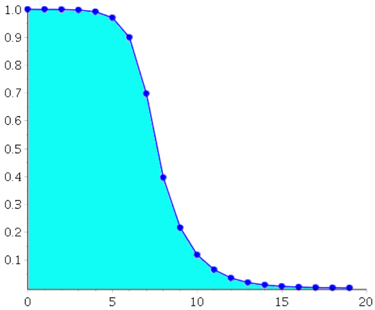
An entropic approach to the cutoff phenomenon

Justin Salez



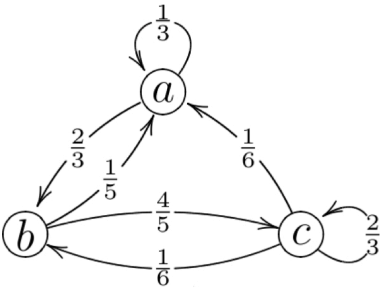
Université Paris-Dauphine & PSL

Part I

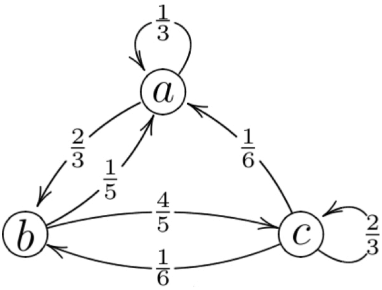
The cutoff phenomenon

Markov chains

Markov chains

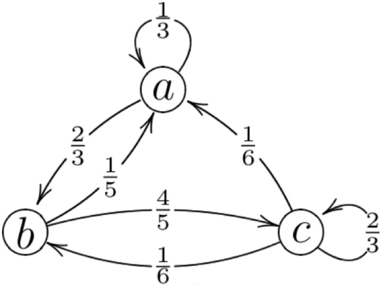


Markov chains



P irreducible, aperiodic transition matrix on a finite space $\mathscr{X}$

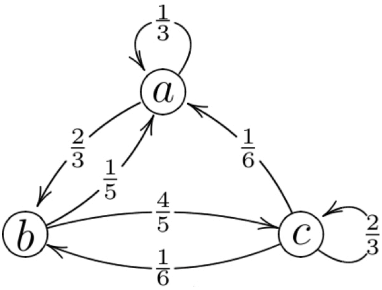
Markov chains



P irreducible, aperiodic transition matrix on a finite space $\mathscr{X}$

- there is a unique invariant law $\pi = \pi P$

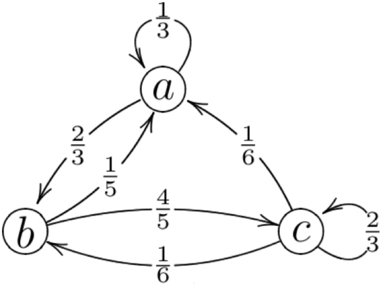
Markov chains



P irreducible, aperiodic transition matrix on a finite space $\mathscr{X}$

- there is a unique invariant law $\pi = \pi P$
- the system mixes: $P^t(x, y) \xrightarrow[t \to \infty]{} \pi(y)$

Markov chains



P irreducible, aperiodic transition matrix on a finite space $\mathscr{X}$

- there is a unique invariant law $\pi = \pi P$
- the system mixes: $P^t(x, y) \xrightarrow[t\to\infty]{} \pi(y)$

Question (crucial for applications): how fast?

Mixing times (Aldous-Diaconis, 80's)

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0, 1]-$valued

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0,1]-$valued
- non-decreasing

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0, 1]-$valued
- non-decreasing
- sub-multiplicative: $\mathfrak{D}(t+s) \leq 2\mathfrak{D}(t)\mathfrak{D}(s)$.

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0, 1]-$valued
- non-decreasing
- sub-multiplicative: $\mathfrak{D}(t+s) \leq 2\mathfrak{D}(t)\mathfrak{D}(s)$.

$$\mathfrak{D}(t)^{\frac{1}{t}} \xrightarrow[t \to \infty]{} \lambda_\star$$

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0, 1]-$valued
- non-decreasing
- sub-multiplicative: $\mathfrak{D}(t + s) \leq 2\mathfrak{D}(t)\mathfrak{D}(s)$.

$$\mathfrak{D}(t)^{\frac{1}{t}} \xrightarrow[t \to \infty]{} \lambda_\star = \max\{|\lambda| \colon \lambda \neq 1 \text{ eigenv. of } P\} < 1$$

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0, 1]-$valued
- non-decreasing
- sub-multiplicative: $\mathfrak{D}(t+s) \leq 2\mathfrak{D}(t)\mathfrak{D}(s)$.

$$\mathfrak{D}(t)^{\frac{1}{t}} \xrightarrow[t \to \infty]{} \lambda_\star = \max\{|\lambda| \colon \lambda \neq 1 \text{ eigenv. of } P\} < 1$$

Relaxation time: $t_{\mathrm{REL}} := \frac{1}{1-\lambda_\star}$

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x,A) - \pi(A) \right|$

- $[0,1]-$valued
- non-decreasing
- sub-multiplicative: $\mathfrak{D}(t+s) \leq 2\mathfrak{D}(t)\mathfrak{D}(s)$.

$$\mathfrak{D}(t)^{\frac{1}{t}} \xrightarrow[t\to\infty]{} \lambda_\star = \max\{|\lambda| \colon \lambda \neq 1 \text{ eigenv. of } P\} < 1$$

Relaxation time: $t_{\mathrm{REL}} := \frac{1}{1-\lambda_\star}$

Mixing time: $t_{\mathrm{MIX}}(\varepsilon) := \min\{t \geq 0 \colon \mathfrak{D}(t) \leq \varepsilon\}$

Mixing times (Aldous-Diaconis, 80's)

Distance to equilibrium: $\mathfrak{D}(t) := \max_{x \in \mathscr{X}} \max_{A \subseteq \mathscr{X}} \left| P^t(x, A) - \pi(A) \right|$

- $[0, 1]-$valued
- non-decreasing
- sub-multiplicative: $\mathfrak{D}(t+s) \leq 2\mathfrak{D}(t)\mathfrak{D}(s)$.

