- Path Coupling: \( W(1_x P, 1_x' P) \leq (1 - c) W(1_x, 1_x') \) 
  only adjacent states

* Coloring: \( q > 2 \Delta \Rightarrow t_{mix}^{(\text{Metropolis})} = O(n \log n) \)

- Dobrushin: \( \mu \) on \( \mathcal{L}_1 x \ldots x \Omega_n \)
  \[ I[j \rightarrow i] = \max \left\{ \sum_{x, x'} d(M_i | x_j, x'_j) \mid x \sim y, x' \sim y' \right\} \]
  Thm: If columns of \( I \) sum to \( 1 - 8 \)
  \[ \Rightarrow t_{mix}^{(\text{Glauber})} = O(n \log n / 8) \]

* Coloring: \( q > 2 \Delta \Rightarrow t_{mix} = O(\Delta n \log n) \)
* Hardcore: \( \lambda \leq \frac{1 - 8}{\Delta} \Rightarrow t_{mix} = O(\Delta n \log n) \)
* Ising: \( \sum_i |\beta_i| \leq 1 - 8 \Rightarrow t_{mix} = O(\Delta n \log n) \)

- Dobrushin \( \mu \):
  \[ C \in \mathbb{R}^n_+ \text{ with } C I \leq (1 - 8) C \]
  \[ \Rightarrow W(u P, v' P) \leq (1 - \frac{\lambda_{\text{max}}(I)}{\lambda_{\text{max}}(C)}) \]

- Spectral Analysis

- Time-Reversible: eigenvalues are real
  \[ -1 \leq \lambda_n \leq \ldots \leq \lambda_2 \leq \lambda_1 = 1 \]
  - Spectral Gap: \( 1 - \lambda_2 \) or \( 1 - \max(\lambda_2, \lambda_{n-1}) \)
  - Relaxation Time: \( t_{rel} = \frac{1}{1 - \lambda_2} \) aka proxy for \( t_{mix} \)

Plan for Today

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

\[ \mu P = \mu \]
Eigenvalues & Mixing Time

Suppose \( \lambda P = \lambda v \) for \( \lambda \neq 0 \in \mathbb{C} \) v \in \mathbb{C}^n.

\[ \lambda v^t = \lambda v P^t = v \lambda^t \Rightarrow v \lambda^t = 0 \]

Def: \( \| v \|_1 = \sum_i |\text{Re}(v_i)| + |\text{Im}(v_i)| \)

Claim: \( t > t_{\text{mix}}(\varepsilon) \Rightarrow \| v P^t \|_1 \leq O(\varepsilon) \| v \|_1 \)

Proof: Let \( v = \alpha v_1 + \alpha v_2 + i \beta v_3 + i \beta v_4 \) for

\( v_1, v_2 \) disjoint support
\( v_3, v_4 \) disjoint support

\[ \Rightarrow \| v \|_1 = 2\alpha + 2\beta \]

\[ \Rightarrow \| v P^t \|_1 \leq \alpha O(\varepsilon) + \beta O(\varepsilon) = O(\varepsilon) \| v \|_1 \]

Def: \( \| v \|_1' = \sqrt{\text{Re}(v^t) v + \text{Im}(v^t) v} \)

Now if \( t > t_{\text{mix}}(\varepsilon) \) we have

\[ \| \lambda \|_1' \leq O(\varepsilon) \| v \|_1' \]

\[ \Rightarrow \| \lambda \|_1' = O(\varepsilon) \]

* Corollary: \( |\lambda| < 1 \) holds even when for eigenvectors other than \( \mu \) if ergodic.

* Corollary: \( 1 - |\lambda| \geq \Omega(\frac{1}{t_{\text{mix}}(\varepsilon)}) \)

\[ |\lambda| \leq 1 - \Omega(\frac{1}{t_{\text{mix}}(\varepsilon)}) \text{ (usually not tight)} \]
Intro to Functional Analysis

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

Idea: Contraction of proxy for $d_{TV}$

Def: ($f$-Divergence) convex function $f$:

$D_f(v \parallel \mu) := (\mathbb{E}_{x \sim \mu}^v[f(x)]) - f(\mathbb{E}_{x \sim \mu}^{\frac{v}{\mu}})$

Proxy for $d_{TV}(v \parallel \mu)$

Note: We define it for any $v, \mu \in \mathbb{R}$ where $v/\mu$ takes values in domain of $f$.

Fact: $f$ convex $\Rightarrow D_f(v \parallel \mu) \geq 0$

Proof: Jensen's inequality! $\square$

Remark: $f$ strongly convex $\Rightarrow D_f = 0 \iff v = \text{const. } \mu$.
Fact: Suppose \( \mathbb{N} \in \mathbb{R}^{d \times d} \) is row stochastic.

