CS 263: Counting and Sampling

Nima Anari

Stanford University

slides for

Continuous Time

Review

ϕ-entropy

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

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

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

- $\phi(x) = x^2$ vs. $\phi(x) = x \log x$
- x^2 leads to Var and χ^2

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Contraction:

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant (1 - \rho)\, \mathcal{D}_\phi(\nu \parallel \mu)$$

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Contraction:

$$\mathcal{D}_{\phi}(\nu N \parallel \mu N) \leqslant (1-\rho)\, \mathcal{D}_{\phi}(\nu \parallel \mu)$$

Lemma: data processing

$$\mathcal{D}_{\phi}(\nu N \parallel \mu N) \leqslant \mathcal{D}_{\phi}(\nu \parallel \mu)$$

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Contraction:

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant (1-\rho)\, \mathcal{D}_\phi(\nu \parallel \mu)$$

Lemma: data processing

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant \mathcal{D}_\phi(\nu \parallel \mu)$$

- Specializing to χ^2, we have

$$\rho = 1 - \lambda_2(N N^\circ)$$

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Contraction:

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant (1 - \rho)\, \mathcal{D}_\phi(\nu \parallel \mu)$$

Lemma: data processing

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant \mathcal{D}_\phi(\nu \parallel \mu)$$

- Specializing to χ^2, we have

$$\rho = 1 - \lambda_2(NN^\circ)$$

- Abelian walks on group G:

$$x \mapsto x + z$$

sampled i.i.d. from π

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Contraction:

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant (1 - \rho)\, \mathcal{D}_\phi(\nu \parallel \mu)$$

Lemma: data processing

$$\mathcal{D}_\phi(\nu N \parallel \mu N) \leqslant \mathcal{D}_\phi(\nu \parallel \mu)$$

- Specializing to χ^2, we have
$$\rho = 1 - \lambda_2(N N^\circ)$$
- Abelian walks on group G:
$$x \mapsto x + z$$
sampled i.i.d. from π
- Eigvecs are characters χ:
$$\chi(x + y) = \chi(x)\chi(y)$$

Review

ϕ-entropy

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

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

ϕ-divergence

For measure ν and dist μ define

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

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

Contraction:

$$\mathcal{D}_{\phi}(\nu N \parallel \mu N) \leqslant (1-\rho)\, \mathcal{D}_{\phi}(\nu \parallel \mu)$$

Lemma: data processing

$$\mathcal{D}_{\phi}(\nu N \parallel \mu N) \leqslant \mathcal{D}_{\phi}(\nu \parallel \mu)$$

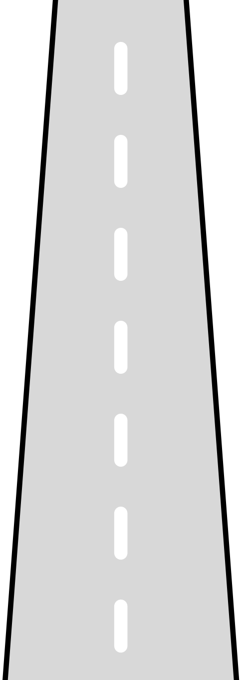
- Specializing to χ^2, we have
$$\rho = 1 - \lambda_2(NN^{\circ})$$
- Abelian walks on group G:
$$x \mapsto x+z$$
sampled i.i.d. from π
- Eigvecs are characters χ:
$$\chi(x+y) = \chi(x)\chi(y)$$
- Eigvals are $\mathbb{E}_{\pi}[\chi]$

