Review

- Relationship between t_{mix} & t_{rel}:

$$\frac{1}{1-|\lambda|} \leqslant O\left(\frac{t_{mix}(\varepsilon)}{\lg\frac{1}{\varepsilon}}\right)$$

t_{rel} when $\lambda=\lambda_2$

← even for nonreversible

← tighter as $\varepsilon\to 0$

$$t_{mix}(\varepsilon) \leqslant O\left(t_{rel}\cdot \lg\left(\frac{1}{\varepsilon\cdot\mu_{min}}\right)\right)$$

← reversible & $\lambda_2 \geqslant |\lambda_n|$

Informally, t_{rel} is the coefficient of $\lg\frac{1}{\varepsilon}$.

- Functional analysis:

$$D_f(\nu\|\mu) := \mathbb{E}_{x\sim\mu}\left[f\left(\frac{\nu(x)}{\mu(x)}\right)\right] - f\left(\mathbb{E}_{x\sim\mu}\left[\frac{\nu(x)}{\mu(x)}\right]\right)$$

this is just $\sum_x \nu(x)$

* $f(x)=x^2$: $\chi^2(\nu\|\mu)$ or $Var_\mu\left[\frac{\nu}{\mu}\right]$
* $f(x)=x\lg x$: $D_{KL}(\nu\|\mu)$ or $Ent_\mu\left[\frac{\nu}{\mu}\right]$

$$d_{TV}(\nu,\mu) \leqslant \frac{1}{2}\sqrt{\chi^2(\nu\|\mu)} \qquad d_{TV}(\nu,\mu) \leqslant \sqrt{\frac{1}{2}D_{KL}(\nu\|\mu)}$$

- Contraction:

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

μ because stationary

- For any stochastic $N\in\mathbb{R}_{\geqslant 0}^{\Omega\times\Omega'}$ we have

$$\sup_\nu \frac{\chi^2(\nu N\|\mu N)}{\chi^2(\nu\|\mu)} = \lambda_2(NN^\circ)$$

time-reversal

contraction in χ^2 ⟷ bound on spectrum

- Fourier analysis

$G=\mathbb{Z}_{n_1}\times\cdots\times\mathbb{Z}_{n_k}$ → additive group

For dist π on G we get Abelian walk:

$$X \xrightarrow{P} X+Z$$

← fresh sample from π

* Eigenvectors ≡ characters:

$\chi: G\to\mathbb{C}-\{0\}$ s.t. $\chi(x+y)=\chi(x)\chi(y)$

* There are $n_1\cdots n_k$ of them

$$\chi((x_1,\dots,x_k)) = \omega_1^{x_1}\cdots\omega_k^{x_k}$$

n_1-th root of unity

Plan for Today

- Examples of Fourier analysis
- Continuous time
- Comparison method
- Trading approx for time ← if time

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

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

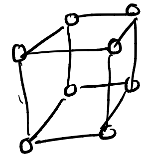
$$\sum_{x \in G} \chi(x)\, \pi(y-x) = \sum_{z} \pi(z)\, \chi(y-z)$$

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

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

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

Example (Hypercube) $G=\mathbb{Z}_2^n$



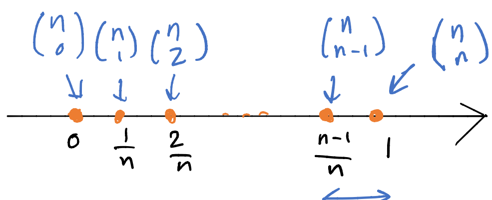
- Characters are

$$\chi(x_1,\dots,x_n)=\omega_1^{x_1}\cdots\omega_n^{x_n}$$
$$\omega_1,\dots,\omega_n\in\{\pm1\}$$

- Glauber dynamics:

$z\sim\Pi$: pick $i\in[n]$ u.a.r. and let $z=e_i$ w.p. $\frac{1}{2}$ and 0 w.p. $\frac{1}{2}$

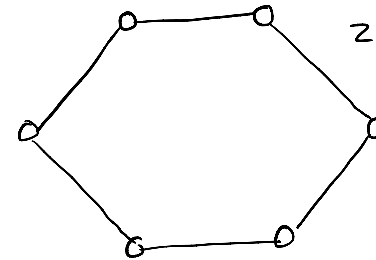
Eigenvalue: $\mathbb{E}_{z\sim\Pi}[\chi(z)]=\mathbb{P}_{i\sim[n]}[\omega_i=1]$



$$\frac{1}{1-\lambda_2}\times\lg\frac{1}{\mu_{min}}$$

$t_{rel}=O(n)\Rightarrow t_{mix}=O(n^2)$ ← not tight

Example (Cycle) $G=\mathbb{Z}_n$

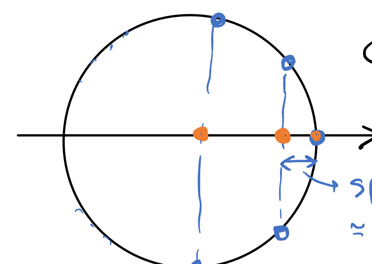


$z\sim\Pi$: pick ±1 w.p. $\frac{1}{2}$

- Characters are

$$\chi(x)=\omega^x \quad \omega\to\exp\left(\frac{2\pi i\cdot t}{n}\right)$$

Eigenvalue: $\mathbb{E}_{z\sim\Pi}[\chi(z)]=\frac{\omega+\omega^{-1}}{2}=\cos\left(\frac{2\pi}{n}\cdot t\right)$



