High-Dimensional Multimodal Sampling: A Pathwise Perspective

和昱辰　Shanghai Jiao Tong University

2026.8

Based on joint work with Tianhui Jiang, Zhehan Lei, Sihan Wang, Jianan Shao, and Chihao Zhang.

Why sampling?

Why sampling?



Complete the story:
“The robot looked at the stars and said, …”

Why sampling?



Complete the story:
“The robot looked at the stars and said, …”

“I wonder what is out there.”



Why sampling?



Complete the story:
“The robot looked at the stars and said, …”

“I wonder what is out there.”



“Maybe one of them is home.”

Why sampling?



Complete the story:
“The robot looked at the stars and said, …”

“I wonder what is out there.”



“Maybe one of them is home.”

“Someday, I will visit them.”

Why sampling?



Complete the story:
“The robot looked at the stars and said, …”

“I wonder what is out there.”



“Maybe one of them is home.”

“Someday, I will visit them.”

Why sampling?

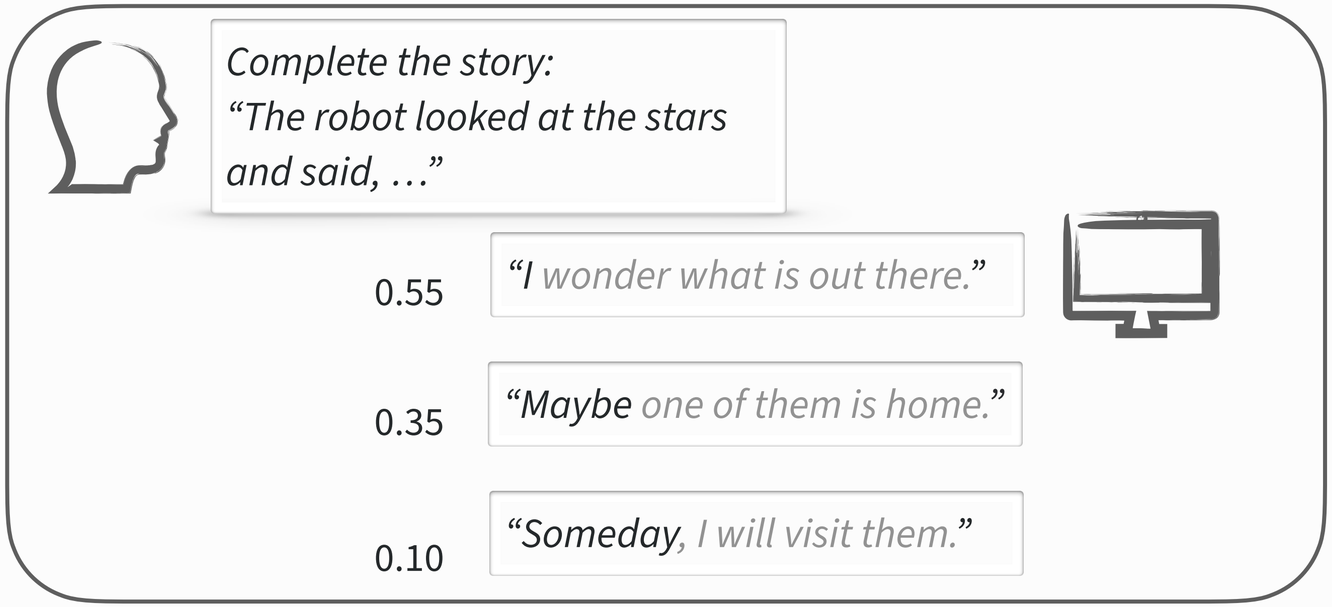


Complete the story:
“The robot looked at the stars and said, …”

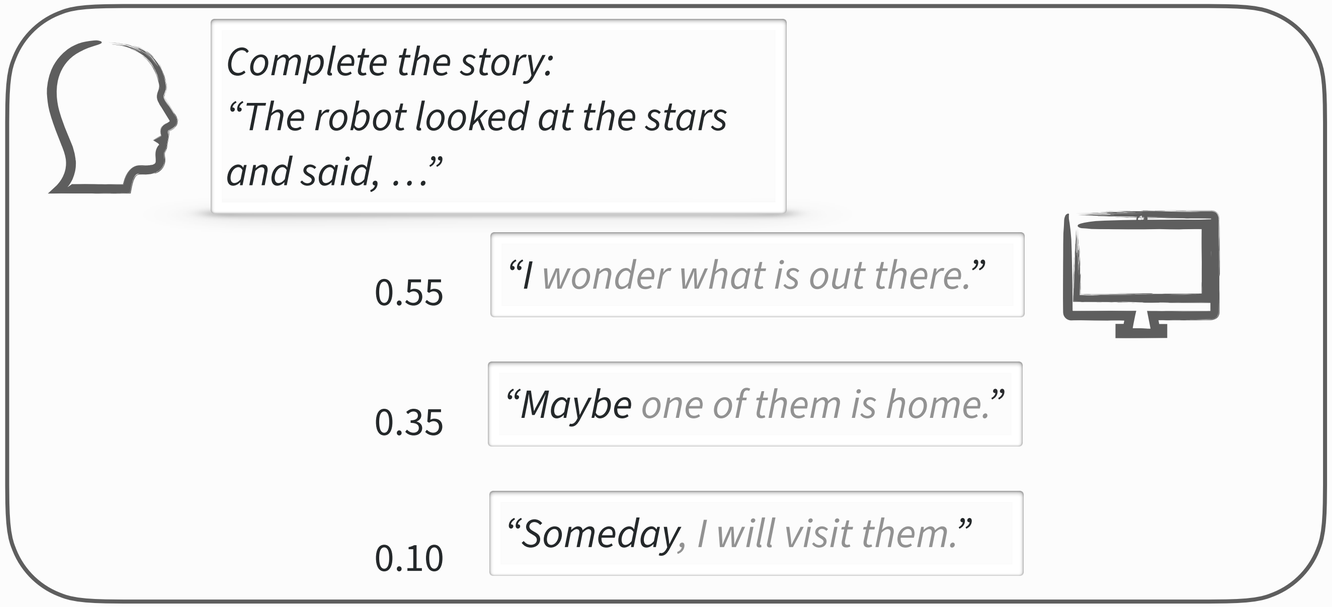


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

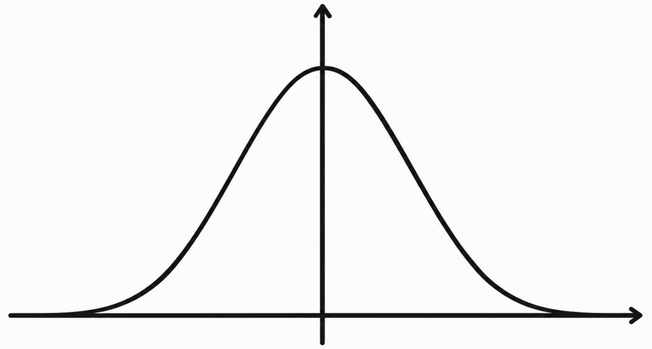
Why sampling?



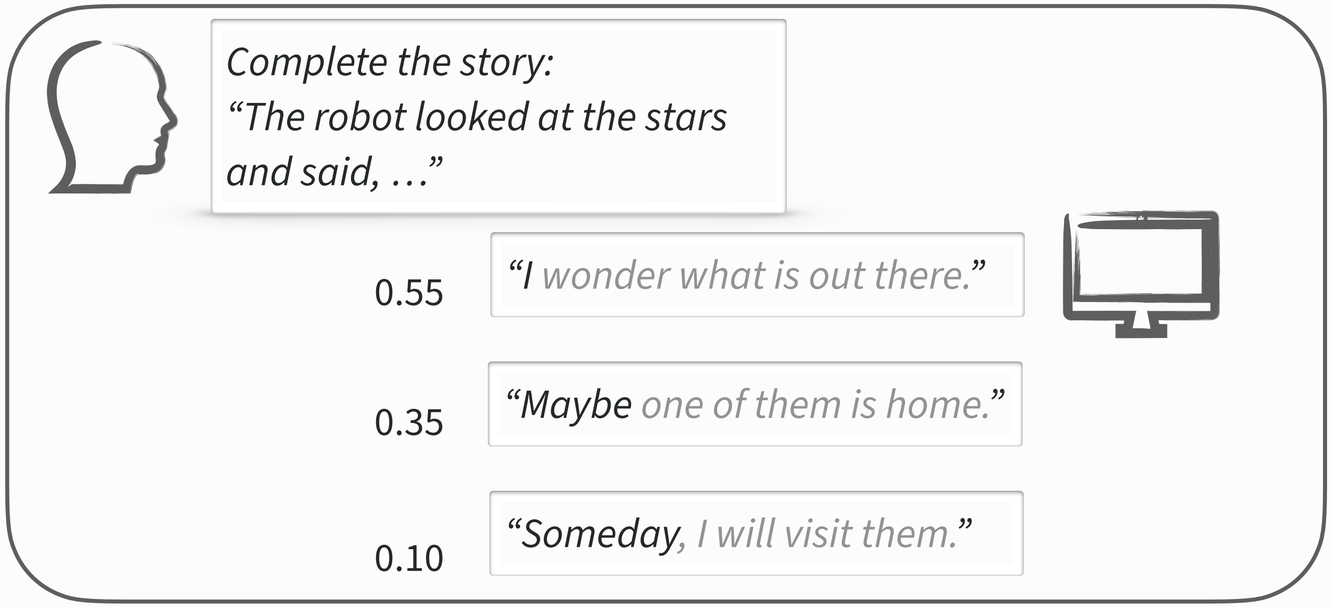
Why sampling?



$\mu(x) \propto e^{-H(x)},\ H(x) = \|x\|^2/2$

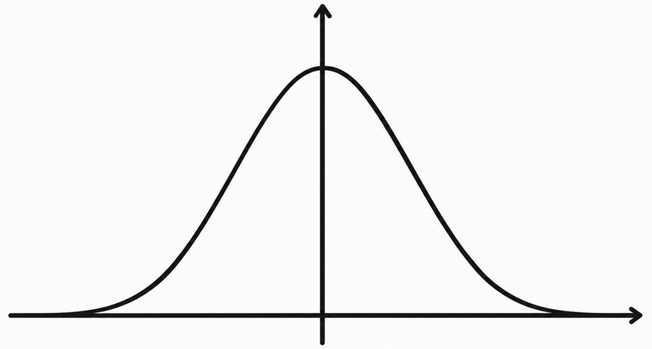


Why sampling?



Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$



Why sampling?



Complete the story:
“The robot looked at the stars and said, …”



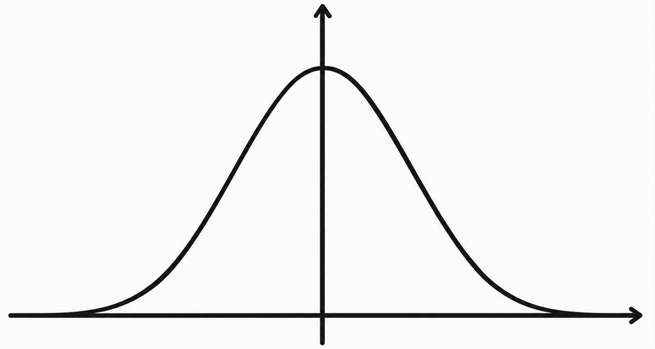
0.55 *“I wonder what is out there.”*

0.35 *“Maybe one of them is home.”*

0.10 *“Someday, I will visit them.”*

Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$



Why sampling?



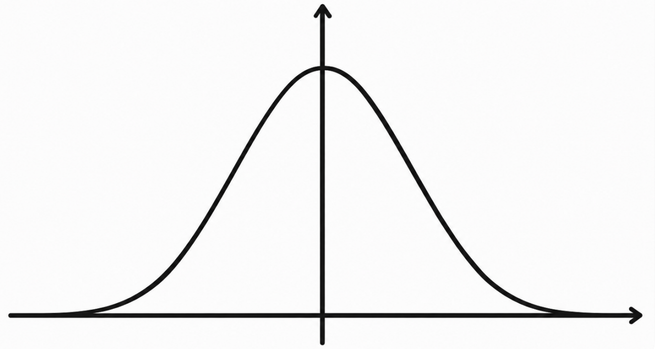
Complete the story:
“The robot looked at the stars and said, …”

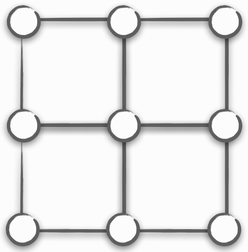


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$

Why sampling?



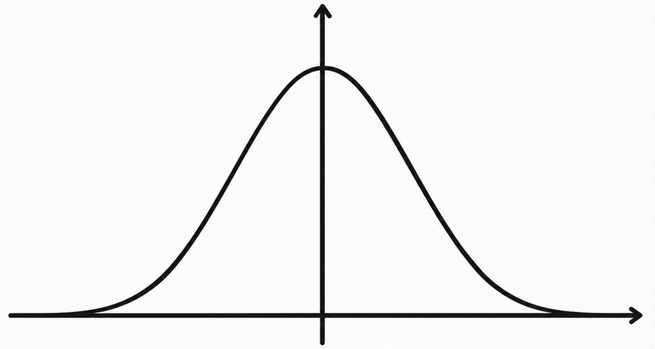
Complete the story:
“The robot looked at the stars and said, …”

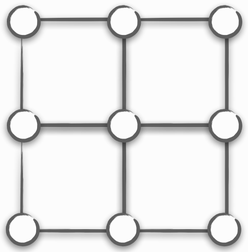


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

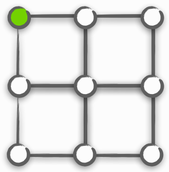
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



I_1

Why sampling?



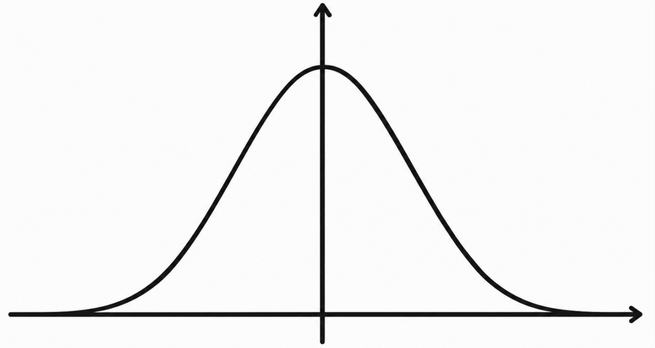
Complete the story:
“The robot looked at the stars and said, …”

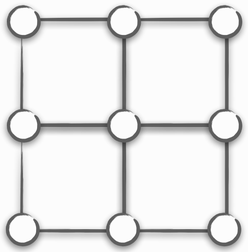


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

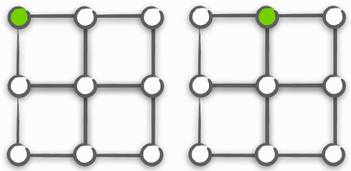
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



I_1 I_2

Why sampling?



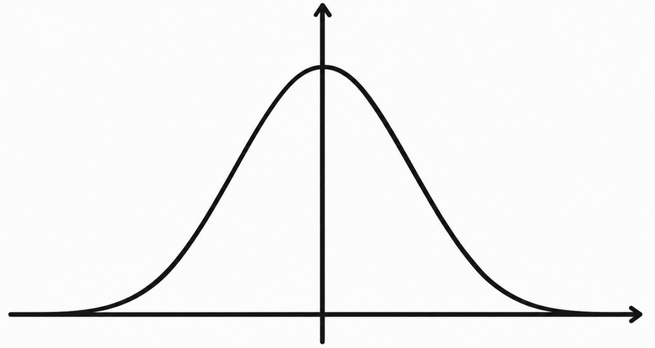
Complete the story:
“The robot looked at the stars and said, …”

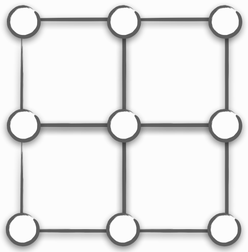


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

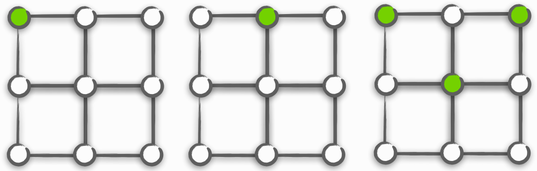
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



I_1 I_2 I_3

Why sampling?



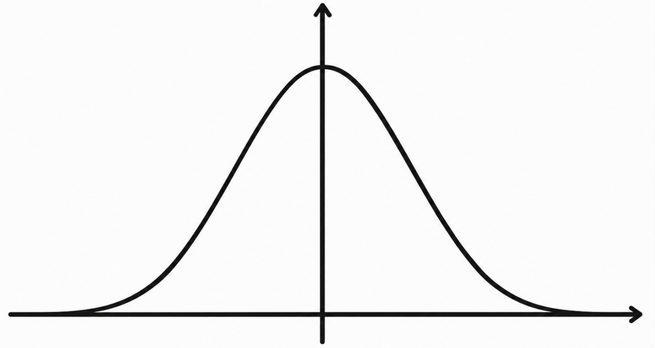
Complete the story:
“The robot looked at the stars and said, …”

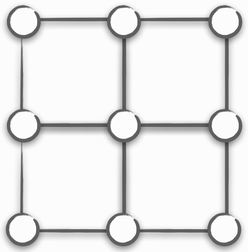


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

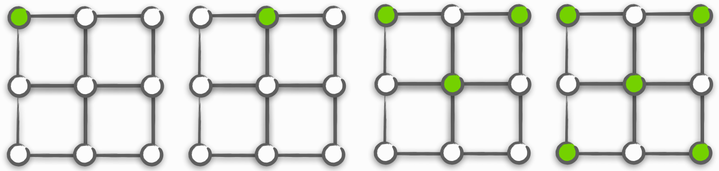
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



I_1 I_2 I_3 I_4

Why sampling?



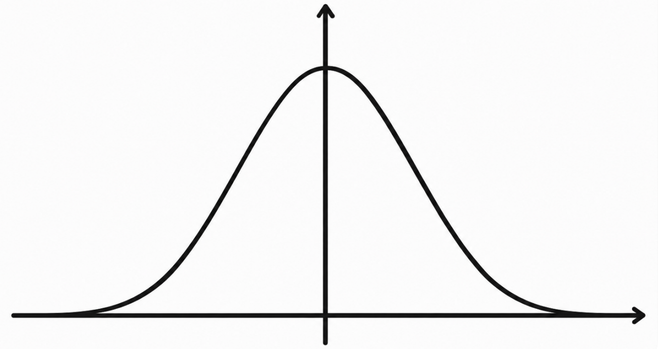
Complete the story:
“The robot looked at the stars and said, …”

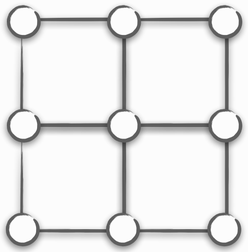


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

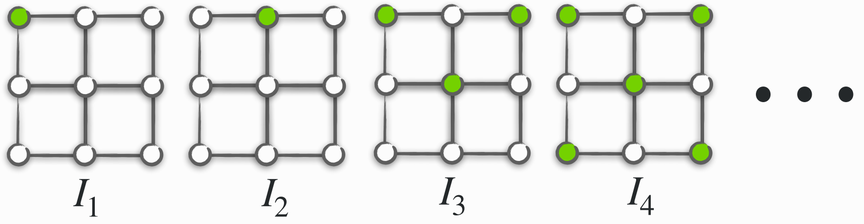
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



Why sampling?



Complete the story:
“The robot looked at the stars and said, …”



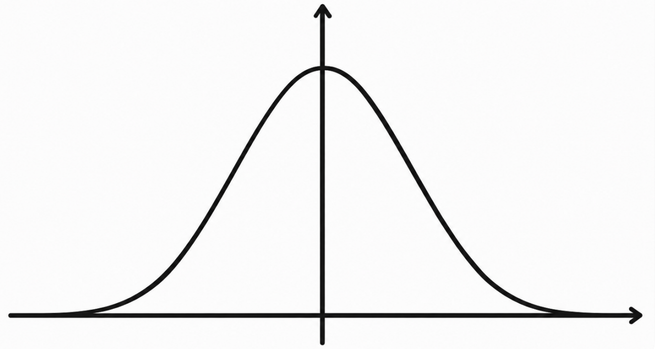
0.55 *“I wonder what is out there.”*

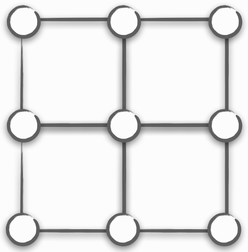
0.35 *“Maybe one of them is home.”*

0.10 *“Someday, I will visit them.”*

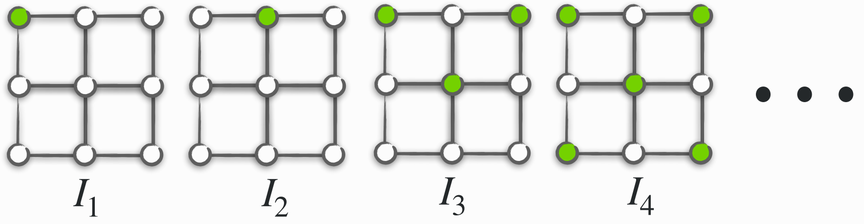
Potential function

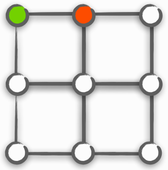
$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$





Why sampling?



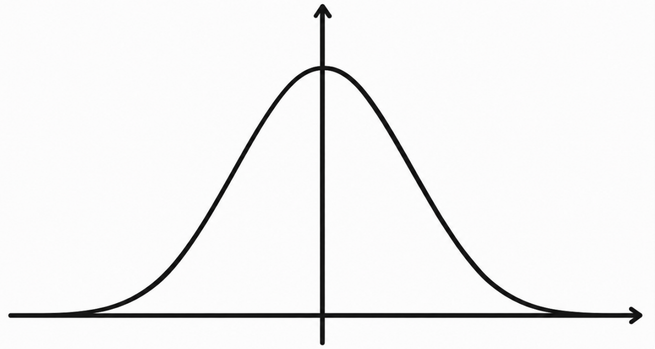
Complete the story:
“The robot looked at the stars and said, …”

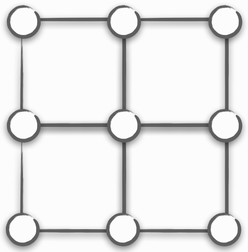


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

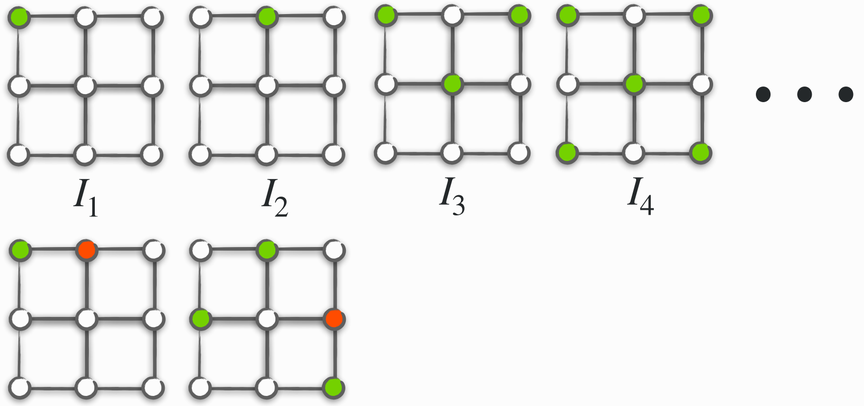
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



I_1 I_2 I_3 I_4 • • •

Why sampling?