Then \( D_f(\mathbb{N} || \mathbb{M}) \leq D_f(\mathbb{M}) \) 

Data processing ineq.

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

\[
\begin{align*}
\frac{\mathbb{N}(y)}{\mu(y)} &= \frac{\sum_x \mathbb{N}(x,y)}{\sum_x \mu(x) \mathbb{N}(x,y)} = \sum_x \frac{\pi(x,y)}{\sum_y \pi(x,y)} \frac{\mathbb{N}(x)}{\mu(x)} \\
\Rightarrow f\left(\frac{\mathbb{N}(y)}{\mu(y)}\right) &\leq \sum_x \pi(x,y) f\left(\frac{\mathbb{N}(x)}{\mu(x)}\right) \Rightarrow \\
\mathbb{E}_{y \sim \mu_N} \left[ f\left(\frac{\mathbb{N}(y)}{\mu_N(y)}\right) \right] &\leq \sum_x \mathbb{E}_{y \sim \mu_N} \left[ \pi(x,y) \right] f\left(\frac{\mathbb{N}(x)}{\mu(x)}\right) = \\
\mathbb{E}_{x \sim \mu} \left[ f\left(\frac{\mathbb{N}(x)}{\mu(x)}\right) \right].
\end{align*}
\]

Note that

\[
\begin{align*}
\mathbb{E}_{x \sim \mu} \left[ \frac{\mathbb{N}(y)}{\mu_N(y)} \right] &= \sum_y \mathbb{N}(y) = \sum_x \mathbb{N}(x) \sum_y \mathbb{N}(x,y) \\
\mathbb{E}_{y \sim \mu_N} \left[ \frac{\mathbb{N}(x)}{\mu(x)} \right].
\end{align*}
\]

Therefore \( D_f(\mathbb{M}) \geq D_f(\mathbb{N} || \mathbb{M}) \). \( \square \)

Remark: This is extremely useful for Markov chains constructed as \( P = \mathbb{N} \).

Contraction of either \( \mathbb{N} \) or \( \mathbb{N}^\circ \) \( \Rightarrow \) contraction of \( P \).

In lots of scenarios \( \mathbb{N} \) by itself easier to analyze!

Remark: We usually want \( f \) such that \( d_{TV} \leq \text{some func of } D_f \).
Popular Choice 1: \( f(x) = x^2 \)
\[
D_f (\nu \| \mu) = \chi^2 (\nu \| \mu)
\]
Also called \( \text{Var}_\mu \left[ \frac{\nu}{\mu} \right] \) variance

- Note that \( \nu \) can take \( <0 \) values too here.

- Alternative formula:
\[
\chi^2 (\nu \| \mu) = E \left[ (\frac{\nu(x)}{\mu(x)} - \frac{\nu(x)}{\mu(x)} )^2 \right]
\]

Thm: For \( \nu \) a dist, we have
\[
d_{TV} (\nu, \mu) \leq \frac{1}{2} \sqrt{\chi^2 (\nu \| \mu)}
\]
Proof: Apply Cauchy-Schwarz:
\[
2d_{TV} (\nu, \mu) = \sum_{x} |\nu(x) - \mu(x)| = E \left[ \frac{\nu(x)}{\mu(x)} - 1 \right] \leq \sqrt{\chi^2 (\nu \| \mu)}
\]

Popular Choice 2: \( f(x) = x \log x \) for \( x > 0 \)
\[
D_f (\nu \| \mu) = D_{KL} (\nu \| \mu)
\]
Also called \( \text{Ent}_\mu \left[ \frac{\nu}{\mu} \right] \) entropy

- Note that \( \nu \) can take \( \geq 0 \) values but doesn't need to sum to 1.

- Formula for dist \( \nu \): \( D_{KL} (\nu \| \mu) = E \left[ \log \frac{\nu(x)}{\mu(x)} \right] \)

Thm (Pinzer's): For \( \nu \) 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 \( \nu(x) < \mu(x) \):
\[
d_{TV} (\nu N, \mu N) = d_{TV} (\nu, \mu), \quad D_{KL} (\nu \| \mu) \leq D_{KL} (\nu N \| \mu N)
\]
\[
\Rightarrow \quad \text{Enough to prove on } \mathcal{S} = \{0,1\}^\mathbb{R}
\]
The rest on HW.
Functional Analysis:

Show contraction of some $D_f$:

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

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

These inequalities are related to Poincare & MLSI

future: cont. time

Spectral gap $\iff$ Contraction of $\nu^2$

Suppose $N \in \mathbb{R}^{n \times D}$ is row-stochastic and $\mu$ dist on $\mathcal{X}$. Let $\mu' = \mu N$ and $N^0$ be time-reversal of $N$ w.r.t $\mu$.

