Review

- Path Coupling: $W(1_x P, 1_{x'} P) \leq (1-c) W(1_x, 1_{x'})$
  only adjacent states
  * Coloring: $q > 2\Delta \Rightarrow t_{mix}(\text{Metropolis}) = O_\Delta(n \lg n)$
- Dobrushin: μ on $\Omega_1 \times \cdots \times \Omega_n$

$$I[j \to i] = \max\left\{ d_{TV}(\mu_i | x_{-i}, \mu_i | x'_{-i}) \mid x \sim_j x' \right\}$$

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

$$\Rightarrow t_{mix}(\text{Glauber}) = O(n \lg n / \delta)$$

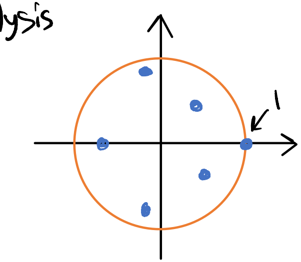
* Coloring: $q > 2\Delta \Rightarrow t_{mix} = O_\Delta(n \lg n)$
* Hardcore: $\lambda \lesssim \frac{1-\delta}{\Delta} \Rightarrow t_{mix} = O_\delta(n \lg n)$
* Ising: $\sum_i |\beta_{ij}| \leq 1-\delta \Rightarrow t_{mix} = O_\delta(n \lg n)$

- Dobrushin ++: $c \in \mathbb{R}^n_{>0}$ with $cI \leq (1-\delta) c$ $\leftarrow \lambda_{max}(I)$

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

c-weighted Hamming distance

- Spectral Analysis



$\mu P = \mu$
$P 1^t = 1^t$

- Time-Reversible: eigenvalues are real

$$-1 \leq \lambda_n \leq \cdots \leq \lambda_2 \leq \lambda_1 = 1$$

- Spectral Gap: $1 - \lambda_2$ or $1 - \max(\lambda_2, |\lambda_n|)$
- Relaxation Time: $t_{rel} := 1 / (1 - \lambda_2)$ ← a proxy for t_{mix}

Plan for Today

- Relationship between t_{mix}, t_{rel}
- Intro to functional analysis
- Fourier analysis

Eigenvalues & Mixing Time

Suppose $vP=\lambda v$ for $1\neq\lambda\in\mathbb{C}$, $v\in\mathbb{C}^n$.

$$\lambda v\mathbb{1}^t = vP\mathbb{1}^t = v\mathbb{1}^t \Rightarrow v\mathbb{1}^t=0$$

Def: $\|v\|_1 = \sum_i \left(|Re(v_i)|+|Im(v_i)|\right)$

Claim: $t\geq t_{mix}(\varepsilon) \Rightarrow \|vP^t\|_1 \leq O(\varepsilon)\cdot\|v\|_1$

Proof: Let $v := \alpha v_1 - \alpha v_2 + i\beta v_3 - i\beta v_4$ for

v_1,v_2,v_3,v_4 dists: $v_i\in\mathbb{R}^n_{\geq 0}$, $v_i\mathbb{1}^t=1$

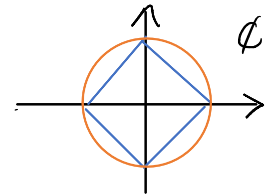
$\left.\begin{array}{l} v_1, v_2 \text{ disjoint support} \\ v_3, v_4 \text{ disjoint support}\end{array}\right\} \Rightarrow \|v\|_1 = 2\alpha+2\beta$

$$vP^t = \alpha\left(\overset{\approx\mu}{v_1P^t} - \overset{\approx\mu}{v_2P^t}\right) + i\beta\left(\overset{\approx\mu}{v_3P^t} - \overset{\approx\mu}{v_4P^t}\right)$$

$$\Rightarrow \|vP^t\|_1 \leq \alpha\cdot O(\varepsilon) + \beta\cdot O(\varepsilon) = O(\varepsilon)\cdot\|v\|_1$$