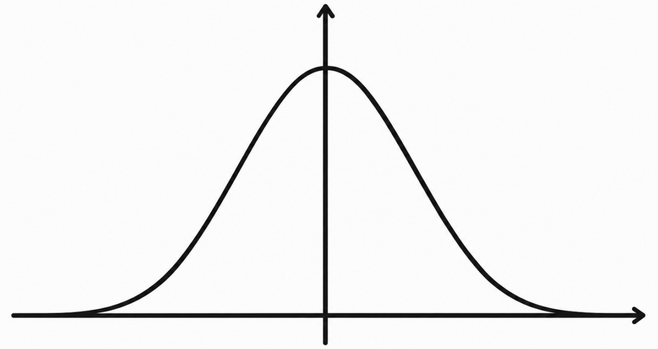
Complete the story:
“The robot looked at the stars and said, …”

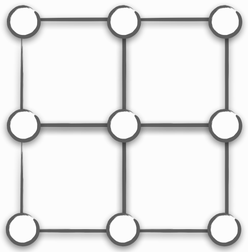


0.55	*“I wonder what is out there.”*
0.35	*“Maybe one of them is home.”*
0.10	*“Someday, I will visit them.”*

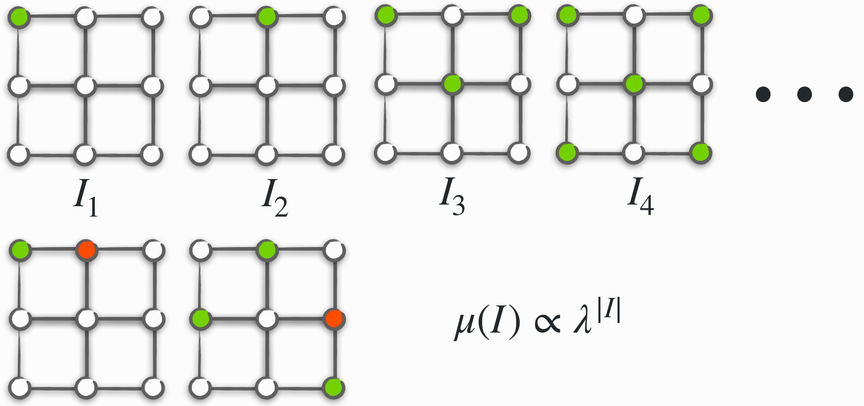
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$





$G = (V, E)$



$\mu(I) \propto \lambda^{|I|}$

Why sampling?



Complete the story:
"The robot looked at the stars and said, …"



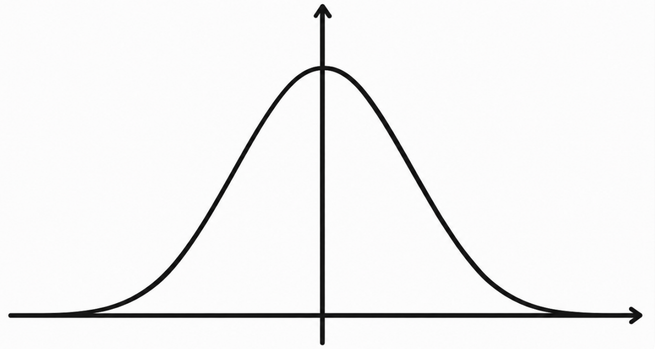
0.55 *"I wonder what is out there."*

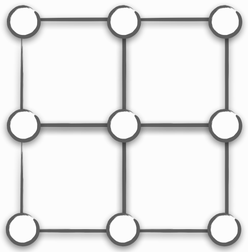
0.35 *"Maybe one of them is home."*

0.10 *"Someday, I will visit them."*

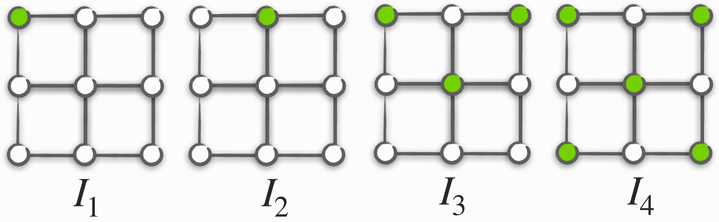
Potential function

$\mu(x) \propto e^{-H(x)}$, $H(x) = \|x\|^2/2$

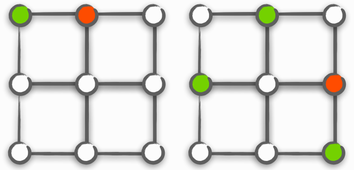




$G = (V, E)$

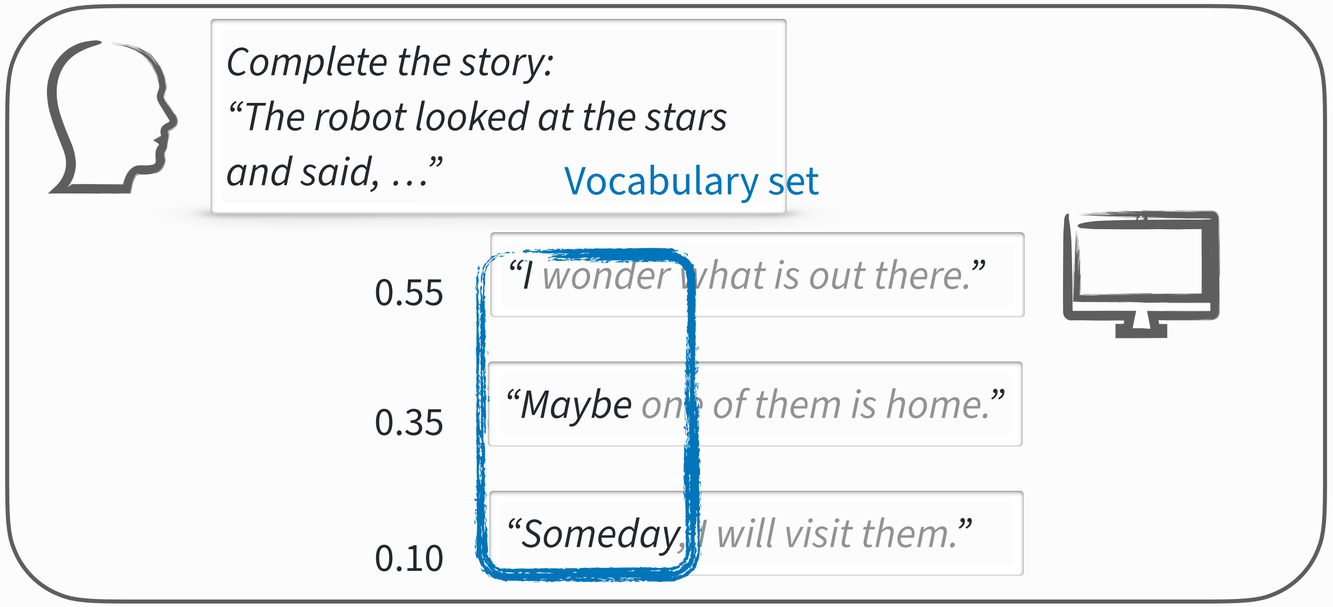


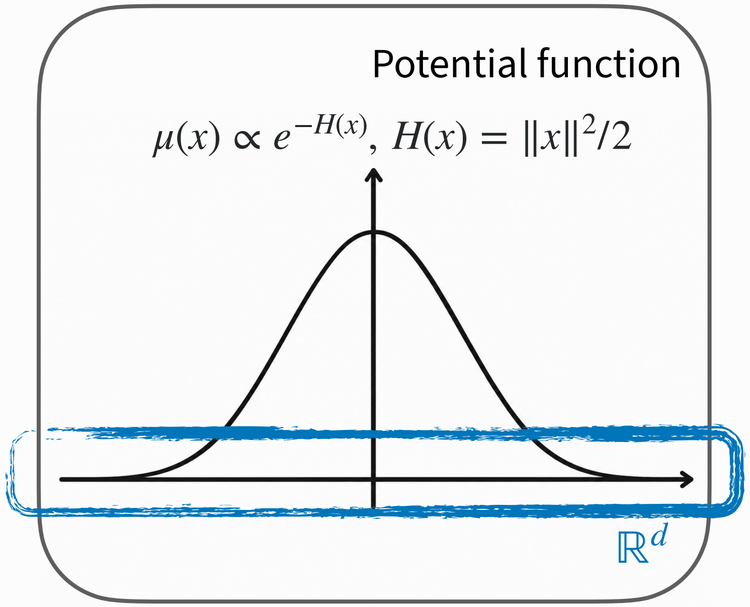
I_1 I_2 I_3 I_4

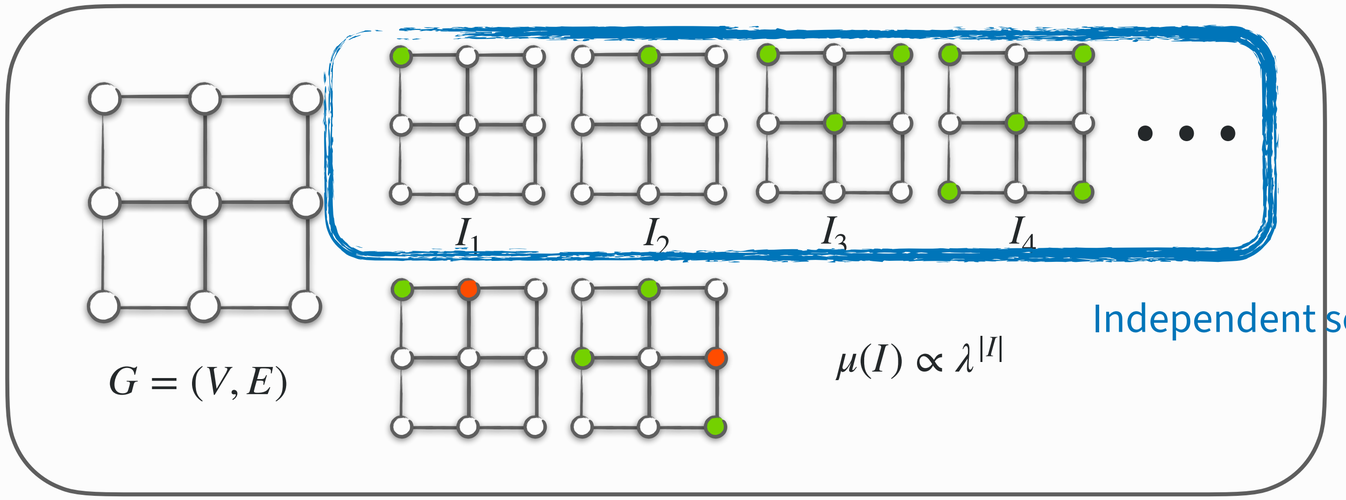


$\mu(I) \propto \lambda^{|I|}$

Why sampling?

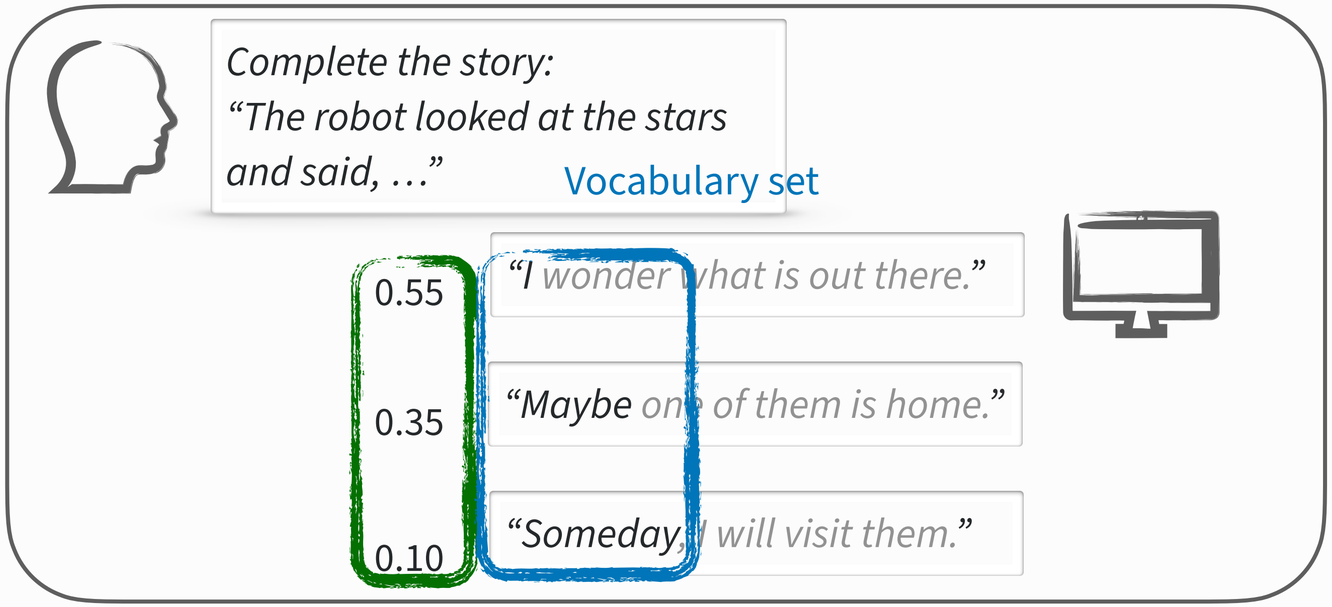


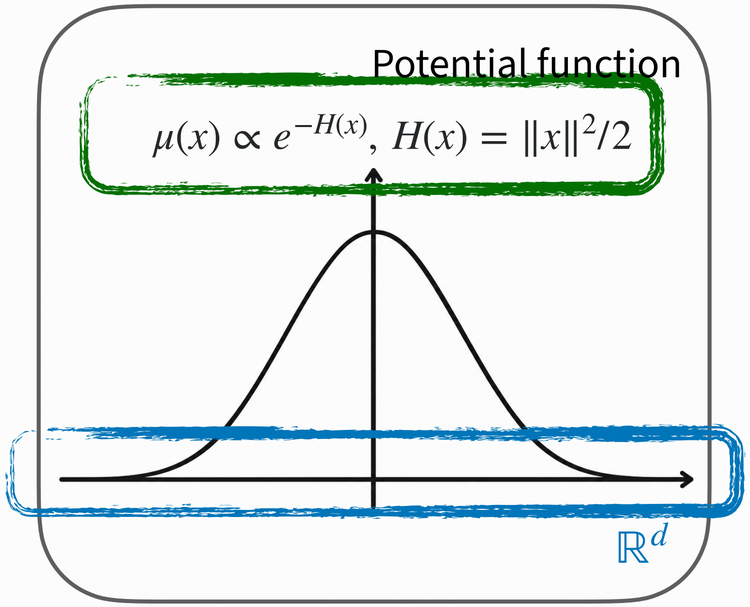


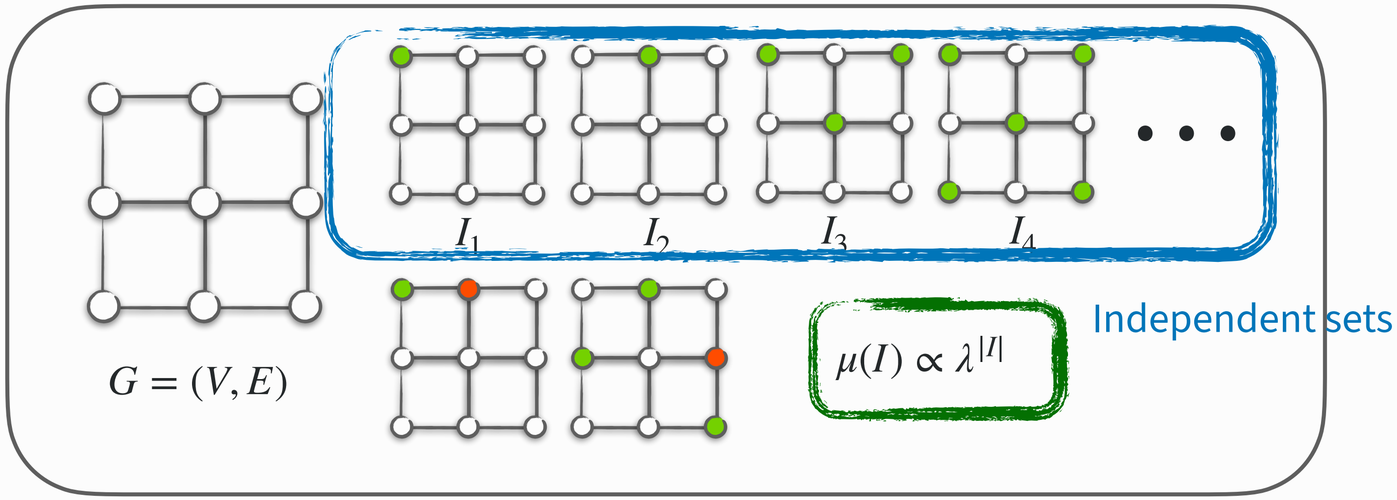


State space

Why sampling?



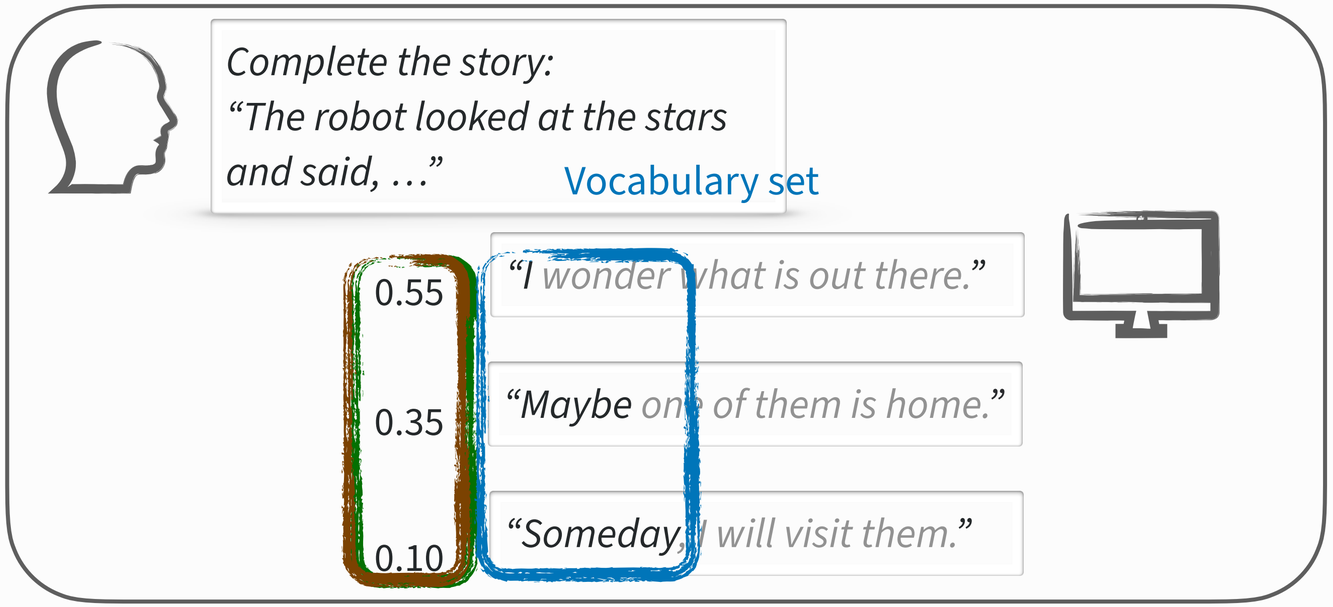


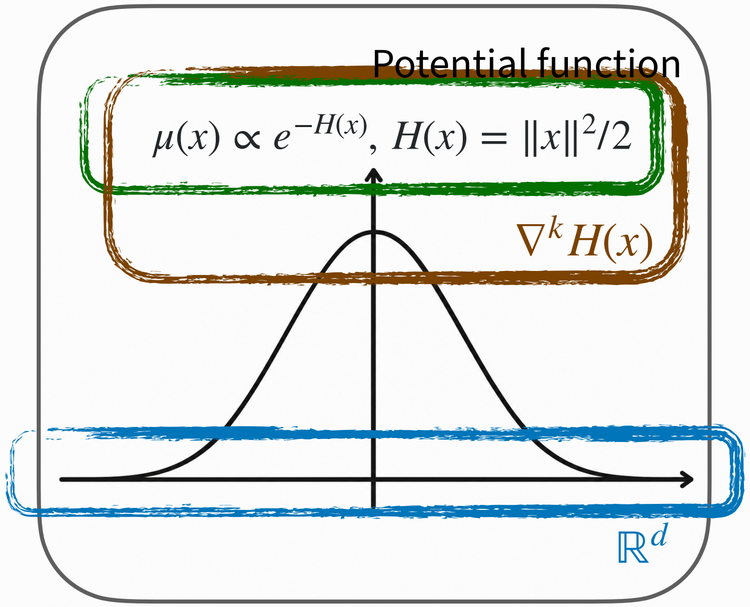


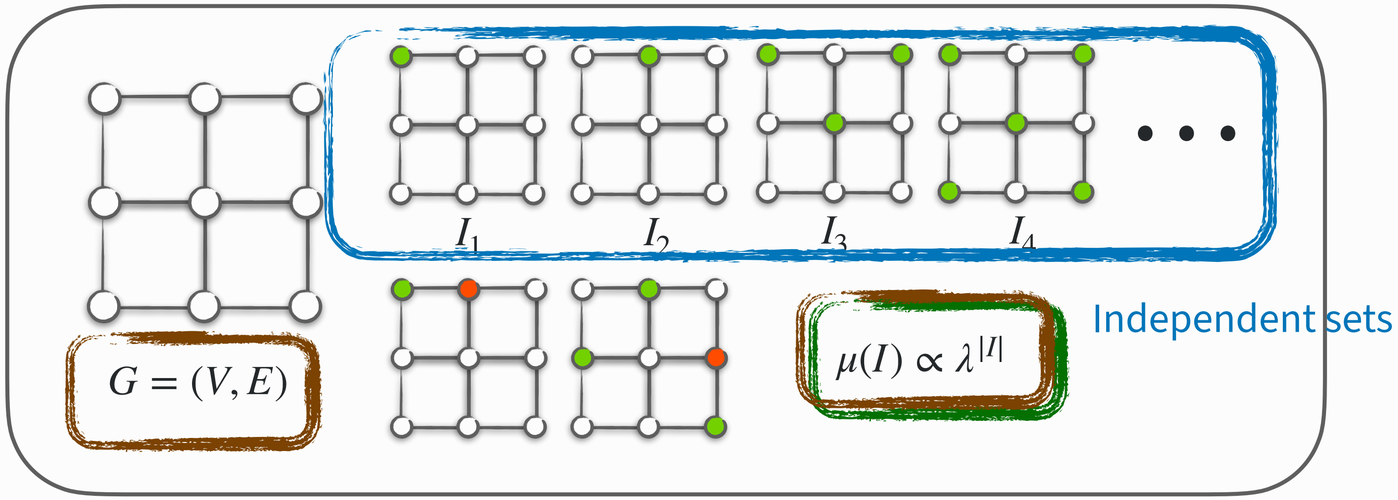
State space

A distribution

Why sampling?