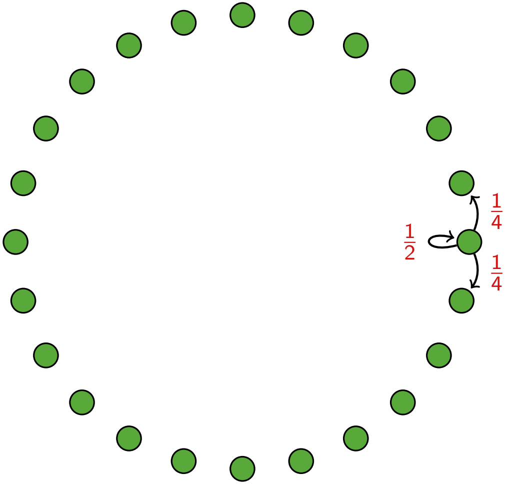
$$\mathfrak{D}(t)^{\frac{1}{t}} \xrightarrow[t\to\infty]{} \lambda_\star = \max\{|\lambda| \colon \lambda \neq 1 \text{ eigenv. of } P\} < 1$$

Relaxation time: $t_{\mathrm{REL}} := \frac{1}{1-\lambda_\star}$

Mixing time: $t_{\mathrm{MIX}}(\varepsilon) := \min\{t \geq 0 \colon \mathfrak{D}(t) \leq \varepsilon\}$

Research program: estimate $t_{\mathrm{MIX}}(\varepsilon)$ (see Levin-Peres-Wilmer)

Lazy random walk on the cycle



Lazy random walk on the cycle

$X_t = \xi_1 + \cdots + \xi_t \mod n$ with (ξ_t) i.i.d. $\begin{cases} +1 & w.p. \quad 1/4 \\ 0 & w.p. \quad 1/2 \\ -1 & w.p. \quad 1/4 \end{cases}$

Lazy random walk on the cycle

$X_t = \xi_1 + \cdots + \xi_t \mod n$ with (ξ_t) i.i.d. $\begin{cases} +1 & w.p. \quad 1/4 \\ 0 & w.p. \quad 1/2 \\ -1 & w.p. \quad 1/4 \end{cases}$

Take $t \sim \lambda n^2$ and write f_λ for the density of $\mathcal{N}(0, \lambda/2) \mod 1$.

Lazy random walk on the cycle

$X_t = \xi_1 + \cdots + \xi_t \mod n$ with (ξ_t) i.i.d. $\begin{cases} +1 & w.p. \quad 1/4 \\ 0 & w.p. \quad 1/2 \\ -1 & w.p. \quad 1/4 \end{cases}$

Take $t \sim \lambda n^2$ and write f_λ for the density of $\mathcal{N}(0, \lambda/2) \mod 1$.

- CLT: $\mathbb{P}(X_t \in [an, bn]) \xrightarrow[n\to\infty]{} \int_a^b f_\lambda(u)\mathrm{d}u$

Lazy random walk on the cycle

$X_t = \xi_1 + \cdots + \xi_t \mod n$ with (ξ_t) i.i.d. $\begin{cases} +1 & w.p. \quad 1/4 \\ 0 & w.p. \quad 1/2 \\ -1 & w.p. \quad 1/4 \end{cases}$

Take $t \sim \lambda n^2$ and write f_λ for the density of $\mathcal{N}(0, \lambda/2) \mod 1$.

- CLT: $\mathbb{P}(X_t \in [an, bn]) \xrightarrow[n\to\infty]{} \int_a^b f_\lambda(u)\mathrm{d}u$
- local CLT: $\mathbb{P}(X_t = \lfloor nu \rfloor) = \dfrac{f_\lambda(u)}{n} + o\left(\dfrac{1}{n}\right)$

Lazy random walk on the cycle

$X_t = \xi_1 + \cdots + \xi_t \mod n$ with (ξ_t) i.i.d. $\begin{cases} +1 & w.p. & 1/4 \\ 0 & w.p. & 1/2 \\ -1 & w.p. & 1/4 \end{cases}$

Take $t \sim \lambda n^2$ and write f_λ for the density of $\mathcal{N}(0, \lambda/2) \mod 1$.

- CLT: $\mathbb{P}\left(X_t \in [an, bn]\right) \xrightarrow[n\to\infty]{} \int_a^b f_\lambda(u)\mathrm{d}u$
- local CLT: $\mathbb{P}\left(X_t = \lfloor nu \rfloor\right) = \dfrac{f_\lambda(u)}{n} + o\left(\dfrac{1}{n}\right)$

This implies $\mathfrak{D}(t) \xrightarrow[n\to\infty]{} \psi(\lambda) := \dfrac{1}{2}\displaystyle\int_0^1 \left|1 - f_\lambda(u)\right| \mathrm{d}u$

Lazy random walk on the cycle

$X_t = \xi_1 + \cdots + \xi_t \mod n$ with (ξ_t) i.i.d. $\begin{cases} +1 & w.p. & 1/4 \\ 0 & w.p. & 1/2 \\ -1 & w.p. & 1/4 \end{cases}$

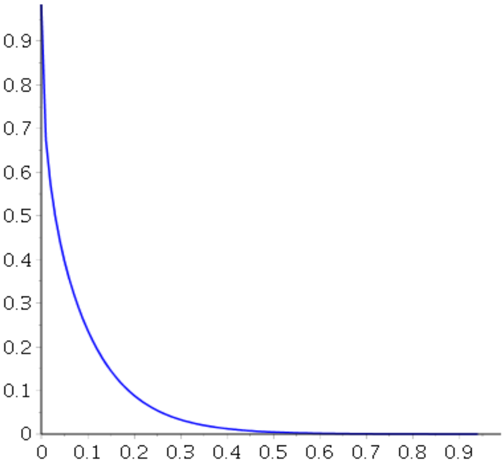
Take $t \sim \lambda n^2$ and write f_λ for the density of $\mathcal{N}(0, \lambda/2) \mod 1$.

