A transport approach to the cutoff phenomenon

Francesco Pedrotti
Based on joint work with Justin Salez [PS25]

ETH Zurich

Oxford Discrete Mathematics and Probability Seminar
November 10, 2026

Set-up: Markov processes and convergence to equilibrium

- Let $(\Omega, \mathcal{F}, \pi)$ be a probability space, and consider a Markov process $(X_t)_{t\geq 0}$ which is reversible with respect to π.
- We denote by L the generator and by $P_t = e^{tL}$ the associated semigroup

$$P_t f(x) := \mathbb{E}[f(X_t) \mid X_0 = x] \qquad \text{for bounded } f : \Omega \to \mathbb{R} \text{ and } x \in \Omega;$$
$$\mu P_t := \mathrm{law}(X_t) \qquad \text{given } X_0 \sim \mu \text{ for } \mu \in \mathcal{P}(\Omega).$$

Set-up: Markov processes and convergence to equilibrium

- Let $(\Omega, \mathcal{F}, \pi)$ be a probability space, and consider a Markov process $(X_t)_{t\geq 0}$ which is reversible with respect to π.
- We denote by L the generator and by $P_t = e^{tL}$ the associated semigroup

$$P_t f(x) := \mathbb{E}[f(X_t) \mid X_0 = x] \qquad \text{for bounded } f : \Omega \to \mathbb{R} \text{ and } x \in \Omega;$$
$$\mu P_t := \mathrm{law}(X_t) \qquad \text{given } X_0 \sim \mu \text{ for } \mu \in \mathcal{P}(\Omega).$$

- **Continuous-time Markov chains:** (Ω, π) is a finite probability space and $Q : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ a stochastic matrix satisfying $\pi(x)Q(x,y) = \pi(y)Q(y,x)$. Then, $L = Q - I$ and the semigroup is the stochastic matrix $P_t = e^{tL}$.

Set-up: Markov processes and convergence to equilibrium

- Let $(\Omega, \mathcal{F}, \pi)$ be a probability space, and consider a Markov process $(X_t)_{t\geq 0}$ which is reversible with respect to π.
- We denote by L the generator and by $P_t = e^{tL}$ the associated semigroup

$$P_t f(x) := \mathbb{E}[f(X_t) \mid X_0 = x] \qquad \text{for bounded } f : \Omega \to \mathbb{R} \text{ and } x \in \Omega;$$
$$\mu P_t := \text{law}(X_t) \qquad \text{given } X_0 \sim \mu \text{ for } \mu \in \mathcal{P}(\Omega).$$

- **Continuous-time Markov chains:** (Ω, π) is a finite probability space and $Q : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ a stochastic matrix satisfying $\pi(x)Q(x,y) = \pi(y)Q(y,x)$. Then, $L = Q - I$ and the semigroup is the stochastic matrix $P_t = e^{tL}$.
- **Langevin dynamics:** $\Omega = \mathbb{R}^d$ and $\pi = e^{-V} \in \mathcal{P}(\mathbb{R}^d)$ is a probability density, for a smooth potential $V : \mathbb{R}^d \to \mathbb{R}$.
  The Langevin dynamics is given by the solution of the SDE

$$dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t.$$

  Its generator acts as $Lf = \Delta f - \nabla V \cdot \nabla f$ for $f \in C_b^\infty(\mathbb{R}^d)$.

Ergodicity, total variation distance and mixing time

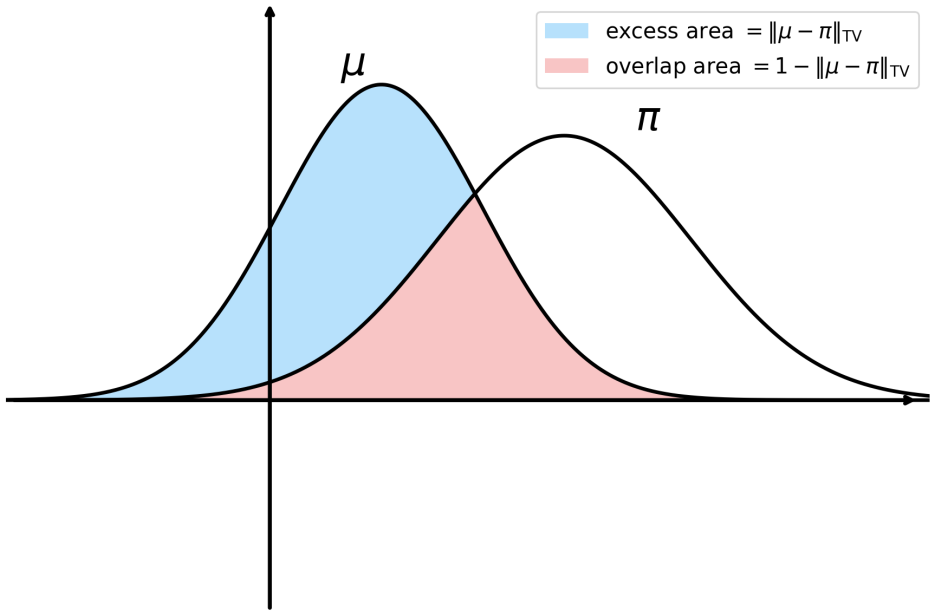
Ergodicity: if $X_0 \sim \mu$, under mild assumptions, $\mu P_t = \text{law}(X_t) \to \pi$ as $t \to \infty$.
Question: how quick is the convergence to equilibrium?

Ergodicity, total variation distance and mixing time

Ergodicity: if $X_0 \sim \mu$, under mild assumptions, $\mu P_t = \text{law}(X_t) \to \pi$ as $t \to \infty$.
Question: how quick is the convergence to equilibrium?

- The **total variation distance** between $\mu, \pi \in \mathcal{P}(\Omega)$ is defined by

$$\begin{aligned} \mathrm{TV}(\mu, \pi) &= \sup_{A \subset \Omega} |\mu(A) - \pi(A)| \\ &= \frac{1}{2} \left\| \frac{d\mu}{d\pi} - 1 \right\|_{L^1(\pi)} \in [0, 1]. \end{aligned}$$

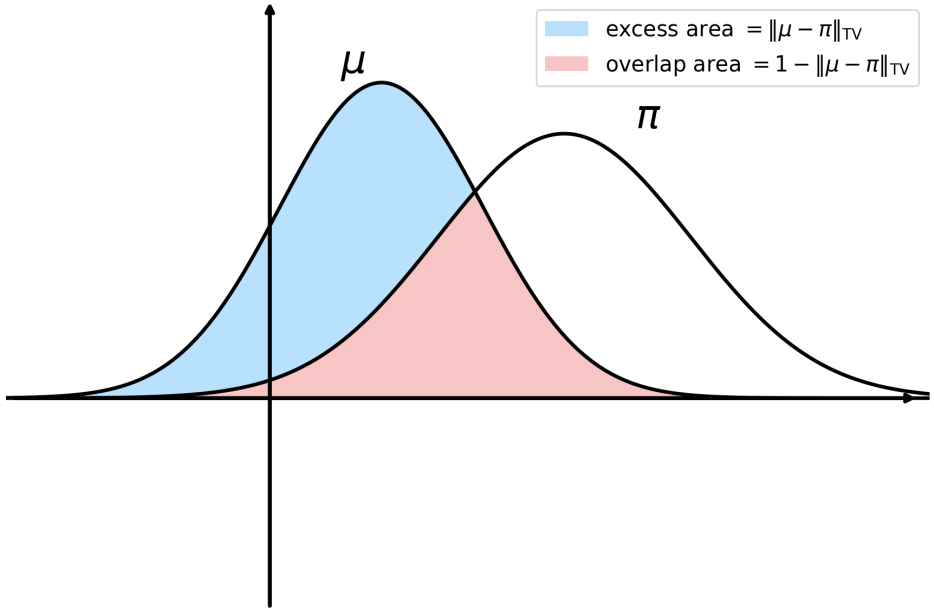


Ergodicity, total variation distance and mixing time

Ergodicity: if $X_0 \sim \mu$, under mild assumptions, $\mu P_t = \text{law}(X_t) \to \pi$ as $t \to \infty$.
Question: how quick is the convergence to equilibrium?