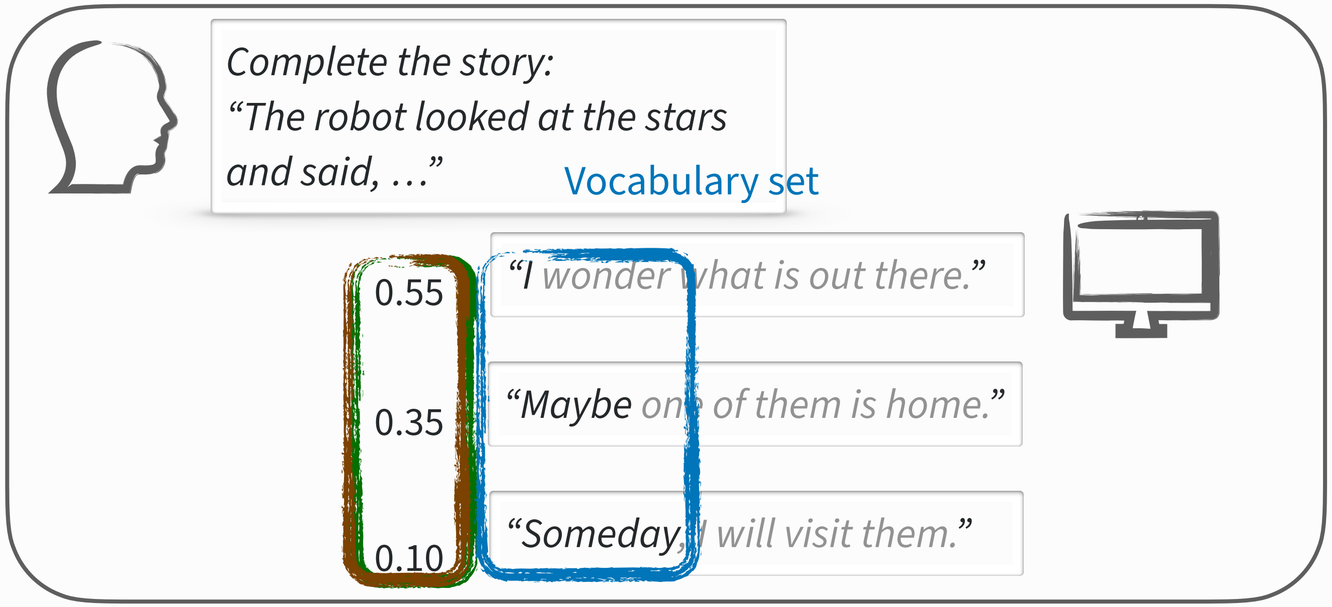


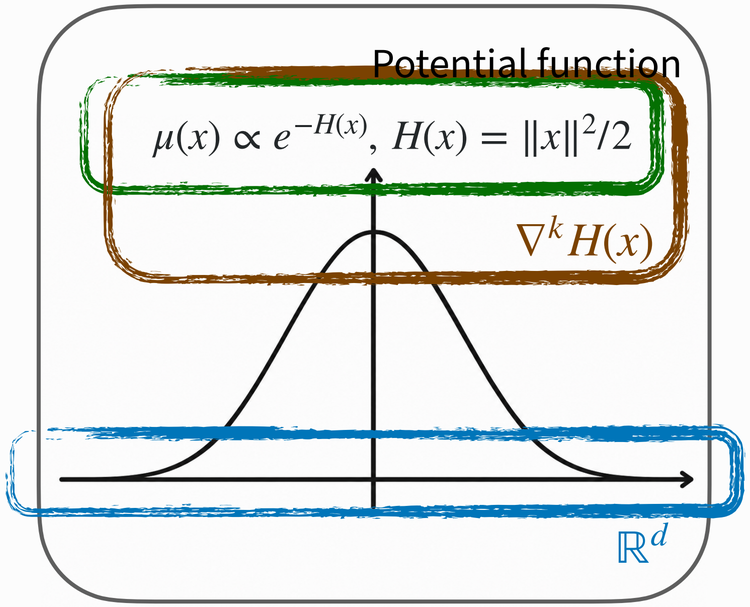
State space

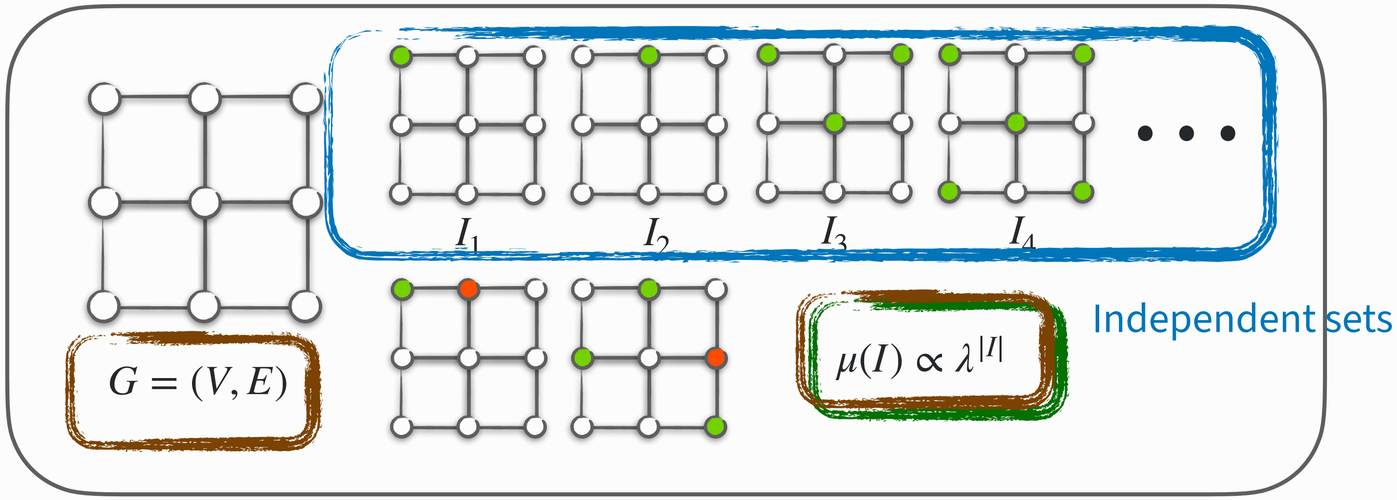
A distribution

What is given

Why sampling?







State space

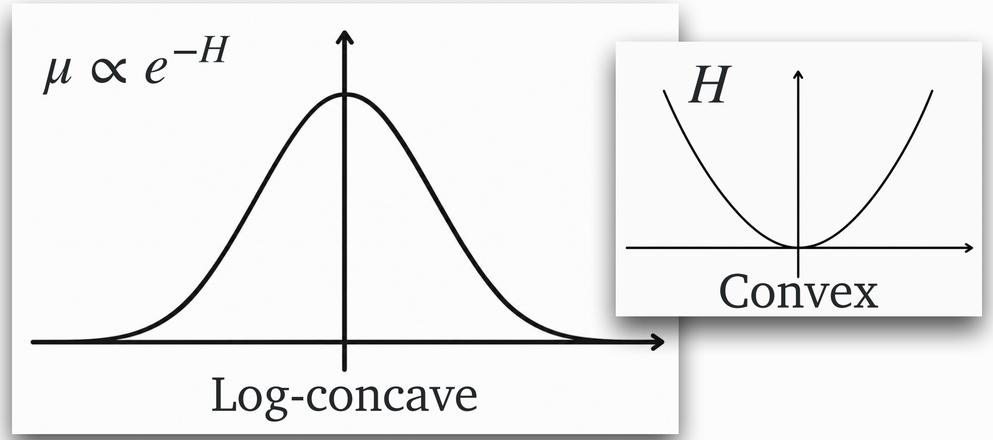
A distribution

What is given

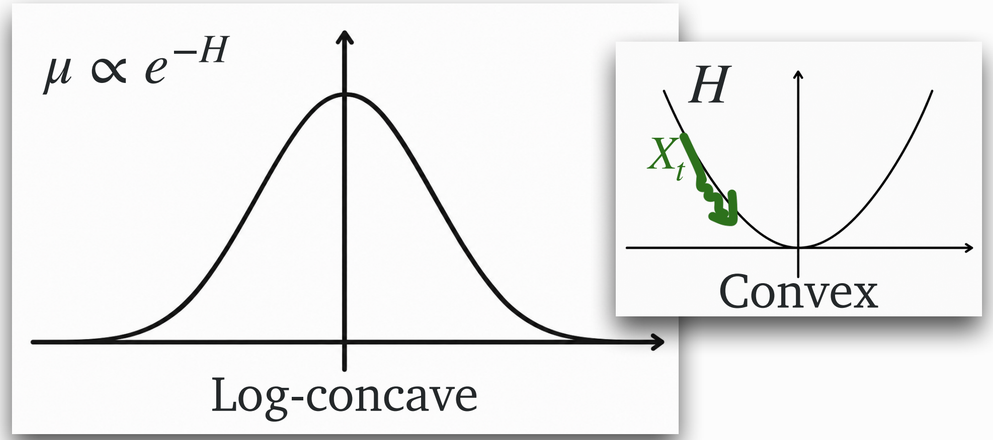
Goal: generate samples from a target distribution.

When is sampling easy/hard?

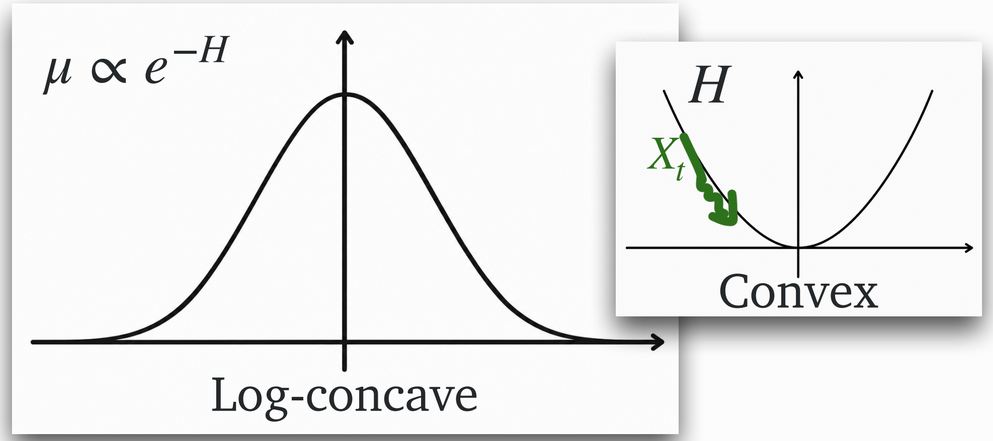
When is sampling easy/hard?

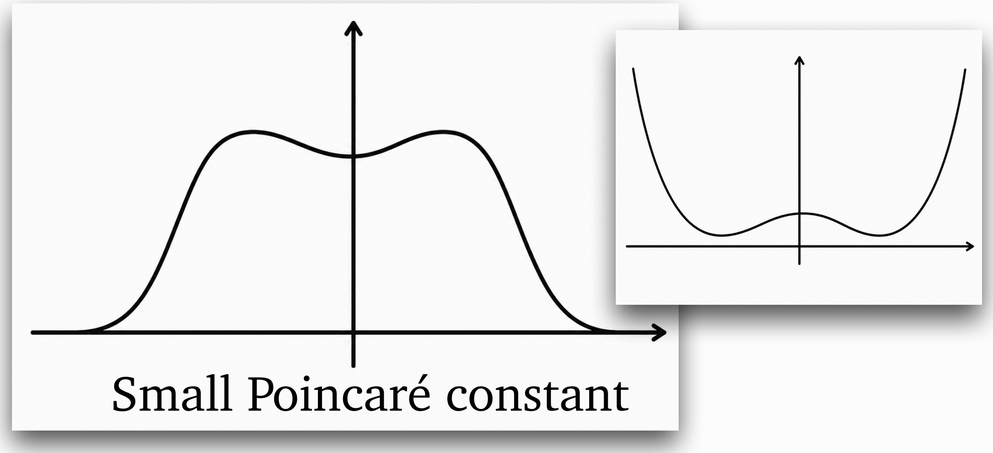


When is sampling easy/hard?

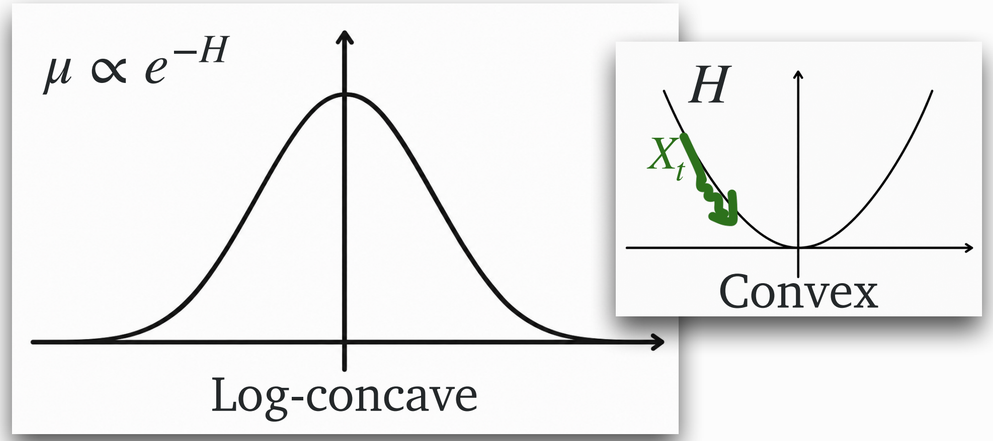


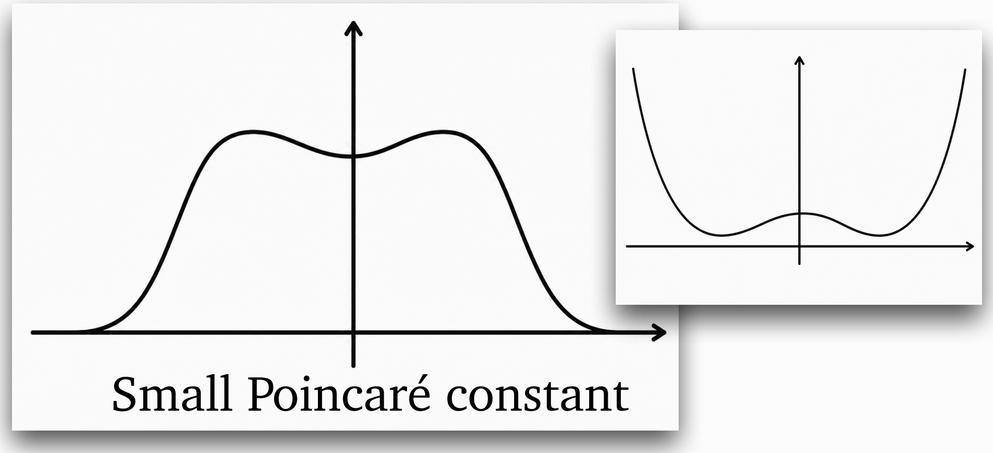
When is sampling easy/hard?





When is sampling easy/hard?



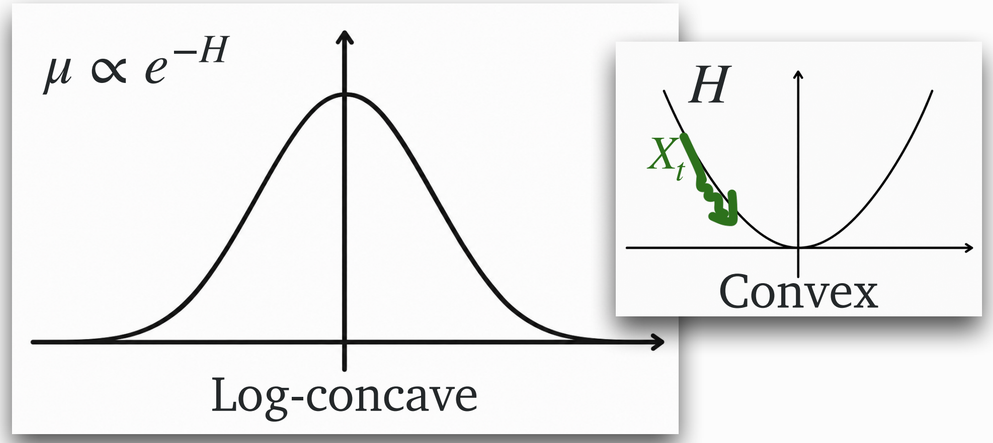


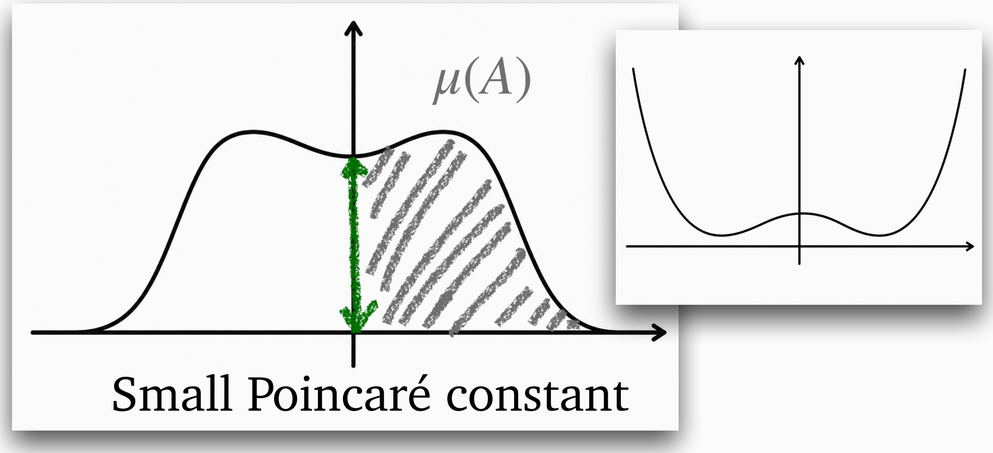
Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \frac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

Small Poincaré constant

When is sampling easy/hard?



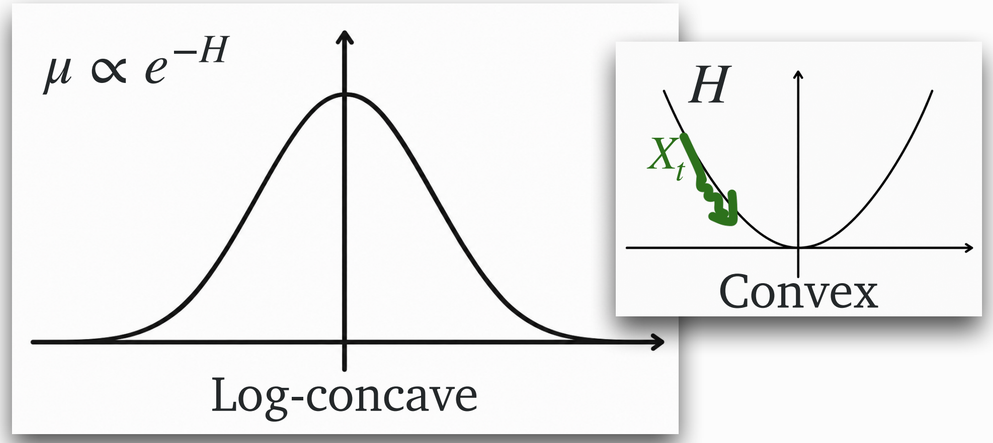


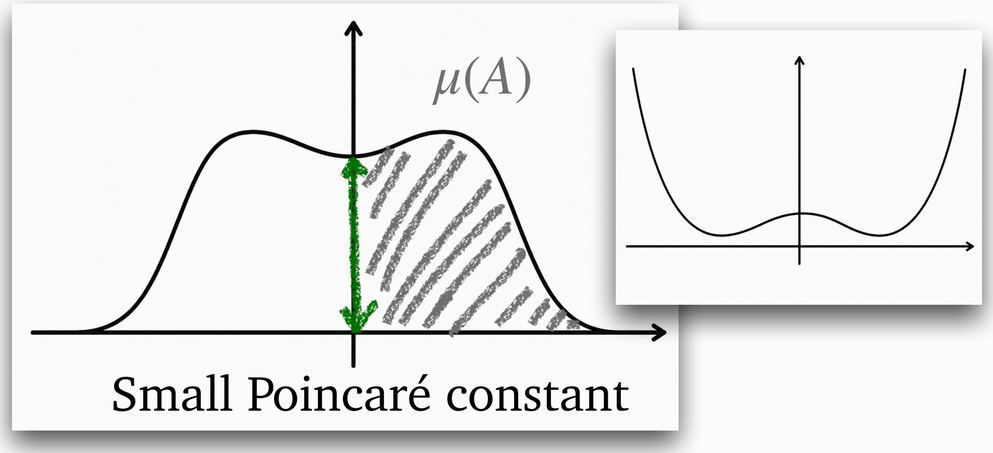
Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \dfrac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

Small Poincaré constant

When is sampling easy/hard?





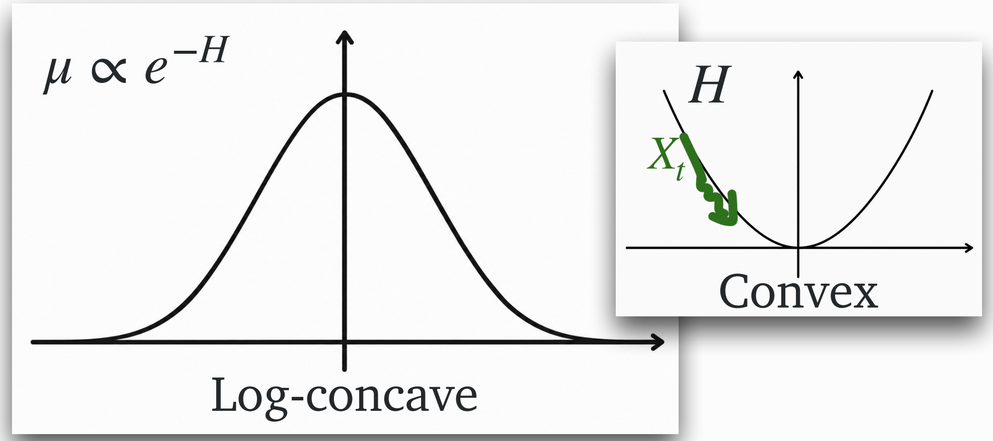
Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \dfrac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

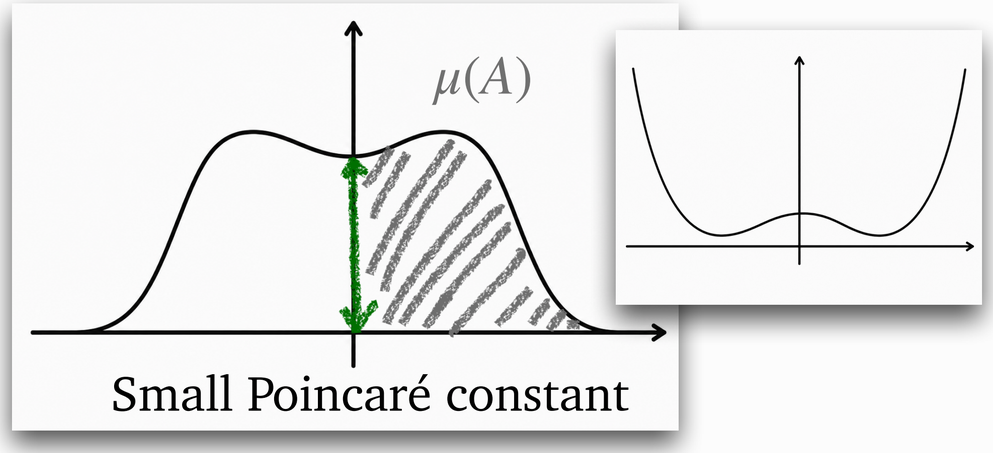
$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

Small Poincaré constant

Langevin dynamics: $\mathrm{d}X_t = -\nabla H(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

When is sampling easy/hard?





Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \dfrac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

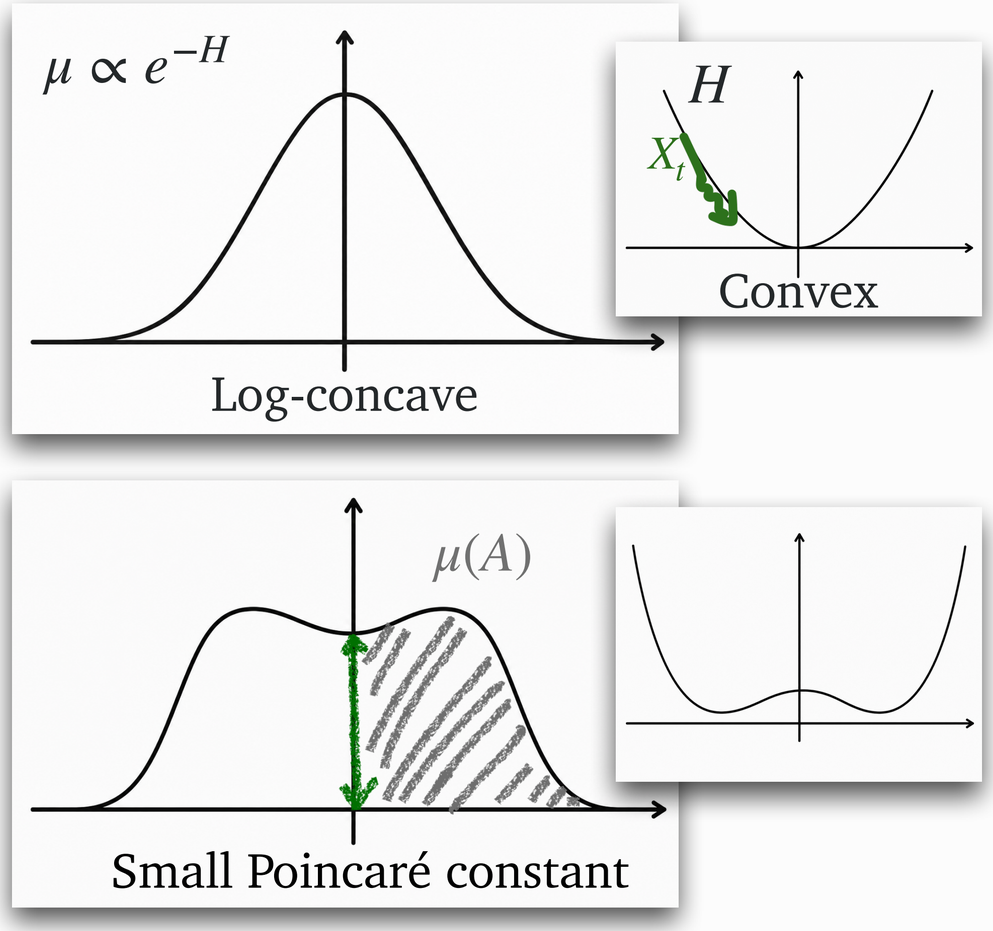
$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

Small Poincaré constant

Langevin dynamics: $\mathrm{d}X_t = -\nabla H(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

Mix in poly(d) time

When is sampling easy/hard?



Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \frac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

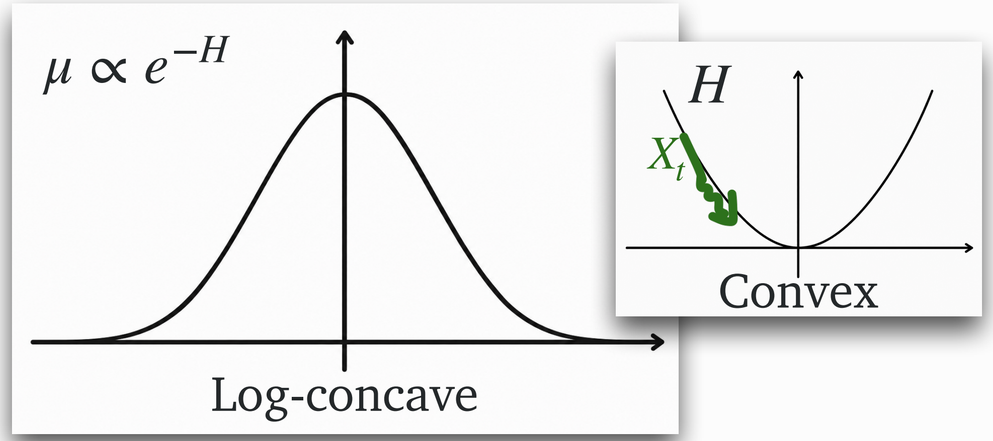
Small Poincaré constant

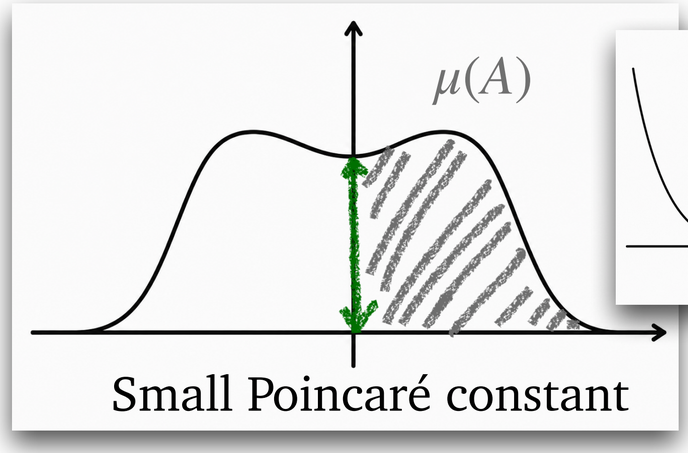
Langevin dynamics: $\mathrm{d}X_t = -\nabla H(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

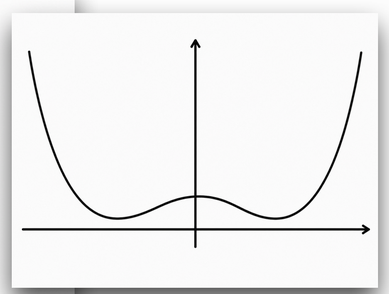
Mix in poly(d) time

Sampling is easy when the distribution has good global connectivity.

When is sampling easy/hard?







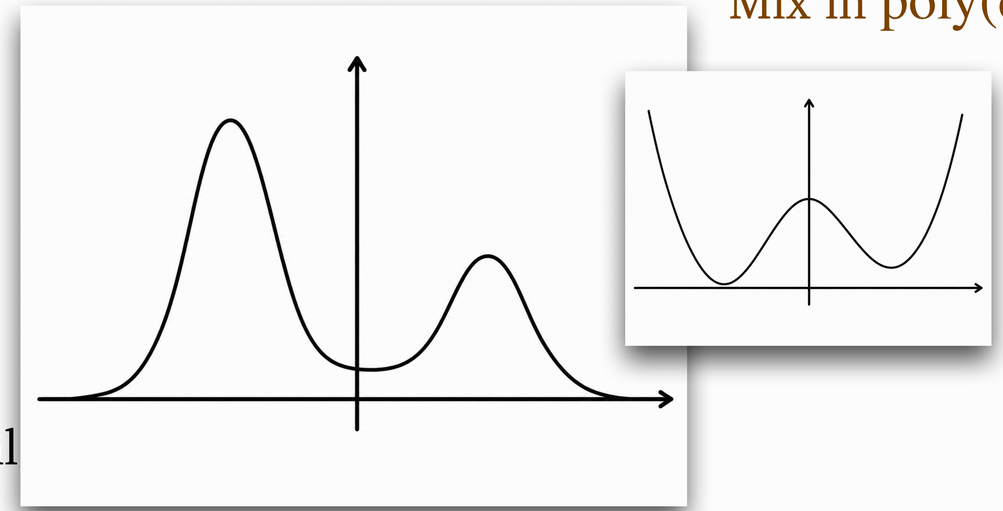
Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \frac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

Small Poincaré constant

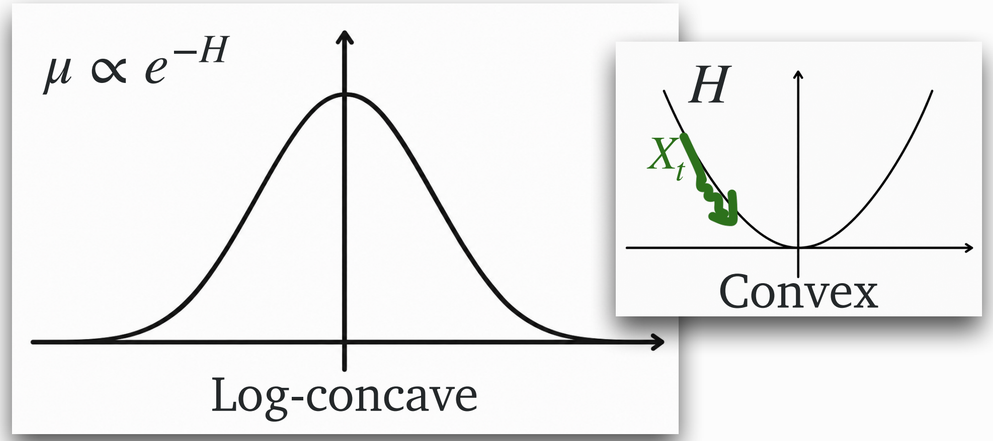
Langevin dynamics: $\mathrm{d}X_t = -\nabla H(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

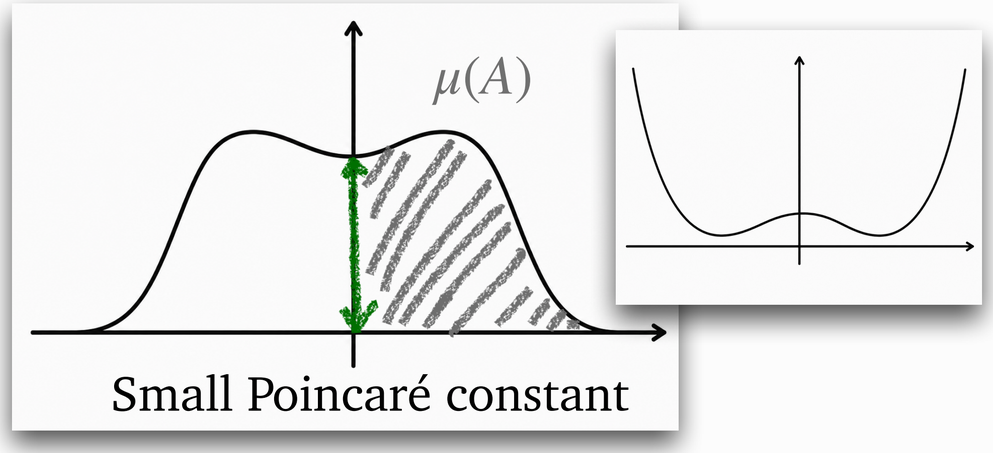
Mix in poly(d) time



Sampling is easy when the distribution has good global

When is sampling easy/hard?





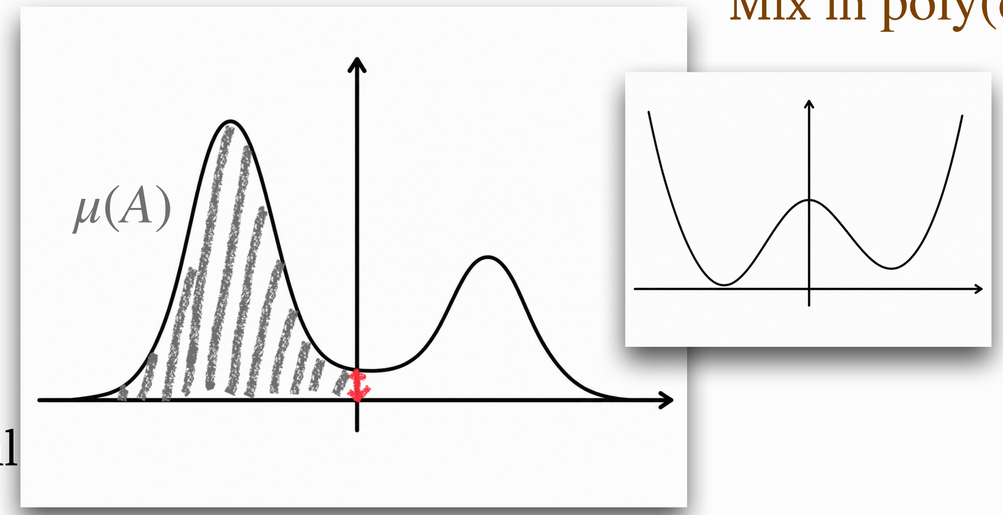
Poincaré constant: $C_{\mathrm{PI}}(\mu) = \sup_{f \not\equiv \mathrm{const}} \frac{\mathrm{Var}_\mu(f)}{\mathscr{E}_\mu(f,f)}$

$\mathscr{E}_\mu(f,f) = \mathbf{E}_\mu[\|\nabla f\|^2]$

Small Poincaré constant

Langevin dynamics: $\mathrm{d}X_t = -\nabla H(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

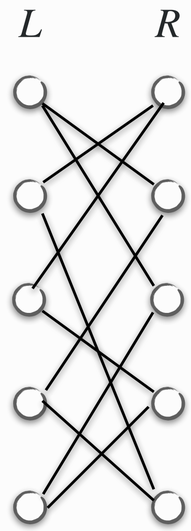
Mix in poly(d) time



Sampling is easy when the distribution has good global

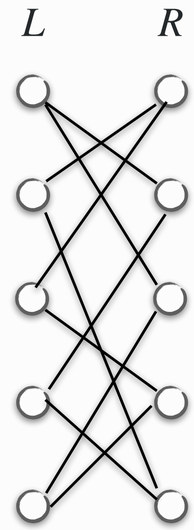
When is sampling easy/hard?

When is sampling easy/hard?



$G = (V, E)$

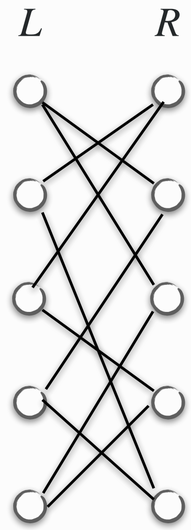
When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

When is sampling easy/hard?

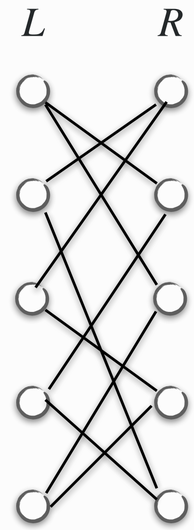


$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

When is sampling easy/hard?



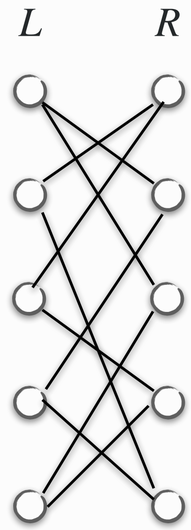
$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

When is sampling easy/hard?



$G = (V, E)$

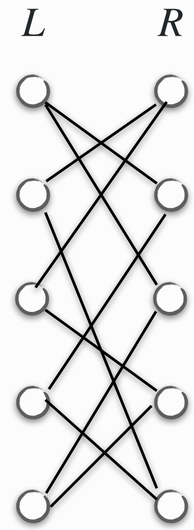
$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

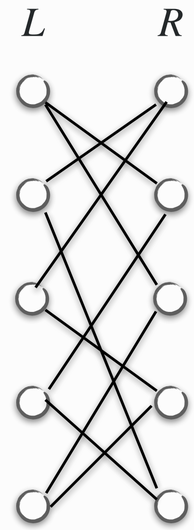
I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta - 1)^{\Delta - 1}}{(\Delta - 2)^{\Delta}} \approx \frac{e}{\Delta}$$

When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

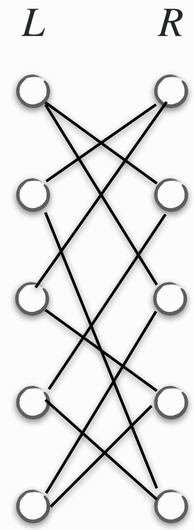
Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta - 1)^{\Delta - 1}}{(\Delta - 2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \leq \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

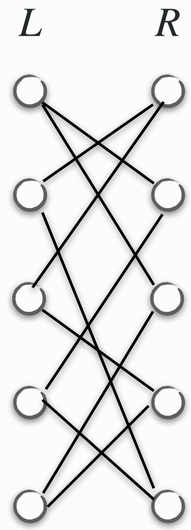
On general graphs:

$\lambda \leq \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

A Markov chain Monte Carlo algorithm.

A local algorithm.

When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \leq \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

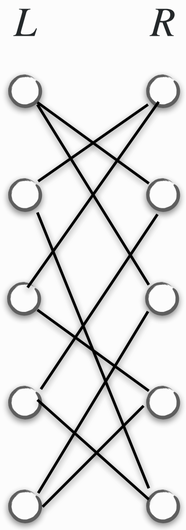
A Markov chain Monte Carlo algorithm.

A local algorithm.

Large λ



When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \leq \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

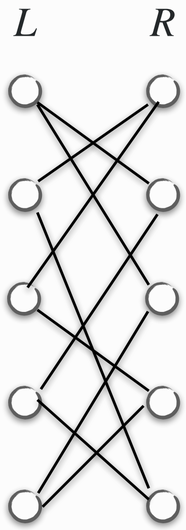
A Markov chain Monte Carlo algorithm.

A local algorithm.

Large λ



When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \leq \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

$\lambda > \lambda_c(\Delta)$: no polynomial-time sampler unless NP = RP ([SS12])

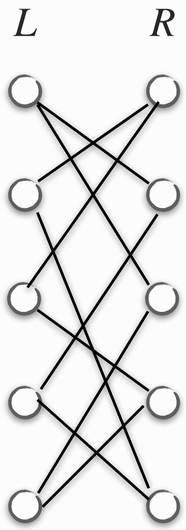
A Markov chain Monte Carlo algorithm.

A local algorithm.

Large λ



When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \le \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

$\lambda > \lambda_c(\Delta)$: no polynomial-time sampler unless NP = RP ([SS12])

On bipartite graphs:

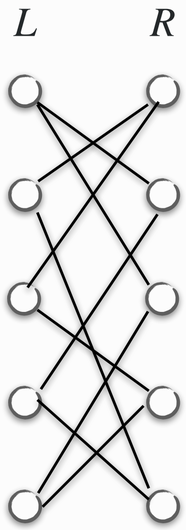
$\lambda > \lambda_c(\Delta)$: open

A Markov chain Monte Carlo algorithm.
A local algorithm.

Large λ



When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \le \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

$\lambda > \lambda_c(\Delta)$: no polynomial-time sampler unless NP = RP ([SS12])

On bipartite graphs:

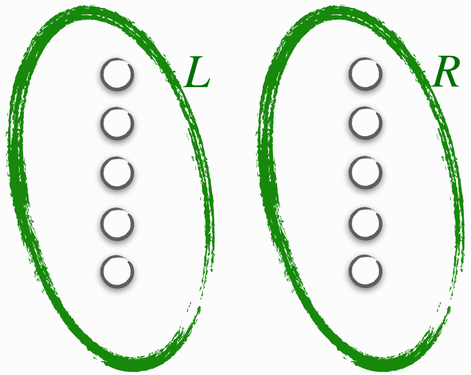
$\lambda > \lambda_c(\Delta)$: open

A Markov chain Monte Carlo algorithm.
A local algorithm.

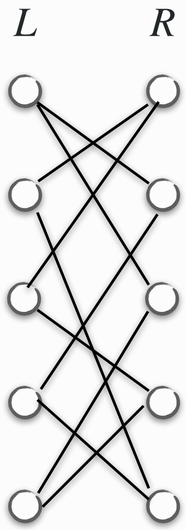
Large λ



$\lambda \to \infty$



When is sampling easy/hard?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}, \lambda > 0$

I: independent set

$n = |V|$

Δ: maximum degree

$$\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta}$$

On general graphs:

$\lambda \le \lambda_c(\Delta)$: Glauber dynamics mix in poly(n) ([CLV23,CCYZ25])

$\lambda > \lambda_c(\Delta)$: no polynomial-time sampler unless NP = RP ([SS12])

On bipartite graphs:

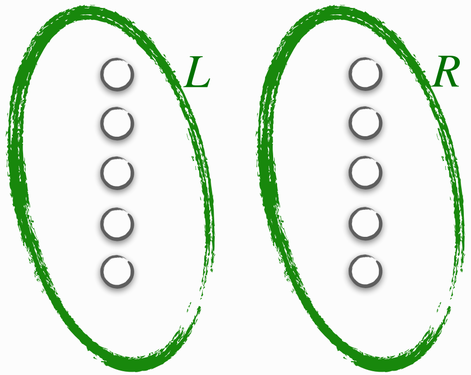
$\lambda > \lambda_c(\Delta)$: open

A Markov chain Monte Carlo algorithm.
A local algorithm.

Large λ



$\lambda \to \infty$

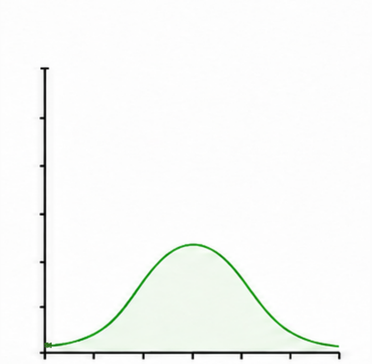


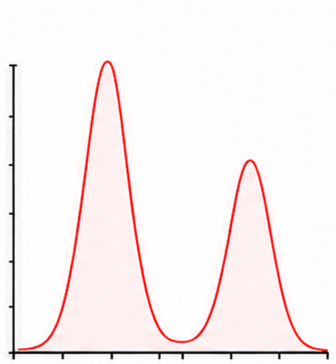
$$\mu = \frac{1}{2}\delta_L + \frac{1}{2}\delta_R$$

A trivial distribution

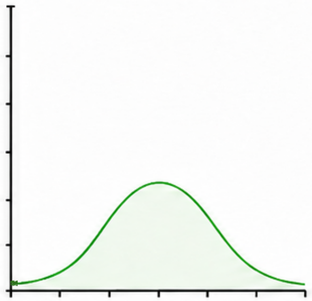
What if the target distribution has poor geometry?

What if the target distribution has poor geometry?

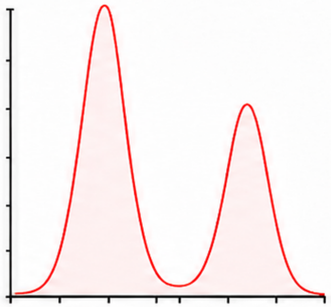




What if the target distribution has poor geometry?