- CLT: $\mathbb{P}\left(X_t \in [an, bn]\right) \xrightarrow[n\to\infty]{} \int_a^b f_\lambda(u)\mathrm{d}u$
- local CLT: $\mathbb{P}\left(X_t = \lfloor nu \rfloor\right) = \dfrac{f_\lambda(u)}{n} + o\left(\dfrac{1}{n}\right)$

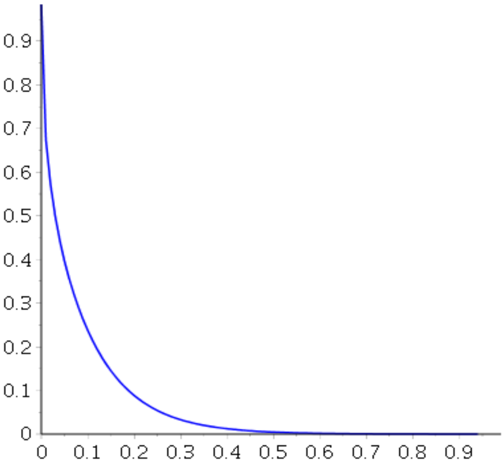
This implies $\mathfrak{D}(t) \xrightarrow[n\to\infty]{} \psi(\lambda) := \dfrac{1}{2}\int_0^1 \left|1 - f_\lambda(u)\right| \mathrm{d}u$

Conclusion: $\mathrm{t}_{\mathrm{MIX}}(\varepsilon) = \psi^{-1}(\varepsilon)n^2 + o(n^2)$

Lazy random walk on the cycle

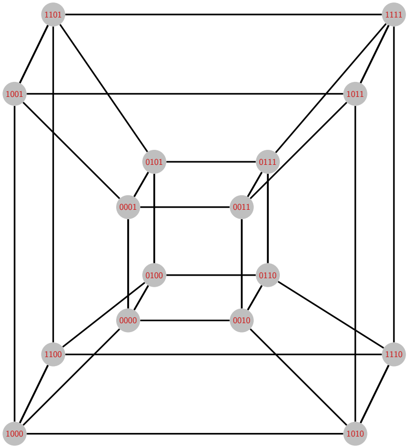


Lazy random walk on the cycle



- ▷ Convergence to stationarity occurs gradually on timescale $\Theta(n^2)$

Lazy random walk on the hypercube



Lazy random walk on the hypercube

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it
- $N_t = \#\{U_1, \ldots, U_t\}$ is a **coupon collector process**

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it
- $N_t = \#\{U_1, \ldots, U_t\}$ is a **coupon collector process**
- Take $t = \frac{1}{2} n \ln n + \lambda n + o(n)$

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it
- $N_t = \#\{U_1, \ldots, U_t\}$ is a **coupon collector process**
- Take $t = \frac{1}{2} n \ln n + \lambda n + o(n)$

$$\triangleright \quad N_t = n - e^{-\lambda}\sqrt{n} + o_{\mathbb{P}}(\sqrt{n})$$

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it
- $N_t = \#\{U_1, \ldots, U_t\}$ is a **coupon collector process**
- Take $t = \frac{1}{2} n \ln n + \lambda n + o(n)$

  ▷ $N_t = n - e^{-\lambda}\sqrt{n} + o_{\mathbb{P}}(\sqrt{n})$

  ▷ $\mathbb{P}(X_t = x | N_t) = 2^{-N_t} \binom{\|x\|}{n - N_t} / \binom{n}{n - N_t}$

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it
- $N_t = \#\{U_1, \ldots, U_t\}$ is a **coupon collector process**
- Take $t = \frac{1}{2} n \ln n + \lambda n + o(n)$

  ▷ $N_t = n - e^{-\lambda}\sqrt{n} + o_{\mathbb{P}}(\sqrt{n})$

  ▷ $\mathbb{P}(X_t = x | N_t) = 2^{-N_t} \binom{\|x\|}{n - N_t} \Big/ \binom{n}{n - N_t}$

  ▷ $\mathfrak{D}\left(t\right) \xrightarrow[n\to\infty]{} \psi(\lambda) := \frac{1}{2\pi} \int_{-\frac{e^{-\lambda}}{2}}^{+\frac{e^{-\lambda}}{2}} e^{-\frac{u^2}{2}} \,\mathrm{d}u$

Lazy random walk on the hypercube

- Start with the all-1 vector: $X_0 = (1, 1, \ldots, 1)$ (w.l.o.g.)
- At each time t, choose a random coordinate U_t and refresh it
- $N_t = \#\{U_1, \ldots, U_t\}$ is a **coupon collector process**
- Take $t = \frac{1}{2} n \ln n + \lambda n + o(n)$

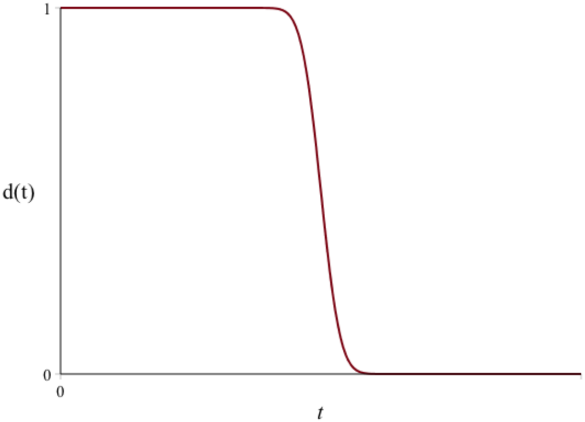
  ▷ $N_t = n - e^{-\lambda}\sqrt{n} + o_{\mathbb{P}}(\sqrt{n})$