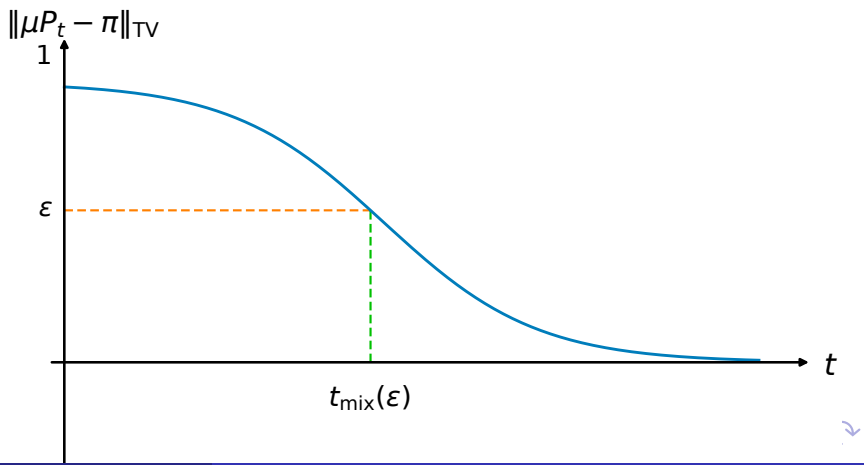
- The **total variation distance** between $\mu, \pi \in \mathcal{P}(\Omega)$ is defined by

$$\begin{aligned} \mathrm{TV}(\mu, \pi) &= \sup_{A \subset \Omega} |\mu(A) - \pi(A)| \\ &= \frac{1}{2} \left\| \frac{d\mu}{d\pi} - 1 \right\|_{L^1(\pi)} \in [0, 1]. \end{aligned}$$



- For $0 < \varepsilon < 1$, the **mixing time** is

$$t_{\mathrm{mix}}(\varepsilon) = \inf_{t \geq 0} \{\mathrm{TV}(\mu P_t, \pi) \leq \varepsilon\}$$



χ^2-divergence, mixing time and Poincaré inequality

- The χ^2-divergence is $\chi^2(\mu \,\|\, \pi) = \left\| \frac{d\mu}{d\pi} - 1 \right\|^2_{L^2(\pi)} = \mathrm{Var}_\pi\left(\frac{d\mu}{d\pi}\right)$.

  By Cauchy–Schwarz, $\chi^2(\mu \,\|\, \pi) \geq 4\mathrm{TV}(\mu, \pi)^2$.

χ^2-divergence, mixing time and Poincaré inequality

- The χ^2-divergence is $\chi^2(\mu \,\|\, \pi) = \left\| \frac{d\mu}{d\pi} - 1 \right\|^2_{L^2(\pi)} = \mathrm{Var}_\pi\left(\frac{d\mu}{d\pi}\right)$.

  By Cauchy–Schwarz, $\chi^2(\mu \,\|\, \pi) \geq 4\mathrm{TV}(\mu, \pi)^2$.
- The **spectral gap/Poincaré constant** of the Markov chain is the best constant $\gamma > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$,

$$\chi^2(\mu P_t \,\|\, \pi) \leq \exp(-2\gamma t)\chi^2(\mu \,\|\, \pi).$$

χ^2-divergence, mixing time and Poincaré inequality

- The χ^2-divergence is $\chi^2(\mu \,\|\, \pi) = \left\| \frac{d\mu}{d\pi} - 1 \right\|^2_{L^2(\pi)} = \mathrm{Var}_\pi\left(\frac{d\mu}{d\pi}\right)$.

  By Cauchy–Schwarz, $\chi^2(\mu \,\|\, \pi) \geq 4\mathrm{TV}(\mu, \pi)^2$.
- The **spectral gap/Poincaré constant** of the Markov chain is the best constant $\gamma > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$,

$$\chi^2(\mu P_t \,\|\, \pi) \leq \exp(-2\gamma t)\chi^2(\mu \,\|\, \pi).$$

- Equivalently, γ is the best constant in the Poincaré inequality for π:

$$\gamma \mathrm{Var}_\pi(f) \leq \int f(-Lf)\, d\pi$$

$$= \begin{cases} \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)Q(x,y)(f(y)-f(x))^2 & \text{for Markov chains,} \\ \int |\nabla f|^2 \, d\pi & \text{for Langevin dynamics.} \end{cases}$$

χ^2-divergence, mixing time and Poincaré inequality

- The χ^2-divergence is $\chi^2(\mu \,\|\, \pi) = \left\| \frac{d\mu}{d\pi} - 1 \right\|^2_{L^2(\pi)} = \mathrm{Var}_\pi\left(\frac{d\mu}{d\pi}\right)$.

  By Cauchy–Schwarz, $\chi^2(\mu \,\|\, \pi) \geq 4\mathrm{TV}(\mu, \pi)^2$.
- The **spectral gap/Poincaré constant** of the Markov chain is the best constant $\gamma > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$,

$$\chi^2(\mu P_t \,\|\, \pi) \leq \exp(-2\gamma t)\chi^2(\mu \,\|\, \pi).$$

- Equivalently, γ is the best constant in the Poincaré inequality for π:

$$\gamma \mathrm{Var}_\pi(f) \leq \int f(-Lf)\, d\pi$$

$$= \begin{cases} \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)Q(x,y)(f(y)-f(x))^2 & \text{for Markov chains,} \\ \int |\nabla f|^2 \, d\pi & \text{for Langevin dynamics.} \end{cases}$$

- We have the general mixing time bound

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\gamma} \log \frac{\chi^2(\mu \,\|\, \pi)}{4\varepsilon^2}.$$

χ^2-divergence, mixing time and Poincaré inequality

- The χ^2-divergence is $\chi^2(\mu \,\|\, \pi) = \left\| \frac{d\mu}{d\pi} - 1 \right\|^2_{L^2(\pi)} = \mathrm{Var}_\pi\left(\frac{d\mu}{d\pi}\right)$.

  By Cauchy–Schwarz, $\chi^2(\mu \,\|\, \pi) \geq 4\mathrm{TV}(\mu, \pi)^2$.
- The **spectral gap/Poincaré constant** of the Markov chain is the best constant $\gamma > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$,

$$\chi^2(\mu P_t \,\|\, \pi) \leq \exp(-2\gamma t)\chi^2(\mu \,\|\, \pi).$$

- Equivalently, γ is the best constant in the Poincaré inequality for π:

$$\gamma \mathrm{Var}_\pi(f) \leq \int f(-Lf)\, d\pi$$

$$= \begin{cases} \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)Q(x,y)(f(y)-f(x))^2 & \text{for Markov chains,} \\ \int |\nabla f|^2 \, d\pi & \text{for Langevin dynamics.} \end{cases}$$

- We have the general mixing time bound

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\gamma} \log \frac{\chi^2(\mu \,\|\, \pi)}{4\varepsilon^2}.$$

Pros: γ is often easy to estimate.
Cons: $\chi^2(\mu \,\|\, \pi)$ is often too large.

Relative entropy and mixing time

- The relative entropy is $\mathcal{H}(\mu \,|\, \pi) = \int \log \frac{d\mu}{d\pi} d\mu \geq 0$.
  By Pinsker's inequality $\mathcal{H}(\mu \,|\, \pi) \geq 2\mathrm{TV}(\mu, \pi)^2$.

Relative entropy and mixing time

- The relative entropy is $\mathcal{H}(\mu \,|\, \pi) = \int \log \frac{d\mu}{d\pi} d\mu \geq 0$.
  By Pinsker's inequality $\mathcal{H}(\mu \,|\, \pi) \geq 2\mathrm{TV}(\mu, \pi)^2$.
- The **modified log-Sobolev (MLSI) constant** of the Markov chain is the best constant $\beta > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$

$$\mathcal{H}(\mu P_t \,|\, \pi) \leq \exp(-2\beta t)\mathcal{H}(\mu \,|\, \pi).$$

Relative entropy and mixing time

- The relative entropy is $\mathcal{H}(\mu \,|\, \pi) = \int \log \frac{d\mu}{d\pi} d\mu \geq 0$.
  By Pinsker's inequality $\mathcal{H}(\mu \,|\, \pi) \geq 2\mathrm{TV}(\mu, \pi)^2$.
- The **modified log-Sobolev (MLSI) constant** of the Markov chain is the best constant $\beta > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$

$$\mathcal{H}(\mu P_t \,|\, \pi) \leq \exp(-2\beta t)\mathcal{H}(\mu \,|\, \pi).$$

- We have the following mixing time bound analogous to $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\gamma} \log \frac{\chi^2(\mu \,\|\, \pi)}{4\varepsilon^2}$:

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\beta} \log \frac{\mathcal{H}(\mu \,|\, \pi)}{2\varepsilon^2}.$$

Relative entropy and mixing time

- The relative entropy is $\mathcal{H}(\mu \,|\, \pi) = \int \log \frac{d\mu}{d\pi} d\mu \geq 0$.
  By Pinsker's inequality $\mathcal{H}(\mu \,|\, \pi) \geq 2\mathrm{TV}(\mu, \pi)^2$.
- The **modified log-Sobolev (MLSI) constant** of the Markov chain is the best constant $\beta > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$

$$\mathcal{H}(\mu P_t \,|\, \pi) \leq \exp(-2\beta t)\mathcal{H}(\mu \,|\, \pi).$$

- We have the following mixing time bound analogous to $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\gamma} \log \frac{\chi^2(\mu \,\|\, \pi)}{4\varepsilon^2}$:

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\beta} \log \frac{\mathcal{H}(\mu \,|\, \pi)}{2\varepsilon^2}.$$

Pros: $\mathcal{H}(\mu \,|\, \pi) \leq \log(1 + \chi^2(\mu \,\|\, \pi))$.
Cons: $\beta \leq \gamma$, and sometimes more difficult to estimate.