Def: $\|v\|_1' = \sum_i \underbrace{|v_i|}_{\sqrt{Re(v_i)^2+Im(v_i)^2}}$

$\|v\|_1' \overset{\sqrt{2}}{\simeq} \|v\|$ but $\|\lambda v\|_1' = |\lambda|\,\|v\|_1'$



Now if $t\geq t_{mix}(\varepsilon)$ we have

$$|\lambda|^t\|v\|_1' = \|vP^t\|_1' \leq O(\varepsilon)\cdot\|v\|_1'$$

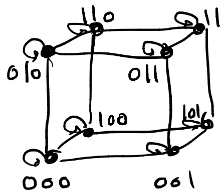
$$\Rightarrow |\lambda|^t = O(\varepsilon)$$

* Corollary: $|\lambda|<1$. $\Leftarrow$ Holds even when for eigenvectors other than μ if ergodic.

* Corollary: $1-|\lambda| \geq \Omega\left(\dfrac{\lg\frac{1}{\varepsilon}+const}{t_{mix}(\varepsilon)}\right)$

$|\lambda| \leq 1-\Omega\left(\frac{1}{t_{mix}}\right)$ $\leftarrow$ usually not tight

Example. (Hypercube)



$$t_{mix}(\varepsilon) = O\left(n\lg n + n\lg\frac{1}{\varepsilon}\right)$$

$$\Rightarrow t_{rel} = O\left(\frac{n\lg n + n\lg\frac{1}{\varepsilon}}{\text{const} + \lg\frac{1}{\varepsilon}}\right) \xRightarrow{\varepsilon\to 0} t_{rel} = O(n)$$

Example. (Dobrushin)

Suppose $\sum_i |I[j\to i]| \leqslant 1-\delta$

Glauber

$$\Rightarrow t_{mix} = O\left(\frac{n\lg n}{\delta} + \frac{n\lg\frac{1}{\varepsilon}}{\delta}\right)$$

$$\Rightarrow t_{rel} = O\left(\frac{n\lg n + n\lg\frac{1}{\varepsilon}}{\delta\cdot\lg\frac{1}{\varepsilon}}\right) \xRightarrow{\varepsilon\to 0} t_{rel} = O(n/\delta)$$

HW: What happens for Dobrushin++?

$$\lambda_{max}(I) \leqslant 1-\delta \Rightarrow t_{rel} = O(\,?\,)$$

Intro to Functional Analysis

Question: Can we bound t_{mix} by t_{rel}?

Idea: Contraction of proxy for d_{TV}

Def: (f-Divergence) convex function f:

$$\underbrace{D_f(\nu\|\mu)}_{\text{proxy for } d_{TV}(\nu,\mu)} := \mathbb{E}_{x\sim\mu}\left[f\left(\frac{\nu(x)}{\mu(x)}\right)\right] - f\Bigg(\underbrace{\mathbb{E}_{x\sim\mu}\left[\frac{\nu(x)}{\mu(x)}\right]}_{1 \text{ when } \nu \text{ is dist}}\Bigg)$$

Note: We define it for any $\nu : \Omega \to \mathbb{R}$ where ν/μ takes values in domain of f.

Fact: f convex $\Rightarrow D_f(\nu\|\mu) \geqslant 0$

Proof: Jensen's inequality! □

Remark: f strongly convex $\Rightarrow D_f = 0 \leftrightarrow \nu = \text{const}\cdot\mu$.