  ▷ $\mathbb{P}(X_t = x | N_t) = 2^{-N_t} \binom{\|x\|}{n - N_t} \Big/ \binom{n}{n - N_t}$

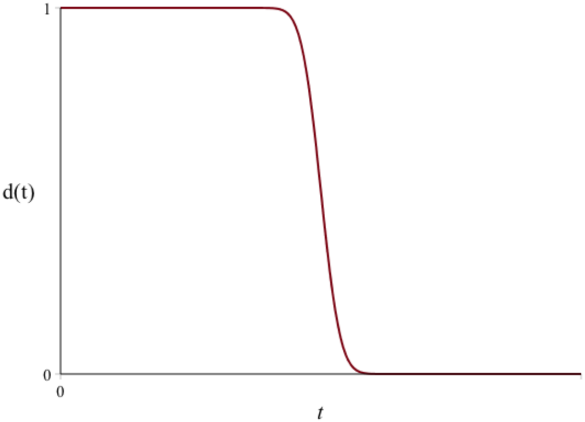
  ▷ $\mathfrak{D}(t) \xrightarrow[n\to\infty]{} \psi(\lambda) := \frac{1}{2\pi} \int_{-\frac{e^{-\lambda}}{2}}^{+\frac{e^{-\lambda}}{2}} e^{-\frac{u^2}{2}} \, \mathrm{d}u$

Conclusion: $\mathrm{t}_{\mathrm{MIX}}(\varepsilon) = \frac{1}{2} n \ln n + \psi^{-1}(\varepsilon) n + o(n)$.

Lazy random walk on the hypercube



Lazy random walk on the hypercube



▷ Convergence to stationarity occurs abruptly at $t \approx \frac{n \log n}{2}$

The cutoff phenomenon (Aldous-Diaconis '86)

The cutoff phenomenon (Aldous-Diaconis '86)

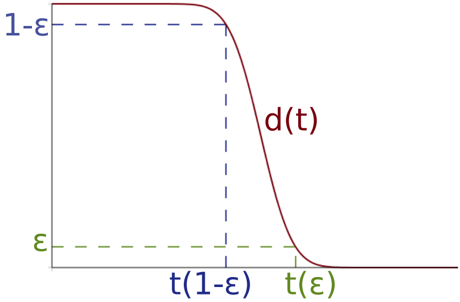
A sequence of Markov chains (indexed by n) exhibits cutoff if

$$\forall \varepsilon \in (0,1), \qquad \frac{\mathrm{t}_{\mathrm{MIX}}^{(n)}(1-\varepsilon)}{\mathrm{t}_{\mathrm{MIX}}^{(n)}(\varepsilon)} \xrightarrow[n\to\infty]{} 1$$

The cutoff phenomenon (Aldous-Diaconis '86)

A sequence of Markov chains (indexed by n) exhibits cutoff if

$$\forall \varepsilon \in (0,1), \qquad \frac{\mathrm{t}_{\mathrm{MIX}}^{(n)}(1-\varepsilon)}{\mathrm{t}_{\mathrm{MIX}}^{(n)}(\varepsilon)} \xrightarrow[n\to\infty]{} 1$$



Ubiquity of the cutoff phenomenon

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)
- Interacting particles (Hermon, Lacoin, Lubetzky, S., Sly...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)
- Interacting particles (Hermon, Lacoin, Lubetzky, S., Sly...)
- Random walks on sparse random graphs (Ben-Hamou, Berestycki, Hermon, Lubetzky, Peres, S., Sly, Sousi...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)
- Interacting particles (Hermon, Lacoin, Lubetzky, S., Sly...)
- Random walks on sparse random graphs (Ben-Hamou, Berestycki, Hermon, Lubetzky, Peres, S., Sly, Sousi...)
- Random walks on random digraphs (Bordenave, Caputo, S...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)
- Interacting particles (Hermon, Lacoin, Lubetzky, S., Sly...)
- Random walks on sparse random graphs (Ben-Hamou, Berestycki, Hermon, Lubetzky, Peres, S., Sly, Sousi...)
- Random walks on random digraphs (Bordenave, Caputo, S...)
- Random random walks on groups (Hermon, Olesker-Taylor...)

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)
- Interacting particles (Hermon, Lacoin, Lubetzky, S., Sly...)
- Random walks on sparse random graphs (Ben-Hamou, Berestycki, Hermon, Lubetzky, Peres, S., Sly, Sousi...)
- Random walks on random digraphs (Bordenave, Caputo, S...)
- Random random walks on groups (Hermon, Olesker-Taylor...)

▷ Still very far from being understood.

Ubiquity of the cutoff phenomenon

Cutoff has been shown to arise in various contexts, including

- Card shuffling (Aldous, Diaconis, Shahshahani...)
- Birth-and-death chains (Diaconis, Saloff-Coste...)
- Random walks on certain groups (Chen, Saloff-Coste...)
- Interacting particles (Hermon, Lacoin, Lubetzky, S., Sly...)
- Random walks on sparse random graphs (Ben-Hamou, Berestycki, Hermon, Lubetzky, Peres, S., Sly, Sousi...)
- Random walks on random digraphs (Bordenave, Caputo, S...)
- Random random walks on groups (Hermon, Olesker-Taylor...)

▷ Still very far from being understood.

▷ Embarrassingly, no effective sufficient condition is known.