Relative entropy and mixing time

- The relative entropy is $\mathcal{H}(\mu \,|\, \pi) = \int \log \frac{d\mu}{d\pi} d\mu \geq 0$.
  By Pinsker's inequality $\mathcal{H}(\mu \,|\, \pi) \geq 2\mathrm{TV}(\mu, \pi)^2$.
- The **modified log-Sobolev (MLSI) constant** of the Markov chain is the best constant $\beta > 0$ such that for all $\mu \in \mathcal{P}(\Omega)$ and $t \geq 0$

$$\mathcal{H}(\mu P_t \,|\, \pi) \leq \exp(-2\beta t)\mathcal{H}(\mu \,|\, \pi).$$

- We have the following mixing time bound analogous to $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\gamma} \log \frac{\chi^2(\mu \,\|\, \pi)}{4\varepsilon^2}$:

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2\beta} \log \frac{\mathcal{H}(\mu \,|\, \pi)}{2\varepsilon^2}.$$

  Pros: $\mathcal{H}(\mu \,|\, \pi) \leq \log(1 + \chi^2(\mu \,\|\, \pi))$.
  Cons: $\beta \leq \gamma$, and sometimes more difficult to estimate.
- Hybrid bound: $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{\mathcal{H}(\mu \,|\, \pi)}{\gamma\varepsilon^2}$.

The cutoff phenomenon

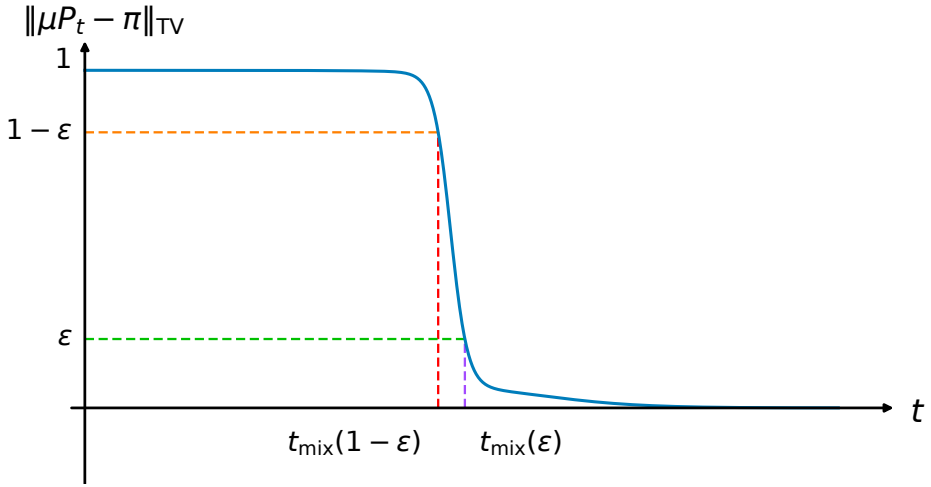
- The cutoff phenomenon is a surprising behavior of many high-dimensional Markov processes about the dependence of $\mathrm{t}_{\mathrm{mix}}(\varepsilon)$ on the precision ε.

The cutoff phenomenon

- The cutoff phenomenon is a surprising behavior of many high-dimensional Markov processes about the dependence of $\mathrm{t}_{\mathrm{mix}}(\varepsilon)$ on the precision ε.

- For fixed $0 < \varepsilon < \frac{1}{2}$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0$$

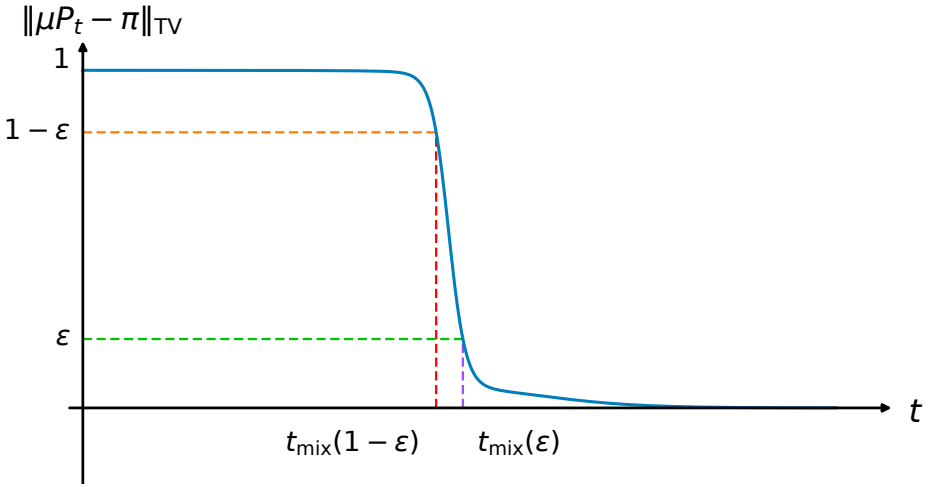


The cutoff phenomenon

- The cutoff phenomenon is a surprising behavior of many high-dimensional Markov processes about the dependence of $\mathrm{t}_{\mathrm{mix}}(\varepsilon)$ on the precision ε.

- For fixed $0 < \varepsilon < \frac{1}{2}$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0$$



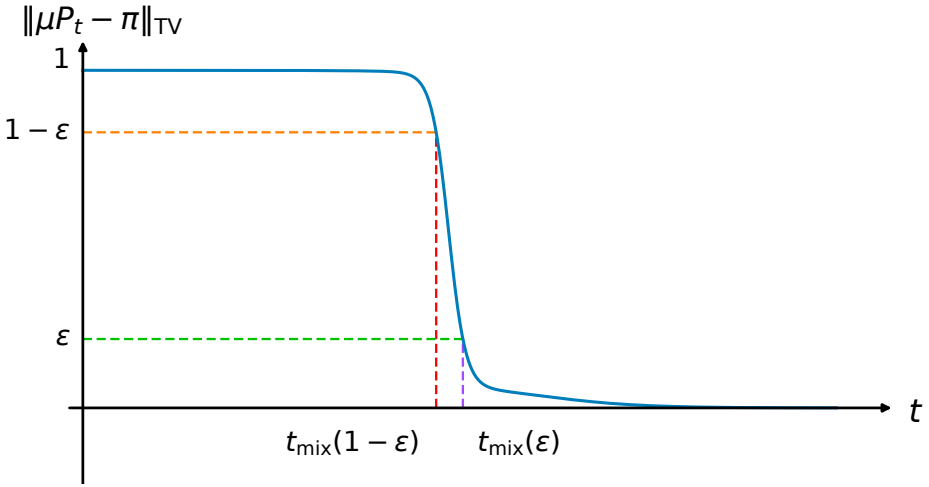
- *The careful work required to prove variation cutoffs often leads to a more or less complete understanding of the chain such that essentially any natural question can be answered.* [Dia96]

The cutoff phenomenon

- The cutoff phenomenon is a surprising behavior of many high-dimensional Markov processes about the dependence of $\mathrm{t}_{\mathrm{mix}}(\varepsilon)$ on the precision ε.

- For fixed $0 < \varepsilon < \frac{1}{2}$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0$$



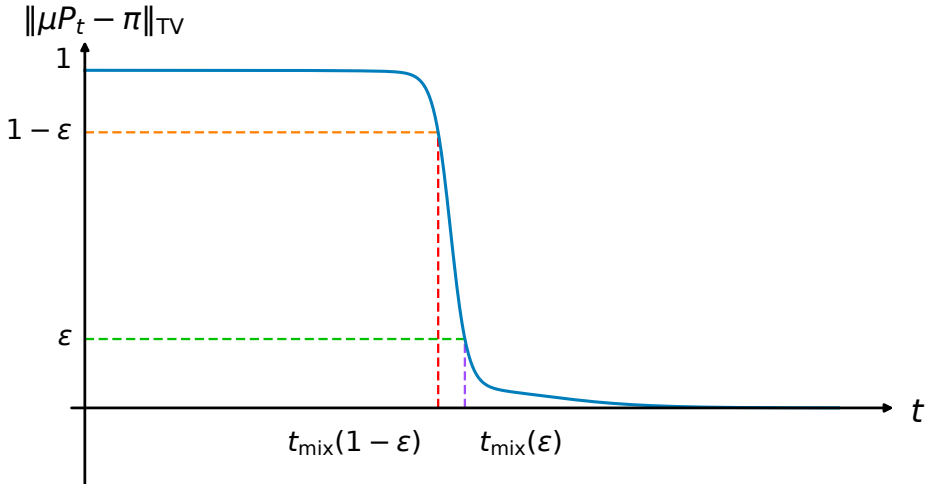
- *The careful work required to prove variation cutoffs often leads to a more or less complete understanding of the chain such that essentially any natural question can be answered.* [Dia96]
- **Peres' product conjecture (2004):** if $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma \to \infty$ for fixed $0 < \varepsilon < 1$, then cutoff occurs

The cutoff phenomenon

- The cutoff phenomenon is a surprising behavior of many high-dimensional Markov processes about the dependence of $\mathrm{t}_{\mathrm{mix}}(\varepsilon)$ on the precision ε.

