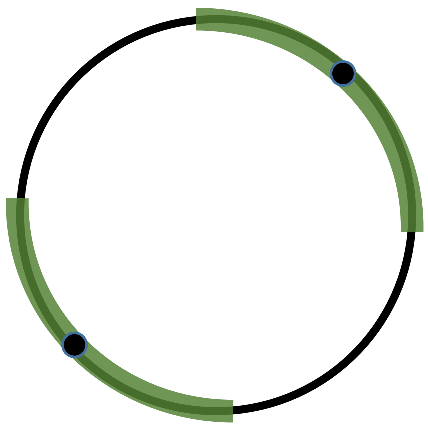
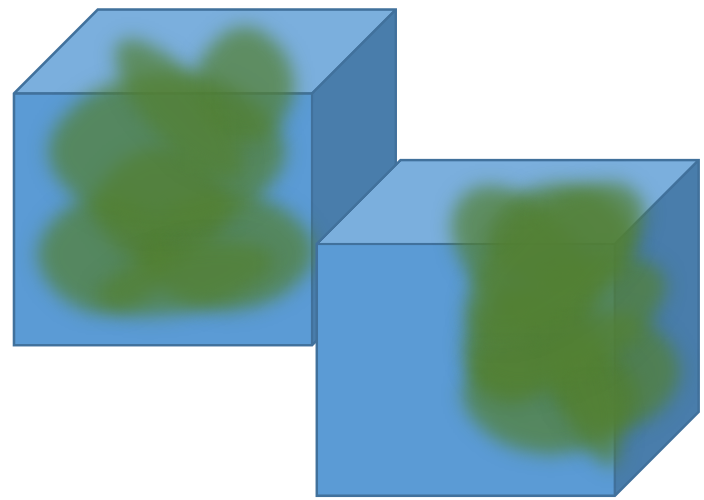
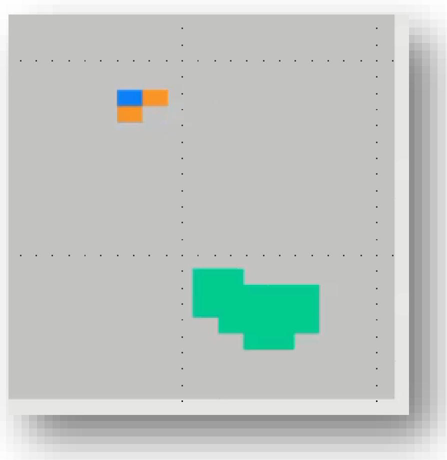
curvature, mixing, and entropic interpolation

Simons Feb-2016 and CSE 599s Lecture 13

James R. Lee
University of Washington







Joint with Ronen Eldan (Weizmann) and Joseph Lehec (Paris-Dauphine)

Markov chain setup

Let $\{X_t\}$ be a reversible Markov chain on a finite state space Ω with stationary measure π.

Denote by $\mathcal{L} = I - P$ the (positive semi-definite) Laplacian, and let $H_t = e^{-t\mathcal{L}}$ be the continuous-time heat semigroup.

Dirichlet form: For $f, g \in L^2(\Omega, \pi)$:

$$\mathcal{E}(f, g) = \frac{1}{2}\mathbb{E}_{X_0 \sim \pi}\left[\big(f(X_1) - f(X_0)\big)\big(g(X_1) - g(X_0)\big)\right]$$

Heat equation: If $\{h_t : t \geq 0\}$ is the time-evolution of a density $h_0 : \Omega \to \mathbb{R}_+$, then

$$\frac{d}{dt}h_t = -\mathcal{L}^* h_t$$

convergence to equilibrium

Spectral gap:

$$\frac{d}{dt}\,\mathrm{Var}_\pi(h_t) = -2\,\mathcal{E}(h_t, h_t)$$

$$\lambda = \inf\left\{\frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)} : f \neq 0\right\} \qquad \mathrm{Var}_\pi(h_t) \le e^{-2\lambda t}\,\mathrm{Var}_\pi(h_0)$$