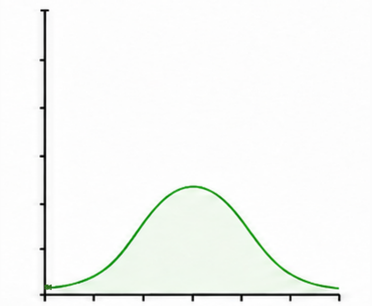


easy

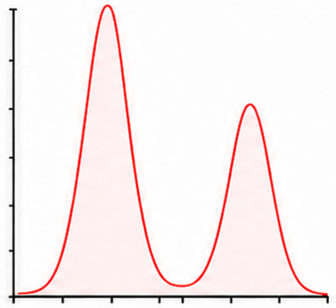


hard distribution

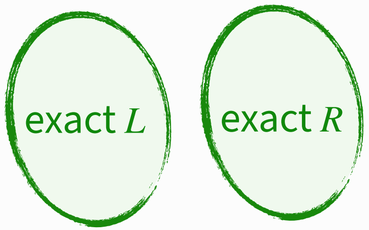
What if the target distribution has poor geometry?

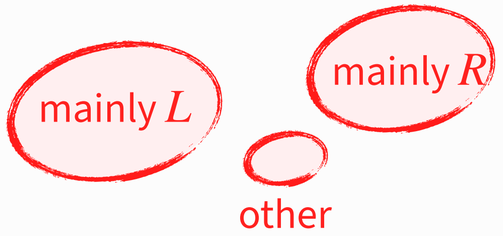


easy



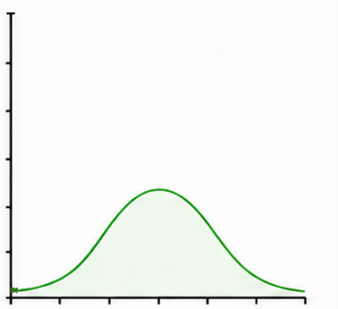
hard distribution



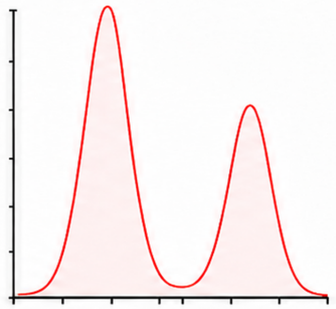


What if the target distribution has poor geometry?

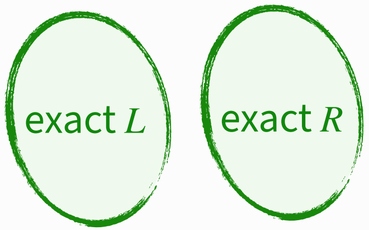
Can we reach a hard target through a well-behaved path?

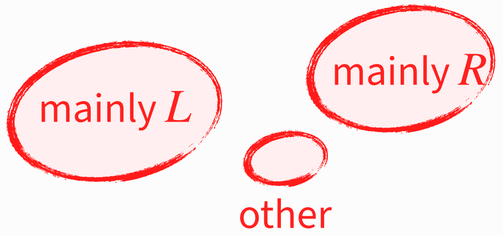


easy



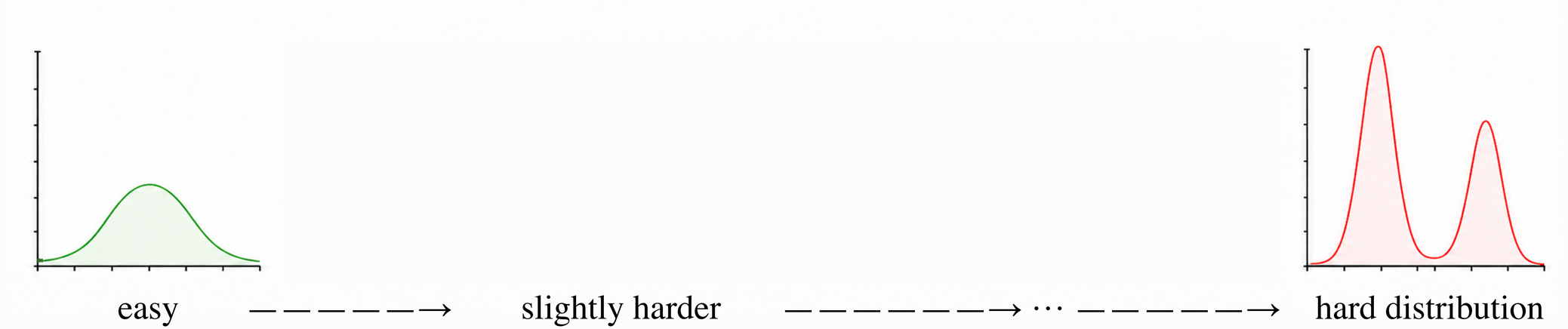
hard distribution

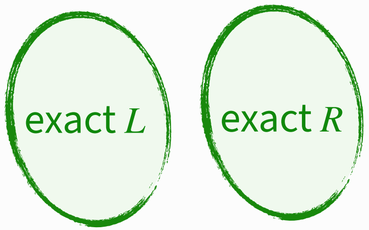


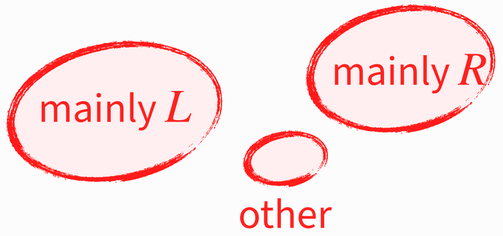


What if the target distribution has poor geometry?

Can we reach a hard target through a well-behaved path?

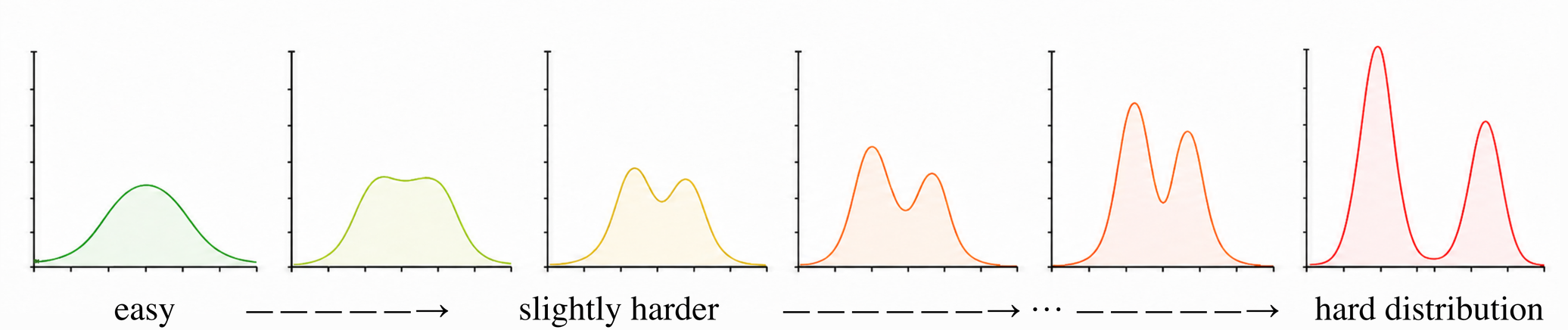






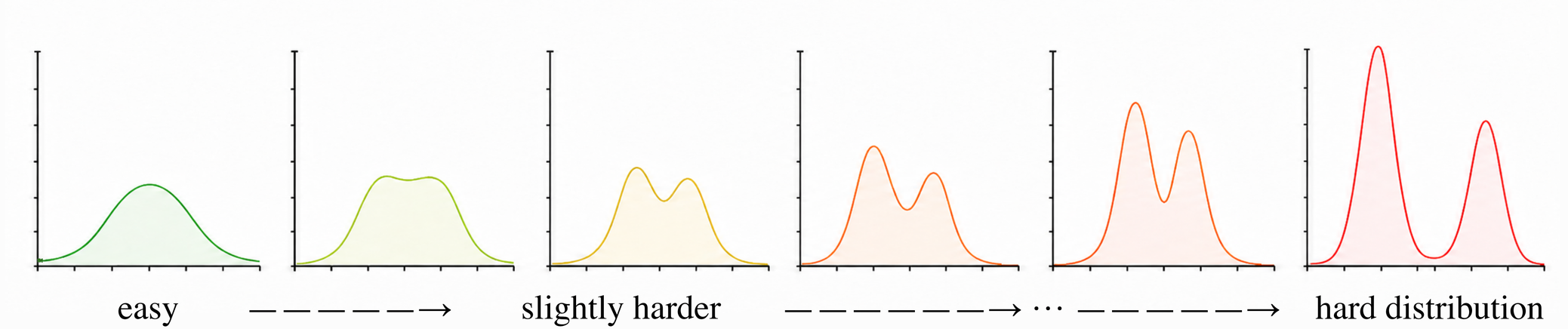
What if the target distribution has poor geometry?

Can we reach a hard target through a well-behaved path?



What if the target distribution has poor geometry?

Can we reach a hard target through a well-behaved path?



What geometric property makes such a path algorithmically useful?

Noisy paths

Noisy paths

Target: $X_0 \sim \mu$

Noisy paths

Target: $X_0 \sim \mu$

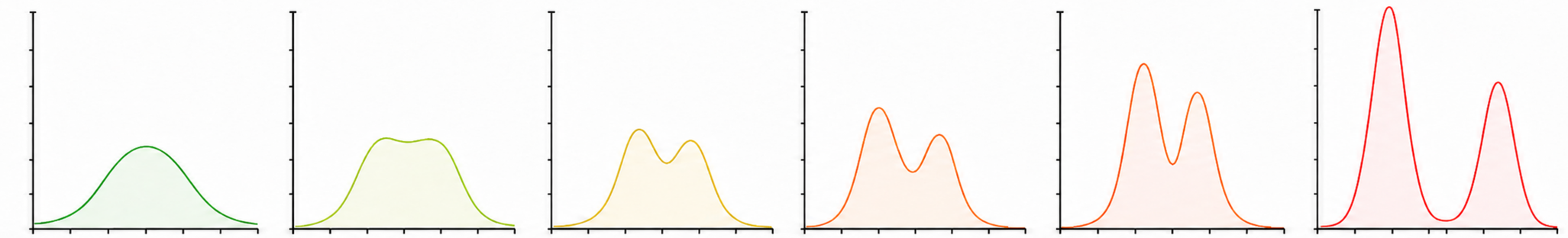
Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

Noisy paths

Target: $X_0 \sim \mu$

Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

This induces a path:

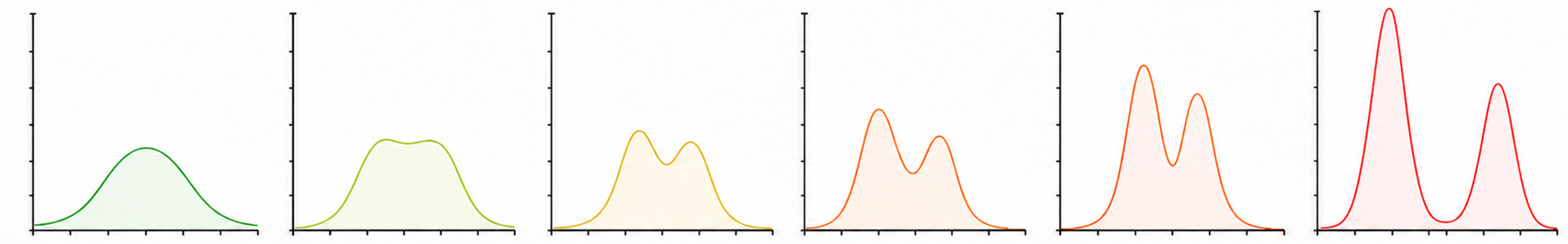


Noisy paths

Target: $X_0 \sim \mu$

Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

This induces a path:

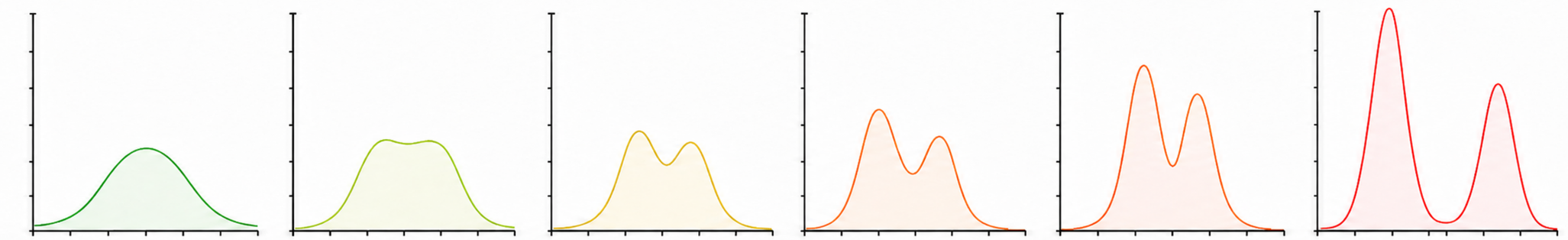


$\mu_\infty = N(0,I)$ $\mu_t = \text{law of } X_t$ $\mu_0 = \mu$

Noisy paths

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

This induces a path:



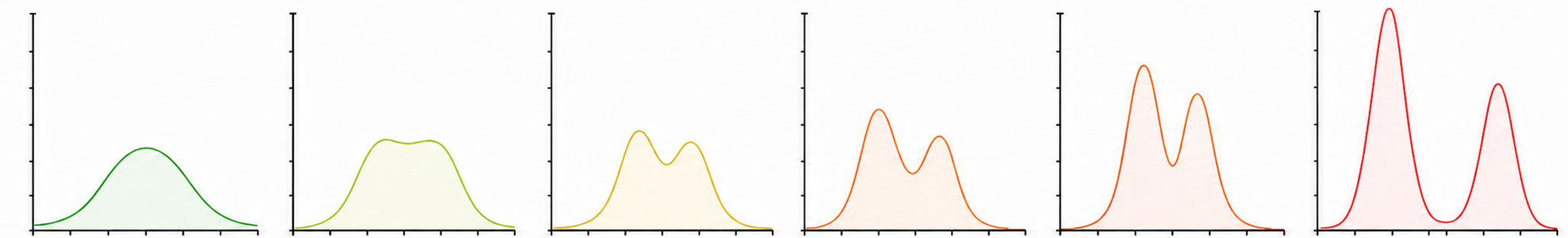
$\mu_\infty = N(0,I)$ $\mu_t = \text{law of } X_t$ $\mu_0 = \mu$

$$Y_0 \sim N(0,I), \quad \mathrm{d}Y_t = \left(-Y_t - 2\nabla \log \mu_t(Y_t)\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

Noisy paths

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

This induces a path:



$\mu_\infty = N(0,I)$ $\mu_t = \text{law of } X_t$ $\mu_0 = \mu$

$$Y_0 \sim N(0,I), \quad \mathrm{d}Y_t = \left(-Y_t - 2\nabla \log \mu_t(Y_t)\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

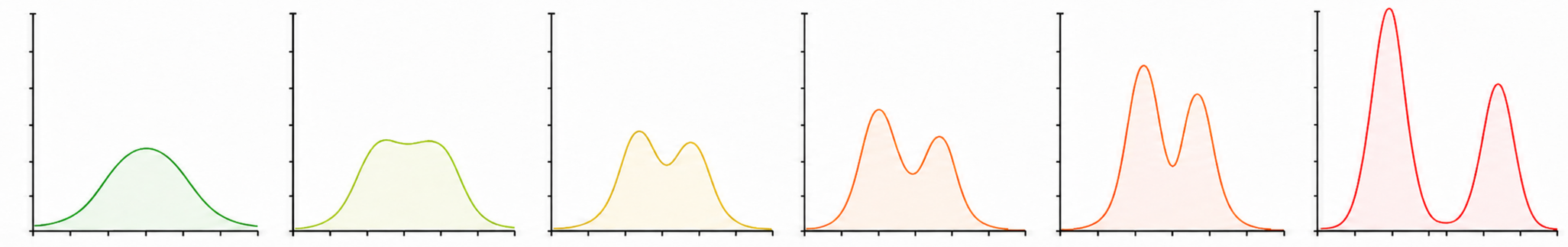
Score-based generative model

Noisy paths

Target: $X_0 \sim \mu$

Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

This induces a path:



$\mu_\infty = N(0,I)$ $\qquad$ $\mu_t = \text{law of } X_t$ $\qquad$ $\mu_0 = \mu$

$$Y_0 \sim N(0,I), \quad \mathrm{d}Y_t = \left(-Y_t - 2\nabla \log \mu_t(Y_t)\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

Score function

Score-based generative model

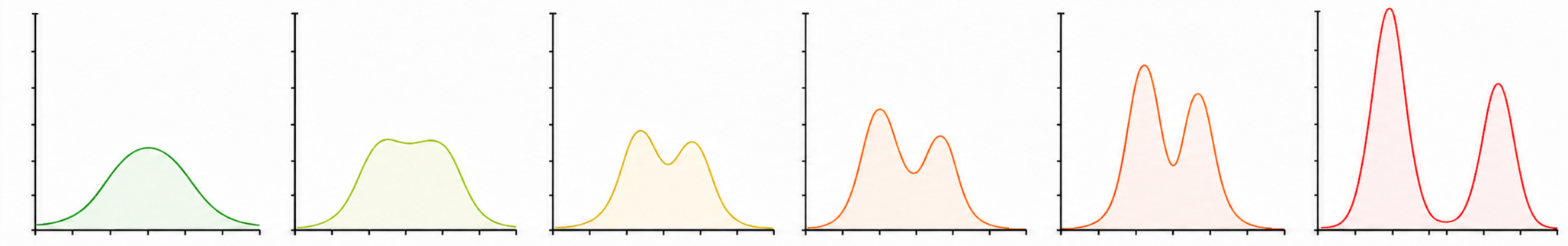
Noisy paths

Target: $X_0 \sim \mu$

Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

The density of μ_t is not easy to compute.

This induces a path:



$\mu_\infty = N(0,I)$ $\qquad$ $\mu_t = \text{law of } X_t$ $\qquad$ $\mu_0 = \mu$

$$Y_0 \sim N(0,I), \quad \mathrm{d}Y_t = \left(-Y_t - 2\nabla \log \mu_t(Y_t)\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

Score function

Score-based generative model

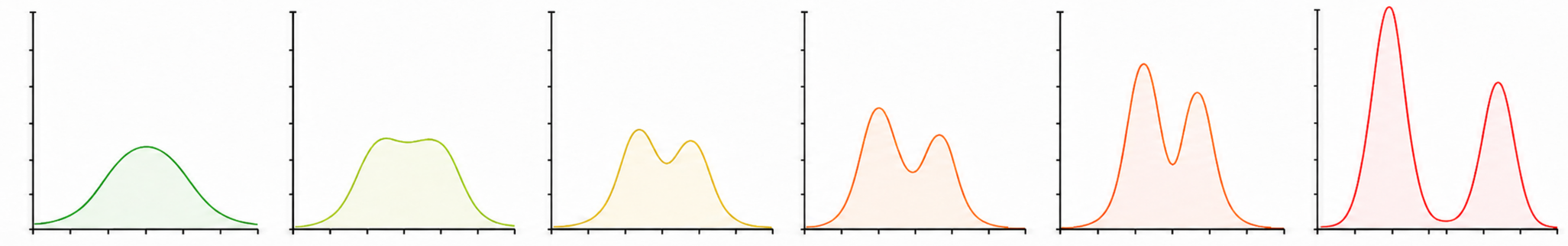
Noisy paths

Target: $X_0 \sim \mu$

Add noise: $X_t = e^{-t} X_0 + \sqrt{1 - e^{-2t}}\, G, \quad G \sim N(0, I)$

The density of μ_t is not easy to compute.

This induces a path:



$\mu_\infty = N(0, I)$ $\qquad$ $\mu_t = \text{law of } X_t$ $\qquad$ $\mu_0 = \mu$

$$Y_0 \sim N(0, I), \quad \mathrm{d}Y_t = \left(-Y_t - 2\nabla \log \mu_t(Y_t)\right) \mathrm{d}t + \sqrt{2}\, \mathrm{d}B_t$$

Score function

Score-based generative model →

If we have access to the score function $\nabla \log \mu_t$, we can simulate the stochastic differential equation (with suitable discretization) in $\mathrm{poly}(d)$ time. ([CCL+23, LLT23…])

Why are noisy paths interesting?

Why are noisy paths interesting?

Target: $X_0 \sim \mu$

Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1 - e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth.

Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

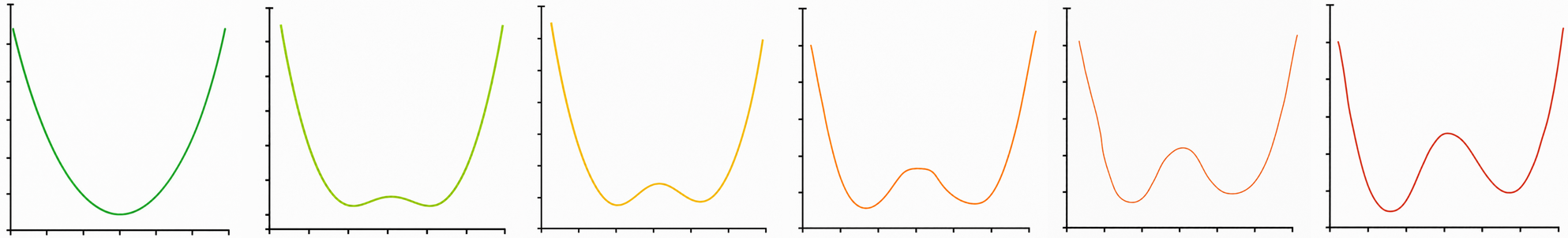
The distributions μ_t are L-log-smooth. $\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([**H**LSZ26])

Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth. $\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([**H**LSZ26])



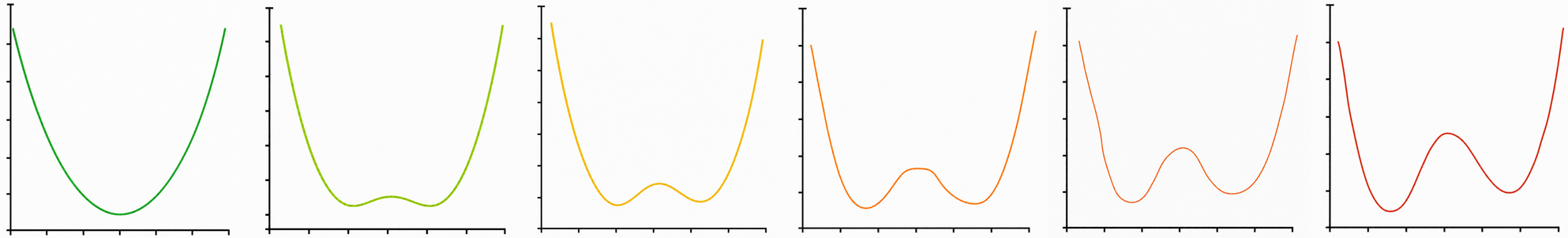
Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth. $\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([**H**LSZ26])

$$-\nabla^2 \log \mu_t(x) \succeq -LI_d$$



$-\log \mu_\infty$ $-\log \mu_t$ $-\log \mu_0$

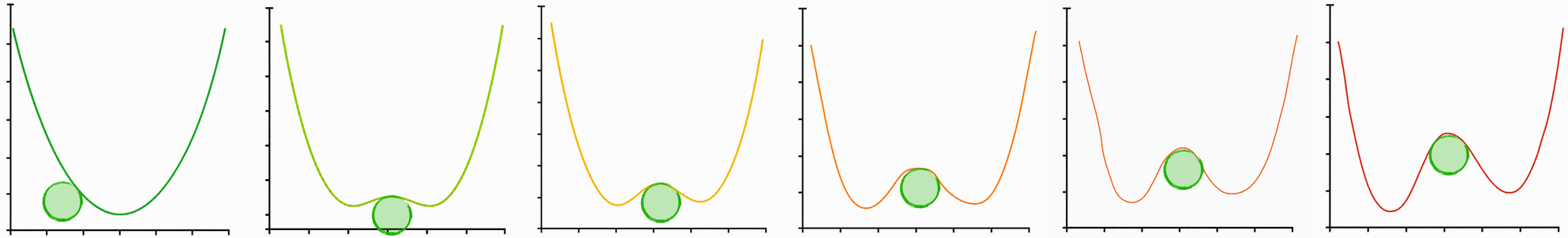
Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth. $\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([**H**LSZ26])

$$-\nabla^2 \log \mu_t(x) \succeq -LI_d$$



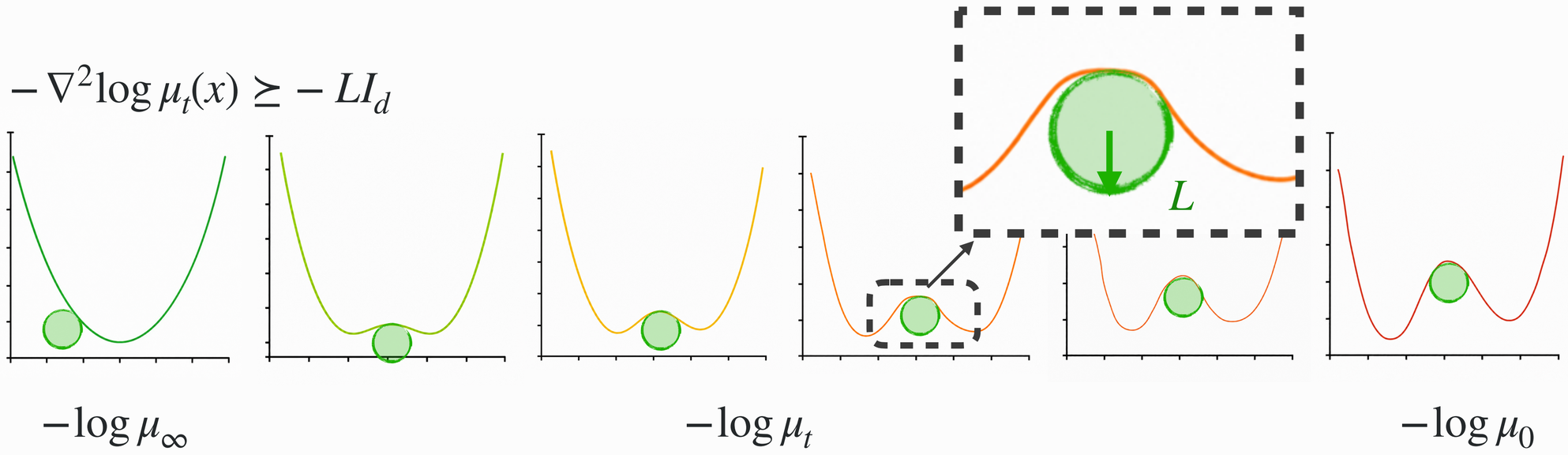
$-\log \mu_\infty$ $-\log \mu_t$ $-\log \mu_0$

Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth. $\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([HLSZ26])



Why are noisy paths interesting?

Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

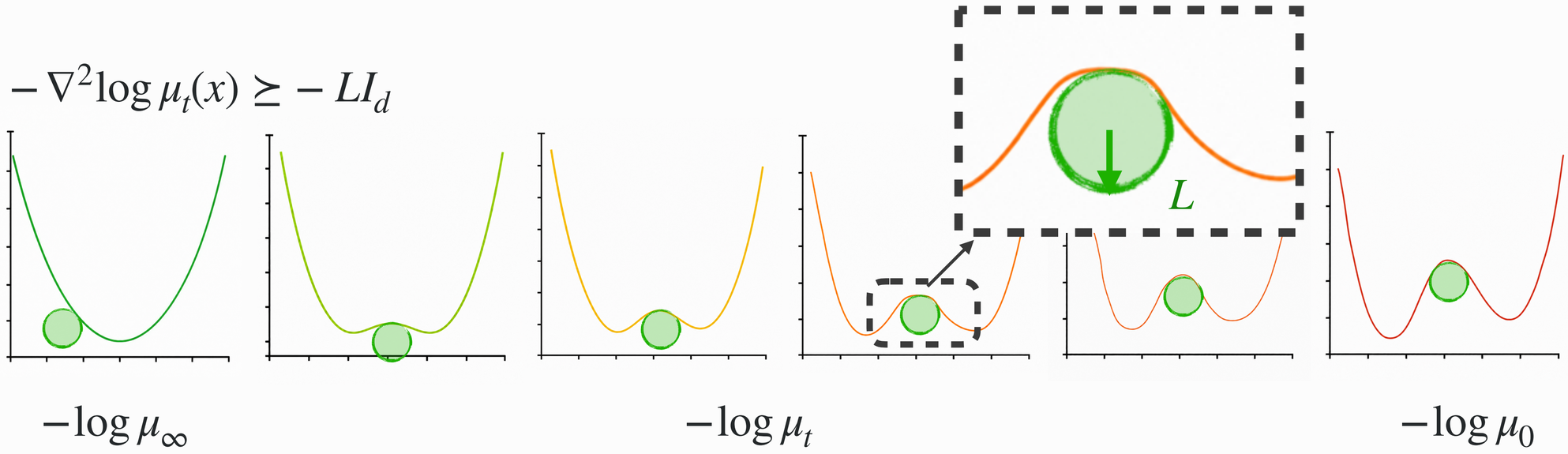
Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth.

$\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([**H**LSZ26])

Only the target distribution μ is L-log-smooth.

$-\nabla^2 \log \mu_t(x) \succeq -LI_d$



Why are noisy paths interesting?

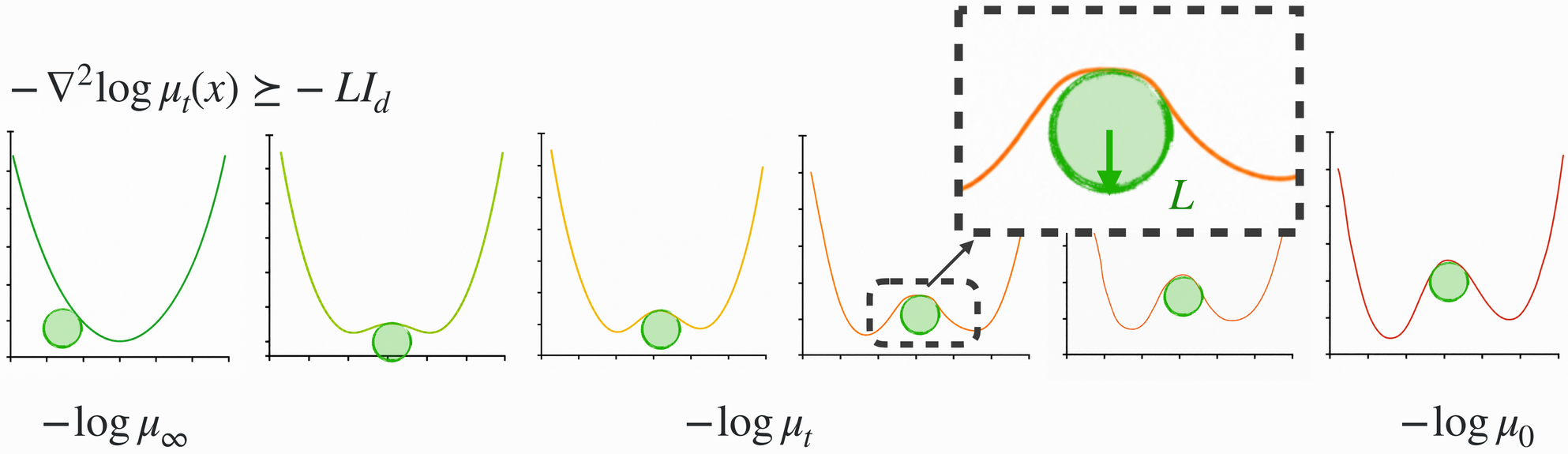
Target: $X_0 \sim \mu$ Add noise: $X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G, \quad G \sim N(0,I)$

Can we exploit the path geometry without assuming access to the score function?

The distributions μ_t are L-log-smooth. $\mathrm{poly}(d)^{\mathcal{O}(L)}$ ([**H**LSZ26])

Only the target distribution μ is L-log-smooth. $e^{\Omega(d)}$ ([**H**Z25])

$-\nabla^2 \log \mu_t(x) \succeq -LI_d$



Annealing paths

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$

Example: $\mu_s(x) \propto e^{-sH(x)}$

Annealing paths

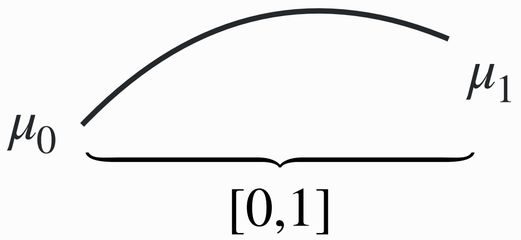
A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $s \in [0,1]$, run Langevin dynamics targeting the current distribution μ_s

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

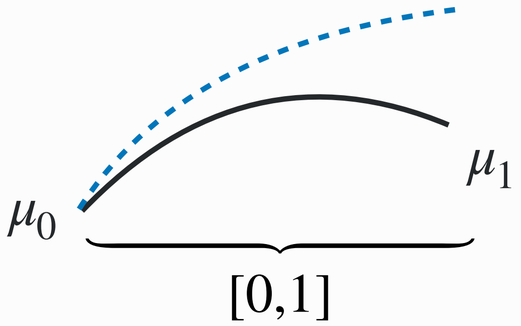
At time $s \in [0,1]$, run Langevin dynamics targeting the current distribution μ_s



Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

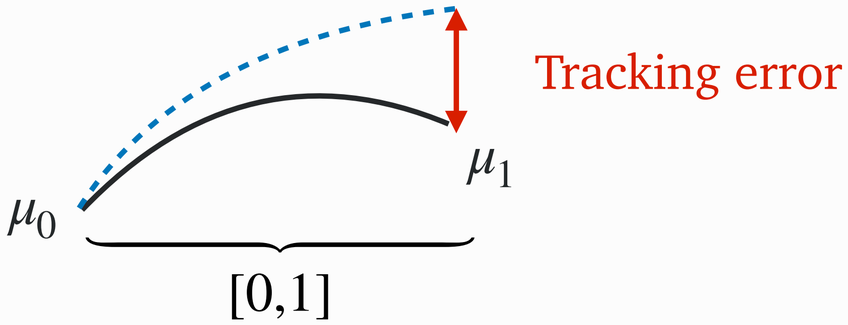
At time $s \in [0,1]$, run Langevin dynamics targeting the current distribution μ_s



Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $s \in [0,1]$, run Langevin dynamics targeting the current distribution μ_s

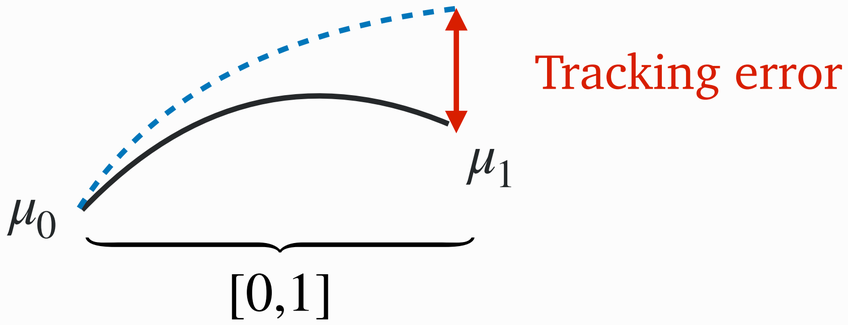


Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $s \in [0,1]$, run Langevin dynamics targeting the current distribution μ_s

How slowly should the distribution path evolve so that Langevin dynamics can track it?

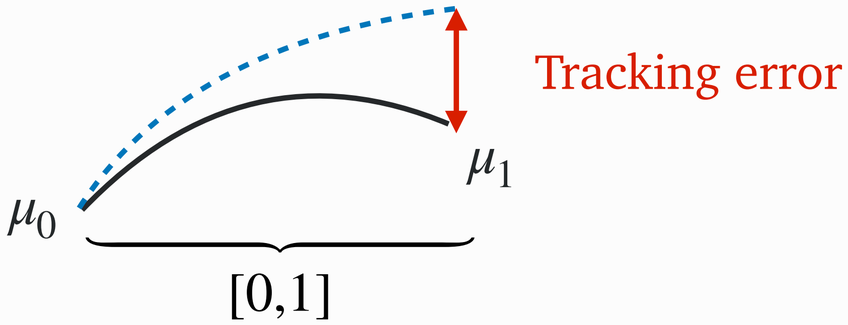


Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?

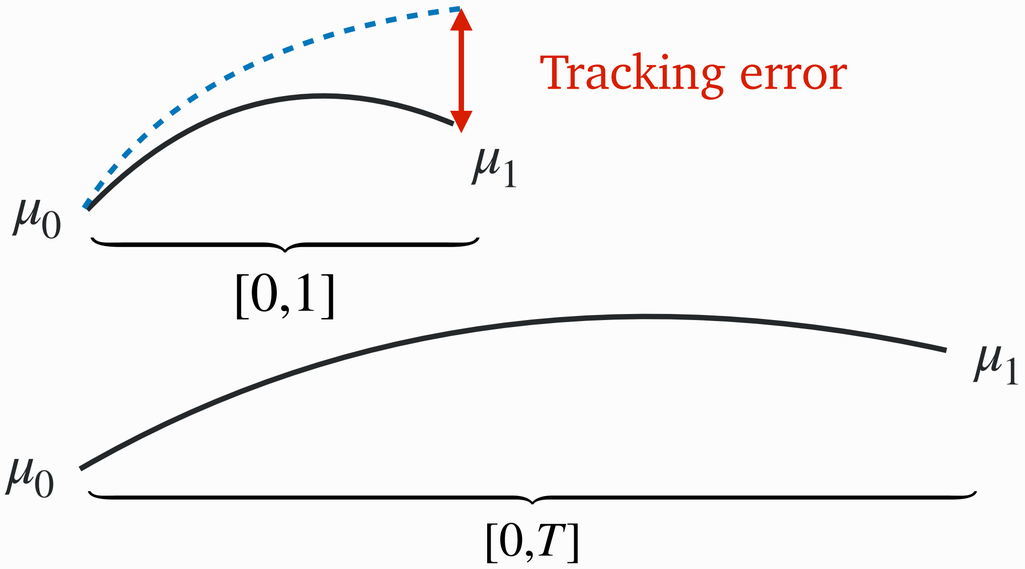


Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?

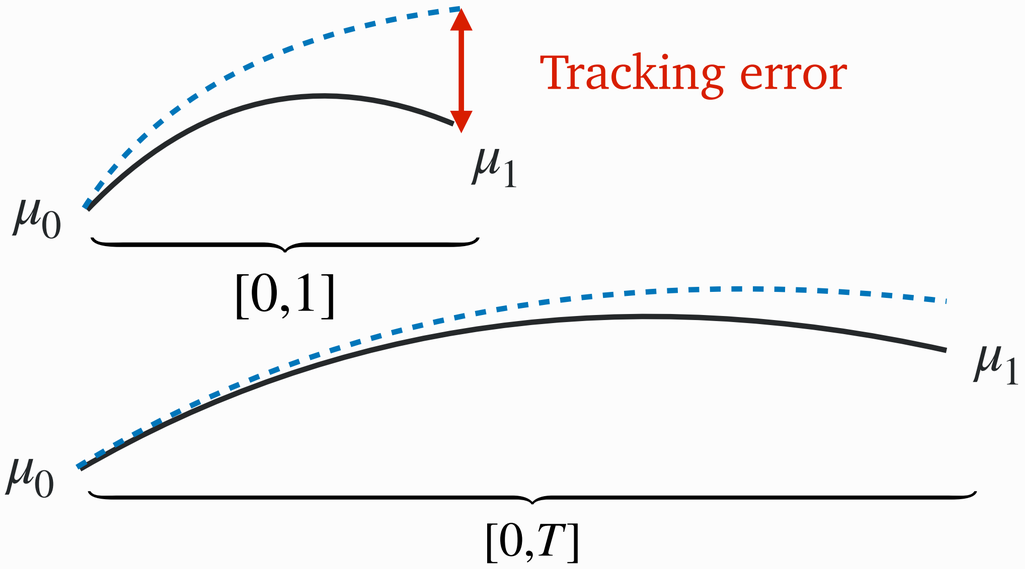


Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?

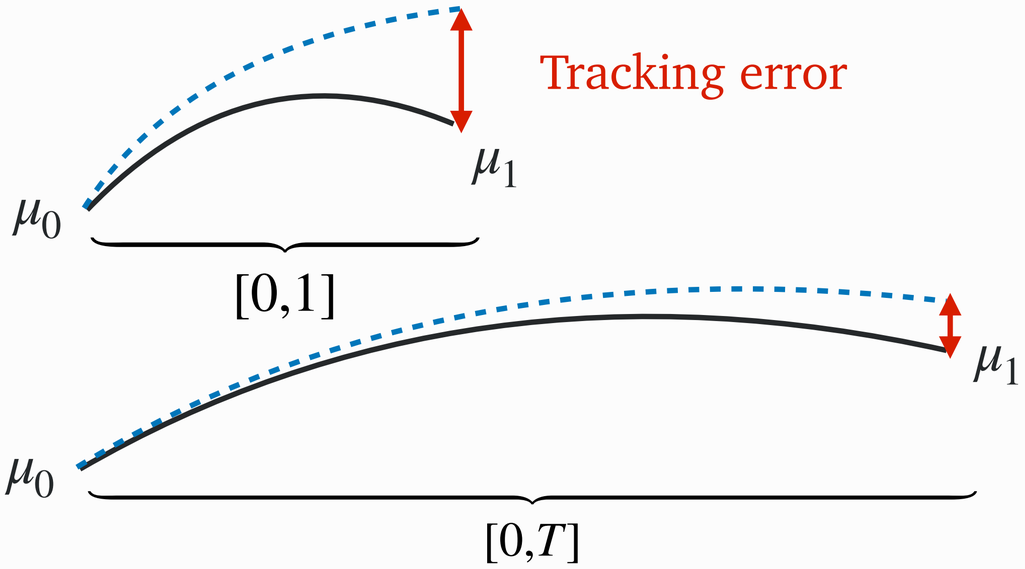


Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?

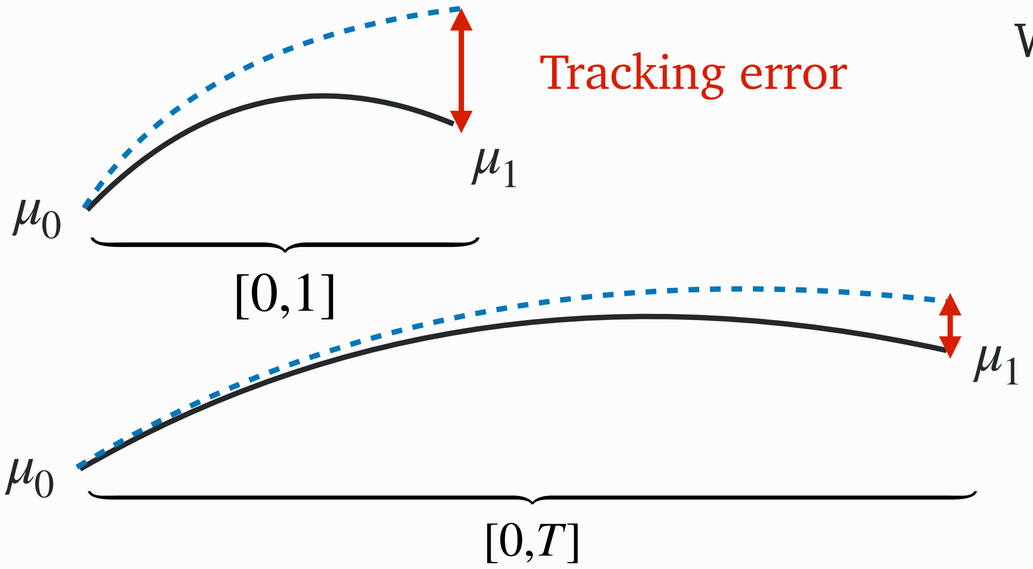


Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



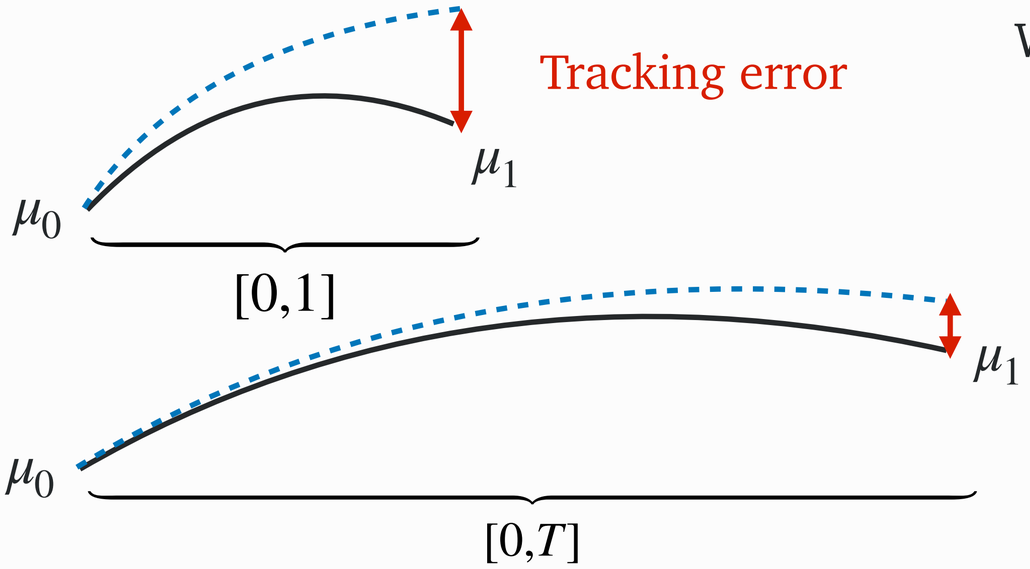
Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