- For fixed $0 < \varepsilon < \frac{1}{2}$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0$$



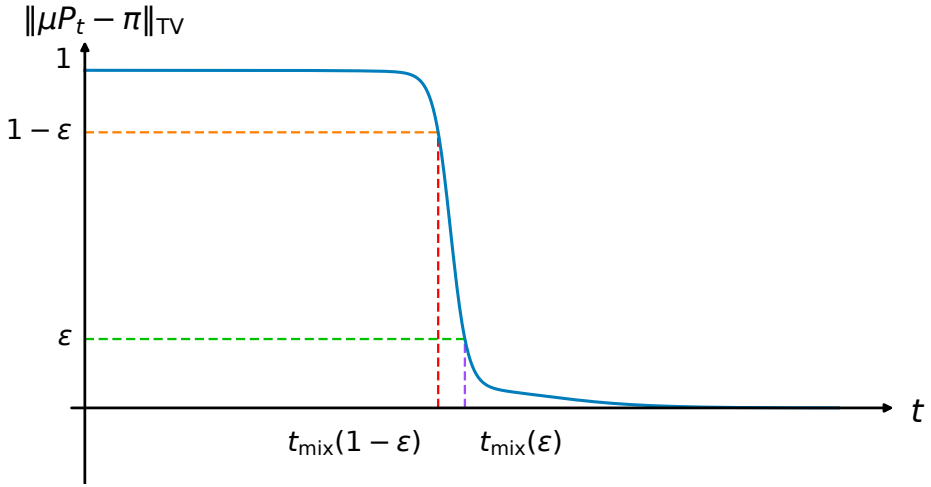
- *The careful work required to prove variation cutoffs often leads to a more or less complete understanding of the chain such that essentially any natural question can be answered.* [Dia96]
- **Peres' product conjecture (2004):** if $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma \to \infty$ for fixed $0 < \varepsilon < 1$, then cutoff occurs
- **Generic counterexamples:** for $\Pi(x, y) = \pi(y)$ and $0 < \theta \ll 1$ appropriately chosen, consider the stochastic matrix $\tilde{Q} = (1-\theta)Q + \theta\Pi$.

The cutoff phenomenon

- The cutoff phenomenon is a surprising behavior of many high-dimensional Markov processes about the dependence of $\mathrm{t}_{\mathrm{mix}}(\varepsilon)$ on the precision ε.

- For fixed $0 < \varepsilon < \frac{1}{2}$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0$$



- *The careful work required to prove variation cutoffs often leads to a more or less complete understanding of the chain such that essentially any natural question can be answered.* [Dia96]
- **Peres' product conjecture (2004):** if $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma \to \infty$ for fixed $0 < \varepsilon < 1$, then cutoff occurs for "typical Markov chains".
- **Generic counterexamples:** for $\Pi(x, y) = \pi(y)$ and $0 < \theta \ll 1$ appropriately chosen, consider the stochastic matrix $\tilde{Q} = (1-\theta)Q + \theta\Pi$.

Cutoff for Langevin dynamics $dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t$

Definition

A probability density $\pi = e^{-V} \in \mathcal{P}(\mathbb{R}^d)$ is log-concave if $V : \mathbb{R}^d \to \mathbb{R}$ is convex.

Cutoff for Langevin dynamics $dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t$

Definition

A probability density $\pi = e^{-V} \in \mathcal{P}(\mathbb{R}^d)$ is log-concave if $V : \mathbb{R}^d \to \mathbb{R}$ is convex.

Theorem ([Sal26], P-Salez 2025)

*Suppose that π is **log-concave** and $X_0 \sim \mu = \delta_x$. Then, for any fixed $0 < \varepsilon < \frac{1}{2}$,*

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon) \leq C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

In particular, the product condition

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma(\pi) \to \infty \qquad \text{for some } 0 < \varepsilon < 1$$

implies cutoff:

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0 \qquad \text{for all } 0 < \varepsilon < \frac{1}{2}.$$

Proof sketch: Wasserstein distances

- Fix $0 < \varepsilon < \frac{1}{2}$ (e.g. $\varepsilon = 0.01$) and set $t_0 := \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$. The goal is to prove

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\left(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\right).$$

Proof sketch: Wasserstein distances

- Fix $0 < \varepsilon < \frac{1}{2}$ (e.g. $\varepsilon = 0.01$) and set $t_0 := \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$. The goal is to prove

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

- Recall the general bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{\mathcal{H}(\mu \,|\, \pi)}{\gamma \varepsilon^2}$. Applying this to μP_{t_0}

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + \frac{C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)}{\gamma(\pi)}.$$

Proof sketch: Wasserstein distances

- Fix $0 < \varepsilon < \frac{1}{2}$ (e.g. $\varepsilon = 0.01$) and set $t_0 := \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$. The goal is to prove

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

- Recall the general bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \frac{\mathcal{H}(\mu \,|\, \pi)}{\gamma \varepsilon^2}$. Applying this to μP_{t_0}

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + \frac{C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)}{\gamma(\pi)}.$$

- **Problem:** $\mathrm{TV}(\mu P_{t_0}, \pi) \leq 1-\varepsilon \implies \mathcal{H}(\mu P_{t_0} \,|\, \pi) < \infty$.

Proof sketch: Wasserstein distances

- Fix $0 < \varepsilon < \frac{1}{2}$ (e.g. $\varepsilon = 0.01$) and set $t_0 := \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$. The goal is to prove

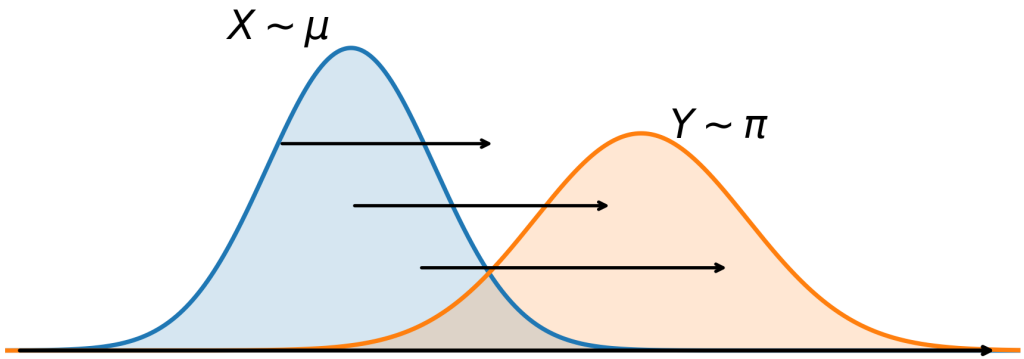
$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \le t_0 + C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

- Recall the general bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \le \frac{\mathcal{H}(\mu \,|\, \pi)}{\gamma \varepsilon^2}$. Applying this to μP_{t_0}

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \le t_0 + \frac{C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)}{\gamma(\pi)}.$$

- **Problem:** $\mathrm{TV}(\mu P_{t_0}, \pi) \le 1-\varepsilon \implies \mathcal{H}(\mu P_{t_0} \,|\, \pi) < \infty$.
- Detour: introduce the **Wasserstein distance** W_2.
  Lift Euclidean distance to $\mathcal{P}(\mathbb{R}^d)$, so that $W_2(\delta_x, \delta_y) = |x-y|$:

$$W_2(\mu, \pi) = \inf_{\substack{X \sim \mu \\ Y \sim \pi}} \sqrt{\mathbb{E}\Big[|X-Y|^2\Big]}.$$



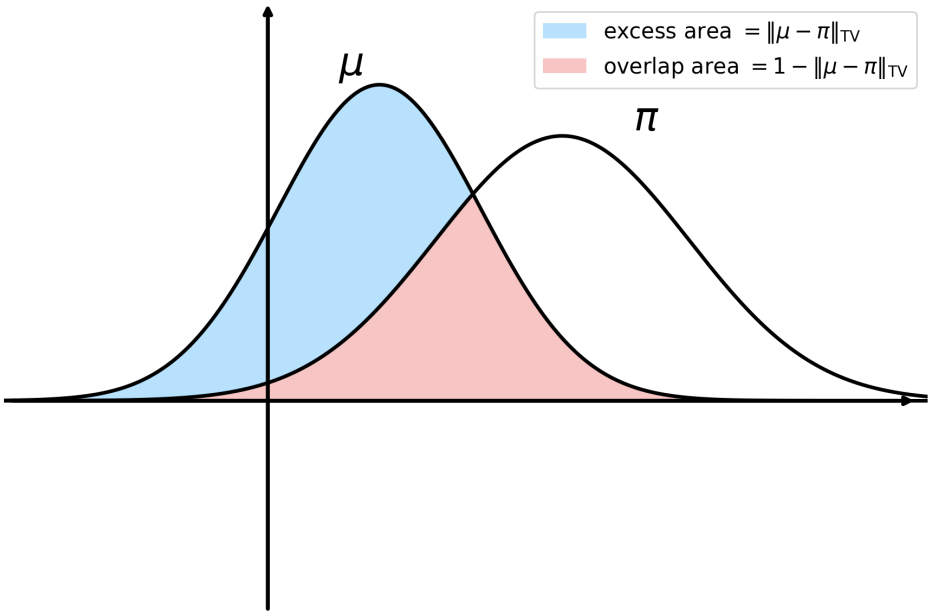
Proof sketch: bounded support and naive approach

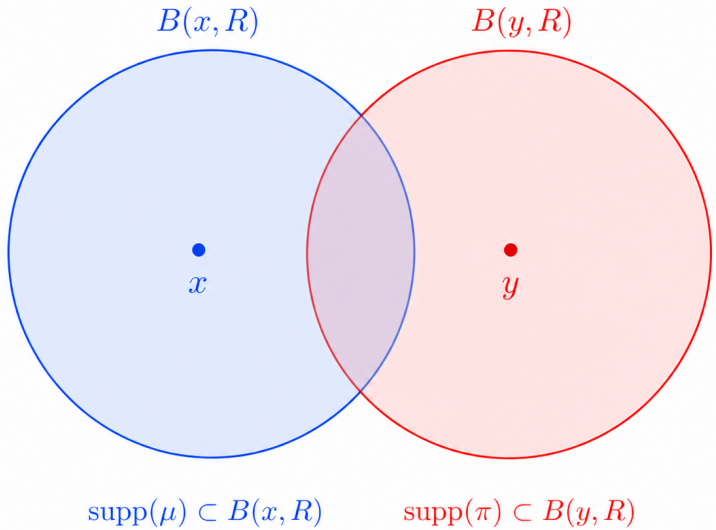
- **Question:** is $W_2(\mu P_{t_0}, \pi)$ small, given $\mathrm{TV}(\mu P_{t_0}, \pi) \leq 1 - \varepsilon$?
  In general, no: consider $\mu = \varepsilon\delta_0 + (1-\varepsilon)\delta_M$, $\pi = \delta_0 \in \mathcal{P}(\mathbb{R})$ for $M \gg 1$.