Modified log-Sobolev (MLS): [Bobkov-Tetali 2006]

$$\mathrm{Ent}_\pi(h_t) = \sum_{x\in\Omega} \pi(x) h_t(x) \log h_t(x)$$

$$\frac{d}{dt}\,\mathrm{Ent}_\pi(h_t) = -\,\mathcal{E}(h_t, \log h_t)$$

$$\rho_0 = \inf\left\{\frac{\mathcal{E}(f,\log f)}{\mathrm{Ent}_\pi(f)} : f \ge 0\right\} \qquad \mathrm{Ent}_\pi(h_t) \le e^{-\rho_0 t}\,\mathrm{Ent}_\pi(h_0)$$

LSI vs. MLSI

Log-Sobolev constant: $\rho = \inf \left\{ \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{\mathrm{Ent}_\pi(f)} : f \geq 0 \right\}$

Modified log-Sobolev: $\rho_0 = \inf \left\{ \frac{\mathcal{E}(f, \log f)}{\mathrm{Ent}_\pi(f)} : f \geq 0 \right\}$

For diffusions: $\mathcal{E}(f, g) = \int \nabla f \nabla g = \int f \Delta g$

$$\mathcal{E}(\sqrt{f}, \sqrt{f}) = \int \left(\nabla \sqrt{f}\right)^2 = \frac{1}{4} \int \frac{|\nabla f|^2}{f}$$

$$= \frac{1}{4} \int \nabla f \; \nabla \log f = \frac{1}{4} \mathcal{E}(f, \log f)$$

LSI vs. MLSI

Log-Sobolev constant: $\rho = \inf \left\{ \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{\mathrm{Ent}_\pi(f)} : f \geq 0 \right\}$

Modified log-Sobolev: $\rho_0 = \inf \left\{ \frac{\mathcal{E}(f, \log f)}{\mathrm{Ent}_\pi(f)} : f \geq 0 \right\}$

$$4\rho \leq \rho_0 \leq \frac{\lambda}{2}$$

$$\frac{1}{2\rho} \leq \ell_2 \text{ mixing time} \leq \frac{1}{\rho}\left(1 + \frac{1}{4} \log \log \frac{1}{\pi_{\min}}\right)$$

LSI vs. MLSI

Log-Sobolev constant: $$\rho = \inf \left\{ \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{\mathrm{Ent}_\pi(f)} : f \geq 0 \right\}$$

Modified log-Sobolev: $$\rho_0 = \inf \left\{ \frac{\mathcal{E}(f, \log f)}{\mathrm{Ent}_\pi(f)} : f \geq 0 \right\}$$

$$4\rho \leq \rho_0 \leq \frac{\lambda}{2}$$

$$\frac{1}{2\rho} \leq \ell_2 \text{ mixing time} \leq \frac{1}{\rho}\left(1 + \frac{1}{4}\log\log\frac{1}{\pi_{\min}}\right)$$

[Diaconis Saloff-Coste 1996]

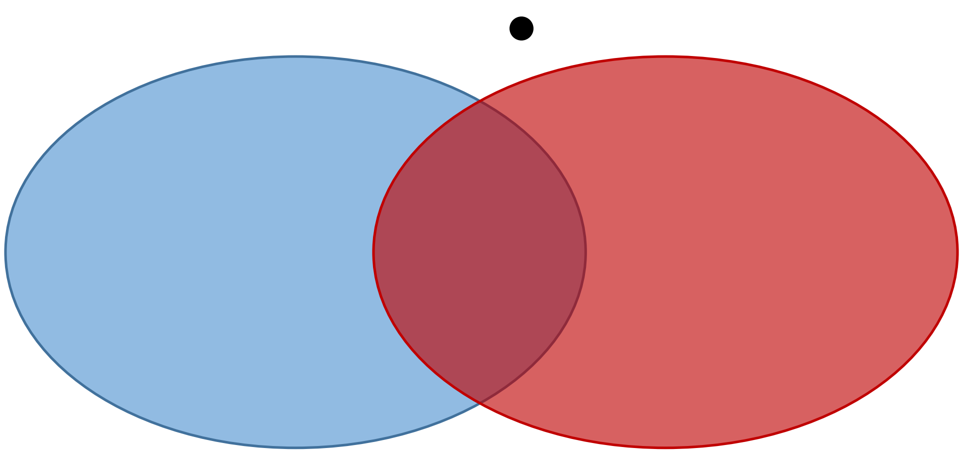
curvature and path couplings

Bakry-Emery (1985) theory: For Markov diffusions,

Positive curvature ⇒ Log-Sob inequality (quantitatively)

