CS 263: Counting and Sampling

Nima Anari

slides for

Continuous Time
For function $\phi$ and $f : \Omega \to \mathbb{R}$ define

$$\text{Ent}^\phi_{\mu}[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$
**Review**

**φ-entropy**
For function \( \phi \) and \( f : \Omega \rightarrow \mathbb{R} \) define

\[
\text{Ent}^{\phi}_{\mu}[f] = \mathbb{E}_{\mu}[\phi \circ f] - \phi(\mathbb{E}_{\mu}[f]).
\]

**φ-divergence**
For measure \( \nu \) and dist \( \mu \) define

\[
\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}^{\phi}_{\mu}\left[\frac{\nu}{\mu}\right].
\]
Review

**Φ-entropy**
For function $\phi$ and $f : \Omega \rightarrow \mathbb{R}$ define

$$\text{Ent}_\mu^\phi[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$

**Φ-divergence**
For measure $\nu$ and dist $\mu$ define

$$\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}_\mu^\phi\left[\frac{\nu}{\mu}\right]$$

▶ $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
### Review

| **ϕ-entropy** |  
|---|---|
| For function $\phi$ and $f : \Omega \rightarrow \mathbb{R}$ define |
| \[
\text{Ent}^\phi_\mu[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).
\] |

| **ϕ-divergence** |  
|---|---|
| For measure $\nu$ and dist $\mu$ define |
| \[
\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}^\phi_\mu \left[ \frac{\nu}{\mu} \right]
\] |

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
- $x^2$ leads to $\text{Var}$ and $\chi^2$
Review

**φ-entropy**

For function $\phi$ and $f: \Omega \to \mathbb{R}$ define

$$\text{Ent}_\mu^\phi[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$

**φ-divergence**

For measure $\nu$ and dist $\mu$ define

$$\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}_\mu^\phi\left[\frac{\nu}{\mu}\right].$$

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
- $x^2$ leads to Var and $\chi^2$
- $x \log x$ leads to Ent and $\mathcal{D}_{KL}$
Review

\( \phi \)-entropy
For function \( \phi \) and \( f : \Omega \to \mathbb{R} \) define

\[
\text{Ent}_\mu^{\phi}[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).
\]

\( \phi \)-divergence
For measure \( \nu \) and dist \( \mu \) define

\[
\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}_\mu^{\phi}\left[\frac{\nu}{\mu}\right]
\]

- \( \phi(x) = x^2 \) vs. \( \phi(x) = x \log x \)
- \( x^2 \) leads to Var and \( \chi^2 \)
- \( x \log x \) leads to Ent and \( \mathcal{D}_{KL} \)

Contraction:

\[
\mathcal{D}_\phi(\nu_N \parallel \mu_N) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu)
\]
**φ-entropy**

For function $\phi$ and $f : \Omega \rightarrow \mathbb{R}$ define

$$\text{Ent}^\phi_\mu[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$

**φ-divergence**

For measure $\nu$ and dist $\mu$ define

$$D_\phi(\nu \parallel \mu) = \text{Ent}^\phi_\mu[\frac{\nu}{\mu}]$$

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
- $x^2$ leads to Var and $\chi^2$
- $x \log x$ leads to Ent and $D_{KL}$

**Contraction:**

$$D_\phi(\nuN \parallel \muN) \leq (1 - \rho) D_\phi(\nu \parallel \mu)$$

**Lemma: data processing**

$$D_\phi(\nuN \parallel \muN) \leq D_\phi(\nu \parallel \mu)$$
**Review**

**ϕ-entropy**
For function $\phi$ and $f : \Omega \rightarrow \mathbb{R}$ define

$$\text{Ent}_{\mu}^{\phi}[f] = \mathbb{E}_{\mu}[\phi \circ f] - \phi(\mathbb{E}_{\mu}[f]).$$

**ϕ-divergence**
For measure $\nu$ and dist $\mu$ define

$$\mathcal{D}_{\phi}(\nu \| \mu) = \text{Ent}_{\mu}^{\phi}\left[\frac{\nu}{\mu}\right].$$

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
- $x^2$ leads to $\text{Var}$ and $\chi^2$
- $x \log x$ leads to $\text{Ent}$ and $\mathcal{D}_{KL}$

**Contraction:**

$$\mathcal{D}_{\phi}(\nuN \| \muN) \leq (1 - \rho) \mathcal{D}_{\phi}(\nu \| \mu)$$

**Lemma: data processing**

$$\mathcal{D}_{\phi}(\nuN \| \muN) \leq \mathcal{D}_{\phi}(\nu \| \mu)$$

- Specializing to $\chi^2$, we have
  $$\rho = 1 - \lambda_2(\mathbb{NN}^\circ)$$

**Abelian walks on group $G$:**

$x \mapsto x + z$ sampled i.i.d. from $\pi$

Eigvecs are characters $\chi$:

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

Eigvals are $\mathbb{E}_{\pi}[\chi]$
**Review**

<table>
<thead>
<tr>
<th><strong>(\phi)-entropy</strong></th>
<th><strong>(\phi)-divergence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>For function (\phi) and (f : \Omega \to \mathbb{R}) define</td>
<td>For measure (\nu) and dist (\mu) define</td>
</tr>
</tbody>
</table>
| \[
\text{Ent}_\mu^\phi[f] = E_\mu[\phi \circ f] - \phi(E_\mu[f]).
\] | \[
\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}_\mu^\phi\left[\frac{\nu}{\mu}\right]
\] |

**Contraction:**
\[
\mathcal{D}_\phi(\nu_N \parallel \mu_N) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu)
\]

**Lemma: data processing**
\[
\mathcal{D}_\phi(\nu_N \parallel \mu_N) \leq \mathcal{D}_\phi(\nu \parallel \mu)
\]

- Specializing to \(\chi^2\), we have \(\rho = 1 - \lambda_2(\mathbb{N}^\circ)\)
- Abelian walks on group \(G\):
  \[
  x \mapsto x + z \quad \text{sampling i.i.d. from} \ \pi
  \]

- \(\phi(x) = x^2\) vs. \(\phi(x) = x \log x\)
- \(x^2\) leads to \(\text{Var}\) and \(\chi^2\)
- \(x \log x\) leads to \(\text{Ent}\) and \(\mathcal{D}_{KL}\)
**Review**