Proof sketch: bounded support and naive approach

- **Question:** is $W_2(\mu P_{t_0}, \pi)$ small, given $\mathrm{TV}(\mu P_{t_0}, \pi) \leq 1 - \varepsilon$?
  In general, no: consider $\mu = \varepsilon\delta_0 + (1-\varepsilon)\delta_M$, $\pi = \delta_0 \in \mathcal{P}(\mathbb{R})$ for $M \gg 1$.
- **Observation:** suppose $\mathrm{Supp}(\mu) \subset B(x, R)$ and $\mathrm{Supp}(\pi) \subset B(y, R)$. Then,

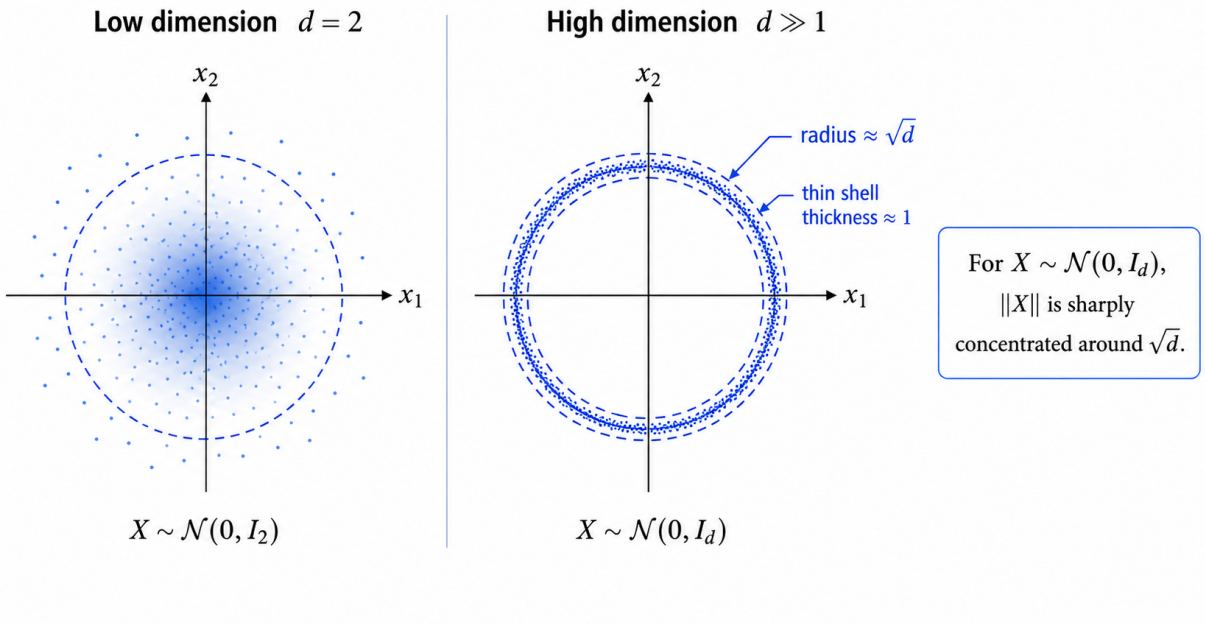
$$\mathrm{TV}(\mu, \pi) < 1 \implies W_2(\mu, \pi) \leq 4R.$$





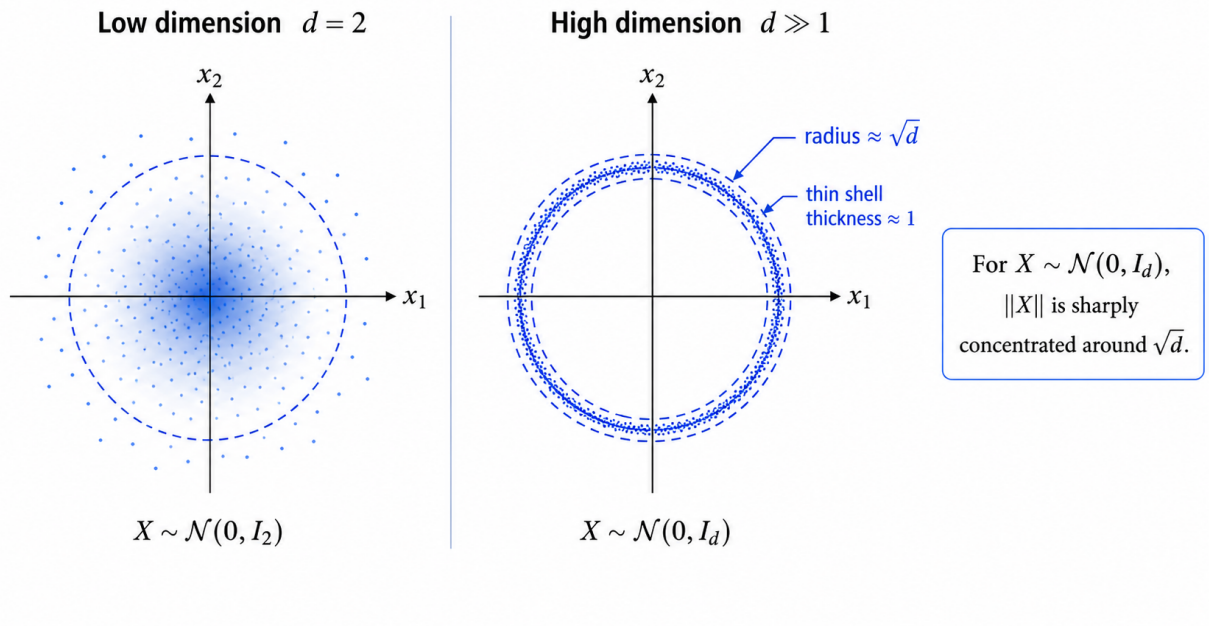
Proof sketch: concentration of measure

- Concentration of measure: for many natural high-dimensional distributions, most of the mass is concentrated in a small region.



Proof sketch: concentration of measure

- Concentration of measure: for many natural high-dimensional distributions, most of the mass is concentrated in a small region.



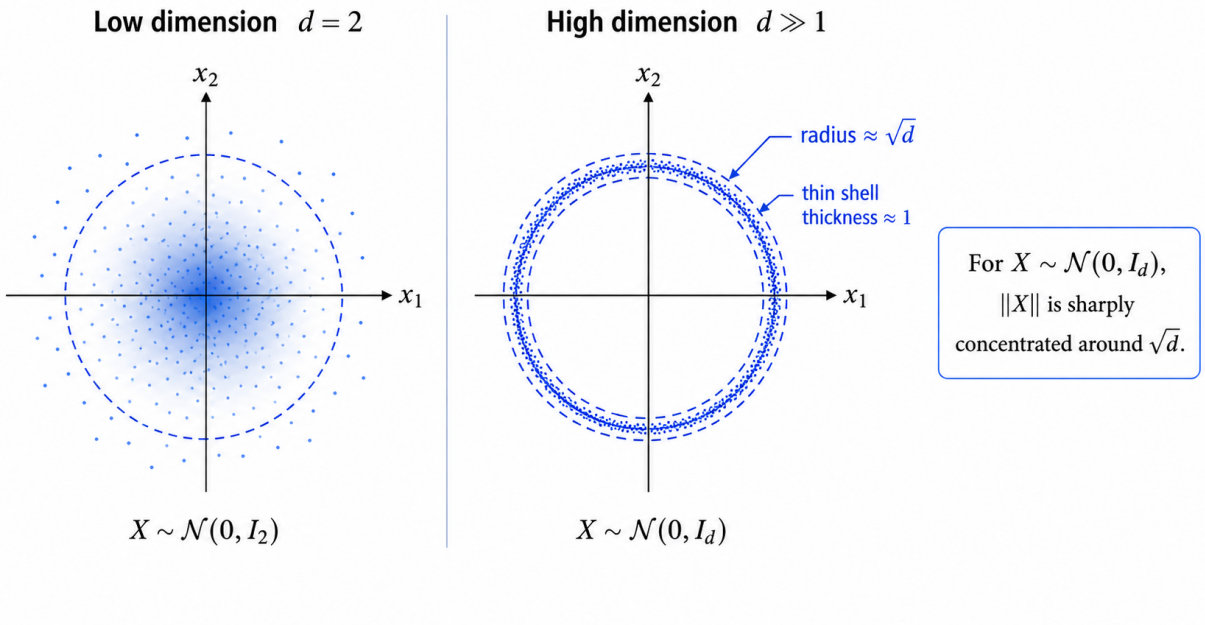
- ~~Bounded support~~ $\rightarrow$ Poincaré inequality

$$\gamma(\pi)\mathrm{Var}_\pi(f) \leq \int |\nabla f|^2 \, d\pi \quad \text{for all } f \in C_b^1(\mathbb{R}^d).$$

Bigger $\gamma(\pi) \iff$ stronger concentration.