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $t_{\mathrm{MIX}}(\varepsilon) \geq t_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)

The “product” condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM’04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $t_{\mathrm{MIX}}(\varepsilon) \geq t_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04’)

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $t_{\mathrm{MIX}}(\varepsilon) \geq t_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04')
- Known to be sufficient on trees (Basu-Hermon-Peres'17)

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $t_{\mathrm{MIX}}(\varepsilon) \geq t_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04')
- Known to be sufficient on trees (Basu-Hermon-Peres'17)

Generic counter-example:

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $t_{\mathrm{MIX}}(\varepsilon) \geq t_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04')
- Known to be sufficient on trees (Basu-Hermon-Peres'17)

Generic counter-example: consider the rank-1 perturbation

$$\widetilde{P}(x,y) := (1-\delta)P(x,y) + \delta\pi(y)$$

The "product" condition $\mathrm{t}_{\mathrm{REL}} \ll \mathrm{t}_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $\mathrm{t}_{\mathrm{MIX}}(\varepsilon) \geq \mathrm{t}_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04')
- Known to be sufficient on trees (Basu-Hermon-Peres'17)

Generic counter-example: consider the rank-1 perturbation

$$\widetilde{P}(x,y) := (1-\delta)P(x,y) + \delta\pi(y) \quad \Rightarrow \quad \widetilde{\mathcal{D}}(t) = (1-\delta)^t \mathcal{D}(t)$$

The “product” condition $\mathrm{t_{REL}} \ll \mathrm{t_{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $\mathrm{t_{MIX}}(\varepsilon) \geq \mathrm{t_{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04')
- Known to be sufficient on trees (Basu-Hermon-Peres'17)

Generic counter-example: consider the rank-1 perturbation

$$\widetilde{P}(x,y) := (1-\delta)P(x,y) + \delta\pi(y) \quad \Rightarrow \quad \widetilde{\mathcal{D}}(t) = (1-\delta)^t \mathcal{D}(t)$$

If $\mathrm{t_{REL}} \ll \frac{1}{\delta} \ll \mathrm{t_{MIX}}$, then $\widetilde{\mathrm{t}}_{\mathrm{REL}} \ll \widetilde{\mathrm{t}}_{\mathrm{MIX}}$ but cutoff is destroyed!

The "product" condition $t_{\mathrm{REL}} \ll t_{\mathrm{MIX}}(\varepsilon)$

- Proposed by Peres (AIM'04) as an effective criterion for cutoff
- Satisfied on the hypercube, not on the cycle
- Always necessary for cutoff (because $t_{\mathrm{MIX}}(\varepsilon) \geq t_{\mathrm{REL}} \log \frac{1}{2\varepsilon}$)
- Fails to be sufficient in general (Aldous 04')
- Known to be sufficient on trees (Basu-Hermon-Peres'17)

Generic counter-example: consider the rank-1 perturbation

$$\widetilde{P}(x,y) := (1-\delta)P(x,y) + \delta\pi(y) \quad \Rightarrow \quad \widetilde{\mathcal{D}}(t) = (1-\delta)^t \mathcal{D}(t)$$

If $t_{\mathrm{REL}} \ll \frac{1}{\delta} \ll t_{\mathrm{MIX}}$, then $\widetilde{t}_{\mathrm{REL}} \ll \widetilde{t}_{\mathrm{MIX}}$ but cutoff is destroyed!

Corollary: the criterion is wrong even for abelian random walks...

Quotes from the Masters (Aldous-Diaconis 86-96)

Quotes from the Masters (Aldous-Diaconis 86-96)

- *At present writing, proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*

Quotes from the Masters (Aldous-Diaconis 86-96)

- *At present writing, proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*

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

Quotes from the Masters (Aldous-Diaconis 86-96)

- *At present writing, proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors. Most of the examples where this can be pushed through arise from random walk on groups, with the walk having a fair amount of symmetry.*
- *The careful work required to prove cutoff often leads to a more or less complete understanding of the chain such that essentially any natural question can be answered.*
- *It occurs in all the examples we can explicitly calculate, but we know no general result which says that the phenomenon must happen for all "reasonable" chains.*

Part II

An entropic approach

Entropic concentration

Entropic concentration

Entropy: $\mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \log \frac{\mu(x)}{\pi(x)}$

Entropic concentration

Entropy: $\mathrm{d}_{\mathrm{KL}}\left(\mu\|\pi\right) := \sum_{x\in\mathscr{X}} \mu(x) \log \frac{\mu(x)}{\pi(x)}$

Varentropy: $\mathscr{V}_{\mathrm{KL}}\left(\mu\|\pi\right) := \sum_{x\in\mathscr{X}} \mu(x) \left(\log \frac{\mu(x)}{\pi(x)} - \mathrm{d}_{\mathrm{KL}}\left(\mu\|\pi\right)\right)^2$

Entropic concentration

Entropy: $\mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \log \frac{\mu(x)}{\pi(x)}$

Varentropy: $\mathscr{V}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \left(\log \frac{\mu(x)}{\pi(x)} - \mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) \right)^2$

Worst-case varentropy at time t: $\mathscr{V}^{\star}_{\mathrm{KL}}(t) := \max_{x \in \mathscr{X}} \mathscr{V}_{\mathrm{KL}}\left(P^t(x, \cdot) | \pi\right)$

Entropic concentration

Entropy: $\mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \log \frac{\mu(x)}{\pi(x)}$

Varentropy: $\mathscr{V}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \left(\log \frac{\mu(x)}{\pi(x)} - \mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) \right)^2$

Worst-case varentropy at time t: $\mathscr{V}^{\star}_{\mathrm{KL}}(t) := \max_{x \in \mathscr{X}} \mathscr{V}_{\mathrm{KL}}\left(P^t(x, \cdot) | \pi\right)$