### \(\phi\)-entropy

For function \(\phi\) and \(f : \Omega \to \mathbb{R}\) define

\[
\text{Ent}_{\mu}^{\phi}[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).
\]

### \(\phi\)-divergence

For measure \(\nu\) and dist \(\mu\) define

\[
\mathcal{D}_{\phi}(\nu \parallel \mu) = \text{Ent}_{\mu}^{\phi}\left[\frac{\nu}{\mu}\right]
\]

- \(\phi(x) = x^2\) vs. \(\phi(x) = x \log x\)
- \(x^2\) leads to \(\text{Var}\) and \(\chi^2\)
- \(x \log x\) leads to \(\text{Ent}\) and \(\mathcal{D}_{KL}\)

### Lemma: data processing

\[
\mathcal{D}_{\phi}(\nu_N \parallel \mu_N) \leq (1 - \rho) \mathcal{D}_{\phi}(\nu \parallel \mu)
\]

- Specializing to \(\chi^2\), we have
  \[
  \rho = 1 - \lambda_2(\mathbb{N} \circ \mathbb{N})
  \]
- Abelian walks on group \(G\):
  \[
  x \mapsto x + z \quad \text{sampled i.i.d. from } \pi
  \]
- Eigvecs are characters \(\chi\):
  \[
  \chi(x + y) = \chi(x)\chi(y)
  \]
Review

**φ-entropy**
For function $\phi$ and $f : \Omega \to \mathbb{R}$ define
\[
\text{Ent}^\phi_{\mu}[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).
\]

**φ-divergence**
For measure $\nu$ and dist $\mu$ define
\[
\mathcal{D}_\phi(\nu \parallel \mu) = \text{Ent}^\phi_{\mu}\left[\frac{\nu}{\mu}\right]
\]

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
- $x^2$ leads to $\text{Var}$ and $\chi^2$
- $x \log x$ leads to $\text{Ent}$ and $\mathcal{D}_{\text{KL}}$

**Contraction:**
\[
\mathcal{D}_\phi(\nu_N \parallel \mu_N) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu)
\]

**Lemma: data processing**
\[
\mathcal{D}_\phi(\nu_N \parallel \mu_N) \leq \mathcal{D}_\phi(\nu \parallel \mu)
\]

- Specializing to $\chi^2$, we have
  \[
  \rho = 1 - \lambda_2(\mathbb{NN}^\circ)
  \]
- Abelian walks on group $G$: $x \mapsto x + z$ sampled i.i.d. from $\pi$
- Eigvecs are characters $\chi$:
  \[
  \chi(x + y) = \chi(x)\chi(y)
  \]
- Eigvals are $\mathbb{E}_\pi[\chi]$
Fourier Analysis
- Characters
- Examples
- Relaxation time

Continuous Time
- Functional analysis in continuous time
- Dirichlet form
Fourier Analysis

- Characters
- Examples
- Relaxation time

Continuous Time

- Functional analysis in continuous time
- Dirichlet form
We know characters of $\mathbb{Z}_n$:

$$\chi(x) = \omega^x$$

for $\omega$ an $n$-th root of unity.
We know characters of $\mathbb{Z}_n$:

$$\chi(x) = \omega^x$$

for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.
- We know characters of $\mathbb{Z}_n$:
  $$\chi(x) = \omega^x$$
  for $\omega$ an $n$-th root of unity.

- There are exactly $n$ of them.

- Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
  $$x \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$
We know characters of $\mathbb{Z}_n$:
\[ \chi(x) = \omega^x \]
for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.

Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
\[ x \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}. \]

For $G$, we get $|G|$ characters.
We know characters of $\mathbb{Z}_n$:
$$\chi(x) = \omega^x$$
for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.

Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
$$\chi \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$

For $G$, we get $|G|$ characters.

We just need to compute
$$E_{x \sim \pi} [\omega_1^{x_1} \cdots \omega_k^{x_k}]$$
for all of these characters.
We know characters of $\mathbb{Z}_n$:

$$\chi(x) = \omega^x$$

for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.

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

$$\chi \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$

For $G$, we get $|G|$ characters.

We just need to compute

$$E_{x \sim \pi} [\omega_1^{x_1} \cdots \omega_k^{x_k}]$$

for all of these characters.

The $\omega_1 = \cdots = \omega_k = 1$ character gives us the special 1 eigval.

If $P$ is Abelian walk, then $P \circ P^\top$ is also Abelian walk. Eigvals are

$$E_{x \sim \pi} [\chi(-x)]$$

Since $P$ and $P \circ P^\top$ commute, we have

$$\lambda_k(P \circ P^\top) = |\lambda_k(P)|^2$$

Mixing: largest $|\cdot|$ of an eig?
We know characters of $\mathbb{Z}_n$:
\[
\chi(x) = \omega^x
\]
for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.

Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
\[
\chi \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.
\]
For $G$, we get $|G|$ characters.

We just need to compute
\[
\mathbb{E}_{x\sim\pi} [\omega_1^{x_1} \cdots \omega_k^{x_k}]
\]
for all of these characters.

The $\omega_1 = \cdots = \omega_k = 1$ character gives us the special 1 eigval.

If $P$ is Abelian walk, then $P^\circ = P^T$ is also Abelian walk. Eigvals are
\[
\mathbb{E}_{x\sim\pi} [\chi(-x)]
\]
Mixing: largest $|\cdot|$ of an eig?
We know characters of $\mathbb{Z}_n$:

$$\chi(x) = \omega^x$$

for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.

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

$$\chi \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$  

For $G$, we get $|G|$ characters.

We just need to compute

$$\mathbb{E}_{x \sim \pi} [\omega_1^{x_1} \cdots \omega_k^{x_k}]$$

for all of these characters.

The $\omega_1 = \cdots = \omega_k = 1$ character gives us the special 1 eigval.

If $P$ is Abelian walk, then $P^\circ = P^T$ is also Abelian walk. Eigvals are

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

Since $P$ and $P^\circ$ commute, we have

$$\lambda_k(PP^\circ) = |\lambda_k(P)|^2$$
We know characters of $\mathbb{Z}_n$:

$$\chi(x) = \omega^x$$

for $\omega$ an $n$-th root of unity.

There are exactly $n$ of them.

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

$$\chi \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$  

For $G$, we get $|G|$ characters.