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

$$D_{\mu'} N^0 = (D_{\mu} N)^T = N^T D_{\mu}$$

Formula for $\nu^2$: Suppose $\nu$ sums to 0.

$$\nu^2(v || \mu) = \| \nu D_{\mu}^{-\frac{1}{2}} \|^2$$
For such a \( \nu \) we have \( (\nu N)^t = 0 \) too, so 
\[
\chi^2(\nu N^t\nu N) = \| \nu N D_{\mu}^{-1/2} \|_2^2.
\]
What is
\[
\sup \left\{ \frac{\| \nu N D_{\mu}^{-1/2} \|_2^2}{\| \nu D_{\mu}^{-1/2} \|_2^2} \bigg| \nu \| \nu^t = 0 \right\}?
\]
This is equal to \( \lambda_2(\nu N^t\nu N) \).

[If time --- ]

Contraction for arbitrary \( \nu \) \( \Longleftrightarrow \) Contraction for \( \nu \| \nu^t = 0 \)

\[\text{P-\text{variance}} = 1 - \lambda_2(\nu N^t\nu N)\]

**Corollary:** Time-reversible:
\[
\lambda_2(P^2) = \max(\lambda_2(\nu N^t\nu N), \lambda_2(\nu N^t\nu N))
\]

**Proof:** Let \( \nu = \nu D_{\mu}^{-1/2} \)

\( \text{change of variable} \)

Let \( A = D_{\mu}^{1/2} N D_{\mu}^{-1/2} \). Then

\[
\sup \left\{ \frac{\nu A^T A \nu}{\nu \| \nu^t} \bigg| \nu D_{\mu}^{1/2} \| \nu^t = 0 \right\}.
\]

Note that
\[
A A^T = D_{\mu}^{1/2} N D_{\mu}^{-1} N^T D_{\mu} D_{\mu}^{-1/2} = D_{\mu}^{1/2} N N^0 D_{\mu}^{-1/2}
\]
is a symmetric matrix and similar to \( NN^0 \), so has same eigenvalues.

\[
A A^T D_{\mu}^{1/2} = D_{\mu}^{1/2} \nu \}
\]
eigenvector for eigenvalue 1.

So we are looking at
\[
\sup \left\{ \frac{\nu A A^T \nu}{\nu \| \nu^t} \bigg| \nu \text{ orthogonal to top} \right\}
\]

\[= \lambda_2(A A^T) = \lambda_2(\nu N^t\nu N) \]

\( \smile \)
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 \( \Pi \) over \( G \).

Think: Sparse support.

We get an Abelian walk:

\[ X \rightarrow X + \mathbb{Z} \]

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

\[ \Pi: \sum_{\omega} \frac{1}{2} \begin{cases} \text{choose } (\omega, \omega) & \text{if } \omega \neq 0 \text{ or } \omega \neq 0 \text{ mod } 2 \text{ \#sup. } \\
\text{choose } (-\omega, -\omega) & \text{otherwise} \end{cases} \]

This is Glauber dynamics.

Example. (Cycle)

\[ \Pi: \sum_{\omega} \frac{1}{2} \begin{cases} \text{choose } +1 & \text{if } \omega \neq 0 \text{ \#sup. } \\
\text{choose } -1 & \text{otherwise} \end{cases} \]

Fact 1: \( \mu = \text{uniform always stationary} \)

Fact 2: \( \Pi \) symmetric \( \iff \) time-reversible

\[ \Pi(2) = \Pi(-2) \]

Fact 3: Support \( \Pi \) generate \( G \) \( \iff \) irreducible.

Thm: Regardless of \( \Pi \), eigenvectors are fixed; they are characters of \( G \).

Def: (Character): \( \chi: G \rightarrow \mathbb{C}^\ast \) s.t.

\[ \chi(x+y) = \chi(x) \chi(y) \]
Example: \((\mathbb{Z}_n)\):

Pick \(w \in \mathbb{C}\) s.t. \(w^n = 1\).

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

All characters are
\[\chi(x) = \exp\left(\frac{2\pi i \cdot t x}{n}\right) = w^{tx}\]

Characters of \(\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}\)
\[\chi(x_1, \ldots, x_k) = w^{tx_1} \cdots w^{tx_k}\]

There are exactly \(n_1 \cdots n_k\) such characters.

Thm: If \(P\) is an Abelian walk, every character \(\chi\) is an eigenvector.

\[(\chi P)(y) = \sum_{x \in G} x(x) P(x, y) = \sum_{x \in G} \chi(x) \Pi(y - x) = \sum_{z \in G} \Pi(z) \chi(y - z)\]

\[= \left(\sum_{z \in G} \Pi(z) \chi(z)^{-1}\right) \chi(y)\]

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