$\cos x\approx 1-\frac{x^2}{2}\pm\cdots$

spectral gap $\approx\Theta\left(\frac{1}{n^2}\right)$

$t_{rel}=O(n^2)\Rightarrow t_{mix}=O(n^2\lg n)$ unnecessary

Continuous time

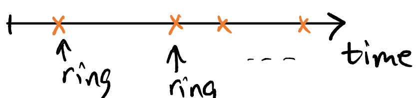
- So far we have been running

$$X_0 \xrightarrow{P} X_1 \xrightarrow{P} X_2 \longrightarrow \cdots$$

discrete time

- We can run a chain in continuous time too:

Std. Poisson Clock:



on average 1 ring per unit of time

Everytime the clock rings we take a step of P.

X_t: location at time t

makes sense for $t \in \mathbb{R}_{\geq 0}$

Remark. Algorithmically to simulate X_t we can just sample $n \sim \mathrm{Pois}(t)$ and run n steps of P.

Remark: For any t, conditioned on X_t, past & future are independent.

Poisson clock is memoryless.

Remark. For any t, $X_t \mid X_0$ follows a Markov chain. Same as $X_{t+s} \mid X_s$.

Think of dividing time into ε-sized intervals & w.p. ε taking a P in each interval.

$$\Big(\underbrace{(1-\varepsilon)I + \varepsilon P}_{\text{transition for an interval}}\Big)^{t/\varepsilon} = \big(I + \varepsilon(P-I)\big)^{\frac{t}{\varepsilon}} \xrightarrow[\text{as } \varepsilon \to 0]{} \exp\big(t(P-I)\big)$$

Markov chain matrix

Remark. Main benefit of continuous time is "periodicity" goes away.

Ultimate lazification: $P \mapsto (1-\varepsilon)I + \varepsilon P$

Functional analysis in continuous time

- In discrete time we wanted

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

- In cont. time the analogous condition is:

$$\frac{d}{dt} D_f(\nu_t \| \mu)\Big|_{t=0} \leq -\rho \cdot D_f(\nu_0 \| \mu)$$

$$\nu_t = \nu_0 \cdot \exp(t \cdot (P - I))$$

* Corollary: $D_f(\nu_t \| \mu) \leq e^{-t\rho} \cdot D_f(\nu \| \mu)$
So we again get mixing time bounds.

- Poincaré: This for χ^2
- Modified Log-Sobolev: This for D_{KL}
Should be called Gross's inequality

Note. Discrete time contraction $\Rightarrow$ cont. time

If P contracts by $1-\rho$ then $(1-\varepsilon)I + \varepsilon P$ contracts by $1 - \varepsilon\rho$.

$$D_f((1-\varepsilon)\nu + \varepsilon \nu P \| \mu) \underset{f \text{ is convex}}{\leq} (1-\varepsilon) D_f(\nu \| \mu) + \varepsilon D_f(\nu P \| \mu)$$

$$\leq (\cancel{(1-\varepsilon) + \varepsilon(1-\rho)})^{1-\varepsilon\rho} D_f(\nu \| \mu)$$

$$\Longrightarrow \frac{d}{dt} D_f(\nu_t \| \mu) \leq -\rho D_f(\nu_0 \| \mu)$$

Mini-proof: $e^{t\rho} \cdot D_f(\nu_t \| \mu) \leftarrow z_t$

$$\frac{d}{dt} z_t = e^{t\rho} \rho \cdot D_f(\nu_t \| \mu) - \left(\frac{d}{dt} D_f(\nu_t \| \mu)\right) \cdot e^{t\rho}$$

Remark: In general the reverse doesn't hold :(

Example: Periodic chains.

However:

For reversible & lazy chains in χ^2:

say $\lambda_n \geq 0$
or $\lambda_n \geq -\lambda_2$

discrete time $\Longleftrightarrow$ continuous time

Sketch:

- Poincare $\longleftrightarrow 2(1-\lambda_2(P))$

  This is because $1-\lambda_2((1-\varepsilon)I+\varepsilon P) = \varepsilon(1-\lambda_2(P))$

  So χ^2 contracts by $\simeq 2\varepsilon(1-\lambda_2(P))$

  Derivative in ε is $2(1-\lambda_2(P))$

- Discrete time χ^2 contraction $\longleftrightarrow 1-\max(\lambda_2^2(P), \lambda_n^2(P))$

  by assumption this dominates

Dirichlet form

Convenient to write $\frac{d}{dt} D_f(\nu_t \| \mu)$ in terms of quantity called Dirichlet form.