We just need to compute

$$E_{x \sim \pi}[\omega_1^{x_1} \cdots \omega_k^{x_k}]$$

for all of these characters.

The $\omega_1 = \cdots = \omega_k = 1$ character gives us the special 1 eigval.

If $P$ is Abelian walk, then $P^\circ = PT$ is also Abelian walk. Eigvals are

$$E_{x \sim \pi}[\chi(-x)]$$

Since $P$ and $P^\circ$ commute, we have

$$\lambda_k(PP^\circ) = |\lambda_k(P)|^2$$

Mixing: largest $|\cdot|$ of an eig?
Example: hypercube

Distribution $\pi$:
- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2^n$

Eigval is $\#\{+1\}/n$ ($n$-th root of $n$)

Spectral gap: $1 - (n-1)/n = 1/n$
$t_{\text{mix}} \leq O(n^2)$

Example: cycle

Distribution $\pi$:
- $+$ w.p. $1/2$
- $-$ w.p. $1/2$

There are $n$ characters: $x \mapsto \omega^x$

Eigval is $($$\omega + \omega - 1$$)/2$

Eigvals are $\cos(2\pi k/n)$

Spectral gap: $1 - \cos(2\pi/n) \approx \Theta(1/n^2)$
$t_{\text{mix}} \leq O(n^2 \log n)$

Not for even $n$. 
Example: hypercube

Distribution $\pi$:

- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2n$

There are $2^n$ characters:

$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
Example: hypercube

Distribution $\pi$:
- 0 w.p. $1/2$
- $1_i$ w.p. $1/2n$

- There are $2^n$ characters:
  \[ x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n} \]

- Eigval is $\#\{+1\}/n$
Example: hypercube

Distribution $\pi$:

- 0 w.p. 1/2
- $1_i$ w.p. 1/2n

There are $2^n$ characters:

$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$

- Eigval is $\#\{+1\}/n$
- $(\binom{n}{k})$ of eigvals are $k/n$

Spectral gap:

$$1 - \left( \frac{n-1}{n} \right) = \frac{1}{n}$$

$t_{\text{mix}} \leq O(n^2)$

Example: cycle

Distribution $\pi$:

+1 w.p. 1/2
-1 w.p. 1/2

There are $n$ characters:

$$x \mapsto \omega^x$$

- Eigvals are $\cos\left(\frac{2\pi k}{n}\right)$
- Spectral gap:
  $$1 - \cos\left(\frac{2\pi}{n}\right) \approx \Theta\left(\frac{1}{n^2}\right)$$
  $$t_{\text{mix}} \leq O(n^2 \log n)$$

Not for even $n$. 

Example: hypercube

Distribution $\pi$:
- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2n$

There are $2^n$ characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$

Eigval is $\#\{+1\}/n$

$(\binom{n}{k})$ of eigvals are $k/n$

Spectral gap:
$$1 - (n - 1)/n = 1/n$$
Example: hypercube

Distribution $\pi$:
- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2n$

There are $2^n$ characters:

$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$

Eigval is $\#\{+1\}/n$

$(\binom{n}{k})$ of eigvals are $k/n$

Spectral gap:

$$1 - (n - 1)/n = 1/n$$

$t_{mix} \leq O(n^2)$
Example: hypercube

Distribution $\pi$:
- 0 w.p. $1/2$
- $1_i$ w.p. $1/2n$

There are $2^n$ characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$

Eigval is $\#\{+1\}/n$

$\binom{n}{k}$ of eigvals are $k/n$

Spectral gap:
$$1 - (n - 1)/n = 1/n$$

$t_{mix} \leq O(n^2)$

Example: cycle

Distribution $\pi$:
- +1 w.p. $1/2$
- −1 w.p. $1/2$

There are $n$ characters:
$$x \mapsto \omega^x$$

Eigval is $(\omega + \omega - 1)/2$

Eigvals are $\cos(2\pi k/n)$

Spectral gap:
$$1 - \cos(2\pi/n) \approx \Theta(1/n^2)$$

$t_{mix} \leq O(n^2 \log n)$

Not for even $n$. 
### Example: hypercube

<table>
<thead>
<tr>
<th>Distribution $\pi$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ 0 w.p. $1/2$</td>
</tr>
<tr>
<td>▶ $1_i$ w.p. $1/2n$</td>
</tr>
</tbody>
</table>

There are $2^n$ characters:

$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$

- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are $k/n$
- Spectral gap:
  $$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leq O(n^2)$

### Example: cycle

<table>
<thead>
<tr>
<th>Distribution $\pi$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ $+1$ w.p. $1/2$</td>
</tr>
<tr>
<td>▶ $-1$ w.p. $1/2$</td>
</tr>
</tbody>
</table>

There are $n$ characters:

$$x \mapsto \omega^x$$

- Eigval is $(\omega + \omega - 1)/2$
- Eigvals are $\cos(2\pi k/n)$
- Spectral gap:
  $$1 - \cos(2\pi/n) \approx \Theta(1/n^2)$$
- $t_{\text{mix}} \leq O(n^2 \log n)$

Not for even $n$. 

---
### Example: hypercube

**Distribution $\pi$:**
- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2n$

- There are $2^n$ characters:
  $$ x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n} $$
- Eigval is $\#\{+1\}/n$
- $(\binom{n}{k})$ of eigvals are $k/n$
- Spectral gap:
  $$ 1 - (n - 1)/n = 1/n $$
- $t_{\text{mix}} \leq O(n^2)$

### Example: cycle

**Distribution $\pi$:**
- $+1$ w.p. $1/2$
- $-1$ w.p. $1/2$

- There are $n$ characters:
  $$ x \mapsto \omega^x $$
- Eigvals are $\cos(2\pi k/n)$
- Spectral gap:
  $$ 1 - \cos(2\pi/n) \approx \Theta(1/n^2) $$
- $t_{\text{mix}} \leq O(n^2 \log n)$

*Note: Not for even $n$.\*
Example: hypercube

Distribution $\pi$:
- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2n$

There are $2^n$ characters:

$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$

Eigval is $\#\{+1\}/n$

$\binom{n}{k}$ of eigvals are $k/n$

Spectral gap:

$1 - (n - 1)/n = 1/n$

$t_{mix} \leq \Theta(n^2)$

Example: cycle

Distribution $\pi$:
- $+1$ w.p. $1/2$
- $-1$ w.p. $1/2$

There are $n$ characters:

$x \mapsto \omega^x$

Eigval is $(\omega + \omega^{-1})/2$

Spectral gap:

$1 - \cos(2\pi/n) \approx \Theta(1/n^2)$

$t_{mix} \leq \Theta(n^2 \log n)$

Not for even $n$. 
Example: hypercube

Distribution $\pi$:
- 0 w.p. $1/2$
- $1_i$ w.p. $1/2n$

- There are $2^n$ characters:
  $x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are $k/n$
- Spectral gap:
  $1 - (n - 1)/n = 1/n$
- $t_{mix} \leq O(n^2)$

Example: cycle

Distribution $\pi$:
- $+1$ w.p. $1/2$
- $-1$ w.p. $1/2$

- There are $n$ characters:
  $x \mapsto \omega^x$
- Eigval is $(\omega + \omega^{-1})/2$
- Eigvals are $\cos(2\pi k/n)$
**Example: hypercube**

Distribution $\pi$:
- $0$ w.p. $1/2$
- $1_i$ w.p. $1/2n$

- There are $2^n$ characters:
  \[ x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n} \]
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are $k/n$
- Spectral gap:
  \[ 1 - (n - 1)/n = 1/n \]
- $t_{mix} \leq O(n^2)$

**Example: cycle**

Distribution $\pi$:
- $+1$ w.p. $1/2$
- $-1$ w.p. $1/2$

- There are $n$ characters:
  \[ x \mapsto \omega^x \]
- Eigval is $\frac{\omega + \omega^{-1}}{2}$
- Eigvals are $\cos(2\pi k/n)$
- Spectral gap:
  \[ 1 - \cos(2\pi/n) \simeq \Theta(1/n^2) \]
Example: hypercube

Distribution $\pi$:
- 0 w.p. 1/2
- $1_i$ w.p. 1/2n

There are $2^n$ characters:
$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$

Eigval is $\#\{+1\}/n$

$n \choose k$ of eigvals are $k/n$

Spectral gap:
$1 - (n - 1)/n = 1/n$

$t_{\text{mix}} \leq O(n^2)$

Example: cycle

Distribution $\pi$:
- +1 w.p. 1/2
- -1 w.p. 1/2

There are $n$ characters:
$x \mapsto \omega^x$

Eigval is $(\omega + \omega^{-1})/2$

Eigvals are $\cos(2\pi k/n)$

Spectral gap:
$1 - \cos(2\pi/n) \simeq \Theta(1/n^2)$?

$t_{\text{mix}} \leq O(n^2 \log n)$?
Example: hypercube

Distribution $\pi$:
- $0$ w.p. 1/2
- $I_i$ w.p. 1/2n

There are $2^n$ characters:
$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$

Eigval is $\#\{+1\}/n$
$(\binom{n}{k})$ of eigvals are $k/n$

Spectral gap:
$1 - (n - 1)/n = 1/n$

$t_{mix} \leq O(n^2)$

Example: cycle

Distribution $\pi$:
- $+1$ w.p. 1/2
- $-1$ w.p. 1/2

There are $n$ characters:
$x \mapsto \omega^x$

Eigval is $(\omega + \omega^{-1})/2$

Eigvals are $\cos(2\pi k/n)$

Spectral gap:
$1 - \cos(2\pi/n) \simeq \Theta(1/n^2)$?

$t_{mix} \leq O(n^2 \log n)$? Not for even $n$. 
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

Relaxation time: $1/(1 - \lambda_2(P))$
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

▶ Relaxation time: $1/(1 - \lambda_2(P))$

▶ Relaxation time does not directly control mixing time 😞
Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- Relaxation time: $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$.
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- **Relaxation time:** $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

**Lemma**

$$t_{\text{mix}}(\epsilon) = O\left(\frac{\log(\chi^2(\nu_0 || \mu)) + \log(1/\epsilon)}{1 - \lambda_2(P)}\right)$$
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- Relaxation time: $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

**Lemma**

$$t_{\text{mix}}(\epsilon) = O\left(\frac{\log(\chi^2(\nu_0 \parallel \mu)) + \log(1/\epsilon)}{1 - \lambda_2(P)}\right)$$

- We have

$$t_{\text{rel}} = \Theta\left(\lim_{\epsilon \to 0} \frac{t_{\text{mix}}(\epsilon)}{\log(1/\epsilon)}\right)$$
Relaxation time

Suppose $P$ is time-reversible and lazy:

\[ \lambda_1(P) \geq 0 \]

- **Relaxation time:** $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

\[
t_{\text{mix}}(\epsilon) = O \left( \frac{\log(\chi^2(\nu_0\|\mu)) + \log(1/\epsilon)}{1 - \lambda_2(P)} \right)
\]

- We have

\[
t_{\rel} = \Theta \left( \lim_{\epsilon \to 0} \frac{t_{\text{mix}}(\epsilon)}{\log(1/\epsilon)} \right)
\]
Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- **Relaxation time:** $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

### Lemma

$$t_{\text{mix}}(\epsilon) = O\left(\frac{\log\left(\chi^2(\nu_0 \| \mu)\right) + \log(1/\epsilon)}{1 - \lambda_2(P)}\right)$$

- We have

$$t_{\text{rel}} = \Theta\left(\lim_{\epsilon \to 0} \frac{t_{\text{mix}}(\epsilon)}{\log(1/\epsilon)}\right)$$
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- **Relaxation time:** $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

**Lemma**

$$t_{\text{mix}}(\epsilon) = O\left(\frac{\log(\chi^2(\nu_0\|\mu)) + \log(1/\epsilon)}{1 - \lambda_2(P)}\right)$$

- We have $t_{\text{rel}} = \Theta\left(\lim_{\epsilon \to 0} \frac{t_{\text{mix}}(\epsilon)}{\log(1/\epsilon)}\right)$

**Proof:**

- Let $\nu$ be left eigvec for $\lambda \neq 1$:
  $$\nu P = \lambda \nu$$
- We have $\langle \nu, 1 \rangle = 0$, so write
  $$\nu = \alpha(\nu_1 - \nu_2) + i\beta(\nu_3 - \nu_4)$$
  for dists $\nu_1, \nu_2, \nu_3, \nu_4$. 
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- **Relaxation time:** $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