Fourier Analysis

- Characters
- Examples
- Relaxation time

Continuous Time

- Functional analysis in continuous time
- Dirichlet form

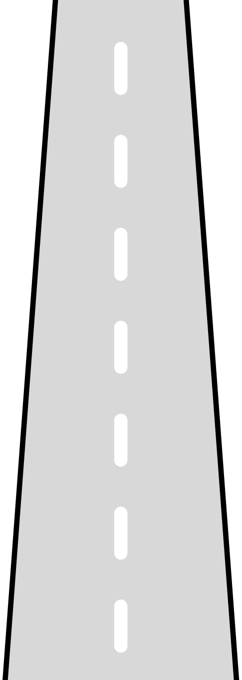


Fourier Analysis

- Characters
- Examples
- Relaxation time

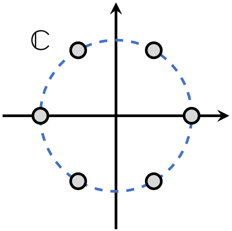
Continuous Time

- Functional analysis in continuous time
- Dirichlet form



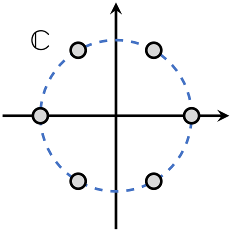
Characters

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



Characters

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



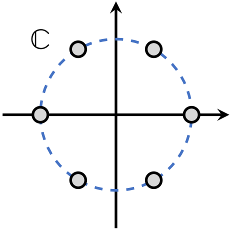
- There are exactly n of them. 🙂

Characters

- We know characters of $\mathbb{Z}_n$:

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

  for ω an n-th root of unity.

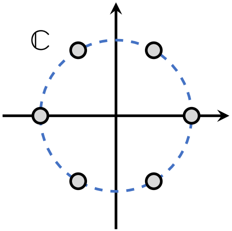


- There are exactly n of them. ☺
- Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:

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

Characters

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



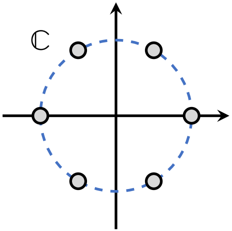
- There are exactly n of them. 🙂
- Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
$$x \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$
- For G, we get $|G|$ characters. 🙂

Characters

- We know characters of $\mathbb{Z}_n$:

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

for ω 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. ☺

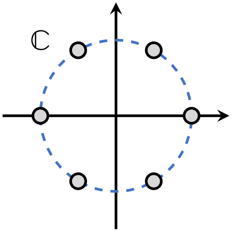
- We just need to compute

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

for all of these characters.

Characters

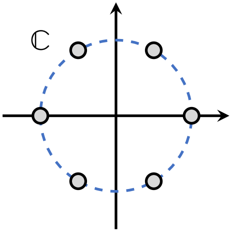
- We know characters of $\mathbb{Z}_n$:
$$\chi(x) = \omega^x$$
for ω an n-th root of unity.



- 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 just need to compute
$$\mathbb{E}_{x\sim\pi}\left[\omega_1^{x_1} \cdots \omega_k^{x_k}\right]$$
for all of these characters.
- The $\omega_1 = \cdots = \omega_k = 1$ character gives us the special 1 eigval.

Characters

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

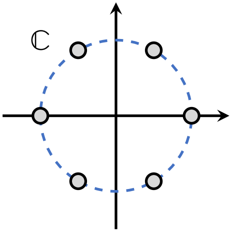


- There are exactly n of them. 😊
- Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
$$x \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$
- For G, we get $|G|$ characters. 😊

- We just need to compute
$$\mathbb{E}_{x \sim \pi}\left[\omega_1^{x_1} \cdots \omega_k^{x_k}\right]$$
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 $P^\circ = P^\intercal$ is also Abelian walk. Eigvals are
$$\mathbb{E}_{x \sim \pi}[\chi(-x)]$$

Characters

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

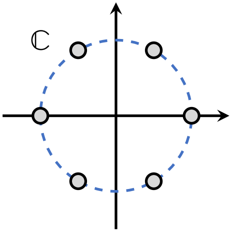


- There are exactly n of them. 🙂
- Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
$$x \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$
- For G, we get $|G|$ characters. 🙂

- We just need to compute
$$\mathbb{E}_{x\sim\pi}\left[\omega_1^{x_1} \cdots \omega_k^{x_k}\right]$$
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 $P^\circ = P^\intercal$ is also Abelian walk. Eigvals are
$$\mathbb{E}_{x\sim\pi}[\chi(-x)]$$
- Since P and P° commute, we have
$$\lambda_k(PP^\circ) = |\lambda_k(P)|^2$$

Characters

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



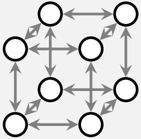
- There are exactly n of them. ☺
- Characters of $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$:
$$x \mapsto \omega_1^{x_1} \cdots \omega_k^{x_k}.$$
- For G, we get $|G|$ characters. ☺