Otto-Villani (2000): Proved this (and stronger versions) using Otto's interpretation of diffusion as the gradient flow of the entropy on an appropriate Riemannian manifold of probability measures.

In recent years, a rather large body of work attempting to define these notions / extend these implications to discrete spaces.



curvature and path couplings

Suppose we have a metric d on the state space Ω.

Y. Ollivier (following Bubley-Dyer'97, etc.):

The metric chain (Ω, P, d) has **coarse Ricci curvature** $\geq \kappa$ if for every pair $u, v \in \Omega$, there is a pair of random variables (U, V) such that

$$U \sim X_1 \mid X_0 = u \qquad\qquad V \sim X_1 \mid X_0 = v$$

and

$$\mathbb{E}[d(U,V)] \leq (1-\kappa)\, d(u,v)$$

curvature and path couplings

Suppose we have a metric d on the state space Ω.

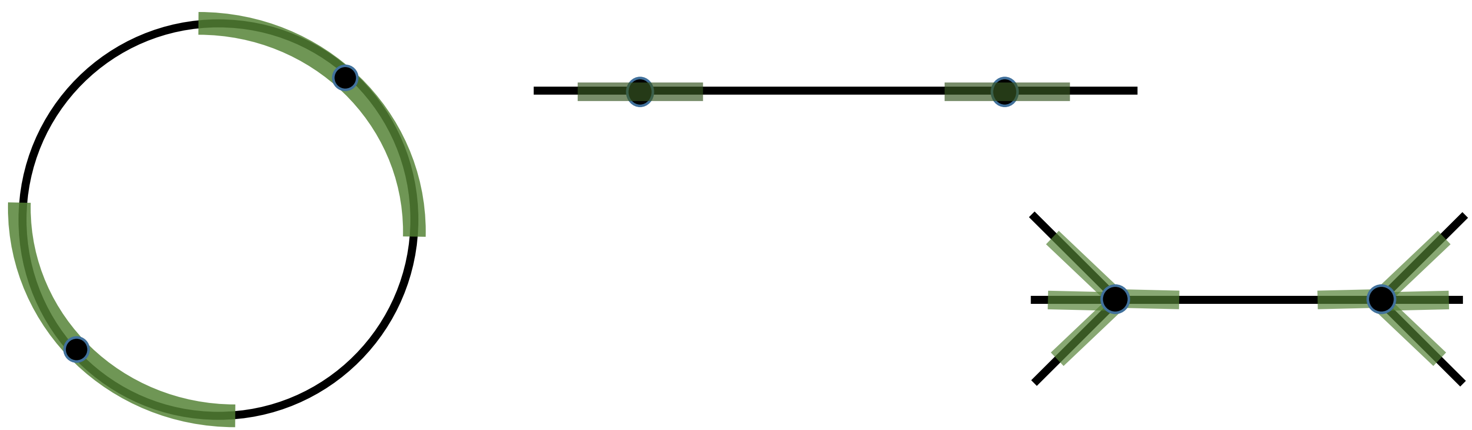
Y. Ollivier (following Bubley-Dyer'97, etc.):

The metric chain (Ω, P, d) has **coarse Ricci curvature** $\geq \kappa$ if for every pair $u, v \in \Omega$, there is a pair of random variables (U, V) such that

$$U \sim X_1 \mid X_0 = u \qquad\qquad V \sim X_1 \mid X_0 = v$$

and

$$\mathbb{E}[d(U,V)] \leq (1-\kappa)\, d(u,v)$$



Peres-Tetali conjecture

Conjecture:

If we metricize the chain so that $d(x, y) = 1$ when $P(x, y) > 1$ and then take the induced path metric, the following holds:

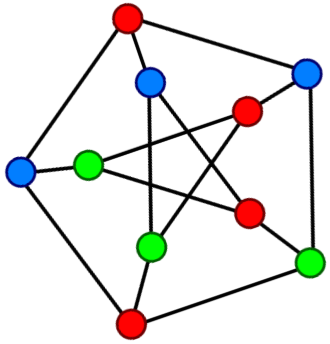
Whenever (Ω, P, d) has coarse Ricci curvature $\geq \kappa$, the chain admits a [modified*] log-Sobolev inequality with constant $O(1/\kappa)$.

Challenge / test chain:

For what values of Δ (maximum degree) and k (# colors) does the Glauber dynamics on k-colorings of a graph admit a (uniform) log-Sobolev inequality?

$k \geq 2\Delta$ [Marton 2015]

$k \geq \frac{11}{6}\Delta$???



The $\boldsymbol{W_p}$ **distance** between densities f and g on a metric measure space (Ω, π, d) is

$$W_p(f,g) = \min_{\mu} \left\{ \left(\int d(u,v)^p \, d\mu(u,v) \right)^{\frac{1}{p}} \right\}$$

where the minimum is over all couplings μ of $(f \; d\pi, g \; d\pi)$.

Inequalities relating transportation distances to entropy were studied by Marton (1996) and Talagrand (1996).

[Bobkov-Götze 1999]: If (Ω, P, π) admits a log-Sobolev inequality with constant $1/\alpha$, then it admits a W_1 entropy-transport inequality:

$$W_1(f, \mathbf{1}) \leq \sqrt{2\alpha \, \mathrm{Ent}_\pi(f)}$$

where Ω is equipped with the graph metric introduced earlier.

entropy-transport

Conjecture: If we metricize the chain so that $d(x,y) = 1$ when $P(x,y) > 1$ and then take the induced path metric, the following holds:

Whenever (Ω, P, d) has coarse Ricci curvature $\geq \kappa$, the chain admits a [modified*] log-Sobolev inequality with constant $O(1/\kappa)$.

[Bobkov-Götze 1999]: If (Ω, P, π) admits a log-Sobolev inequality with constant $1/\alpha$, then it admits a W_1 entropy-transport inequality:

$$W_1(f, \mathbf{1}) \leq \sqrt{2\alpha\, \mathrm{Ent}_\pi(f)}$$

Theorem [Eldan-L-Lehec 2015]:

Coarse Ricci curvature $\geq \kappa$ implies a W_1 entropy-transport inequality with constant $\alpha = \kappa^{-1}/(1 - \frac{\kappa}{2})$.

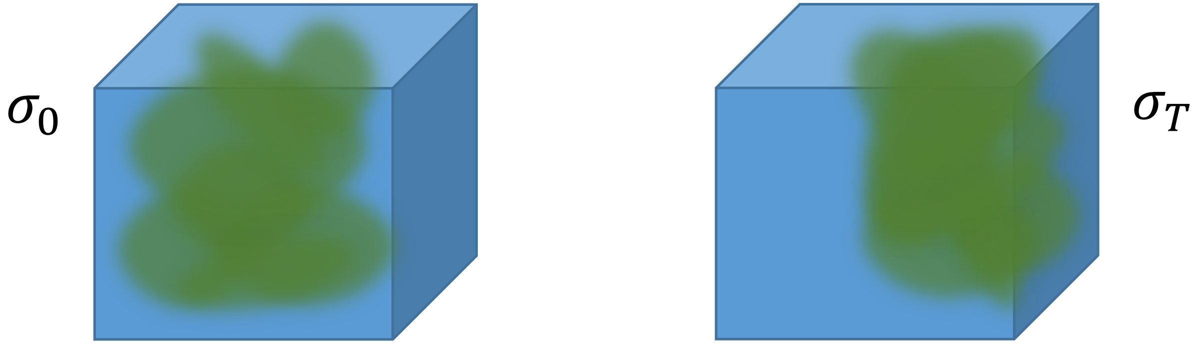
[See also work of Fathi and Shu, 2015]

Consider a space $\mathcal{P}$ of paths $\gamma : [0, T] \to \Omega$ equipped with a background measure μ (e.g., trajectories of continuous-time random walk), and also two measures σ_0 and σ_T on Ω.