**Lemma**

$$t_{\text{mix}}(\epsilon) = O\left(\frac{\log(\chi^2(\nu_0 \| \mu)) + \log(1/\epsilon)}{1 - \lambda_2(P)}\right)$$

- We have

$$t_{\text{rel}} = \Theta\left(\lim_{\epsilon \to 0} \frac{t_{\text{mix}}(\epsilon)}{\log(1/\epsilon)}\right)$$

**Proof:**

- Let $\nu$ be left eigvec for $\lambda \neq 1$:
  $$\nu P = \lambda \nu$$
- We have $\langle \nu, 1 \rangle = 0$, so write
  $$\nu = \alpha(\nu_1 - \nu_2) + i\beta(\nu_3 - \nu_4)$$
  for dists $\nu_1, \nu_2, \nu_3, \nu_4$.
- For $t \geq t_{\text{mix}}(\epsilon)$ we get
  $$\|\nu P^t\|_1 \leq O(\epsilon) \cdot \|\nu\|_1$$
Relaxation time

Suppose $P$ is time-reversible and lazy:

$$\lambda_i(P) \geq 0$$

- Relaxation time: $1/(1 - \lambda_2(P))$
- Relaxation time does not directly control mixing time 😞
- But it controls $t_{\text{mix}}(\epsilon)$ for tiny $\epsilon$:

**Lemma**

$$t_{\text{mix}}(\epsilon) = O\left(\frac{\log(\chi^2(\nu_0\|\mu)) + \log(1/\epsilon)}{1 - \lambda_2(P)}\right)$$

- We have

$$t_{\text{rel}} = \Theta\left(\lim_{\epsilon \to 0} \frac{t_{\text{mix}}(\epsilon)}{\log(1/\epsilon)}\right)$$

**Proof:**

- Let $\nu$ be left eigvec for $\lambda \neq 1$:

$$\nu P = \lambda \nu$$

- We have $\langle \nu, 1 \rangle = 0$, so write

$$\nu = \alpha (\nu_1 - \nu_2) + i\beta (\nu_3 - \nu_4)$$

for dists $\nu_1, \nu_2, \nu_3, \nu_4$.

- For $t \geq t_{\text{mix}}(\epsilon)$ we get

$$\|\nu P^t\|_1 \leq O(\epsilon) \cdot \|\nu\|_1$$

- But this means

$$\lambda^t = O(\epsilon)$$

which means

$$1 - |\lambda| \geq \Omega\left(\frac{\log(1/\epsilon)}{t_{\text{mix}}(\epsilon)}\right)$$
Corollary

Under Dobrushin, we have $t_{rel} = O(n)$; in other words

$$\lambda_2 \leq 1 - \Omega(1/n).$$
Corollary

Under Dobrushin, we have $t_{\text{rel}} = O(n)$; in other words

$$\lambda_2 \leq 1 - \Omega(1/n).$$

Another proof that hypercube has

$$\lambda_2 \leq 1 - \Omega(1/n)$$
Corollary

Under Dobrushin, we have $t_{\text{rel}} = O(n)$; in other words

$$\lambda_2 \leq 1 - \Omega(1/n).$$

- Another proof that hypercube has
  $$\lambda_2 \leq 1 - \Omega(1/n)$$

- First proof that Glauber for coloring with $> 2\Delta$ colors has
  $$\lambda_2 \leq 1 - \Omega(1/n)$$
Corollary
Under Dobrushin, we have \( t_{\text{rel}} = O(n) \); in other words
\[
\lambda_2 \leq 1 - \Omega(1/n).
\]

- Another proof that hypercube has
  \[
  \lambda_2 \leq 1 - \Omega(1/n)
  \]

- First proof that Glauber for coloring with \( > 2\Delta \) colors has
  \[
  \lambda_2 \leq 1 - \Omega(1/n)
  \]

- Note: going back from \( \lambda_2 \) to \( t_{\text{mix}} \) gives us non-tight bound of \( O(n^2) \). 😞
Fourier Analysis
- Characters
- Examples
- Relaxation time

Continuous Time
- Functional analysis in continuous time
- Dirichlet form
Fourier Analysis

- Characters
- Examples
- Relaxation time

Continuous Time

- Functional analysis in continuous time
- Dirichlet form
Continuous time

So far, we have been running Markov chains in discrete time:

\[ X_0 \rightarrow X_1 \rightarrow \cdots \rightarrow X_t \rightarrow \cdots \]

\( t \) is integer
Continuous time

- So far, we have been running Markov chains in **discrete time**:
  \[ X_0 \rightarrow X_1 \rightarrow \cdots \rightarrow X_t \rightarrow \cdots \]
  
  \[ t \text{ is integer} \]

- We can run a chain in **continuous time** via Poisson clock:

  ![Diagram showing time progression with Poisson clock](image)

  \[ \text{draw } n \sim \text{Poisson}(t) \text{ and take } n \text{ discrete steps} \]

  \[ \text{How is } X_t \text{ distributed given } X_0? \]

  \[ \text{Approximate the process as time } \epsilon \text{ where in each interval we take transition of } P \text{ w.p. } \epsilon. \]

  \[ \text{Result at time } t: (1 - \epsilon I + \epsilon P)^{t/\epsilon} \]

  \[ \text{transition matrix} \rightarrow \exp(t(P - I)) \]

  \[ \text{Ultimate lazification!} \]
Continuous time

So far, we have been running Markov chains in discrete time:

\[ X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \ldots \]

\( t \) is integer

We can run a chain in continuous time via Poisson clock:

Every ring, take one step of \( P \).

Approximate the process as time \( \epsilon \) where in each interval we take transition of \( P \) w.p. \( \epsilon \).

Result at time \( t \):

\[ ( (1 - \epsilon) I + \epsilon P )^{t/\epsilon} \]
Continuous time

- So far, we have been running Markov chains in **discrete time**:
  \[ X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \cdots \]
  \( t \) is integer

- We can run a chain in **continuous time** via Poisson clock:

  - Every ring, take one step of \( P \).
  - \( X_t \): position at time \( t \in \mathbb{R}_{\geq 0} \)
Continuous time

So far, we have been running Markov chains in **discrete time**:
\[ X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \cdots \]

\( t \) is integer

We can run a chain in **continuous time** via Poisson clock:

To algorithmically simulate \( X_t \): draw \( n \sim \text{Poisson}(t) \) and take \( n \) discrete steps

Every ring, take one step of \( P \).