- We just need to compute
$$\mathbb{E}_{x\sim\pi}\left[\omega_1^{x_1} \cdots \omega_k^{x_k}\right]$$
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 $P^\circ = P^\mathsf{T}$ is also Abelian walk. Eigvals are
$$\mathbb{E}_{x\sim\pi}[\chi(-x)]$$
- Since P and P° commute, we have
$$\lambda_k(PP^\circ) = |\lambda_k(P)|^2$$
- Mixing: largest $|\cdot|$ of an eig?

Example: hypercube

Distribution π:

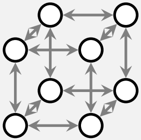
- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



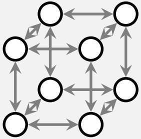
- There are 2^n characters:

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

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



- There are 2^n characters:

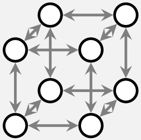
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$

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

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



- There are 2^n characters:

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

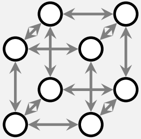
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are

$$k/n$$

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ 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

$$k/n$$

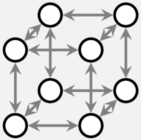
- Spectral gap:

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

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ 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:

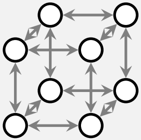
$$1 - (n-1)/n = 1/n$$

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

Example: hypercube

Distribution π:

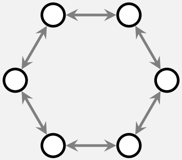
- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



Example: cycle

Distribution π:

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

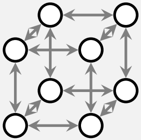


- There are 2^n characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are
$$k/n$$
- Spectral gap:
$$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leqslant O(n^2)$

Example: hypercube

Distribution π:

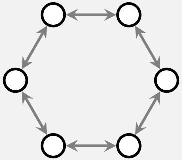
- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



Example: cycle

Distribution π:

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

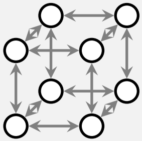


- There are 2^n characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are
$$k/n$$
- Spectral gap:
$$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leqslant O(n^2)$

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$

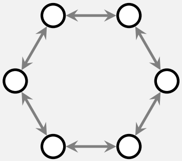


- There are 2^n characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are
$$k/n$$
- Spectral gap:
$$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leqslant O(n^2)$

Example: cycle

Distribution π:

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

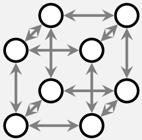


- There are n characters:
$$x \mapsto \omega^x$$

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$

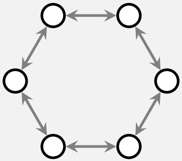


- There are 2^n characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are
$$k/n$$
- Spectral gap:
$$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leqslant O(n^2)$

Example: cycle

Distribution π:

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

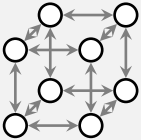


- There are n characters:
$$x \mapsto \omega^x$$
- Eigval is $(\omega + \omega^{-1})/2$

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$

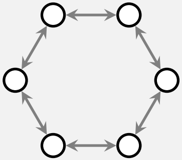


- There are 2^n characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are
$$k/n$$
- Spectral gap:
$$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leqslant O(n^2)$

Example: cycle

Distribution π:

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

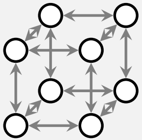


- There are n characters:
$$x \mapsto \omega^x$$
- Eigval is $(\omega + \omega^{-1})/2$
- Eigvals are
$$\cos(2\pi k/n)$$

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ 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:

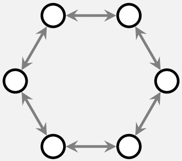
$$1 - (n-1)/n = 1/n$$

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

Example: cycle

Distribution π:

- $+1$ 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

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

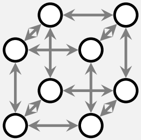
- Spectral gap:

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

Example: hypercube

Distribution π:

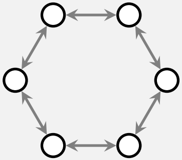
- 0 w.p. $1/2$
- $\mathbb{1}_i$ w.p. $1/2n$



Example: cycle

Distribution π:

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



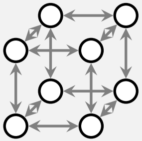
- There are 2^n characters:
$$x \mapsto (\pm 1)^{x_1} \cdots (\pm 1)^{x_n}$$
- Eigval is $\#\{+1\}/n$
- $\binom{n}{k}$ of eigvals are
$$k/n$$
- Spectral gap:
$$1 - (n-1)/n = 1/n$$
- $t_{\text{mix}} \leqslant O(n^2)$

- There are n characters:
$$x \mapsto \omega^x$$
- Eigval is $(\omega + \omega^{-1})/2$
- Eigvals are
$$\cos(2\pi k/n)$$
- Spectral gap:
$$1 - \cos(2\pi/n) \simeq \Theta(1/n^2)?$$
- $t_{\text{mix}} \leqslant O(n^2 \log n)?$

Example: hypercube

Distribution π:

- 0 w.p. $1/2$
- $\mathbb{1}_i$ 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:

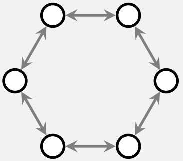
$$1 - (n-1)/n = 1/n$$

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

Example: cycle

Distribution π:

- $+1$ 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:

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

- $t_{\text{mix}} \leqslant O(n^2 \log n)$? Not for even n.

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

- Relaxation time: $1/(1-\lambda_2(P))$
- Relaxation time does not directly control mixing time ☹

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Lemma

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

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Lemma

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

- We have

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

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Lemma

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

- We have

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

Proof:

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Lemma

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

- We have

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

Proof:

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

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

Lemma

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

- We have

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

Proof:

- Let ν be left eigvec for $\lambda \neq 1$: $\nu P = \lambda \nu$
- We have $\langle \nu, \mathbb{1} \rangle = 0$, 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:

$$\lambda_i(P) \geqslant 0$$

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

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 ν be left eigvec for $\lambda \neq 1$: $\nu P = \lambda\nu$
- We have $\langle \nu, \mathbb{1}\rangle = 0$, 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 $t \geqslant t_{\text{mix}}(\epsilon)$ we get

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

Relaxation time

Suppose P is time-reversible and lazy:

$$\lambda_i(P) \geqslant 0$$

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

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 ν be left eigvec for $\lambda \neq 1$: $\nu P = \lambda \nu$
- We have $\langle \nu, \mathbb{1}\rangle = 0$, 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 $t \geqslant t_{\text{mix}}(\epsilon)$ we get

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

- But this means

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

which means

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

Corollary

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

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

Corollary

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

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

- Another proof that hypercube has

$$\lambda_2 \leqslant 1 - \Omega(1/n)$$

Corollary

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

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

- Another proof that hypercube has

$$\lambda_2 \leqslant 1 - \Omega(1/n)$$

- First proof that Glauber for coloring with $> 2\Delta$ colors has

$$\lambda_2 \leqslant 1 - \Omega(1/n)$$

Corollary

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

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

- Another proof that hypercube has

$$\lambda_2 \leqslant 1 - \Omega(1/n)$$

- First proof that Glauber for coloring with $> 2\Delta$ colors has

$$\lambda_2 \leqslant 1 - \Omega(1/n)$$

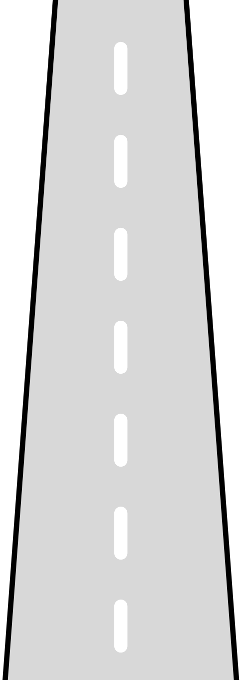
- Note: going back from λ_2 to t_{mix} gives us non-tight bound of $O(n^2)$. ☹