Proof sketch: concentration of measure

- Concentration of measure: for many natural high-dimensional distributions, most of the mass is concentrated in a small region.



- ~~Bounded support~~ $\rightarrow$ Poincaré inequality

$$\gamma(\pi)\mathrm{Var}_\pi(f) \leq \int |\nabla f|^2 \, d\pi \quad \text{for all } f \in C_b^1(\mathbb{R}^d).$$

Bigger $\gamma(\pi) \iff$ stronger concentration.
Transport-variance inequality: $\gamma(\pi) W_2^2(\mu, \pi) \leq 2\chi^2(\mu \,\|\, \pi)$ for all $\mu \in \mathcal{P}(\mathbb{R}^d)$ [Liu20].

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Lemma (W_2-TV inequality under Poincaré inequalities)

For all $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ we have $W_2^2(\mu,\pi) \leq 4\big(\gamma(\mu)^{-1} + \gamma(\pi)^{-1}\big)\frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.

Proof sketch: local concentration and W_2-bound at t_{mix}

Lemma (W_2-TV inequality under Poincaré inequalities)

For all $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ *we have* $\quad W_2^2(\mu,\pi) \le 4\big(\gamma(\mu)^{-1} + \gamma(\pi)^{-1}\big)\frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.

Proof.

1. For $\nu = \frac{\mu \wedge \pi}{1-\mathrm{TV}(\mu,\pi)}$ the common mass of μ, π, we have $\chi^2(\nu \,\|\, \cdot) \le \frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.
2. Use $W_2(\mu,\pi) \le W_2(\mu,\nu) + W_2(\nu,\pi)$ and apply transport-variance to both.

□

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Lemma (W_2-TV inequality under Poincaré inequalities)

For all $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ we have $W_2^2(\mu,\pi) \le 4\big(\gamma(\mu)^{-1} + \gamma(\pi)^{-1}\big)\frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.

Proof.

1. For $\nu = \frac{\mu\wedge\pi}{1-\mathrm{TV}(\mu,\pi)}$ the common mass of μ,π, we have $\chi^2(\nu\,\|\,\cdot) \le \frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.
2. Use $W_2(\mu,\pi) \le W_2(\mu,\nu) + W_2(\nu,\pi)$ and apply transport-variance to both.

□

Recall we want to upper bound $W_2(\mu P_{t_0},\pi)$, where $t_0 = \mathrm{t_{mix}}(1-\varepsilon)$.

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Lemma (W_2-TV inequality under Poincaré inequalities)

For all $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ we have $W_2^2(\mu,\pi) \le 4\big(\gamma(\mu)^{-1} + \gamma(\pi)^{-1}\big)\frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.

Proof.

1. For $\nu = \frac{\mu\wedge\pi}{1-\mathrm{TV}(\mu,\pi)}$ the common mass of μ,π, we have $\chi^2(\nu\,\|\,\cdot) \le \frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}$.
2. Use $W_2(\mu,\pi) \le W_2(\mu,\nu) + W_2(\nu,\pi)$ and apply transport-variance to both.

□

Recall we want to upper bound $W_2(\mu P_{t_0},\pi)$, where $t_0 = \mathrm{t_{mix}}(1-\varepsilon)$.

Lemma (Local Poincaré inequality [BÉ85])

If $\pi \propto e^{-V}$ is log-concave, then $\gamma(\delta_x P_t) \ge \frac{1}{2t}$.

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Lemma (Local Poincaré inequality)

If $\pi \propto e^{-V}$ is log-concave, then $\gamma(\delta_x P_t) \geq \frac{1}{2t}$.

Heuristics.

Let $\begin{cases} X_0 \sim \mu, & dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t \\ Y_0 \sim \nu, & dY_t = -\nabla V(Y_t)dt + \sqrt{2}dB_t \end{cases}$ with $\mathbb{E}\left[|X_0 - Y_0|^2\right] = W_2^2(\mu, \nu)$.

- $\frac{\mathrm{d}}{\mathrm{d}t}|X_t - Y_t|^2 = -2\langle \nabla V(X_t) - \nabla V(Y_t), X_t - Y_t \rangle \leq 0$ by convexity of V
  $\implies W_2(\mu P_t, \nu P_t) \leq W_2(\mu, \nu)$.

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Lemma (Local Poincaré inequality)

If $\pi \propto e^{-V}$ is log-concave, then $\gamma(\delta_x P_t) \geq \frac{1}{2t}$.

Heuristics.

Let $\begin{cases} X_0 \sim \mu, & dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t \\ Y_0 \sim \nu, & dY_t = -\nabla V(Y_t)dt + \sqrt{2}dB_t \end{cases}$ with $\mathbb{E}\left[|X_0 - Y_0|^2\right] = W_2^2(\mu, \nu)$.

- $\frac{\mathrm{d}}{\mathrm{d}t}|X_t - Y_t|^2 = -2\langle \nabla V(X_t) - \nabla V(Y_t), X_t - Y_t\rangle \leq 0$ by convexity of V $\implies W_2(\mu P_t, \nu P_t) \leq W_2(\mu, \nu)$.
- If $V = 0$, then $\delta_x P_t = \mathrm{law}(X_t) = \mathcal{N}(x, 2t)$ and the conclusion holds.

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Lemma (Local Poincaré inequality)

If $\pi \propto e^{-V}$ is log-concave, then $\gamma(\delta_x P_t) \geq \frac{1}{2t}$.

Heuristics.

Let $\begin{cases} X_0 \sim \mu, & dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t \\ Y_0 \sim \nu, & dY_t = -\nabla V(Y_t)dt + \sqrt{2}dB_t \end{cases}$ with $\mathbb{E}\left[|X_0 - Y_0|^2\right] = W_2^2(\mu, \nu)$.

- $\frac{\mathrm{d}}{\mathrm{d}t}|X_t - Y_t|^2 = -2\langle \nabla V(X_t) - \nabla V(Y_t), X_t - Y_t\rangle \leq 0$ by convexity of V $\implies W_2(\mu P_t, \nu P_t) \leq W_2(\mu, \nu)$.
- If $V = 0$, then $\delta_x P_t = \mathrm{law}(X_t) = \mathcal{N}(x, 2t)$ and the conclusion holds.
- By comparison, the claim holds for convex V too.

□

Proof sketch: local concentration and W_2-bound at $\mathrm{t_{mix}}$

Question: is $W_2(\delta_x P_{t_0}, \pi)$ small at $t_0 = \mathrm{t_{mix}}(1-\varepsilon)$?

Lemma (W_2-TV inequality)

For all $\mu, \pi \in \mathcal{P}(\mathbb{R}^d)$ we have the bound

$$W_2^2(\mu,\pi) \le 4\big(\gamma(\mu)^{-1} + \gamma(\pi)^{-1}\big)\frac{\mathrm{TV}(\mu,\pi)}{1-\mathrm{TV}(\mu,\pi)}.$$

Lemma (Local Poincaré inequality [BÉ85])

If $\pi \propto e^{-V}$ is log-concave, then $\gamma(\delta_x P_t) \ge \frac{1}{2t}$.

Corollary (W_2-bound at $\mathrm{t_{mix}}$)

If π is log-concave, $W_2^2(\delta_x P_{t_0}, \pi) \le \frac{8(t_0+1/\gamma(\pi))}{\varepsilon} = C(\varepsilon)(t_0 + 1/\gamma(\pi))$.

Proof sketch: Entropy regularization and conclusion

Theorem

*If π is **log-concave** and $X_0 \sim \mu = \delta_x$, then, for any $0 < \varepsilon < \frac{1}{2}$,*

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon) \leq C(\varepsilon)\left(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\right).$$

- Recall the bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)/\gamma(\pi)$ for $t_0 = \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$.
- We have $W_2^2(\mu P_{t_0}, \pi) \leq C(\varepsilon)(t_0 + 1/\gamma(\pi))$

Proof sketch: Entropy regularization and conclusion

Theorem

*If π is **log-concave** and $X_0 \sim \mu = \delta_x$, then, for any $0 < \varepsilon < \frac{1}{2}$,*

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon) \leq C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

- Recall the bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)/\gamma(\pi)$ for $t_0 = \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$.
- We have $W_2^2(\mu P_{t_0}, \pi) \leq C(\varepsilon)(t_0 + 1/\gamma(\pi))$ but $\not\Longrightarrow$ $\mathcal{H}(\mu P_{t_0} \,|\, \pi) < \infty$.