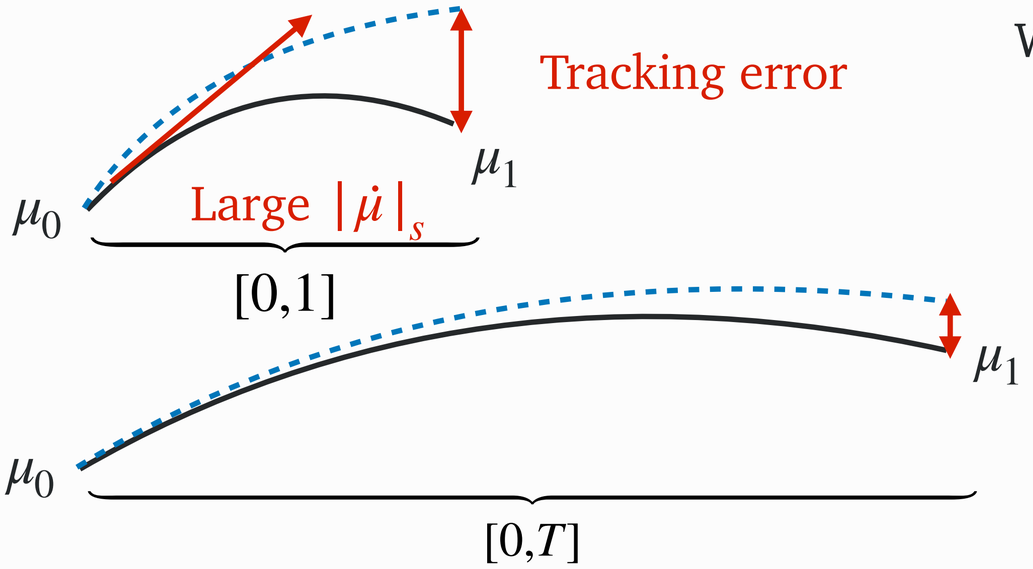
$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W_2(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

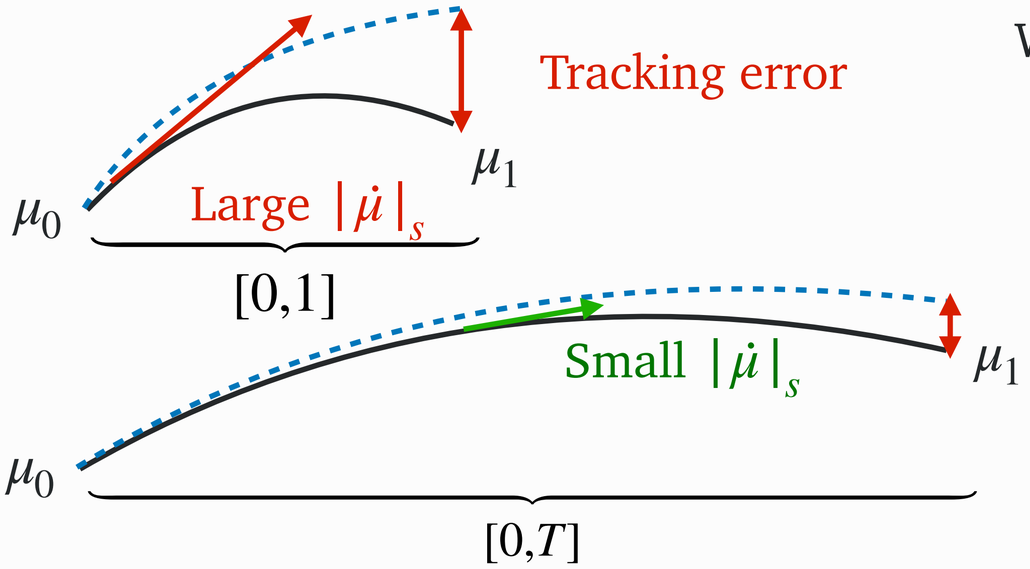
$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W_2(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

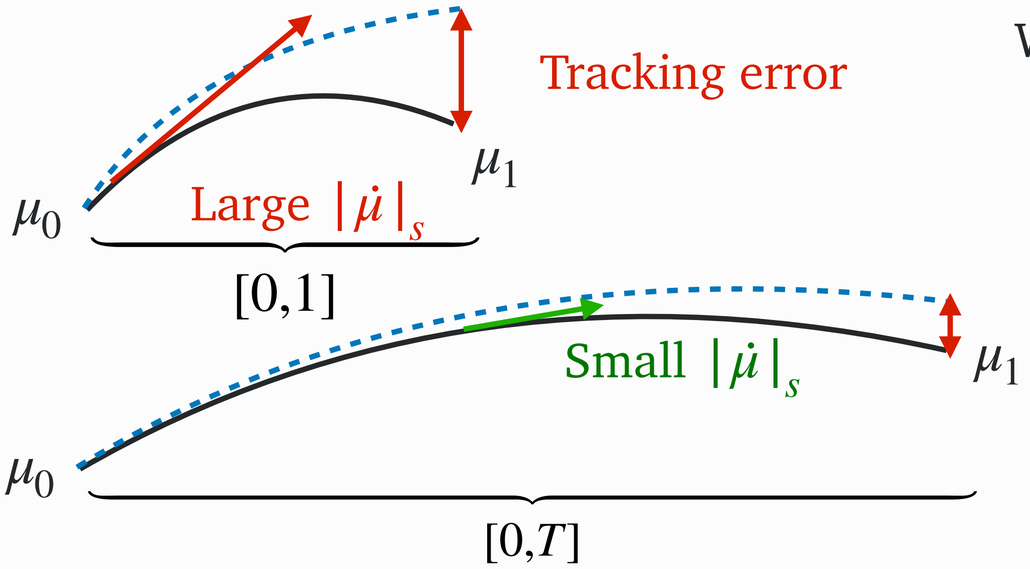
$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W_2(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W_2(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

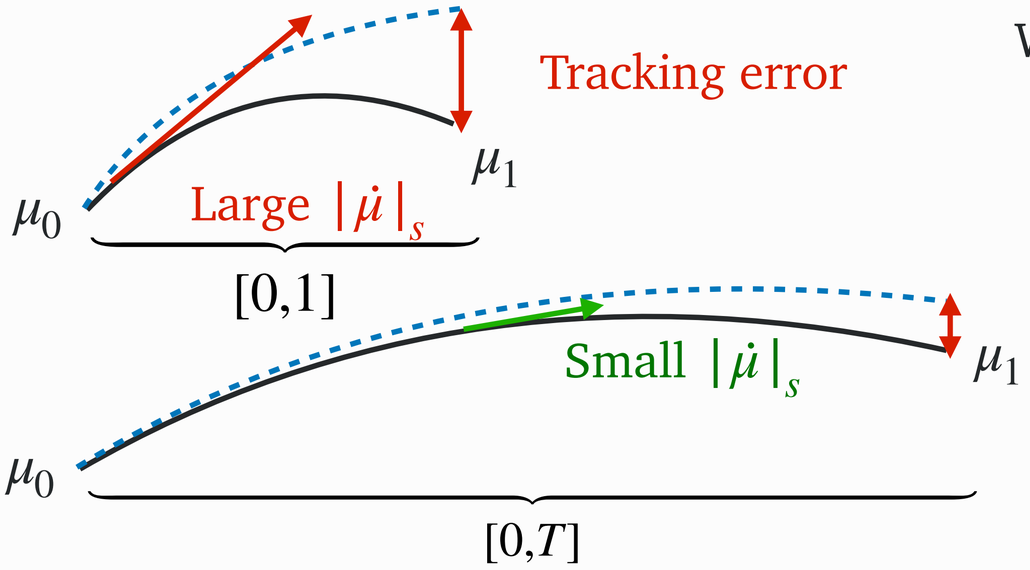
Action: $\mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W_2(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

Action: $\mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds$

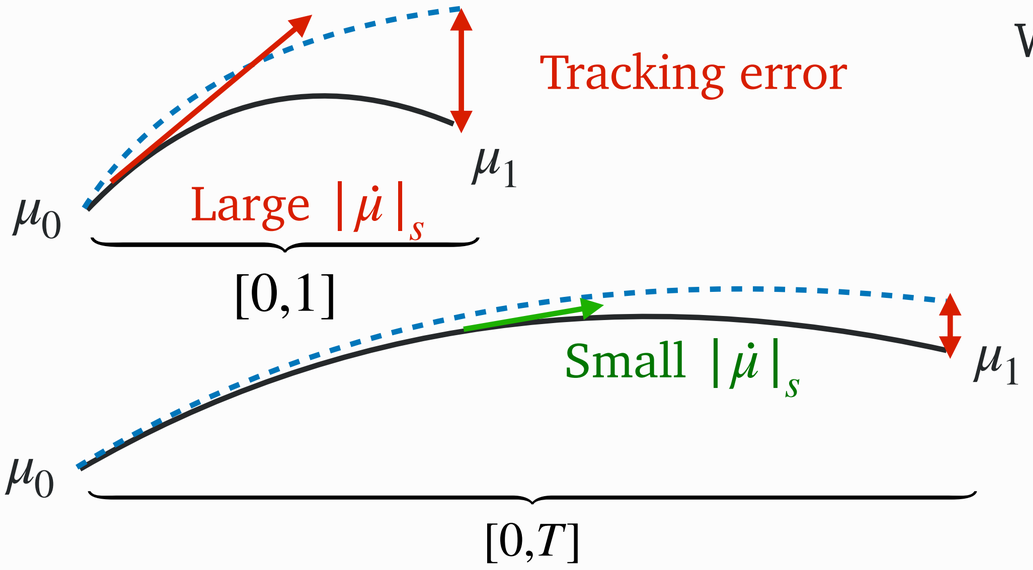
$$\Rightarrow \quad D_{\mathrm{KL}}\left(\mu \| \tilde{\mu}_T\right) \le \frac{\mathscr{A}(\{\mu_s\})}{4T} \text{ ([GTC25])}$$

Annealing paths

A distribution path: $\mu_0 \longrightarrow \mu_s \longrightarrow \mu_{(s+\Delta s)} \longrightarrow \mu_1 = \mu$ Example: $\mu_s(x) \propto e^{-sH(x)}$

At time $t \in [0,T]$, run Langevin dynamics targeting the current distribution $\mu_{t/T}$

How slowly should the distribution path evolve so that Langevin dynamics can track it?



Wasserstein distance: $W_2^2(\pi,\nu) = \min_{(X,Y)} \mathbb{E}\|X-Y\|^2, \quad X \sim \pi,\ Y \sim \nu$

$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W_2(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

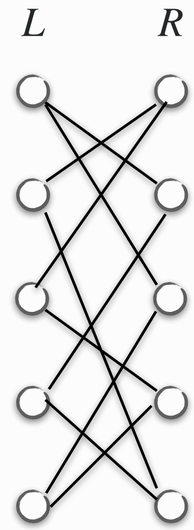
Action: $\mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds$

$$\Rightarrow \quad D_{\mathrm{KL}}\left(\mu \| \tilde{\mu}_T\right) \le \frac{\mathscr{A}(\{\mu_s\})}{4T} \text{ ([GTC25])}$$

What is the analogue in discrete state spaces?

Action: what is the discrete space picture?

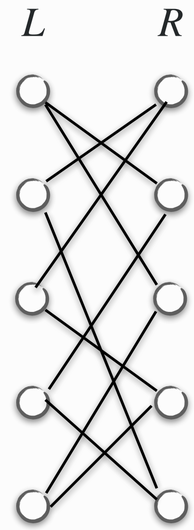
Action: what is the discrete space picture?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|},\ \lambda > 0,\ I$ independent set

Action: what is the discrete space picture?

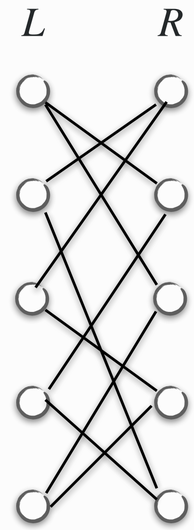


$G = (V, E)$

$\mu(I) \propto \lambda^{|I|},\ \lambda > 0,\ I$ independent set

A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

Action: what is the discrete space picture?



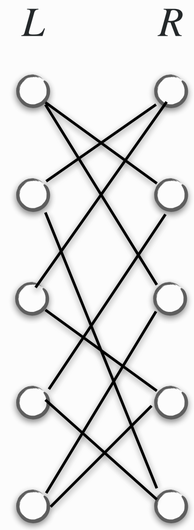
$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

Action: what is the discrete space picture?

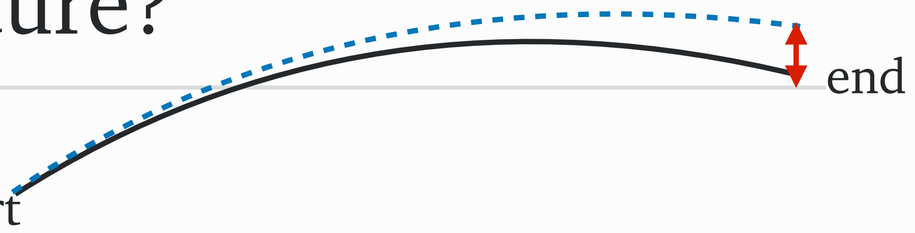


$G = (V, E)$

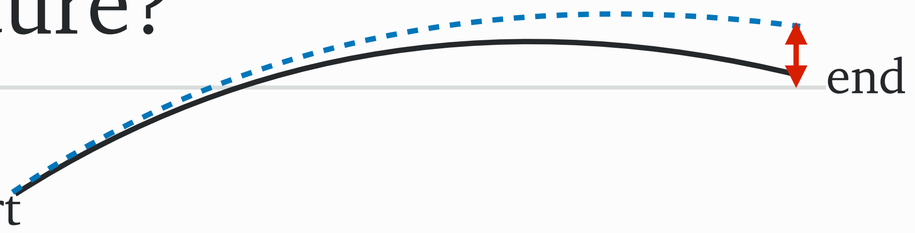
$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

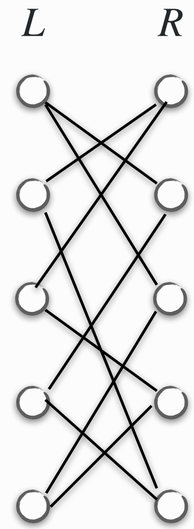
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$



Action: what is the discrete space picture?





$G = (V, E)$

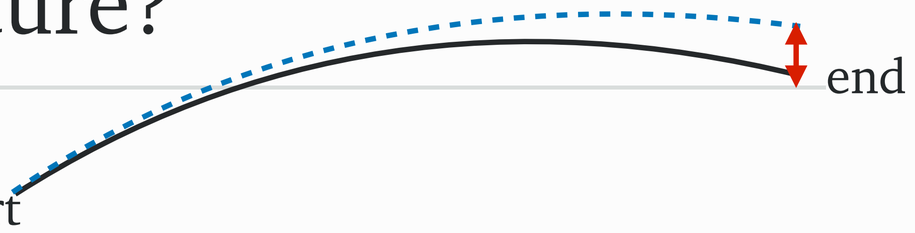
$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

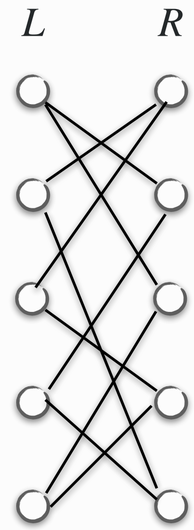
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

π $I \backslash \{w\}$ 0.11 I 0.08 $I \cup \{u\}$ 0.12 J 0.17 • • •

Action: what is the discrete space picture?





$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

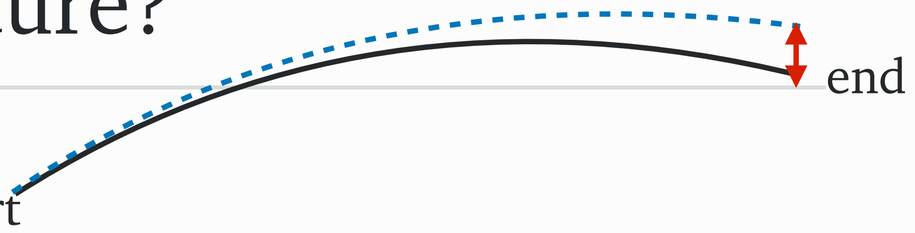
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

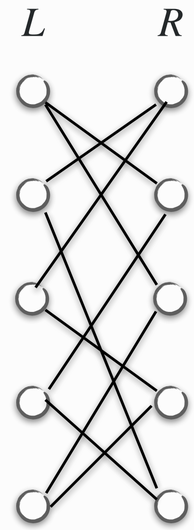
At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

π $I\backslash\{w\}$ 0.11 I 0.08 $I \cup \{u\}$ 0.12 J 0.17 • • •

ν $I\backslash\{w\}$ 0.08 I 0.11 $I \cup \{u\}$ 0.12 J 0.17 • • •

Action: what is the discrete space picture?





$G = (V, E)$

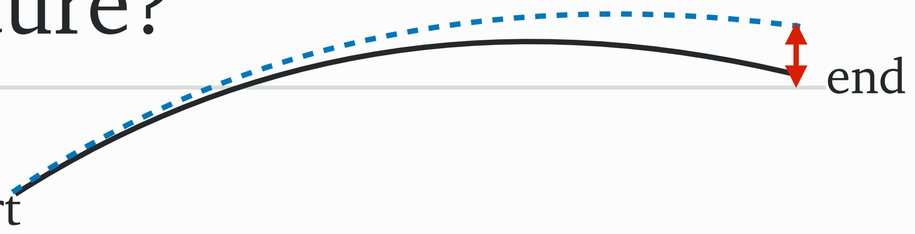
$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

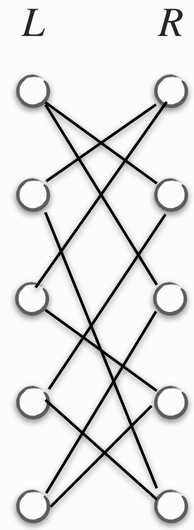
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

π	$I\backslash\{w\}$ 0.11	I 0.08	$I \cup \{u\}$ 0.12	J 0.17	• • •
ν	$I\backslash\{w\}$ 0.08	I 0.11	$I \cup \{u\}$ 0.12	J 0.17	• • •

Action: what is the discrete space picture?



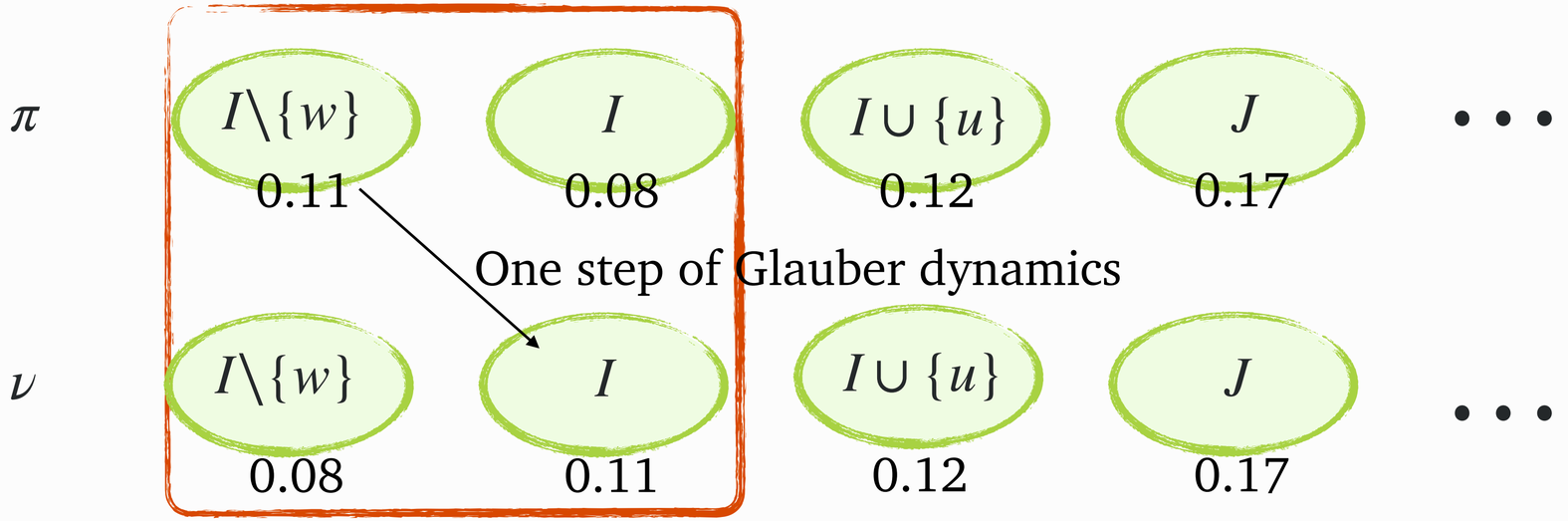


$G = (V, E)$

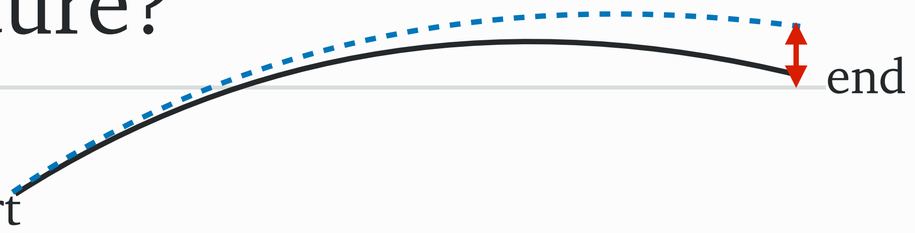
$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

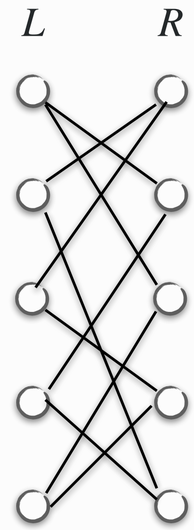
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$



Action: what is the discrete space picture?



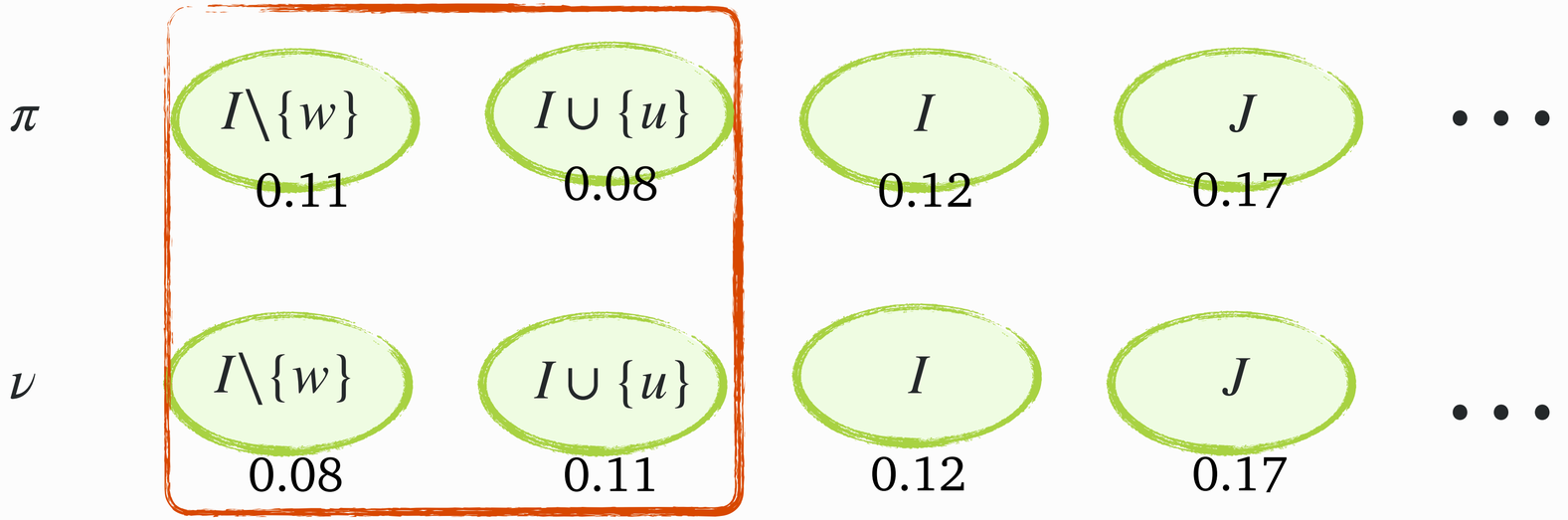


$G = (V, E)$

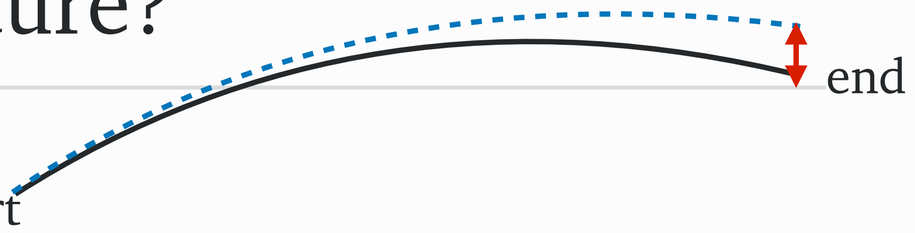
$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

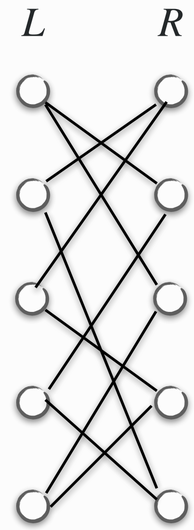
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$



Action: what is the discrete space picture?