\( X_t \): position at time \( t \in \mathbb{R}_{\geq 0} \)
Continuous time

- So far, we have been running Markov chains in discrete time:
  \[ X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \cdots \]
  \( t \) is integer

- We can run a chain in continuous time via Poisson clock:
  - Every ring, take one step of \( P \).
  - \( X_t \): position at time \( t \in \mathbb{R}_{\geq 0} \)

- To algorithmically simulate \( X_t \):
  - draw \( n \sim \text{Poisson}(t) \) and take \( n \) discrete steps

- How is \( X_t \) distributed given \( X_0 \)?
Continuous time

- So far, we have been running Markov chains in **discrete time**:
  \[ X_0 \mapsto X_1 \mapsto \ldots \mapsto X_t \mapsto \ldots \]
  \( t \) is integer

- We can run a chain in **continuous time** via Poisson clock:
  
  ![Diagram showing a Poisson clock with rings and steps](image)

  - Every ring, take one step of \( P \).
  - \( X_t \): position at time \( t \in \mathbb{R}_{\geq 0} \)

- To algorithmically simulate \( X_t \):
  - draw \( n \sim \text{Poisson}(t) \) and take \( n \) discrete steps

- How is \( X_t \) distributed given \( X_0 \)?

- Approximate the process as
  
  \[ \epsilon \epsilon \epsilon \epsilon \epsilon \epsilon \epsilon \epsilon \text{ time} \]
  where in each interval we take transition of \( P \) w.p. \( \epsilon \).
Continuous time

- So far, we have been running Markov chains in **discrete time**: 
  \[ X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \cdots \]
  \( t \) is integer

- We can run a chain in **continuous time** via Poisson clock:

  - Every ring, take one step of \( P \).
  - \( X_t \): position at time \( t \in \mathbb{R}_{\geq 0} \)

- To algorithmically simulate \( X_t \):
  - draw \( n \sim \text{Poisson}(t) \) and take \( n \) discrete steps
  - How is \( X_t \) distributed given \( X_0 \)?

- Approximate the process as

  \[
  \text{result at time } t: \quad \underbrace{((1 - \epsilon)I + \epsilon P)^{t/\epsilon}}_{\text{transition matrix}} \rightarrow \exp(t(P - I))
  \]

  where in each interval we take transition of \( P \) w.p. \( \epsilon \).
Continuous time

- So far, we have been running Markov chains in **discrete time**: 
  \[ X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \cdots \]
  \( t \) is integer

- We can run a chain in **continuous time** via Poisson clock:
  - Every ring, take one step of \( P \).
  - \( X_t \): position at time \( t \in \mathbb{R}_{\geq 0} \)

- To algorithmically simulate \( X_t \):
  - draw \( n \sim \text{Poisson}(t) \) and take \( n \) discrete steps

- How is \( X_t \) distributed given \( X_0 \)?

- Approximate the process as
  
  \[
  \begin{align*}
  \text{time} & : \epsilon \epsilon \epsilon \epsilon \epsilon \epsilon \epsilon \\
  \text{transition of } P & \text{ w.p. } \epsilon.
  \end{align*}
  \]

  where in each interval we take transition of \( P \) w.p. \( \epsilon \).

- Result at time \( t \):
  \[
  ((1 - \epsilon)I + \epsilon P)^{t/\epsilon} \rightarrow \exp(t(P-I))
  \]

- **Ultimate lazification! 😊**
What happens to functional analysis in continuous time?

In discrete time we want

\[ D\phi(\nu P \parallel \mu) \leq (1 - \rho) D\phi(\nu \parallel \mu) \]

Analogue in continuous time:

\[ \frac{d}{dt} D\phi(\nu_t \parallel \mu) \leq -\rho D\phi(\nu_t \parallel \mu) \]

where \( \nu_t = \nu_0 \exp(t(P - I)) \).

Corollary: we get

\[ D\phi(\nu_t \parallel \mu) \leq e^{-t\rho} \cdot D\phi(\nu_0 \parallel \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds.

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof: will show \((1 - \epsilon) I + \epsilon P\) contracts \( D\phi \) by \( 1 - \epsilon \rho \). Taking \( \epsilon \to 0 \) gives what we want.

Because \( \phi \) is convex:

\[ D\phi((1 - \epsilon)\nu + \epsilon \nu P \parallel \mu) \leq (1 - \epsilon) D\phi(\nu \parallel \mu) + \epsilon D\phi(\nu P \parallel \mu) \]

But this is \( \leq (1 - \epsilon \rho) D\phi(\nu \parallel \mu) \).
What happens to functional analysis in continuous time?

In discrete time we want

$$\mathcal{D}_\phi(\nu P \parallel \mu) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu)$$

Analogue in continuous time:

$$\frac{d}{dt} \mathcal{D}_\phi(\nu_t \parallel \mu) \leq -\rho \mathcal{D}_\phi(\nu_t \parallel \mu)$$

where

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

Corollary: we get

$$\mathcal{D}_\phi(\nu_t \parallel \mu) \leq e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$

By comparing to \(dTV\) we get continuous mixing time bounds.

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof: Will show \((1 - \epsilon) I + \epsilon P\) contracts \(\mathcal{D}_\phi\) by \(1 - \epsilon \rho\). Taking \(\epsilon \to 0\) gives what we want.

Because \(\phi\) is convex:

$$\mathcal{D}_\phi((1 - \epsilon) \nu + \epsilon \nu P \parallel \mu) \leq (1 - \epsilon) \mathcal{D}_\phi(\nu \parallel \mu) + \epsilon \mathcal{D}_\phi(\nu P \parallel \mu)$$

But this is \(\leq (1 - \epsilon \rho) \mathcal{D}_\phi(\nu \parallel \mu)\).
What happens to functional analysis in continuous time?

In discrete time we want
\[ \mathcal{D}_\phi(\nu P \parallel \mu) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu) \]

Analogue in continuous time:
\[ \frac{d}{dt} \mathcal{D}_\phi(\nu_t \parallel \mu) \leq -\rho \mathcal{D}_\phi(\nu_t \parallel \mu) \]
where \( \nu_t = \nu_0 \exp(t(P - I)) \).
Functional analysis in continuous time

- What happens to functional analysis in continuous time?