Proof sketch: Entropy regularization and conclusion

Theorem

*If π is **log-concave** and $X_0 \sim \mu = \delta_x$, then, for any $0 < \varepsilon < \frac{1}{2}$,*

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon) \leq C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

- Recall the bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)/\gamma(\pi)$ for $t_0 = \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$.
- We have $W_2^2(\mu P_{t_0}, \pi) \leq C(\varepsilon)(t_0 + 1/\gamma(\pi))$ but $\not\Longrightarrow$ $\mathcal{H}(\mu P_{t_0} \,|\, \pi) < \infty$.

Lemma (Entropy-Wasserstein regularization [BGL01; OV01])

If π is log-concave, $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $s \geq 0$, we have $\mathcal{H}(\mu P_s \,|\, \pi) \leq \frac{W_2^2(\mu,\pi)}{4s}$.

Proof sketch: Entropy regularization and conclusion

Theorem

*If π is **log-concave** and $X_0 \sim \mu = \delta_x$, then, for any $0 < \varepsilon < \frac{1}{2}$,*

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon) \leq C(\varepsilon)\Big(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\Big).$$

- Recall the bound $\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + C(\varepsilon)\mathcal{H}(\mu P_{t_0} \,|\, \pi)/\gamma(\pi)$ for $t_0 = \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)$.
- We have $W_2^2(\mu P_{t_0}, \pi) \leq C(\varepsilon)(t_0 + 1/\gamma(\pi))$ but $\not\Longrightarrow$ $\mathcal{H}(\mu P_{t_0} \,|\, \pi) < \infty$.

Lemma (Entropy-Wasserstein regularization [BGL01; OV01])

If π is log-concave, $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $s \geq 0$, we have $\mathcal{H}(\mu P_s \,|\, \pi) \leq \frac{W_2^2(\mu,\pi)}{4s}$.

For $s > 0$, use the mixing estimate below, and optimize over s to conclude

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq t_0 + s + C(\varepsilon)\frac{\mathcal{H}(\mu P_{t_0+s} \,|\, \pi)}{\gamma(\pi)} \leq t_0 + s + \frac{C(\varepsilon) W_2^2(\mu P_{t_0}, \pi)}{s\gamma(\pi)}.$$

Wasserstein gradient flows and entropy regularization [OV01]

Fact (Classical convex optimization)

Let $F\colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ *be smooth, with* $F(x^*) = 0 = \min F$. *Consider a gradient flow curve for* F, *given by* $\frac{\mathrm{d}}{\mathrm{d}t} y_t = -\nabla F(y_t)$.

- *If* F *is convex, then* $F(y_t) \leq \frac{|y_0 - x^*|^2}{4t}$.
- F *is convex* $\iff$ *for all gradient flow curves* $(y_t)_{t\geq 0}, (\tilde{y}_t)_{t\geq 0}$ $\frac{\mathrm{d}}{\mathrm{d}t}|y_t - \tilde{y}_t|^2 = -2\langle \nabla F(y_t) - \nabla F(\tilde{y}_t), y_t - \tilde{y}_t\rangle \leq 0$

Wasserstein gradient flows and entropy regularization [OV01]

Fact (Classical convex optimization)

Let $F\colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be smooth, with $F(x^*) = 0 = \min F$. Consider a gradient flow curve for F, given by $\frac{\mathrm{d}}{\mathrm{d}t} y_t = -\nabla F(y_t)$.

- *If F is convex, then $F(y_t) \leq \frac{|y_0 - x^*|^2}{4t}$.*
- *F is convex $\iff$ for all gradient flow curves $(y_t)_{t\geq 0}, (\tilde{y}_t)_{t\geq 0}$ $\frac{\mathrm{d}}{\mathrm{d}t}|y_t - \tilde{y}_t|^2 = -2\langle \nabla F(y_t) - \nabla F(\tilde{y}_t), y_t - \tilde{y}_t\rangle \leq 0$*

Theorem ([JKO98])

For the Langevin dynamics, the curve $(\mu P_t)_{t\geq 0}$ is the gradient flow of the relative entropy functional $\mathcal{H}(\cdot \,|\, \pi)$ in the Wasserstein space $(\mathcal{P}(\mathbb{R}^d), W_2)$.

Wasserstein gradient flows and entropy regularization [OV01]

Fact (Classical convex optimization)

Let $F\colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be smooth, with $F(x^*) = 0 = \min F$. Consider a gradient flow curve for F, given by $\frac{\mathrm{d}}{\mathrm{d}t} y_t = -\nabla F(y_t)$.

- *If F is convex, then $F(y_t) \leq \frac{|y_0 - x^*|^2}{4t}$.*
- *F is convex $\iff$ for all gradient flow curves $(y_t)_{t\geq 0}, (\tilde{y}_t)_{t\geq 0}$ $\frac{\mathrm{d}}{\mathrm{d}t}|y_t - \tilde{y}_t|^2 = -2\langle \nabla F(y_t) - \nabla F(\tilde{y}_t), y_t - \tilde{y}_t\rangle \leq 0$*

Theorem ([JKO98])

For the Langevin dynamics, the curve $(\mu P_t)_{t\geq 0}$ is the gradient flow of the relative entropy functional $\mathcal{H}(\cdot\,|\,\pi)$ in the Wasserstein space $(\mathcal{P}(\mathbb{R}^d), W_2)$.

π log-concave $\implies$ $\frac{\mathrm{d}}{\mathrm{d}t} W_2(\mu P_t, \nu P_t) \leq 0$ $\implies$ $\mathcal{H}(\cdot\,|\,\pi)$ convex in $(\mathcal{P}(\mathbb{R}^d), W_2)$.

Corollary ([OV01])

If π is log-concave, $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $s \geq 0$, we have $\mathcal{H}(\mu P_s\,|\,\pi) \leq \frac{W_2^2(\mu,\pi)}{4s}$.

Recap

Theorem

If $\pi = e^{-V}$ *is* ***log-concave*** *and* $X_0 \sim \mu = \delta_x$, *then, for any* $0 < \varepsilon < \frac{1}{2}$,

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon) \leq C(\varepsilon)\left(1/\gamma(\pi) + \sqrt{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)/\gamma(\pi)}\right)$$

Hence, the product condition $\mathrm{t}_{\mathrm{mix}}(\delta) \cdot \gamma(\pi) \to \infty$ *for* $0 < \delta < 1$ *implies cutoff:*

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) - \mathrm{t}_{\mathrm{mix}}(1-\varepsilon)}{\mathrm{t}_{\mathrm{mix}}(1-\varepsilon)} \to 0 \quad \text{for all } 0 < \varepsilon < \frac{1}{2}.$$

$$\begin{aligned}
\mathrm{TV}(\mu P_{t_0}, \pi) \leq 1-\varepsilon \quad &\Longrightarrow \quad W_2(\mu P_{t_0}, \pi) < \infty \quad \text{if both measures are concentrated} \\
&\Longrightarrow \quad \mathcal{H}(\mu P_{t_0+s} \,|\, \pi) \quad \text{becomes small if } s > 0 \text{ large enough} \\
&\Longrightarrow \quad \mathrm{TV}(\mu P_{t_0+s+r}, \pi) \leq \varepsilon \quad \text{quickly, after extra time } r.
\end{aligned}$$

Discrete Markov chains and entropic curvature

- **Setting:** Ω is a finite space and Q a stochastic matrix reversible with respect to a probability measure $\pi \in \mathcal{P}(\Omega)$. We consider the continuous-time Markov chain with generator $L = Q - I$ and semigroup $P_t = e^{tL}$.
- **Key ingredients for LD:** log-concavity of $\pi = e^{-V}$, i.e. $\nabla^2 V \succeq 0$:

$$W_2(\mu P_t, \nu P_t) \leq W_2(\mu, \nu) \implies \begin{cases} W_2^2(\nu, \delta_x P_t) \leq 2t\chi^2(\nu \,\|\, \delta_x P_t), \\ \mathcal{H}(\mu P_t \,|\, \pi) \leq \frac{W_2^2(\mu,\pi)}{4t}. \end{cases}$$

Discrete Markov chains and entropic curvature

- **Setting:** Ω is a finite space and Q a stochastic matrix reversible with respect to a probability measure $\pi \in \mathcal{P}(\Omega)$. We consider the continuous-time Markov chain with generator $L = Q - I$ and semigroup $P_t = e^{tL}$.
- **Key ingredients for LD:** log-concavity of $\pi = e^{-V}$, i.e. $\nabla^2 V \succeq 0$:

$$W_2(\mu P_t, \nu P_t) \leq W_2(\mu, \nu) \implies \begin{cases} W_2^2(\nu, \delta_x P_t) \leq 2t\chi^2(\nu \,\|\, \delta_x P_t), \\ \mathcal{H}(\mu P_t \,|\, \pi) \leq \frac{W_2^2(\mu,\pi)}{4t}. \end{cases}$$

Theorem (/Definition [Maa11; EM12])

- *There exists a metric $\mathscr{W}$ on $\mathcal{P}(\Omega)$ such that $(\mu P_t)_{t\geq 0}$ is the gradient flow of $\mathcal{H}(\cdot\,|\,\pi)$ in $(\mathcal{P}(\Omega), \mathscr{W})$.*
- *The Markov chain has **non-negative entropic curvature** if*

$$\mathscr{W}(\mu P_t, \nu P_t) \leq \mathscr{W}(\mu, \nu) \quad \text{for all } \mu, \nu \in \mathcal{P}(\Omega), t \geq 0.$$

Cutoff and entropic curvature

Let (Ω, π, Q) be a Markov chain with non-negative entropic curvature.