Fact: Suppose $N \in \mathbb{R}_{\geq 0}^{\Omega \times \Omega'}$ is row stochastic.

noise or Markov chain

Then $\underbrace{D_f(\nu N \| \mu N) \leq D_f(\nu \| \mu)}_{\text{Data processing ineq.}}$

Proof: We apply Jensen's again: $\pi(x,y) := \mu(x) N(x,y)$

$$\frac{\nu N(y)}{\mu N(y)} = \frac{\sum_x \nu(x) N(x,y)}{\sum_{x'} \mu(x') N(x',y)} = \sum_x \underbrace{\frac{\pi(x,y)}{\sum_{x'} \pi(x',y)}}_{\pi(x|y)} \frac{\nu(x)}{\mu(x)}$$

$$\Rightarrow f\left(\frac{\nu N(y)}{\mu N(y)}\right) \leq \sum_x \pi(x|y) f\left(\frac{\nu(x)}{\mu(x)}\right) \Rightarrow$$

$$\mathbb{E}_{y \sim \underbrace{\mu N}_{y\text{-marginal of } \pi}}\left[f\left(\frac{\nu N(y)}{\mu N(y)}\right)\right] \leq \sum_x \underbrace{\mathbb{E}_{y \sim \mu N}[\pi(x|y)]}_{x\text{-marginal of } \pi, \text{ i.e. } \mu(x)} f\left(\frac{\nu(x)}{\mu(x)}\right) =$$

$$\mathbb{E}_{x \sim \mu}\left[f\left(\frac{\nu(x)}{\mu(x)}\right)\right].$$

Note that

$$\mathbb{E}_{y \sim \mu N}\left[\frac{\nu N(y)}{\mu N(y)}\right] = \sum_y \nu N(y) = \sum_x \nu(x) \cancel{\sum_y N(x,y)} =$$

$$\mathbb{E}_{x \sim \mu}\left[\frac{\nu(x)}{\mu(x)}\right].$$

Therefore $D_f(\nu \| \mu) \geq D_f(\nu N \| \mu N)$ $\square$

Remark: This is extremely useful for Markov chains constructed as $P = NN^{\circ}$.

Contraction of either N or N° $\Longrightarrow$ contraction of P

In lots of scenarios N by itself easier to analyze!

Remark: We usually want fs where $d_{TV} \leq$ some func of D_f

Popular Choice 1: ($f(x) = x^2$)

$$D_f(\nu \| \mu) = \chi^2(\nu \| \mu)$$

χ^2-divergence

Also called $\mathrm{Var}_\mu\left[\frac{\nu}{\mu}\right]$

variance

— Note that ν can take <0 values too here.

— Alternative formula:

$$\chi^2(\nu\|\mu) = \mathbb{E}_{x \sim \mu}\left[\left(\frac{\nu(x)}{\mu(x)} - \mathbb{E}_{x' \sim \mu}\left[\frac{\nu(x')}{\mu(x')}\right]\right)^2\right]$$

Thm: For ν a dist, we have

$$d_{TV}(\nu, \mu) \leq \frac{1}{2}\sqrt{\chi^2(\nu \| \mu)}$$

Proof: Apply Cauchy-Schwarz:

$$2d_{TV}(\nu,\mu) = \sum_x |\nu(x) - \mu(x)| = \mathbb{E}_{x\sim\mu}\left[\left|\frac{\nu(x)}{\mu(x)} - 1\right|\right] \leq \sqrt{\chi^2(\nu\|\mu)}$$

$\square$

Popular Choice 2: ($f(x) = x \lg x$ for $x \geq 0$)

$$D_f(\nu \| \mu) = D_{KL}(\nu \| \mu)$$

KL-divergence

Also called $\mathrm{Ent}_\mu\left[\frac{\nu}{\mu}\right]$

entropy

— Note that ν can take ≥ 0 values but doesn't need to sum to 1.

— Formula for dist ν: $D_{KL}(\nu\|\mu) = \mathbb{E}_{x\sim\nu}\left[\lg\left(\frac{\nu(x)}{\mu(x)}\right)\right]$

Thm (Pinsker's): For ν a dist, we have

$$d_{TV}(\nu,\mu) \leq \sqrt{\frac{1}{2} D_{KL}(\nu\|\mu)}.$$

Proof: Define noise operator that maps x to 0 if $\nu(x) \geq \mu(x)$ and 1 if $\nu(x) < \mu(x)$.

$$d_{TV}(\nu N, \mu N) = d_{TV}(\nu, \mu), \quad D_{KL}(\nu N \| \mu N) \leq D_{KL}(\nu \| \mu)$$

$\Longrightarrow$ Enough to prove on $\Omega = \{0, 1\}$

The rest on HW.