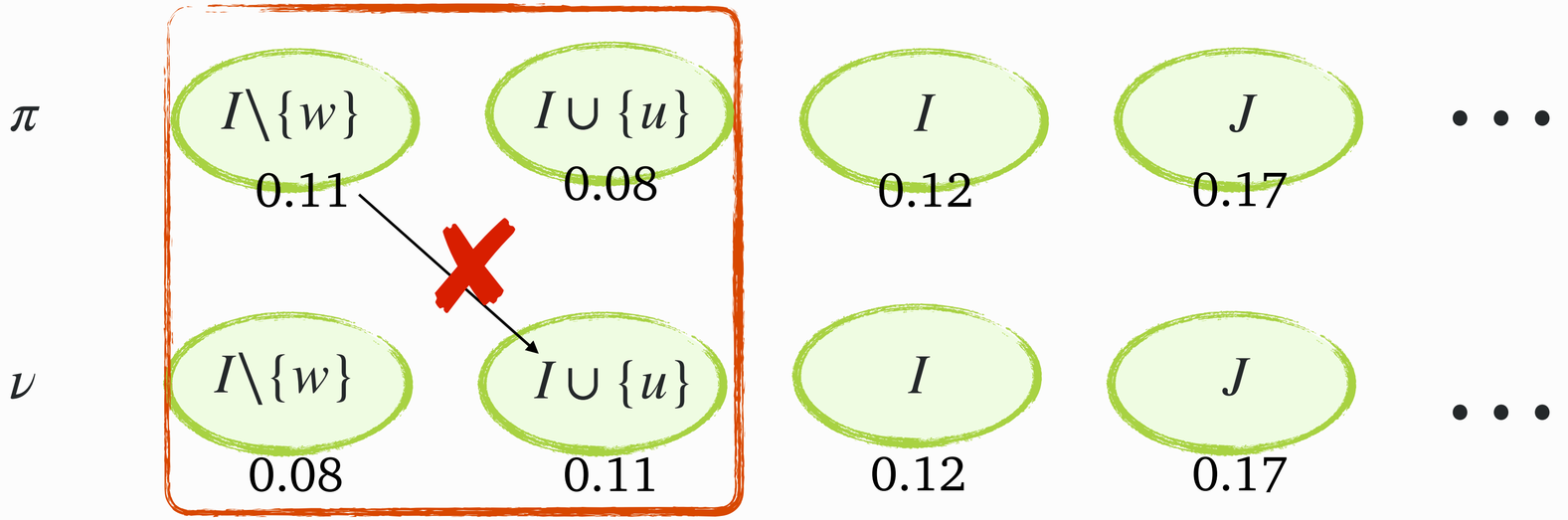


$G = (V, E)$

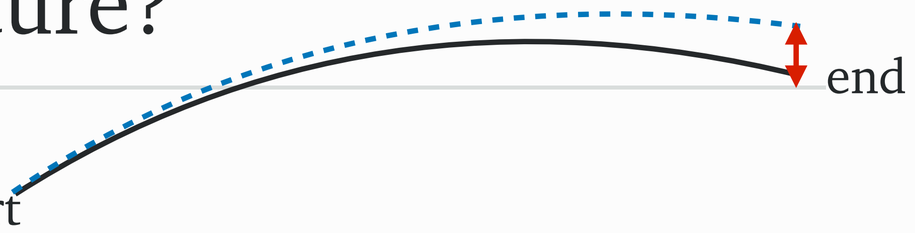
$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

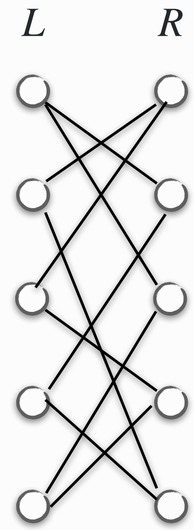
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$



Action: what is the discrete space picture?



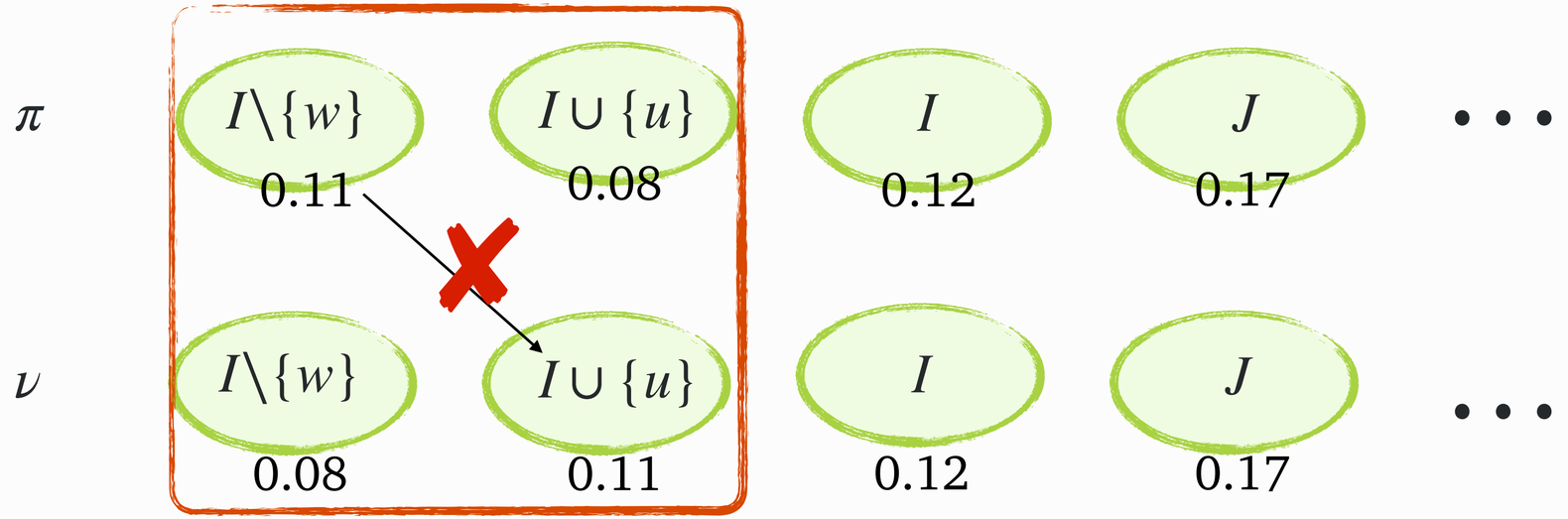


$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

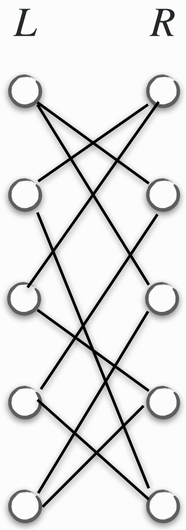
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$



$W(\pi, \nu)$ = distance (under a Markov-chain-induced geometry) between π and ν

Action: what is the discrete space picture?

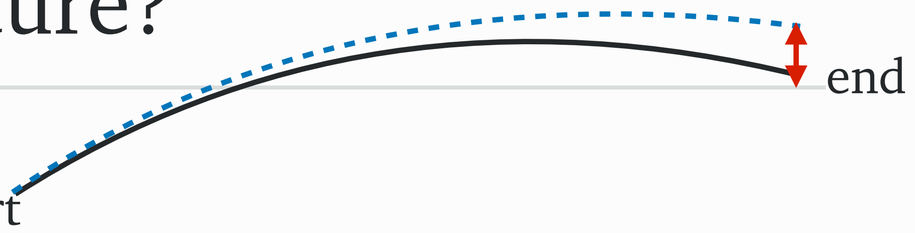


$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

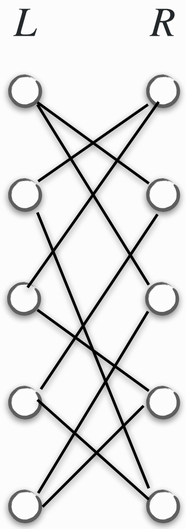
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$



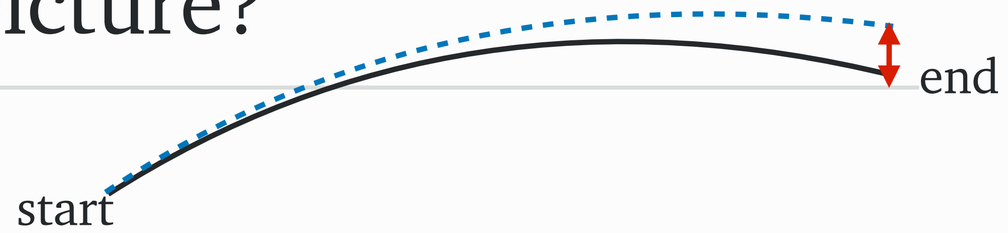
$W(\pi, \nu)$ = distance (under a Markov-chain-induced geometry) between π and ν

Action: what is the discrete space picture?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

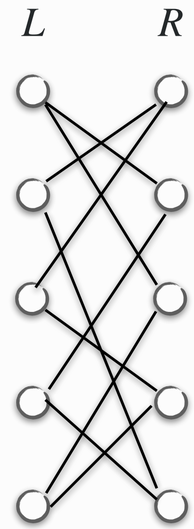


A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

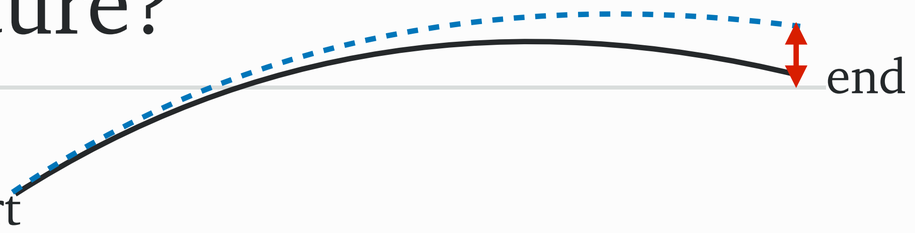
$W(\pi, \nu) =$ distance (under a Markov-chain-induced geometry) between π and ν

Action: what is the discrete space picture?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set



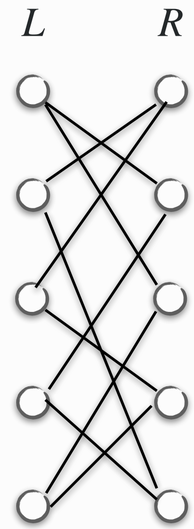
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

$W(\pi, \nu)$ = distance (under a Markov-chain-induced geometry) between π and ν

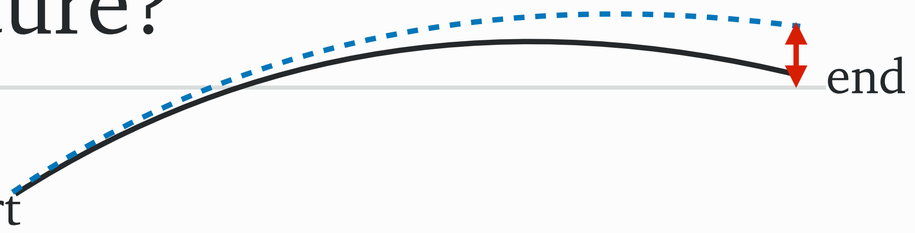
$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|}$$

Action: what is the discrete space picture?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set



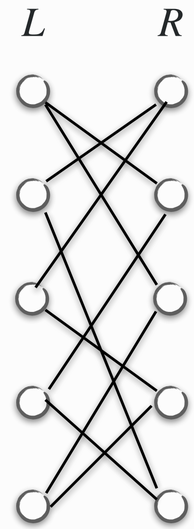
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

$W(\pi, \nu) =$ distance (under a Markov-chain-induced geometry) between π and ν

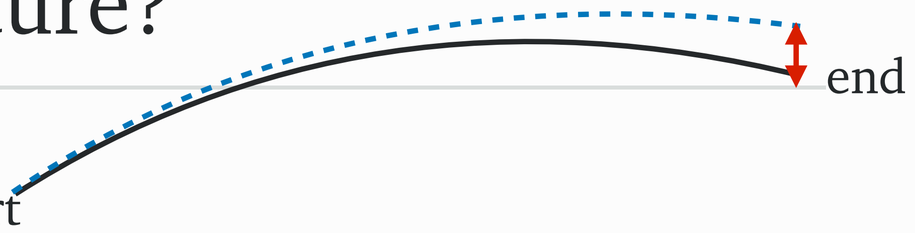
$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|} \quad \text{Action: } \mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds$$

Action: what is the discrete space picture?



$G = (V, E)$

$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set



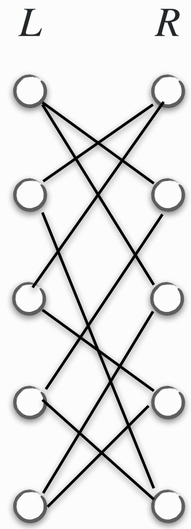
A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

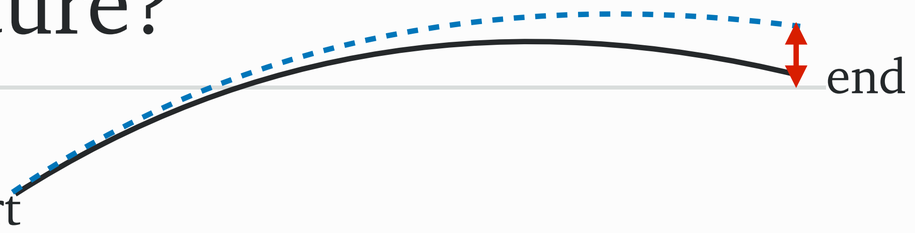
$W(\pi, \nu) =$ distance (under a Markov-chain-induced geometry) between π and ν

$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|} \quad \text{Action: } \mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds \Rightarrow \; D_{\mathrm{KL}}\left(\mu \| \tilde{\mu}_T\right) \leq \frac{\mathscr{A}(\{\mu_s\})}{4T}$$ ([HJWZ26])

Action: what is the discrete space picture?



$G = (V, E)$



$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

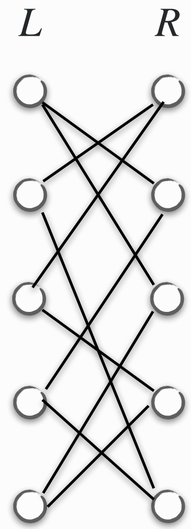
At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

$W(\pi, \nu)$ = distance (under a Markov-chain-induced geometry) between π and ν

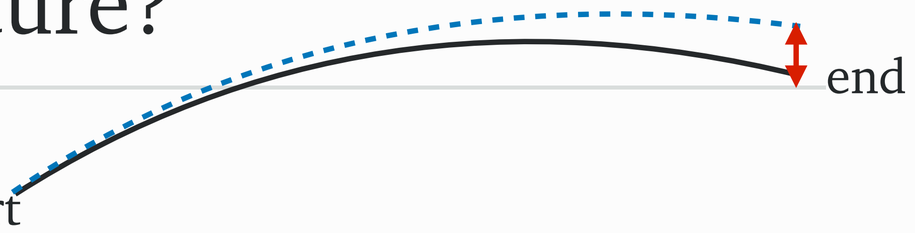
$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|} \quad \text{Action: } \mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds \Rightarrow \; D_{\mathrm{KL}}\left(\mu \| \tilde{\mu}_T\right) \le \frac{\mathscr{A}(\{\mu_s\})}{4T}$$ ([HJWZ26])



Action: what is the discrete space picture?



$G = (V, E)$



$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

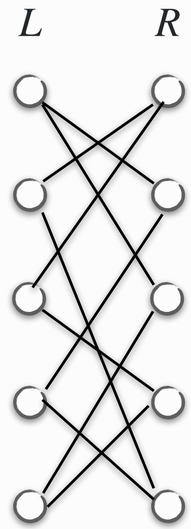
$W(\pi, \nu)$ = distance (under a Markov-chain-induced geometry) between π and ν

$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|} \quad \text{Action: } \mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds \Rightarrow D_{\mathrm{KL}}\left(\mu \| \tilde{\mu}_T\right) \le \frac{\mathscr{A}(\{\mu_s\})}{4T}$$ ([HJWZ26])

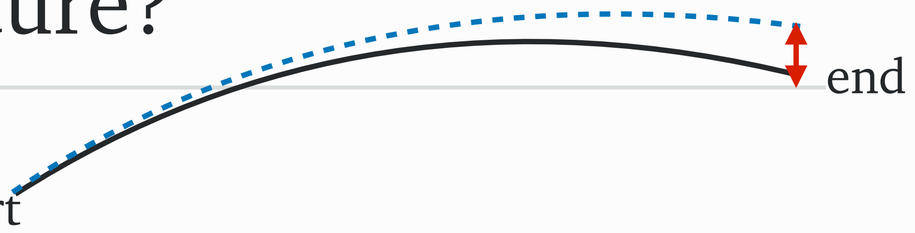
$s_0 \approx \infty$ ——→ smaller s ——→ ⋯ ——→ $s = 1$



Action: what is the discrete space picture?



$G = (V, E)$



$\mu(I) \propto \lambda^{|I|}$, $\lambda > 0$, I independent set

A distribution path: $\mu_0 \longrightarrow \mu_s(I) \propto \lambda^{s|I|} \longrightarrow \mu_1 = \mu$

At time $t \in [0,T]$, run Glauber dynamics with stationary distribution $\mu_{t/T}$

$W(\pi, \nu)$ = distance (under a Markov-chain-induced geometry) between π and ν

$$|\dot{\mu}|_s = \lim_{\Delta s \to 0} \frac{W(\mu_{s+\Delta s}, \mu_s)}{|\Delta s|} \quad \text{Action: } \mathscr{A}(\{\mu_s\}) = \int_0^1 |\dot{\mu}|_s^2 \, ds \Rightarrow \; D_{\mathrm{KL}}\left(\mu \| \tilde{\mu}_T\right) \le \frac{\mathscr{A}(\{\mu_s\})}{4T} \text{ ([HJWZ26])}$$

$s_0 \approx \infty$ ——————→ smaller s ——————→ ⋯ ——————→ $s = 1$



Bounded degree + α-expansion + symmetric $\Rightarrow \mathscr{A} \lesssim \mathrm{poly}(n)$ ([HJWZ26])

Conclusion

Conclusion

1. Why do we need pathwise geometry?

Conclusion

1. Why do we need pathwise geometry?

   To explore distributions with potentially poor geometry.

Conclusion

1. Why do we need pathwise geometry?

   To explore distributions with potentially poor geometry.

2. What do we know about pathwise geometry?

Conclusion

1. Why do we need pathwise geometry?

   To explore distributions with potentially poor geometry.

2. What do we know about pathwise geometry?

	What does the path provide?	What is not guaranteed?	What makes it good?
Noisy paths	$X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G$	Density of μ_t	Log-smoothness
Annealing paths	Density of μ_t	How to move along the path	Action

Conclusion

1. Why do we need pathwise geometry?

   To explore distributions with potentially poor geometry.

2. What do we know about pathwise geometry?

	What does the path provide?	What is not guaranteed?	What makes it good?
Noisy paths	$X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G$	Density of μ_t	Log-smoothness
Annealing paths	Density of μ_t	How to move along the path	Action

3. Future directions

Conclusion

1. Why do we need pathwise geometry?

   To explore distributions with potentially poor geometry.

2. What do we know about pathwise geometry?

	What does the path provide?	What is not guaranteed?	What makes it good?
Noisy paths	$X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G$	Density of μ_t	Log-smoothness
Annealing paths	Density of μ_t	How to move along the path	Action

3. Future directions

   Can we design low-action paths automatically (via flow matching / Schrödinger bridges)?

Conclusion

1. Why do we need pathwise geometry?

   To explore distributions with potentially poor geometry.

2. What do we know about pathwise geometry?

	What does the path provide?	What is not guaranteed?	What makes it good?
Noisy paths	$X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}G$	Density of μ_t	Log-smoothness
Annealing paths	Density of μ_t	How to move along the path	Action

3. Future directions

   Can we design low-action paths automatically (via flow matching / Schrödinger bridges)?

   Can action characterize the difficulty of other sampling dynamics, such as tempering methods?

References

[CLV23]Chen, L., Liu, K., and Vigoda, E. Rapid Mixing of Glauber Dynamics up to Uniqueness via Contraction. SIAM Journal on Computing, 2023

[CCYZ25] Chen, X., Chen, Z., Yin, Y., and Zhang, X. Rapid Mixing at the Uniqueness Threshold. STOC 2025.

[SS12] Sly, A. and Sun, N. The Computational Hardness of Counting in Two-Spin Models on d-Regular Graphs. FOCS 2012.

[CCL+23] Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., and Zhang, A. R. Sampling Is as Easy as Learning the Score: Theory for Diffusion Models with Minimal Data Assumptions. ICLR 2023.

[LLT23] Lee, H., Lu, J., and Tan, Y. Convergence of Score-Based Generative Modeling for General Data Distributions. ALT 2023.

[HZ25]**He, Y.,** and Zhang, C. On the query complexity of sampling from non-log-concave distributions. COLT 2025.

[HLSZ26] **He, Y.,** Lei, Z., Shao, J., and Zhang, C. Improved sampling algorithms and functional inequalities for non-log-concave distributions. Manuscript, 2026.

[GTC25] Guo, H., Tao, M., and Chen, Y. Provable Benefit of Annealed Langevin Monte Carlo for Non-log-concave Sampling. ICLR 2025.

[HJWZ26] **He, Y.**, Jiang, Y., Wang, X., and Zhang, C. Discrete Optimal Transport: Rapid Convergence of Simulated Annealing Algorithms. Manuscript, 2026.