Fourier Analysis

- Characters
- Examples
- Relaxation time

Continuous Time

- Functional analysis in continuous time
- Dirichlet form

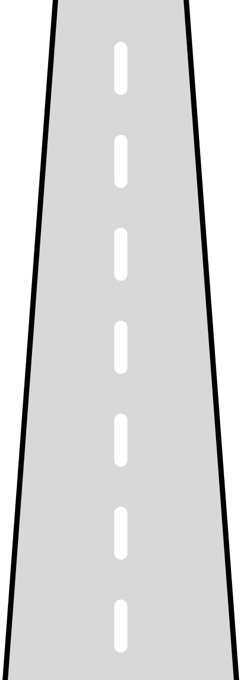


Fourier Analysis

- 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:

$$X_0 \mapsto X_1 \mapsto \cdots \mapsto \underset{\uparrow}{X_t} \mapsto \ldots$$

t is integer

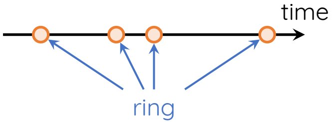
Continuous time

- 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:



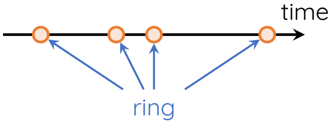
Continuous time

- 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.

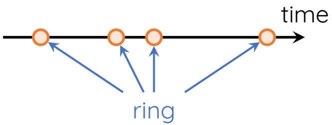
Continuous time

- 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.
- X_t: position at time $t \in \mathbb{R}_{\geqslant 0}$

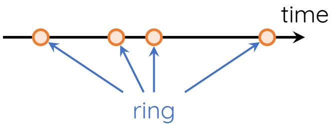
Continuous time

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

$$X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \ldots$$

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 $t \in \mathbb{R}_{\geqslant 0}$

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

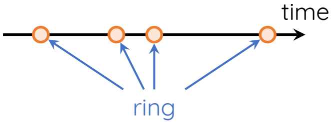
Continuous time

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

$$X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \dots$$

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 $n \sim \mathrm{Poisson}(t)$ and take n discrete steps
- How is X_t distributed given X_0?

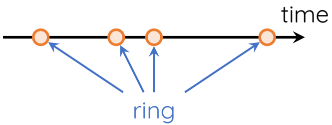
Continuous time

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

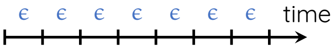
$$X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \ldots$$

t is integer

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



- Every ring, take one step of P.
- X_t: position at time $t \in \mathbb{R}_{\geqslant 0}$
- To algorithmically simulate X_t: draw $n \sim \mathrm{Poisson}(t)$ and take n discrete steps
- How is X_t distributed given X_0?
- Approximate the process as



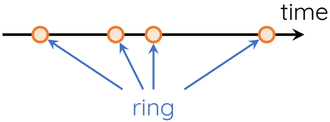
where in each interval we take transition of P w.p. ϵ.

Continuous time

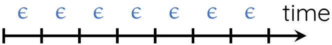
- So far, we have been running Markov chains in discrete time:

$$X_0 \mapsto X_1 \mapsto \cdots \mapsto \underset{\substack{\uparrow \\ t \text{ is integer}}}{X_t} \mapsto \ldots$$

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



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



where in each interval we take transition of P w.p. ϵ.

- Result at time t:

$$\underbrace{((1-\epsilon)I + \epsilon P)^{t/\epsilon}}_{\text{transition matrix}} \to \exp(t(P - I))$$

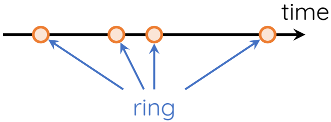
Continuous time

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

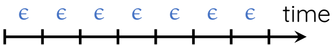
$$X_0 \mapsto X_1 \mapsto \cdots \mapsto X_t \mapsto \ldots$$

t is integer

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



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



where in each interval we take transition of P w.p. ϵ.
- Result at time t:

$$\underbrace{((1-\epsilon)I + \epsilon P)^{t/\epsilon}}_{\text{transition matrix}} \to \exp(t(P - I))$$

- Ultimate lazification! ☺

Functional analysis in continuous time

- What happens to functional analysis in continuous time?