Functional Analysis:

Show contraction of some D_f:

$$D_f(\nu P \| \mu) = D_f(\nu P \| \mu P) \leq (1-\rho) D_f(\nu \| \mu)$$

$$\Rightarrow D_f(\nu P^t \| \mu) \leq (1-\rho)^t D_f(\nu \| \mu)$$

Variance / χ^2

$$t \gtrsim \frac{\lg(\chi^2(\nu_0 \| \mu)/\varepsilon^2)}{\rho}$$

$$\Rightarrow d_{TV}(\nu_0 P^t, \mu) \leq \varepsilon$$

How large can $\chi^2(\nu_0 \| \mu)$ be? $\frac{\nu_0(x)}{\mu(x)} \leq \frac{1}{\mu_{min}}$

$$t_{mix} = O\left(\frac{\lg(\frac{1}{\mu_{min}})}{\rho}\right)$$

Entropy / D_{KL}

$$t \gtrsim \frac{\lg(D_{KL}(\nu_0 \| \mu)/\varepsilon^2)}{\rho}$$

$$\Rightarrow d_{TV}(\nu_0 P^t, \mu) \leq \varepsilon$$

How large can $D_{KL}(\nu_0 \| \mu)$ be? $\lg \frac{\nu_0(x)}{\mu(x)} \leq \lg \frac{1}{\mu_{min}}$

$$t_{mix} = O\left(\frac{\lg\lg(\frac{1}{\mu_{min}})}{\rho}\right)$$

These inequalities are related to

Poincare & MLSI

future: cont. time

Spectral gap $\iff$ Contraction of χ^2

Suppose $N \in \mathbb{R}_{\geq 0}^{\Omega \times \Omega'}$ is row-stochastic and μ dist on Ω. Let $\mu' = \mu N$ and N^o be time-reversal of N w.r.t μ.

D_μ: diagonal matrix with μ on diag. $D_{\mu'}$: diagonal matrix with μ' on diag.

$$D_{\mu'} N^o = (D_\mu N)^T = N^T D_\mu$$

Formula for χ^2: Suppose ν sums to 0.

$$\chi^2(\nu \| \mu) = \| \nu D_\mu^{-\frac{1}{2}} \|_2^2$$

For such a ν we have $(\nu N)1^t=0$ too,

So $\chi^2(\nu N \| \mu N) = \| \nu N D_{\mu'}^{-\frac{1}{2}} \|_2^2$.

What is

$$\sup \left\{ \frac{\| \nu N D_{\mu'}^{-\frac{1}{2}} \|_2^2}{\| \nu D_{\mu}^{-\frac{1}{2}} \|_2^2} \;\middle|\; \nu 1^t = 0 \right\} ?$$

This is equal to $\lambda_2(NN^\circ)$.

[if time ...]

Contraction for arbitrary ν $\longleftrightarrow$ Contraction for $\nu 1^t = 0$

$$P_{\text{variance}} = 1 - \lambda_2(NN^\circ)$$

Corollary: Time-reversible:

$$\lambda_2(P^2) = \max(\lambda_2(N)^2, \lambda_n(N)^2)$$

Proof. Let $v = \nu D_{\mu}^{-1/2}$ ← change of variable

Let $A = D_{\mu}^{\frac{1}{2}} N D_{\mu'}^{-\frac{1}{2}}$. Then

$$\sup \left\{ \frac{v A A^T v^T}{v v^T} \;\middle|\; v D_{\mu}^{\frac{1}{2}} 1^t = 0 \right\}.$$

Note that

$$AA^T = D_{\mu}^{\frac{1}{2}} N \underbrace{D_{\mu'}^{-1} N^T D_{\mu}}_{N^\circ} D_{\mu}^{-\frac{1}{2}} = D_{\mu}^{\frac{1}{2}} N N^\circ D_{\mu}^{-\frac{1}{2}}$$

is a symmetric matrix and similar to NN°, so has same eigenvalues.

$$AA^T D_{\mu}^{\frac{1}{2}} 1^t = D_{\mu}^{\frac{1}{2}} 1^t$$ ← eigenvector for eigenvalue 1.