Schrödinger problem:

Find the unique measure ν on $\mathcal{P}$ that interpolates between σ_0 and σ_T: If $\boldsymbol{\gamma} \sim \nu$, then $\boldsymbol{\gamma}(0) \sim \sigma_0$ and $\boldsymbol{\gamma}(T) \sim \sigma_T$ and minimizes the relative entropy to the background measure:

$$\text{minimize } D(\nu \mid \mu) = \int d\nu(\gamma) \log\left(\frac{d\nu(\gamma)}{d\mu(\gamma)}\right)$$



Now let $\{X_t : t = 0, 1, \dots, T\}$ be discrete-time random walk.
Our initial measure will be concentrated on a fixed point $X_0 = x_0$,
and the final measure will have density $f\mu_T$ where $f : \Omega \to \mathbb{R}_+$
is given and μ_T is the law of $X_T \mid X_0 = x_0$.

The optimal entropic interpolation is the process $\{Z_t\}$ given by
$Z_0 = x_0$ and for $t \leq T$,

$$\mathbb{P}(Z_t = y \mid Z_{t-1}) = p(Z_{t-1}, y) \frac{P_{T-t} f(y)}{P_{T-t+1} f(Z_{t-1})}$$

where $P_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x]$ is the discrete-time heat
semi-group.

entropic interpolation

The optimal entropic interpolation is the process $\{Z_t\}$ given by $Z_0 = x_0$ and for $t \le T$,

$$\mathbb{P}(Z_t = y \mid Z_{t-1}) = p(Z_{t-1}, y) \frac{P_{T-t} f(y)}{P_{T-t+1} f(Z_{t-1})}$$

where $P_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x]$ is the discrete-time heat semi-group.

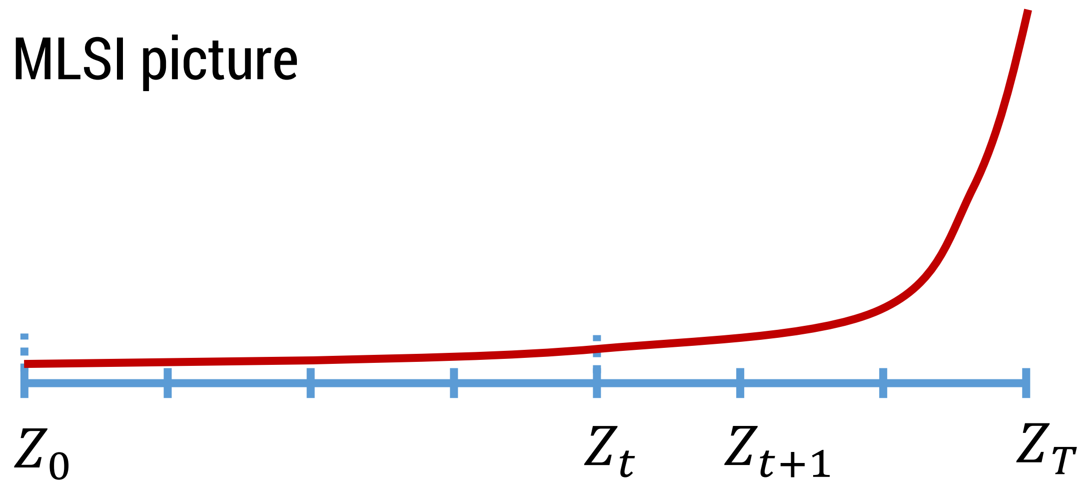
Moreover, one has:

$$D(\{Z_0, \dots, Z_T\} \mid \{X_0, \dots, X_T\}) = D(Z_T \mid X_T) = \mathrm{Ent}_{\mu_T}(f)$$