* Assume time-reversible

$$\nu_{dt} \simeq (1-dt)\nu + dt \cdot \nu P = \nu + dt \cdot \nu(P-I)$$

$$\frac{d}{dt}\mathbb{E}_{\mu}\left[f\left(\frac{\nu_{dt}}{\mu}\right)\right]\Bigg|_{t=0} = \mathbb{E}_{\mu}\left[f'\left(\frac{\nu}{\mu}\right)\cdot\frac{\frac{d}{dt}\nu_t\big|_{t=0}}{\mu}\right]$$

$\nu(P-I)$

$$-\frac{1}{2}\sum_{x,y} Q(x,y)\left(f'\left(\frac{\nu(x)}{\mu(x)}\right)-f'\left(\frac{\nu(y)}{\mu(y)}\right)\right)\left(\frac{\nu(x)}{\mu(x)}-\frac{\nu(y)}{\mu(y)}\right)$$

ergodic flow: $\mu(x)P(x,y) = \mu(y)P(y,x)$

Dirichlet form:

$$\mathcal{E}(g,h) := \frac{1}{2}\mathbb{E}_{(x,y)\sim Q}\left[(g(x)-g(y))(h(x)-h(y))\right]$$

$$\sum \cancel{Q(x,y)}\ \mu(x)P(x,y)\ f'\left(\frac{\nu(x)}{\mu(x)}\right)\cdot\frac{\nu(y)}{\mu(y)}$$

- For χ^2, we get $f(x)=x^2$, $f'=2x$

$$\frac{d}{dt}\left(\chi^2(\nu_t \| \mu)\right) = \underline{-2\mathcal{E}\left(\frac{\nu_t}{\mu}, \frac{\nu_t}{\mu}\right)}$$

Poincaré: $2\mathcal{E}\left(\frac{\nu_t}{\mu}, \frac{\nu_t}{\mu}\right) \geq \rho \operatorname{Var}_\mu\left[\frac{\nu_t}{\mu}\right]$

- For D_{KL}, we get $f(x)=x\lg x$, $f'=\lg x+1$

$$\frac{d}{dt}\left(D_{KL}(\nu_t \| \mu)\right) = -\mathcal{E}\left(\frac{\nu_t}{\mu}, \lg\frac{\nu_t}{\mu}\right)$$

MLSI: $\mathcal{E}\left(\frac{\nu_t}{\mu}, \lg\frac{\nu_t}{\mu}\right) \geq \rho \operatorname{Ent}_\mu\left[\frac{\nu_t}{\mu}\right]$

Remark: There is something called LSI too:

$$\mathcal{E}\left(\sqrt{\frac{\nu_t}{\mu}}, \sqrt{\frac{\nu_t}{\mu}}\right) \geq \rho \operatorname{Ent}_\mu\left[\frac{\nu_t}{\mu}\right]$$

Thm: LSI $\Longrightarrow$ MLSI $\leftarrow$ HW

strictly stronger

$(\sqrt{a}-\sqrt{b})^2 \leq c\cdot(a-b)(\lg a - \lg b)$

Remark: For continuous-space MLSI $\Longleftrightarrow$ LSI, but absolutely NOT in discrete space.

Comparison of Markov chains

Suppose we have two chains P, P' that "look similar" and have same stationary dist μ, but we can only analyze P. How do we transfer to P'.

Example. (Coloring) Metropolis $\longleftrightarrow$ Glauber

Idea 1: Suppose $P(x,y) \overset{c}{\simeq} P'(x,y)$ for $x \neq y$. (multiplicative)

For colorings $\frac{1}{qn} \simeq \frac{1}{n\cdot(q-t)}$ up to $\frac{q}{q-\Delta}$.

Metropolis; Glauber; $t \in [0,\Delta]$

Comparison method (easy version)

$$\mathcal{E}\left(f'\left(\frac{\nu}{\mu}\right), \frac{\nu}{\mu}\right) =$$

$$\frac{1}{2}\sum_{x,y} Q(x,y)\underbrace{\left(f'\left(\frac{\nu(x)}{\mu(x)}\right) - f'\left(\frac{\nu(y)}{\mu(y)}\right)\right)\left(\frac{\nu(x)}{\mu(x)} - \frac{\nu(y)}{\mu(y)}\right)}_{\geq 0 \text{ because } f' \text{ is monotone}}$$

Thm: If $P \overset{c}{\simeq} P'$ and P, P' have the same stationary dist, for any D_f,

$$\rho(P) \overset{c}{\simeq} \rho(P')$$

continuous time contraction factors

Comparison method (hard version)

What if P, P' are not this similar?

Routing: Suppose we have dist π over paths $x_0 \sim x_1 \sim x_2 \sim \cdots \sim x_\ell$, s.t. $(x_0, x_\ell) \sim Q$

ergodic flow for P

We call π a routing.

Congestion: $\dfrac{\mathbb{P}_{\text{path}\sim\pi}[x \sim y \text{ in path}]}{Q'(x,y)}$

think of as capacities.

Length: Maximum ℓ.

Low congestion + Low length $\Longrightarrow$

$$c \cdot \mathcal{E}_P(g,g) \leqslant \mathcal{E}_{P'}(g,g)$$

have to be equal!