So we are looking at

$$\sup \left\{ \frac{v A A^T v}{v v^T} \;\middle|\; v \text{ orthogonal to top eigenvector} \right\}$$

$$= \lambda_2(AA^T) = \lambda_2(NN^\circ)$$ ☺

Fourier Analysis

Finite Abelian group: $G = \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$.

with $+$: component wise addition. mod

Take dist π over G.

think sparse support.

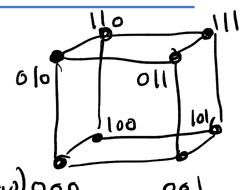
We get an Abelian walk:

$$X \longmapsto X + Z$$

random sample from π

Example. $G = \mathbb{Z}_2^n$

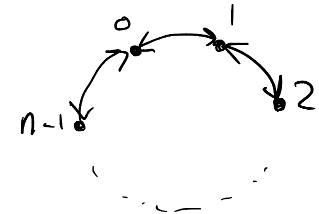
$$\pi: \begin{cases} \text{w.p. } \frac{1}{2} & \text{choose } (0,\dots,0) \\ \text{w.p. } \frac{1}{2n} & \text{choose } e_i = (0,\dots,\overset{i}{1},\dots,0) \end{cases}$$



This is Glauber dynamics.

Example. (Cycle)

$$\pi: \begin{cases} \text{w.p. } \frac{1}{2} & \text{choose } +1 \\ \text{w.p. } \frac{1}{2} & \text{choose } -1 \end{cases}$$



Fact 1: μ = uniform always stationary

Fact 2: π symmetric $\Longleftrightarrow$ time-reversible

$\pi(2) = \pi(-2)$

Fact 3: support π generate $G \Longleftrightarrow$ irreducible.

Thm: Regardless of π, eigenvectors are fixed; they are characters of G.

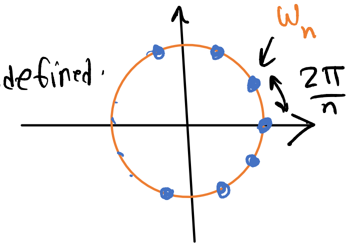
Def: (Character): $\chi: G \to \mathbb{C}^*$ s.t.

$$\chi(x+y) = \chi(x)\chi(y).$$

Example $(\mathbb{Z}_n)$:

Pick $\omega \in \mathbb{C}$ s.t. $\omega^n = 1$.

Then $\chi(x) = \omega^x$ is well-defined.



All characters are

$$\chi(x) = \exp\left(\frac{2\pi i \cdot tx}{n}\right) = \omega_n^{tx}$$

Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$

$$\chi(x_1, \dots, x_k) = \omega_{n_1}^{t_1 x_1} \cdots \omega_{n_k}^{t_k x_k}$$

There are exactly $n_1 \cdots n_k$ such characters.

Thm: If P is an Abelian walk, every character χ is an eigenvector.

$$(\chi P)(y) = \sum_{x \in G} \chi(x) P(x,y) =$$

$$\sum_{x \in G} \chi(x) \pi(y-x) = \sum_{z} \pi(z) \chi(y-z)$$

$$= \underbrace{\left(\sum_{z} \pi(z) \chi(z)^{-1}\right)}_{\text{Corresponding eigenvalue.}} \chi(y).$$

Since there are $n_1 \cdots n_k$ many, these are an eigenbasis!

$$\mathbb{E}_{z \sim \pi}[\chi(-z)]$$