Theorem (S. 21): for any $\varepsilon \in (0,1)$,

$$\mathrm{t}_{\mathrm{MIX}}(\varepsilon) - \mathrm{t}_{\mathrm{MIX}}(1-\varepsilon) \quad \leq \quad \frac{2\mathrm{t}_{\mathrm{REL}}}{\varepsilon^2} \left[1 + \sqrt{\mathscr{V}^{\star}_{\mathrm{KL}}\left(\mathrm{t}_{\mathrm{MIX}}\left(1-\varepsilon\right)\right)} \right].$$

Entropic concentration

Entropy: $\mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \log \frac{\mu(x)}{\pi(x)}$

Varentropy: $\mathscr{V}_{\mathrm{KL}}\left(\mu \| \pi\right) := \sum_{x \in \mathscr{X}} \mu(x) \left(\log \frac{\mu(x)}{\pi(x)} - \mathrm{d}_{\mathrm{KL}}\left(\mu \| \pi\right) \right)^2$

Worst-case varentropy at time t: $\mathscr{V}^{\star}_{\mathrm{KL}}(t) := \max_{x \in \mathscr{X}} \mathscr{V}_{\mathrm{KL}}\left(P^t(x, \cdot) | \pi\right)$

Theorem (S. 21): for any $\varepsilon \in (0,1)$,

$$\mathrm{t}_{\mathrm{MIX}}(\varepsilon) - \mathrm{t}_{\mathrm{MIX}}(1-\varepsilon) \quad \leq \quad \frac{2\mathrm{t}_{\mathrm{REL}}}{\varepsilon^2} \left[1 + \sqrt{\mathscr{V}^{\star}_{\mathrm{KL}}\left(\mathrm{t}_{\mathrm{MIX}}\left(1-\varepsilon\right)\right)} \right].$$

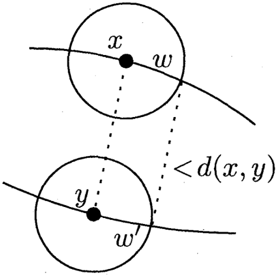
Corollary: a sufficient condition for cutoff is

$$\boxed{\frac{\mathrm{t}_{\mathrm{MIX}}(\varepsilon)}{\mathrm{t}_{\mathrm{REL}}} \gg 1 + \sqrt{\mathscr{V}^{\star}_{\mathrm{KL}}\left(\mathrm{t}_{\mathrm{MIX}}\left(\varepsilon\right)\right)}}$$

Non-negative curvature (Ollivier'10)

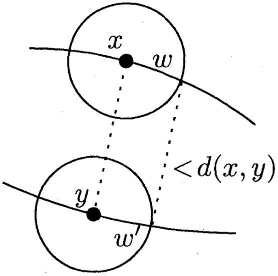
Non-negative curvature (Ollivier'10)

A metric space has non-negative curvature if small balls are closer to each other than their centers are:



Non-negative curvature (Ollivier'10)

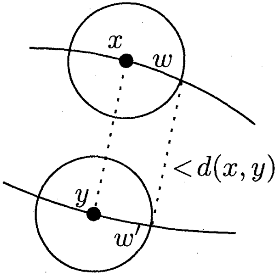
A metric space has non-negative curvature if small balls are closer to each other than their centers are:



- applies, in particular, to the discrete setting of Markov chains

Non-negative curvature (Ollivier'10)

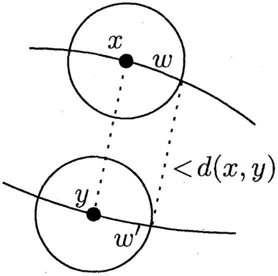
A metric space has non-negative curvature if small balls are closer to each other than their centers are:



- applies, in particular, to the discrete setting of Markov chains
- has remarkable impact on geometry, concentration & mixing

Non-negative curvature (Ollivier'10)

A metric space has non-negative curvature if small balls are closer to each other than their centers are:



- applies, in particular, to the discrete setting of Markov chains
- has remarkable impact on geometry, concentration & mixing
- turns out to provide an effective varentropy estimate

Ollivier-Ricci curvature of Markov chains

Ollivier-Ricci curvature of Markov chains

The curvature between two states x and y is defined as

$$\boxed{\kappa(x, y) \ := \ 1 - \frac{\mathcal{W}_1\left(P(x, \cdot), P(y, \cdot)\right)}{\mathrm{dist}(x, y)}}$$

Ollivier-Ricci curvature of Markov chains

The curvature between two states x and y is defined as

$$\boxed{\kappa(x,y) \; := \; 1 - \frac{\mathcal{W}_1\left(P(x,\cdot),P(y,\cdot)\right)}{\mathrm{dist}(x,y)}}$$

- $\mathrm{dist}(\cdot,\cdot)$ is the graph distance on $G = (\mathscr{X}, \mathrm{supp}(P))$

$$\mathrm{dist}(x,y) \quad := \quad \min\{t \geq 0 \colon P^t(x,y) > 0\}$$

Ollivier-Ricci curvature of Markov chains

The curvature between two states x and y is defined as

$$\boxed{\kappa(x,y) \;:=\; 1 - \frac{\mathcal{W}_1\left(P(x,\cdot),P(y,\cdot)\right)}{\mathrm{dist}(x,y)}}$$

- $\mathrm{dist}(\cdot,\cdot)$ is the graph distance on $G = (\mathscr{X}, \mathrm{supp}(P))$