- **Entropy regularization [EF18]:** $\mathcal{H}(\mu P_t \,|\, \pi) \leq \frac{\mathscr{W}(\mu,\pi)^2}{4t}$ for $\mu \in \mathcal{P}(\Omega), t > 0$.

Cutoff and entropic curvature

Let (Ω, π, Q) be a Markov chain with non-negative entropic curvature.

- **Entropy regularization [EF18]:** $\mathcal{H}(\mu P_t \,|\, \pi) \leq \frac{\mathscr{W}(\mu,\pi)^2}{4t}$ for $\mu \in \mathcal{P}(\Omega), t > 0$.
- Write $x \sim y$ if $Q(x,y) > 0$ and denote by diam the diameter of this graph, and by $\Delta = \max_{x\sim y} \frac{1}{Q(x,y)}$ the "maximum degree".
- **Local concentration:** $\mathscr{W}^2(\nu, \delta_x P_t) \lesssim t(\log \Delta)\chi^2(\nu \,\|\, \delta_x P_t)$ for $t \geq \text{diam}/4$.

Cutoff and entropic curvature

Let (Ω, π, Q) be a Markov chain with non-negative entropic curvature.

- **Entropy regularization [EF18]:** $\mathcal{H}(\mu P_t \,|\, \pi) \leq \frac{\mathscr{W}(\mu,\pi)^2}{4t}$ for $\mu \in \mathcal{P}(\Omega), t > 0$.
- Write $x \sim y$ if $Q(x,y) > 0$ and denote by diam the diameter of this graph, and by $\Delta = \max_{x\sim y} \frac{1}{Q(x,y)}$ the "maximum degree".
- **Local concentration:** $\mathscr{W}^2(\nu, \delta_x P_t) \lesssim t(\log \Delta)\chi^2(\nu \,\|\, \delta_x P_t)$ for $t \geq \text{diam}/4$.

Theorem

Let (Ω, π, Q) be a sequence of Markov chains with non-negative entropic curvature. Denote by $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \max_{x\in\Omega} \mathrm{t}_{\mathrm{mix}}(\delta_x, \varepsilon)$ the worst-initialization mixing time. Then, ***cutoff occurs*** *if the following modified product condition holds:*

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma}{\log \Delta} \to \infty. \tag{1}$$

Cutoff and entropic curvature

Let (Ω, π, Q) be a Markov chain with non-negative entropic curvature.

- **Entropy regularization [EF18]:** $\mathcal{H}(\mu P_t \,|\, \pi) \leq \frac{\mathscr{W}(\mu,\pi)^2}{4t}$ for $\mu \in \mathcal{P}(\Omega), t > 0$.
- Write $x \sim y$ if $Q(x,y) > 0$ and denote by diam the diameter of this graph, and by $\Delta = \max_{x\sim y} \frac{1}{Q(x,y)}$ the "maximum degree".
- **Local concentration:** $\mathscr{W}^2(\nu, \delta_x P_t) \lesssim t(\log \Delta)\chi^2(\nu \,\|\, \delta_x P_t)$ for $t \geq \text{diam}/4$.

Theorem

Let (Ω, π, Q) be a sequence of Markov chains with non-negative entropic curvature. Denote by $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \max_{x\in\Omega} \mathrm{t}_{\mathrm{mix}}(\delta_x, \varepsilon)$ the worst-initialization mixing time. Then, ***cutoff occurs*** *if the following modified product condition holds:*

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma}{\log \Delta} \to \infty. \tag{1}$$

Alternatively, using $\mathcal{H}(\mu P_t \,|\, \pi) \leq \exp(-2\beta t)\mathcal{H}(\mu \,|\, \pi)$, it suffices that

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \beta}{\log\log(\Delta + 2)} \to \infty. \tag{2}$$

Cutoff and entropic curvature - Remarks

Theorem

Let (Ω, π, Q) be a sequence of Markov chain with non-negative entropic curvature. Then, ***cutoff occurs*** *if any of the following modified product condition holds:*

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma}{\log \Delta} \to \infty \tag{3}$$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \beta}{\log \log(\Delta + 2)} \to \infty \tag{4}$$

- The counterexample $\tilde{Q} = (1 - \theta)Q + \theta\Pi$ for $0 < \theta \ll 1$ has $\Delta \gg 1$.

Cutoff and entropic curvature - Remarks

Theorem

Let (Ω, π, Q) be a sequence of Markov chain with non-negative entropic curvature. Then, ***cutoff occurs*** *if any of the following modified product condition holds:*

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma}{\log \Delta} \to \infty \tag{3}$$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \beta}{\log \log(\Delta + 2)} \to \infty \tag{4}$$

- The counterexample $\tilde{Q} = (1-\theta)Q + \theta\Pi$ for $0 < \theta \ll 1$ has $\Delta \gg 1$.
- Biased birth and death chain on $\{0, \dots N\}$ with birth rate $p > \frac{1}{2}$ fixed: $\mathrm{t}_{\mathrm{mix}} \gtrsim N$, $\Delta = (1-p)^{-1}$, $\gamma \geq f(p) > 0$ and (3) predicts cutoff.

Cutoff and entropic curvature - Remarks

Theorem

Let (Ω, π, Q) be a sequence of Markov chain with non-negative entropic curvature. Then, ***cutoff occurs*** *if any of the following modified product condition holds:*

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \gamma}{\log \Delta} \to \infty \tag{3}$$

$$\frac{\mathrm{t}_{\mathrm{mix}}(\varepsilon) \cdot \beta}{\log \log(\Delta + 2)} \to \infty \tag{4}$$

- The counterexample $\tilde{Q} = (1-\theta)Q + \theta\Pi$ for $0 < \theta \ll 1$ has $\Delta \gg 1$.
- Biased birth and death chain on $\{0, \dots N\}$ with birth rate $p > \frac{1}{2}$ fixed: $\mathrm{t}_{\mathrm{mix}} \gtrsim N$, $\Delta = (1-p)^{-1}$, $\gamma \geq f(p) > 0$ and (3) predicts cutoff.
- Examples on $\Omega = \{-1, 1\}^n$, where (4) applies:
  1. the simple random walk ($\Delta = n, \beta \gtrsim \frac{1}{n}, \mathrm{t}_{\mathrm{mix}} \gtrsim n \log n$);
  2. the Ising model in the regime $d\left(1 - e^{-2\beta}\right)e^{2\beta d} \leq 1$;
  3. the hardcore model in the regime $\lambda d \leq 1$.

References

[BÉ85] D. Bakry and Michel Émery. "Diffusions hypercontractives". In: *Séminaire de probabilités, XIX, 1983/84*. Vol. 1123. Lecture Notes in Math. Springer, Berlin, 1985, pp. 177–206.

[BGL01] Sergey G. Bobkov, Ivan Gentil, and Michel Ledoux. "Hypercontractivity of Hamilton-Jacobi equations". In: *J. Math. Pures Appl. (9)* 80.7 (2001), pp. 669–696.

[Dia96] Persi Diaconis. "The cutoff phenomenon in finite Markov chains". In: *Proc. Natl. Acad. Sci. USA* 93.4 (1996), pp. 1659–1664.

[EF18] Matthias Erbar and Max Fathi. "Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature". In: *Journal of Functional Analysis* 274.11 (2018), pp. 3056–3089.

[EM12] Matthias Erbar and Jan Maas. "Ricci curvature of finite Markov chains via convexity of the entropy". In: *Arch. Ration. Mech. Anal.* 206.3 (2012), pp. 997–1038.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. "The variational formulation of the Fokker-Planck equation". In: *SIAM J. Math. Anal.* 29.1 (1998), pp. 1–17.

[Liu20] Yuan Liu. "The Poincaré inequality and quadratic transportation-variance inequalities". In: *Electron. J. Probab.* 25 (2020), Paper No. 1, 16.

[Maa11] Jan Maas. "Gradient flows of the entropy for finite Markov chains". In: *J. Funct. Anal.* 261.8 (2011), pp. 2250–2292.

[OV01] Felix Otto and Cédric Villani. "Comment on: "Hypercontractivity of Hamilton–Jacobi equations", by S. Bobkov, I. Gentil and M. Ledoux". In: *Journal de Mathématiques Pures et Appliquées* 80.7 (2001), pp. 697–700.

[PS25] Francesco Pedrotti and Justin Salez. "A transport approach to the cutoff phenomenon". In: *arXiv preprint arXiv:2509.08560* (2025).

[Sal26] Justin Salez. "Cutoff for non-negatively curved diffusions". In: *Inventiones mathematicae* 246.1 (2026), pp. 1–11.

[Vil09] Cédric Villani. *Optimal transport*. Vol. 338. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009, pp. xxii+973.