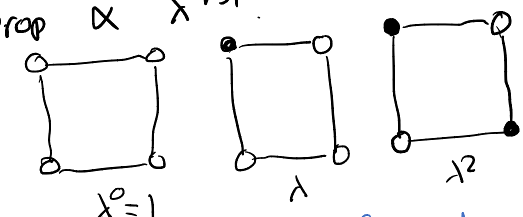
one factor per edge →



Example. (Hardcore model)

μ: independent sets S of graph ($\simeq \{0,1\}^n$)

w. prop $\propto \lambda^{|S|}$



$$\mu \propto \underbrace{\prod_i f(X(i))}_{1/\lambda} \cdot \prod_{i \sim j} \underbrace{g(X(i), X(j))}_{0/1}$$

$$g: \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 1 \\ 1 & 1 & 0 \end{array}$$

- Large λ is hard ← max indep. sets
- Small λ is easy:

$$\mathcal{I}[j \to i] = \begin{cases} i \not\sim j: & 0 \\ i \sim j: & \frac{\lambda}{1+\lambda} \end{cases}$$

Worst case: neighbors unoccupied

$I \leq \frac{\lambda}{1+\lambda} \cdot \underbrace{A}_{\text{adj}}$, so column sums are

$$\frac{\lambda \Delta}{1+\lambda} \leq \lambda \Delta.$$

$$\lambda \leq (1-\delta)/\Delta \Rightarrow t_{mix} = O\left(\frac{n}{\delta} \cdot \lg n\right).$$

Remark: This is not the correct threshold!

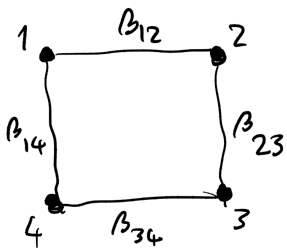
$\lambda \leq (1-\delta)\lambda_c(\Delta) \simeq \frac{e}{\Delta} \Rightarrow$ fast mixing ← will see later

$\lambda \geq (1+\delta)\lambda_c(\Delta) \Rightarrow$ NP-hard! [Sly]

Example. (Ising model)

μ on $\{\pm 1\}^n$ with

$$\mu(x(1),\dots,x(n)) \propto \exp\Big(\underbrace{\sum_{i\sim j}\beta_{ij}\,x(i)x(j)+\sum_i h_i x(i)}_{+\text{ in }\exp\,\equiv\,\text{factors}}\Big)$$



$\beta_{ij} > 0$: ferromagnetic

$\beta_{ij} < 0$: anti-ferromagnetic

$t \longrightarrow t + 2\beta_{ij}$

$$I[j\to i] \leqslant \frac{\tanh(|\beta_{ij}|) - \tanh(-|\beta_{ij}|)}{2} \leqslant |\beta_{ij}|$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Dist on $\{\pm 1\}$ w.p. $\propto \exp(tx)$ has mean $\tanh(t)$.

$d_{TV}(\nu,\nu') = \left|\mathbb{E}_{x\sim\nu}[x] - \mathbb{E}_{x\sim\nu'}[x]\right|/2$ for ν,ν' on $\{\pm 1\}$

$\{\underset{1-q}{-1}, \underset{q}{+1}\} \Longleftarrow \{\underset{1-p}{-1}, \underset{p}{+1}\}$ $\quad 2|p-q| =$ $\quad 1-2p \quad 1-2q$

Corollary. If ℓ_1 norm of β matrix $\leqslant 1-\delta$

$$\Longrightarrow t_{mix} = O\left(\frac{n}{\delta}\cdot \lg n\right)$$

Corollary. If β_{ij} are supported on Δ-max-deg graph and the same β:

$$|\beta| \leqslant \frac{1-\delta}{\Delta} \Longrightarrow \text{fast mixing.}$$

this is up to lower-order terms the correct threshold

Remark: Any spin system on $\{\pm 1\}^n$ with "soft" binary factors is an Ising model.