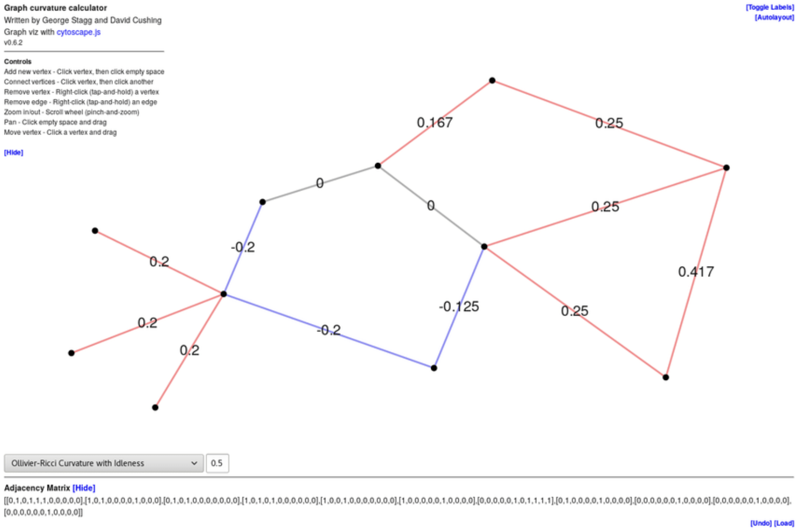
$$\mathrm{dist}(x,y) \quad := \quad \min\{t \geq 0 \colon P^t(x,y) > 0\}$$

- $\mathcal{W}_1(\cdot,\cdot)$ is the L^1-Wassertein metric:

$$\mathcal{W}_1(\mu,\nu) \quad := \quad \min\left\{\mathbb{E}\left[\mathrm{dist}(X,Y)\right] : X \sim \mu, Y \sim \nu\right\}$$

Online Graph Curvature Calculator (Stagg-Cushing)

Online Graph Curvature Calculator (Stagg-Cushing)



Non-negatively curved chains

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \quad \leq \quad \mathrm{dist}(x,y)$$

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \quad \leq \quad \mathrm{dist}(x,y)$$

- Enough to check this on neighbours, i.e. when $P(x,y) > 0$

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \leq \mathrm{dist}(x,y)$$

- Enough to check this on neighbours, i.e. when $P(x,y) > 0$
- Starting point of the path coupling method (Bubley-Dyer'97)

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \quad \leq \quad \mathrm{dist}(x,y)$$

- Enough to check this on neighbours, i.e. when $P(x,y) > 0$
- Starting point of the path coupling method (Bubley-Dyer'97)
- Equivalent to $\|Pf\|_{\mathrm{LIP}} \leq \|f\|_{\mathrm{LIP}}$ for all $f \colon \mathscr{X} \to \mathbb{R}$

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \quad \leq \quad \mathrm{dist}(x,y)$$

- Enough to check this on neighbours, i.e. when $P(x,y) > 0$
- Starting point of the path coupling method (Bubley-Dyer'97)
- Equivalent to $\|Pf\|_{\mathrm{LIP}} \leq \|f\|_{\mathrm{LIP}}$ for all $f\colon \mathscr{X} \to \mathbb{R}$
- Remarkable consequences on geometry and functional analysis (Ollivier'09, Joulin-Ollivier'10, Lin-Lu-Yau'11, Eldan-Lee-Lehec'17, Jost-Münch-Rose '19, Münch'19, Cushing-Kamtue-Koolen-Liu-Münch-Peyerimhoff'20).

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \;\; \leq \;\; \mathrm{dist}(x,y)$$

- Enough to check this on neighbours, i.e. when $P(x,y) > 0$
- Starting point of the path coupling method (Bubley-Dyer'97)
- Equivalent to $\|Pf\|_{\mathrm{LIP}} \leq \|f\|_{\mathrm{LIP}}$ for all $f \colon \mathscr{X} \to \mathbb{R}$
- Remarkable consequences on geometry and functional analysis (Ollivier'09, Joulin-Ollivier'10, Lin-Lu-Yau'11, Eldan-Lee-Lehec'17, Jost-Münch-Rose '19, Münch'19, Cushing-Kamtue-Koolen-Liu-Münch-Peyerimhoff'20).
- Implies concentration at any time: $\mathrm{Var}(f(X_t)) \leq 2t\|f\|_{\mathrm{LIP}}^2$

Non-negatively curved chains

P is non-negatively curved if $\kappa \geq 0$ everywhere, i.e.

$$\forall x, y \in \mathscr{X}, \quad \mathcal{W}_1\left(P(x,\cdot), P(y,\cdot)\right) \quad \leq \quad \mathrm{dist}(x,y)$$

- Enough to check this on neighbours, i.e. when $P(x,y) > 0$
- Starting point of the path coupling method (Bubley-Dyer'97)
- Equivalent to $\|Pf\|_{\mathrm{LIP}} \leq \|f\|_{\mathrm{LIP}}$ for all $f\colon \mathscr{X} \to \mathbb{R}$
- Remarkable consequences on geometry and functional analysis (Ollivier'09, Joulin-Ollivier'10, Lin-Lu-Yau'11, Eldan-Lee-Lehec'17, Jost-Münch-Rose '19, Münch'19, Cushing-Kamtue-Koolen-Liu-Münch-Peyerimhoff'20).
- Implies concentration at any time: $\mathrm{Var}(f(X_t)) \leq 2t\|f\|_{\mathrm{LIP}}^2$
- Also true under non-negative Bakry-Émery curvature

Some examples of non-negatively curved chains

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains
- Random walks on abelian groups

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains
- Random walks on abelian groups
- Conjugacy-invariant random walks on symmetric groups

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains
- Random walks on abelian groups
- Conjugacy-invariant random walks on symmetric groups
- Mean-field Zero-Range dynamics with non-decreasing rates

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains
- Random walks on abelian groups
- Conjugacy-invariant random walks on symmetric groups
- Mean-field Zero-Range dynamics with non-decreasing rates
- Glauber dynamics at high temperature

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains
- Random walks on abelian groups
- Conjugacy-invariant random walks on symmetric groups
- Mean-field Zero-Range dynamics with non-decreasing rates
- Glauber dynamics at high temperature
- Noisy Voter models

Some examples of non-negatively curved chains

- Random walks on complete graphs, paths, stars
- Monotone birth-and-death chains
- Random walks on abelian groups
- Conjugacy-invariant random walks on symmetric groups
- Mean-field Zero-Range dynamics with non-decreasing rates
- Glauber dynamics at high temperature
- Noisy Voter models
- ...