In particular, one can examine the “information burn” at each time:

$$\mathrm{Ent}_{\mu_T}(f) = \sum_{t=1}^{T} \mathbb{E}\left[\log \frac{P_{T-t}(Z_t)}{P_{T-t+1}(Z_{t-1})}\right] = \sum_{t=1}^{T} \mathbb{E}[D(\mathbb{P}_Z(Z_{t-1}, \cdot) \mid p(Z_{t-1}, \cdot))]$$

MLSI picture



Moreover, one has:

$$D(\{Z_0,\dots,Z_T\} \mid \{X_0,\dots,X_T\}) = D(Z_T \mid X_T) = \mathrm{Ent}_{\mu_T}(f)$$

In particular, one can examine the "information burn" at each time:

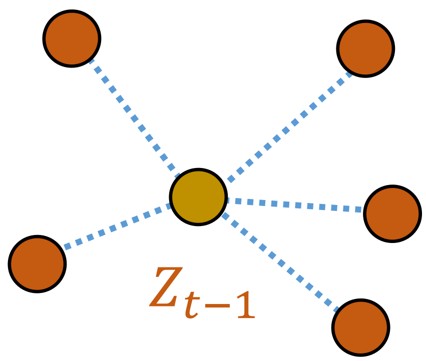
$$\mathrm{Ent}_{\mu_T}(f) = \sum_{t=1}^{T} \mathbb{E}\left[\log \frac{P_{T-t}(Z_t)}{P_{T-t+1}(Z_{t-1})}\right] = \sum_{t=1}^{T} \mathbb{E}[D(\mathbb{P}_Z(Z_{t-1},\cdot) \mid p(Z_{t-1},\cdot))]$$

interpolated random walk

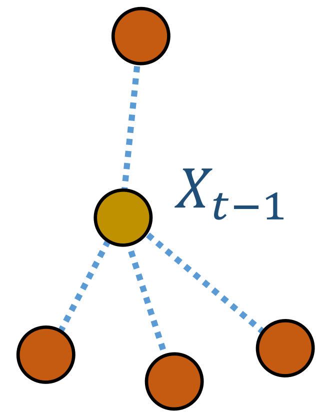
random walk

interpolated random walk

the coupling and contraction



TV-optimal coupling of Z_t and $X_1 \mid (X_0 = Z_{t-1})$

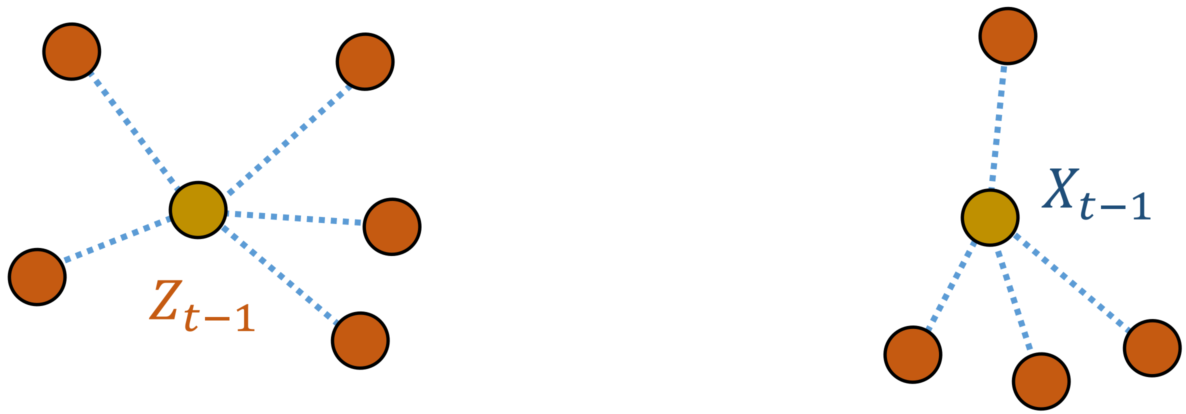


W_1-optimal coupling of X_t and $X_1 \mid (X_0 = Z_{t-1})$

Competing factors:

(i) separation decays at rate $(1-\kappa)$ because of the contraction (spend information at the end)

(ii) Pinsker's inequality $d_{\mathrm{TV}}(\mu, \nu) \leq \sqrt{D(\,\mu \mid \nu\,)}$ (spend information slowly)



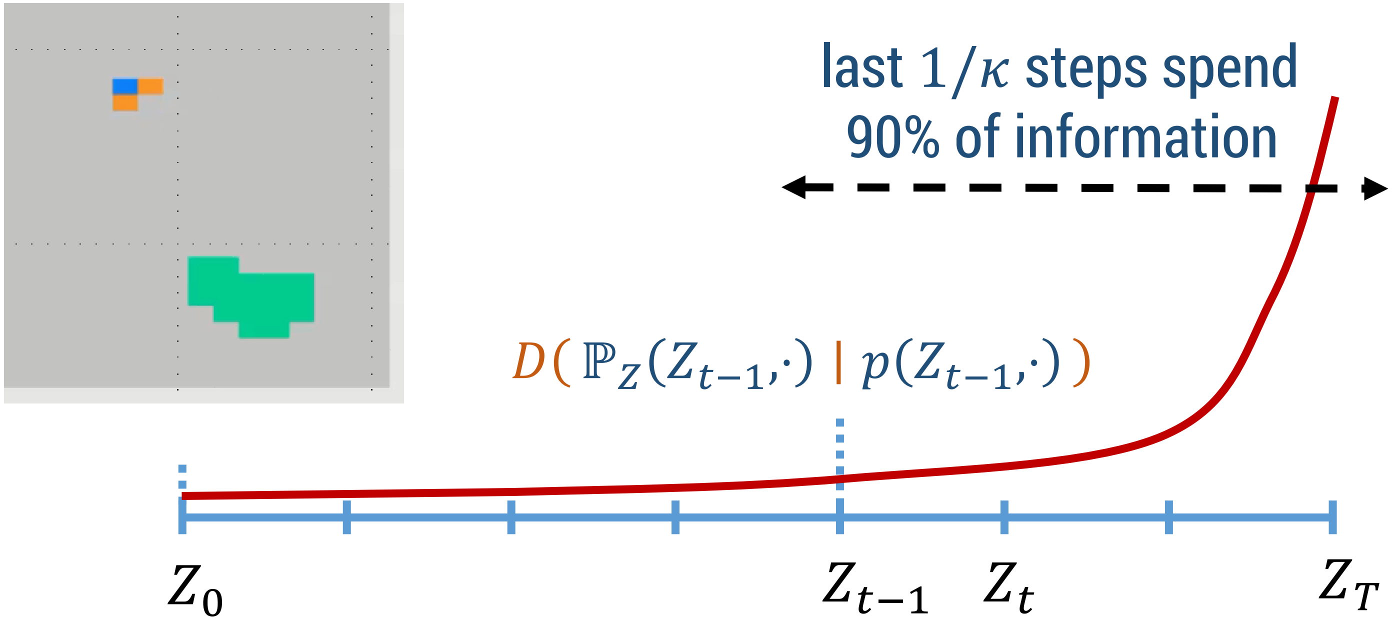
Cauchy-Schwarz of (i) and (ii): $W_1(f, \mathbf{1}) \leq \sqrt{2\kappa^{-1}\mathrm{Ent}_\pi(f)}$

Competing factors:

(i) separation decays at rate $(1-\kappa)$ because of the contraction
(spend information at the end)

(ii) Pinsker's inequality $d_{\mathrm{TV}}(\mu,\nu) \leq \sqrt{D(\mu \mid \nu)}$
(spend information slowly)

back to (modified) log-Sobolev



Question:

Is this curve monotone in time (on average, $Z_0 \sim \pi$), $T \to \infty$?
(open even for diffusion on a compact manifold)

Strategy for modified log-Sobolev:

Duality formula for relative entropy [following Borell 2000, Eldan-L 2015]

questions?