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want

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

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want

  $$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\, \mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:

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

  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
  $$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\, \mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:
  $$\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant -\rho\, \mathcal{D}_\phi(\nu_t \parallel \mu)$$
  where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get
  $$\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want
  $$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\,\mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:
  $$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant -\rho\,\mathcal{D}_\phi(\nu_t \parallel \mu)$$
  where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get
  $$\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$
- By comparing to d_{TV} we get continuous mixing time bounds. ☺

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want
  $$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\,\mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:
  $$\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant -\rho\,\mathcal{D}_\phi(\nu_t \parallel \mu)$$
  where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get
  $$\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$
- By comparing to d_{TV} we get continuous mixing time bounds. ☺

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want

  $$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\,\mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:

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

  where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get

  $$\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$
- By comparing to d_{TV} we get continuous mixing time bounds. ☺

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof:

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want

$$\mathcal{D}_\phi(\nu P \,\|\, \mu) \leqslant (1-\rho)\,\mathcal{D}_\phi(\nu \,\|\, \mu)$$

- Analogue in continuous time:

$$\frac{d}{dt}\mathcal{D}_\phi(\nu_t \,\|\, \mu) \leqslant -\rho\,\mathcal{D}_\phi(\nu_t \,\|\, \mu)$$

  where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get

$$\mathcal{D}_\phi(\nu_t \,\|\, \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \,\|\, \mu)$$

- 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 $\mathcal{D}_\phi$ by $1-\epsilon\rho$. Taking $\epsilon \to 0$ gives what we want.

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want
  $$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\,\mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:
  $$\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant -\rho\,\mathcal{D}_\phi(\nu_t \parallel \mu)$$
  where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get
  $$\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$
- By comparing to d_{TV} we get continuous mixing time bounds. ☺

Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

Proof:

- Will show $(1-\epsilon)I + \epsilon P$ contracts $\mathcal{D}_\phi$ by $1-\epsilon\rho$. Taking $\epsilon \to 0$ gives what we want.
- Because ϕ is convex:
  $$\mathcal{D}_\phi((1-\epsilon)\nu + \epsilon\nu P \parallel \mu) \leqslant (1-\epsilon)\,\mathcal{D}_\phi(\nu \parallel \mu) + \epsilon\,\mathcal{D}_\phi(\nu P \parallel \mu)$$

Functional analysis in continuous time

- What happens to functional analysis in continuous time?
- In discrete time we want
$$\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\rho)\,\mathcal{D}_\phi(\nu \parallel \mu)$$
- Analogue in continuous time:
$$\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant -\rho\,\mathcal{D}_\phi(\nu_t \parallel \mu)$$
where $\nu_t = \nu_0 \exp(t(P-I))$.
- Corollary: we get
$$\mathcal{D}_\phi(\nu_t \parallel \mu) \leqslant e^{-t\rho} \cdot \mathcal{D}_\phi(\nu_0 \parallel \mu)$$
- By comparing to d_{TV} we get continuous mixing time bounds. ☺

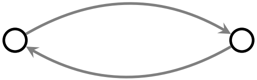
Fact: discrete is stronger

Discrete-time contraction implies continuous-time contraction.

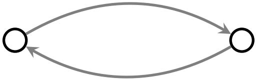
Proof:

- Will show $(1-\epsilon)I + \epsilon P$ contracts $\mathcal{D}_\phi$ by $1-\epsilon\rho$. Taking $\epsilon \to 0$ gives what we want.
- Because ϕ is convex:
$$\mathcal{D}_\phi((1-\epsilon)\nu + \epsilon\nu P \parallel \mu) \leqslant (1-\epsilon)\,\mathcal{D}_\phi(\nu \parallel \mu) + \epsilon\,\mathcal{D}_\phi(\nu P \parallel \mu)$$
- But this is $\leqslant (1-\epsilon\rho)\,\mathcal{D}_\phi(\nu \parallel \mu)$. ☺

- Discrete can be strictly stronger:



- Discrete can be strictly stronger:

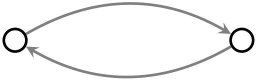


- But, for time-reversible and lazy chains in χ^2:

  say eigs $\geqslant 0$ or $\lambda_n \geqslant -\lambda_2$

  discrete time $\leftrightarrow$ continuous time

- Discrete can be strictly stronger:



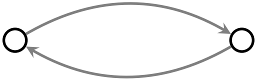
- But, for time-reversible and lazy chains in χ^2:   say eigs $\geqslant 0$ or $\lambda_n \geqslant -\lambda_2$

  discrete time $\leftrightarrow$ continuous time

- χ^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 χ^2:

  say eigs $\geqslant 0$ or $\lambda_n \geqslant -\lambda_2$

  discrete time $\leftrightarrow$ continuous time

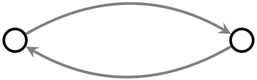
- χ^2 contraction in continuous time is dictated by eigs of

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

- Sketch:

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

- Discrete can be strictly stronger:



- But, for time-reversible and lazy chains in χ^2:

say eigs $\geqslant 0$ or $\lambda_n \geqslant -\lambda_2$

discrete time $\leftrightarrow$ continuous time

- χ^2 contraction in continuous time is dictated by eigs of

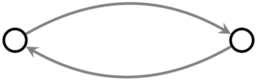
$$(P + P^\circ)/2$$

- 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° is approximately gap of $(P + P^\circ)/2$. ☺

- Discrete can be strictly stronger:



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

  discrete time $\leftrightarrow$ continuous time

- χ^2 contraction in continuous time is dictated by eigs of

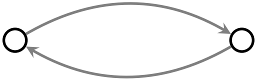
$$(P + P^\circ)/2$$

- 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° is approximately gap of $(P + P^\circ)/2$. 🙂
- 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 χ^2: say eigs $\geqslant 0$ or $\lambda_n \geqslant -\lambda_2$

$$\text{discrete time} \leftrightarrow \text{continuous time}$$

- χ^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° 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.

Dirichlet form

- Assume P is time-reversible.
- Let's expand $\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu)$. We have $\frac{d}{dt}\mathbb{E}_\mu[\phi(\nu_t/\mu)] =$

$$\mathbb{E}_\mu\left[\phi'\left(\frac{\nu_t}{\mu}\right)\frac{d}{dt}\frac{\nu_t}{\mu}\right]$$

Dirichlet form

- Assume P is time-reversible.
- Let's expand $\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu)$. We have $\frac{d}{dt}\mathbb{E}_\mu[\phi(\nu_t/\mu)] =$

$$\mathbb{E}_\mu\left[\phi'\left(\frac{\nu_t}{\mu}\right)\frac{d}{dt}\frac{\nu_t}{\mu}\right]$$

- 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

- Assume P is time-reversible.
- Let's expand $\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu)$. We have $\frac{d}{dt}\mathbb{E}_\mu[\phi(\nu_t/\mu)] =$
$$\mathbb{E}_\mu\left[\phi'\left(\frac{\nu_t}{\mu}\right)\frac{d}{dt}\frac{\nu_t}{\mu}\right]$$
- But $\frac{d}{dt}\nu_t = \nu_t(P - I)$, and we can write above as
$$-\tfrac{1}{2}\sum_{x,y} Q(x,y)\left(\phi'\left(\frac{\nu_t(x)}{\mu(x)}\right) - \phi'\left(\frac{\nu_t(y)}{\mu(y)}\right)\right)\left(\frac{\nu_t(x)}{\mu(x)} - \frac{\nu_t(y)}{\mu(y)}\right)$$

Dirichlet form

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

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

Dirichlet form

- Assume P is time-reversible.
- Let's expand $\frac{d}{dt}\mathcal{D}_\phi(\nu_t \parallel \mu)$. We have $\frac{d}{dt}\mathbb{E}_\mu[\phi(\nu_t/\mu)] =$
$$\mathbb{E}_\mu\left[\phi'\left(\frac{\nu_t}{\mu}\right)\frac{d}{dt}\frac{\nu_t}{\mu}\right]$$
- 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 $f, g : \Omega \to \mathbb{R}$ as

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

Poincaré: $2\,\mathcal{E}(f,f) \geqslant \rho\,\mathrm{Var}[f]$ MLSI: $\mathcal{E}(f, \log f) \geqslant \rho\,\mathrm{Ent}[f]$

Just need to lower bound $\mathcal{E}$