Varentropy estimate under non-negative curvature

Varentropy estimate under non-negative curvature

Theorem (S. '21): If P has non-negative curvature, then

$$\mathscr{V}^{\star}_{\mathrm{KL}}(t) \quad \lesssim \quad (\log \Delta)^2 t,$$

where $\Delta = \max\left\{\frac{1}{P(x,y)} : x \sim y\right\}$ is the "maximum degree".

Varentropy estimate under non-negative curvature

Theorem (S. '21): If P has non-negative curvature, then

$$\mathscr{V}^{\star}_{\mathrm{KL}}(t) \quad \lesssim \quad (\log \Delta)^2 t,$$

where $\Delta = \max\left\{\frac{1}{P(x,y)} : x \sim y\right\}$ is the "maximum degree".

Corollary: non-negatively curved chains exhibit cutoff whenever

$$\mathrm{t}_{\mathrm{MIX}}(\varepsilon) \quad \gg \quad (\mathrm{t}_{\mathrm{REL}} \log \Delta)^2$$

Varentropy estimate under non-negative curvature

Theorem (S. '21): If P has non-negative curvature, then

$$\mathscr{V}^{\star}_{\mathrm{KL}}(t) \quad \lesssim \quad (\log \Delta)^2 t,$$

where $\Delta = \max\left\{\frac{1}{P(x,y)} : x \sim y\right\}$ is the "maximum degree".

Corollary: non-negatively curved chains exhibit cutoff whenever

$$\mathrm{t}_{\mathrm{MIX}}(\varepsilon) \quad \gg \quad (\mathrm{t}_{\mathrm{REL}} \log \Delta)^2$$

Since $\mathrm{t}_{\mathrm{MIX}}(\varepsilon) \gtrsim \frac{\log N}{\log \Delta}$, we obtain the simpler sufficient condition

$$\mathrm{t}_{\mathrm{REL}} \quad \ll \quad \frac{(\log N)^{1/2}}{(\log \Delta)^{3/2}}$$

Varentropy estimate under non-negative curvature

Theorem (S. '21): If P has non-negative curvature, then

$$\mathscr{V}^{\star}_{\mathrm{KL}}(t) \quad \lesssim \quad (\log \Delta)^2 t,$$

where $\Delta = \max\left\{\frac{1}{P(x,y)} : x \sim y\right\}$ is the "maximum degree".

Corollary: non-negatively curved chains exhibit cutoff whenever

$$\mathrm{t}_{\mathrm{MIX}}(\varepsilon) \quad \gg \quad (\mathrm{t}_{\mathrm{REL}} \log \Delta)^2$$

Since $\mathrm{t}_{\mathrm{MIX}}(\varepsilon) \gtrsim \frac{\log N}{\log \Delta}$, we obtain the simpler sufficient condition

$$\mathrm{t}_{\mathrm{REL}} \quad \ll \quad \frac{(\log N)^{1/2}}{(\log \Delta)^{3/2}}$$

Remark: the presence of Δ is crucial (dense counter-examples)

Application: cutoff on almost all abelian Cayley graphs

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$, where

- $(\mathscr{X}, +)$ is an abelian group with N elements

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$, where

- $(\mathscr{X}, +)$ is an abelian group with N elements
- $S \subseteq \mathscr{X}$ is a symmetric subset with d elements

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$, where

- $(\mathscr{X}, +)$ is an abelian group with N elements
- $S \subseteq \mathscr{X}$ is a symmetric subset with d elements

$\triangleright$ Cutoff as soon as $\mathrm{t}_{\mathrm{REL}} \ll \frac{(\log N)^{1/2}}{(\log d)^{3/2}}$

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$, where

- $(\mathscr{X}, +)$ is an abelian group with N elements
- $S \subseteq \mathscr{X}$ is a symmetric subset with d elements

▷ Cutoff as soon as $\mathrm{t}_{\mathrm{REL}} \ll \frac{(\log N)^{1/2}}{(\log d)^{3/2}}$

▷ Alon-Roichman'94: $\mathrm{t}_{\mathrm{REL}} \lesssim 1$ w.h.p. if S is random & $d \gtrsim \log N$

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$, where

- $(\mathscr{X}, +)$ is an abelian group with N elements
- $S \subseteq \mathscr{X}$ is a symmetric subset with d elements

▷ Cutoff as soon as $\mathrm{t}_{\mathrm{REL}} \ll \frac{(\log N)^{1/2}}{(\log d)^{3/2}}$

▷ Alon-Roichman'94: $\mathrm{t}_{\mathrm{REL}} \lesssim 1$ w.h.p. if S is random & $d \gtrsim \log N$

Conclusion: "almost all" abelian Cayley graphs exhibit cutoff!

Application: cutoff on almost all abelian Cayley graphs

Consider simple random walk on $G = \mathrm{Cay}(\mathscr{X}, \mathrm{S})$, where

- $(\mathscr{X}, +)$ is an abelian group with N elements
- $S \subseteq \mathscr{X}$ is a symmetric subset with d elements

▷ Cutoff as soon as $\mathrm{t}_{\mathrm{REL}} \ll \frac{(\log N)^{1/2}}{(\log d)^{3/2}}$

▷ Alon-Roichman'94: $\mathrm{t}_{\mathrm{REL}} \lesssim 1$ w.h.p. if S is random & $d \gtrsim \log N$

Conclusion: "almost all" abelian Cayley graphs exhibit cutoff!

This long-standing conjecture (Aldous-Diaconis'86) was settled very recently (Hermon–Olesker-Taylor'21) via hard computations...

Thanks!