- In discrete time we want
  \[ D_\phi(\nu P \parallel \mu) \leq (1 - \rho) D_\phi(\nu \parallel \mu) \]

- Analogue in continuous time:
  \[ \frac{d}{dt} D_\phi(\nu_t \parallel \mu) \leq -\rho D_\phi(\nu_t \parallel \mu) \]
  where \( \nu_t = \nu_0 \exp(t(P - I)) \).

- Corollary: we get
  \[ D_\phi(\nu_t \parallel \mu) \leq e^{-t\rho} \cdot D_\phi(\nu_0 \parallel \mu) \]
What happens to functional analysis in continuous time?

In discrete time we want
\[ D_\phi(\nu P \parallel \mu) \leq (1 - \rho) D_\phi(\nu \parallel \mu) \]

Analogue in continuous time:
\[ \frac{d}{dt} D_\phi(\nu_t \parallel \mu) \leq -\rho D_\phi(\nu_t \parallel \mu) \]
where \( \nu_t = \nu_0 \exp(t(P - I)) \).

Corollary: we get
\[ D_\phi(\nu_t \parallel \mu) \leq e^{-t\rho} \cdot D_\phi(\nu_0 \parallel \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds. 😊
What happens to functional analysis in continuous time?

In discrete time we want
\[ \mathcal{D}_\phi(\nu P \parallel \mu) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu) \]

Analogue in continuous time:
\[ \frac{d}{dt} \mathcal{D}_\phi(\nu_t \parallel \mu) \leq -\rho \mathcal{D}_\phi(\nu_t \parallel \mu) \]
where \( \nu_t = \nu_0 \exp(t(P - I)) \).

Corollary: we get
\[ \mathcal{D}_\phi(\nu_t \parallel \mu) \leq e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds.
What happens to functional analysis in continuous time?

In discrete time we want
\[ \mathcal{D}_\phi(\nu P \parallel \mu) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu) \]

Analogue in continuous time:
\[ \frac{d}{dt} \mathcal{D}_\phi(\nu_t \parallel \mu) \leq -\rho \mathcal{D}_\phi(\nu_t \parallel \mu) \]
where \( \nu_t = \nu_0 \exp(t(P - I)) \).

Corollary: we get
\[ \mathcal{D}_\phi(\nu_t \parallel \mu) \leq e^{-t \rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds. 😊

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof:
What happens to functional analysis in continuous time?

In discrete time we want
\[ D_\phi(\nu P \parallel \mu) \leq (1 - \rho) D_\phi(\nu \parallel \mu) \]

Analogue in continuous time:
\[ \frac{d}{dt} D_\phi(\nu_t \parallel \mu) \leq -\rho D_\phi(\nu_t \parallel \mu) \]
where \( \nu_t = \nu_0 \exp(t(P - I)). \)

Corollary: we get
\[ D_\phi(\nu_t \parallel \mu) \leq e^{-\rho t} \cdot D_\phi(\nu_0 \parallel \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds.

Fact: discrete is stronger
Discrete-time contraction implies continuous-time contraction.

Proof:
Will show \((1 - \epsilon)I + \epsilon P\) contracts \( D_\phi \) by \(1 - \epsilon \rho\). Taking \( \epsilon \to 0 \) gives what we want.
What happens to functional analysis in continuous time?

In discrete time we want

\[ \mathcal{D}_\phi(\nu P \parallel \mu) \leq (1 - \rho) \mathcal{D}_\phi(\nu \parallel \mu) \]

Analogue in continuous time:

\[ \frac{d}{dt} \mathcal{D}_\phi(\nu_t \parallel \mu) \leq -\rho \mathcal{D}_\phi(\nu_t \parallel \mu) \]

where \( \nu_t = \nu_0 \exp(t(P - I)) \).

Corollary: we get

\[ \mathcal{D}_\phi(\nu_t \parallel \mu) \leq e^{-\rho t} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds.

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof:

- Will show \( (1 - \epsilon)I + \epsilon P \) contracts \( \mathcal{D}_\phi \) by \( 1 - \epsilon \rho \). Taking \( \epsilon \to 0 \) gives what we want.

- Because \( \phi \) is convex:

\[ \mathcal{D}_\phi((1 - \epsilon)\nu + \epsilon \nu P \parallel \mu) \leq (1 - \epsilon) \mathcal{D}_\phi(\nu \parallel \mu) + \epsilon \mathcal{D}_\phi(\nu P \parallel \mu) \]
Functional analysis in continuous time

What happens to functional analysis in continuous time?

In discrete time we want

\[ D_\phi(\nu P \| \mu) \leq (1 - \rho) D_\phi(\nu \| \mu) \]

Analogue in continuous time:

\[ \frac{d}{dt} D_\phi(\nu_t \| \mu) \leq -\rho D_\phi(\nu_t \| \mu) \]

where \( \nu_t = \nu_0 \exp(t(P - I)) \).

Corollary: we get

\[ D_\phi(\nu_t \| \mu) \leq e^{-\rho t} \cdot D_\phi(\nu_0 \| \mu) \]

By comparing to \( d_{TV} \) we get continuous mixing time bounds.

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof:

Will show \( (1 - \epsilon)I + \epsilon P \) contracts \( D_\phi \) by \( 1 - \epsilon \rho \). Taking \( \epsilon \to 0 \) gives what we want.

Because \( \phi \) is convex:

\[ D_\phi((1 - \epsilon)\nu + \epsilon \nu P \| \mu) \leq (1 - \epsilon) D_\phi(\nu \| \mu) + \epsilon D_\phi(\nu P \| \mu) \]

But this is \( \leq (1 - \epsilon \rho) D_\phi(\nu \| \mu) \).
Discrete can be strictly stronger:
Discrete can be strictly stronger:

But, for time-reversible and lazy chains in $\chi^2$:

- say eigs $\geq 0$ or $\lambda_n \geq -\lambda_2$
- discrete time $\leftrightarrow$ continuous time

Sketch:

$$(I + \epsilon(P - I))(I + \epsilon(P \circ - I)) = I + \epsilon(P + P \circ - 2I) + O(\epsilon^2)$$

For lazy reversible $P$, we have gap of $P$ is approximately gap of $(P + P \circ) / 2$.