– Hardcore model is a limit

Dobrushin ++

If vector $c \in \mathbb{R}^n_{\geq 0}$ has $cI \leqslant (1-\delta)c$

so far $\mathbf{1}$

$\Longrightarrow$ c-Hamming has contraction

$$d(X,X') := \sum_i c_i \mathbf{1}[X(i) \neq X'(i)].$$

Proof: $X_0 \underset{j}{\sim} X'_0$: Couple the same way as before.

$$\mathbb{E}[d(X_1,X_1')] = \frac{1}{n} \times 0 + \sum_{i \neq j} \frac{1}{n} \times \left(c_j + c_i \cdot \mathbb{I}[j \sim i]\right)$$

$$= \frac{n-1}{n} c_j + \frac{1}{n} \cdot (cI)_j \leqslant \left(1 - \frac{\delta}{n}\right) c_j = d(X_0, X_0').$$

Remark: Every $\geqslant 0$ matrix has $\geqslant 0$ eigenvector. $cI = \lambda_{max} c$. This scheme "works" if $\lambda_{max} \leqslant 1-\delta$. [We need c_{min}/c_{max} to be not so small]

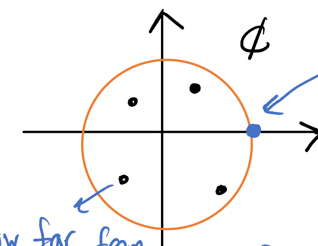
Intro to Spectral Analysis

Perron-Frobenius [for ergodic chains]:

P has a unique eigenvalue 1 and all others are <1 in magnitude.

Eigenvalues of P $\Longrightarrow$ Bounds on t_{mix}

will quantify



$P \cdot \mathbf{1} = \mathbf{1}$
$\mu \cdot P = \mu$

how far from boundary are they?

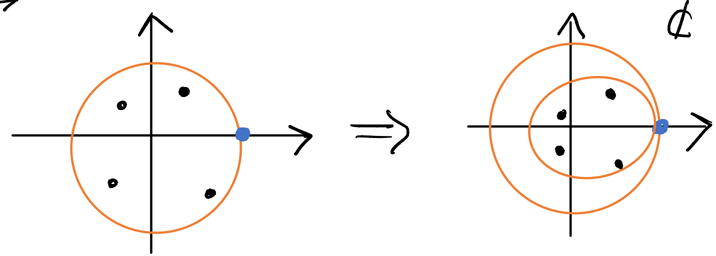
$$v_i P = \lambda_i v_i$$

$$\nu = \sum_i c_i v_i \Longrightarrow \nu P^t = \sum_i c_i \lambda_i^t v_i$$

if <1 converges to 0.

Lazification $\qquad APA^{-1} \sim P$

$P \Rightarrow \lambda P + (1-\lambda) I$



Intuition: It only matters that λ_i are far from 1. If $|\lambda_i| \simeq 1$ but far from 1, its norm shrinks in lazification

Time-reversible Chains

When P is time-reversible w.r.t. μ we have

$$\operatorname{diag}(\mu) P = Q \leftarrow \text{symmetric matrix}$$

$$\underline{\operatorname{diag}(\mu)^{\frac{1}{2}}}\, P\, \underline{\operatorname{diag}(\mu)^{-1/2}} = \underbrace{\operatorname{diag}(\mu)^{-\frac{1}{2}} Q \operatorname{diag}(\mu)^{-\frac{1}{2}}}_{\text{still sym}}$$

Linear Algebra Fact: Symmetric matrices have real eigenvalues, are diagonalizable, and have orthogonal eigenbasis.

Corollary. Time-reversible $\Rightarrow$ real eigenvalues.

Can order: $\quad 1 = \lambda_1 \geq \underline{\lambda_2} \geq \lambda_3 \geq \cdots \geq \lambda_n \geq -1$

Def (spectral gap): $\underline{1-\lambda_2}$ or $1-\max(\lambda_2, |\lambda_n|)$

Def (relaxation time): $1/(1-\lambda_2)$

↙ intuitively how fast λ_2^t decrease.

We will see later that in fact λ_2 dictatetes contraction of some quantity called χ^2-divergence.