Corollary: prove continuous-time contraction if easier, and don’t worry about it. Easier because of Dirichlet form!
Discrete can be strictly stronger:

But, for time-reversible and lazy chains in $\chi^2$: say eigs $\geq 0$ or $\lambda_n \geq -\lambda_2$

discrete time $\leftrightarrow$ continuous time

$\chi^2$ contraction in continuous time is dictated by eigs of

$$(P + P^\circ)/2$$
Discrete can be strictly stronger:

But, for time-reversible and lazy chains in $\chi^2$: say $\text{eigs} \succeq 0$ or $\lambda_n \succeq -\lambda_2$

discrete time $\leftrightarrow$ continuous time

$\chi^2$ contraction in continuous time is dictated by eigs of

$$\frac{(P + P^\circ)}{2}$$

Sketch:

$$(I + \epsilon(P - I))(I + \epsilon(P^\circ - I)) = I + \epsilon(P + P^\circ - 2I) + O(\epsilon^2)$$
Discrete can be strictly stronger:

But, for time-reversible and lazy chains in $\chi^2$:
- say eigs $\geq 0$ or $\lambda_n \geq -\lambda_2$
- discrete time $\leftrightarrow$ continuous time

$\chi^2$ contraction in continuous time is dictated by eigs of

$$(P + P^\circ)/2$$

Sketch:

$$(I + \epsilon(P - I))(I + \epsilon(P^\circ - I)) = I + \epsilon(P + P^\circ - 2I) + O(\epsilon^2)$$

For lazy reversible $P$, we have gap of $PP^\circ$ is approximately gap of $(P + P^\circ)/2$. 😊
Discrete can be **strictly stronger**:

But, for time-reversible and lazy chains in $\chi^2$:

- say $\text{eigs} \geq 0$ or $\lambda_n \geq -\lambda_2$

  
  discrete time $\leftrightarrow$ continuous time

$\chi^2$ contraction in continuous time is dictated by eigs of

$$\frac{(P + P^\circ)}{2}$$

Sketch:

$$\left(I + \epsilon(P - I)\right) \left(I + \epsilon(P^\circ - I)\right) = I + \epsilon(P + P^\circ - 2I) + O(\epsilon^2)$$

For lazy reversible $P$, we have gap of $PP^\circ$ is approximately gap of $\frac{(P + P^\circ)}{2}$. 😊

Corollary: prove **continuous-time contraction** if easier, and don’t worry about it.
Discrete can be strictly stronger:

But, for time-reversible and lazy chains in $\chi^2$:
- say eigs $\geq 0$ or $\lambda_n \geq -\lambda_2$
- discrete time $\leftrightarrow$ continuous time

$\chi^2$ contraction in continuous time is dictated by eigs of

$$(P + P^\circ)/2$$

Sketch:

$$(I + \epsilon(P - I))(I + \epsilon(P^\circ - I)) = I + \epsilon(P + P^\circ - 2I) + O(\epsilon^2)$$

For lazy reversible $P$, we have gap of $PP^\circ$ is approximately gap of $(P + P^\circ)/2$.

Corollary: prove continuous-time contraction if easier, and don’t worry about it.

Easier because of Dirichlet form!
Dirichlet form

Assume $P$ is time-reversible.
Assume $P$ is time-reversible.

Let’s expand $\frac{d}{dt} D_\phi (\nu_t \parallel \mu)$. We have $\frac{d}{dt} E_\mu [\phi(\nu_t/\mu)] =$

$$E_\mu \left[ \phi' \left( \frac{\nu_t}{\mu} \right) \frac{d}{dt} \frac{\nu_t}{\mu} \right]$$
Assume $P$ is time-reversible.

Let's expand $\frac{d}{dt} D_{\phi}(\nu_t \parallel \mu)$. We have $\frac{d}{dt} E_\mu[\phi(\nu_t/\mu)] = E_\mu[\phi'(\frac{\nu_t}{\mu}) \frac{d}{dt} \frac{\nu_t}{\mu}]$.

But $\frac{d}{dt} \nu_t = \nu_t(P - I)$, and we can write above as

$$-\frac{1}{2} \sum_{x, y} Q(x, y) \left( \phi'(\frac{\nu_t(x)}{\mu(x)}) - \phi'(\frac{\nu_t(y)}{\mu(y)}) \right) \left( \frac{\nu_t(x)}{\mu(x)} - \frac{\nu_t(y)}{\mu(y)} \right)$$
Dirichlet form

Assume $P$ is time-reversible.

Let's expand $\frac{d}{dt} D\phi(\nu_t \parallel \mu)$. We have $\frac{d}{dt} E_\mu[\phi(\nu_t/\mu)] =
\mathbb{E}_\mu\left[\phi'\left(\frac{\nu_t}{\mu}\right) \frac{d}{dt} \frac{\nu_t}{\mu}\right]$.

But $\frac{d}{dt} \nu_t = \nu_t(P - I)$, and we can write above as

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

Dirichlet form

Define $\mathcal{E}(f, g)$ for functions $f, g : \Omega \rightarrow \mathbb{R}$ as

$$\frac{1}{2} \mathbb{E}_{(x,y) \sim Q} [(f(x) - f(y)) (g(x) - g(y))] .$$
Dirichlet form

Assume $P$ is time-reversible.

Let’s expand $\frac{d}{dt} D_\phi(\nu_t \parallel \mu)$. We have $\frac{d}{dt} E_\mu[\phi(\nu_t/\mu)] = E_\mu[\phi'(\frac{\nu_t}{\mu}) \frac{d}{dt} \frac{\nu_t}{\mu}]$

But $\frac{d}{dt} \nu_t = \nu_t (P - I)$, and we can write above as

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

Dirichlet form

Define $\mathcal{E}(f, g)$ for functions $f, g: \Omega \rightarrow \mathbb{R}$ as

$$\frac{1}{2} \mathbb{E}_{(x,y) \sim Q} [(f(x) - f(y))(g(x) - g(y))] .$$

Poincaré: $2 \mathcal{E}(f, f) \geq \rho \text{Var}[f]$  
MLSI: $\mathcal{E}(f, \log f) \geq \rho \text{Ent}[f]$
Just need to lower bound $\varepsilon$