CPSC 661: Sampling Algorithms in ML

Andre Wibisono

April 12, 2021

Yale University

Last time

- Wasserstein W_2 metric
- Otto calculus
- Langevin dynamics as gradient flow of relative entropy

Today:

- Convexity of relative entropy
- Convergence rate of Langevin dynamics

References

- Villani, *Topics in Optimal Transportation*, Springer, 2003, §9
- Vempala & Wibisono, *Rapid Convergence of the Unadjusted Langevin Algorithm: Isoperimetry Suffices*, NeurIPS 2019

Recap 1: Optimization on Manifold

Strong convexity and gradient domination

Recall we say a function $f\colon \mathcal{X} \to \mathbb{R}$ on a manifold $\mathcal{X}$ is

1. **geodesically α-strongly convex** if

$$\operatorname{Hess} f(x) \succeq \alpha I$$

2. **α-gradient dominated** if

$$\|\operatorname{grad} f(x)\|_x^2 \geq 2\alpha\left(f(x) - \min f\right)$$

3. has **α-quadratic growth** if

$$f(x) - \min f \geq \frac{\alpha}{2} d(x, x^*)^2$$

where $x^* = \arg\min_{x \in \mathcal{X}} f(x)$

Lemma: (1) $\Rightarrow$ (2) $\Rightarrow$ (3)

Convergence rate of gradient flow

Gradient flow:

$$\dot{X}_t = -\mathrm{grad}\, f(X_t)$$

1. If f is α-strongly convex, then for coevolving solutions X_t, Y_t:

$$d(X_t, Y_t)^2 \leq e^{-2\alpha t} d(X_0, Y_0)^2$$

2. If f is α-gradient dominated, then

$$f(X_t) - \min f \leq e^{-2\alpha t}(f(X_0) - \min f)$$

Recap 2: Sampling as Optimization

Relative entropy

Space of probability distributions:

$$\mathcal{P}(\mathbb{R}^n) = \left\{ \rho \colon \mathbb{R}^n \to \mathbb{R} \;\middle|\; \int_{\mathbb{R}^n} \rho(x)\, dx = 1, \int_{\mathbb{R}^n} \|x\|^2 \rho(x) dx < \infty \right\}$$

Wasserstein metric:

$$W_2(\rho, \nu)^2 = \inf_{\pi \in \Pi(\rho,\nu)} \mathbb{E}_\pi[\|X - Y\|^2]$$

Objective function is **relative entropy:**

$$H_\nu(\rho) = \mathbb{E}_\rho \left[\log \frac{\rho}{\nu} \right] = \int_{\mathbb{R}^n} \rho(x) \log \frac{\rho(x)}{\nu(x)}\, dx$$

Minimizer is the **target distribution:**

$$\nu = \arg \min_{\rho \in \mathcal{P}(\mathbb{R}^n)} H_\nu(\rho)$$

Langevin dynamics

Gradient:

$$\operatorname{grad} H_\nu(\rho) = -\nabla \cdot \left(\rho_t \nabla \log \frac{\rho_t}{\nu}\right)$$

Gradient flow is the **Fokker-Planck equation**:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot \left(\rho_t \nabla \log \frac{\rho_t}{\nu}\right)$$

Implemented by the **Langevin dynamics**:

$$dX_t = -\nabla f(X_t)\, dt + \sqrt{2}\, dW_t$$

where $\nu \propto e^{-f}$

- Will see: 1. Convergence proof via *coupling* technique

Relative Fisher information

Squared norm of gradient is **relative Fisher information:**

$$\|\mathrm{grad}\, H_\nu(\rho)\|_\rho^2 = \mathbb{E}_\rho\left[\left\|\nabla \log \frac{\rho}{\nu}\right\|^2\right] = J_\nu(\rho)$$

Rate of decrease is given by **de Bruijn's identity:**

$$\frac{d}{dt} H_\nu(\rho_t) = -J_\nu(\rho_t)$$

- Will see: 2. Convergence proof via Log-Sobolev Inequality

What we will see

Properties of ν

We say a probability distribution $\nu \propto e^{-f}$ on $\mathbb{R}^n$ satisfies:

1. α-**strongly log-concave** if $f = -\log \nu$ is α-strongly convex

$$\nabla^2 f(x) \succeq \alpha I \qquad \forall\ x \in \mathbb{R}^n$$

2. α-**log-Sobolev inequality** if

$$J_\nu(\rho) \geq 2\alpha H_\nu(\rho) \qquad \forall\ \rho \in \mathcal{P}(\mathbb{R}^n)$$

3. α-**Talagrand inequality** if

$$H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2 \qquad \forall\ \rho \in \mathcal{P}(\mathbb{R}^n)$$

4. α-**Poincaré inequality** if

$$\mathbb{E}_\nu[\|\nabla h\|^2] \geq \alpha \operatorname{Var}_\nu(h) \qquad \forall\ h : \mathbb{R}^n \to \mathbb{R}$$

Properties of ν

Theorem: α-**SLC** $\Rightarrow$ α-**LSI** $\Rightarrow$ α-**TI** $\Rightarrow$ α-**PI**

- Characterizes nice properties of ν
- Implies isoperimetric inequalities
- Implies concentration of measure
- Implies fast convergence of Langevin dynamics
- Has geometric / optimization interpretation

[Otto & Villani, *Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality*, Journal of Functional Analysis, 2000]

Properties of H_ν in terms of ν

Relative entropy on $\mathcal{P}(\mathbb{R}^n)$

$$H_\nu(\rho) = \mathbb{E}_\rho\left[\log \frac{\rho}{\nu}\right]$$

Strong convexity:

$$\mathrm{Hess}\, H_\nu(\rho) \succeq \alpha I$$

Gradient dominated:

$$\|\mathrm{grad}\, H_\nu(\rho)\|_\rho^2 \geq 2\alpha\, H_\nu(\rho)$$

Quadratic growth:

$$H_\nu(\rho) \geq \frac{\alpha}{2} d(\rho,\nu)^2$$

Target distribution on $\mathbb{R}^n$

$$\nu \propto e^{-f}$$

$\Leftarrow$ Strong log-concavity:

$$\nabla^2 f(x) \succeq \alpha I$$

$\Leftrightarrow$ Log-Sobolev inequality:

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

$\Leftrightarrow$ Talagrand inequality:

$$H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho,\nu)^2$$

Convergence rate of Langevin dynamics

Along the Langevin dynamics for $\nu \propto e^{-f}$:

1. If ν is α-SLC, then

$$W_2(\rho_t, \nu)^2 \le e^{-2\alpha t} W_2(\rho_0, \nu)^2$$

2. If ν satisfies α-LSI, then

$$H_\nu(\rho_t) \le e^{-2\alpha t} H_\nu(\rho_0)$$

3. If ν satisfies α-PI, then

$$\chi^2_\nu(\rho_t) \le e^{-2\alpha t} \chi^2_\nu(\rho_0)$$

Langevin dynamics under isoperimetry

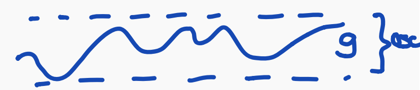
α-LSI/α-PI implies exponential convergence of Langevin dynamics

- Equivalent to isoperimetric inequalities
- Implied by α-SLC, but more general
- Stable under bounded perturbation, Lipschitz map

Lemma: (Holley-Stroock perturbation lemma)

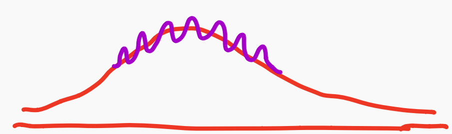
Suppose ν satisfies α-LSI (resp. α-PI). Let $\tilde{\nu} = \nu \cdot e^{-g}$ with

$$\operatorname{osc}(g) := \sup_x g(x) - \inf_x g(x) < \infty.$$



Then $\tilde{\nu}$ satisfies $\tilde{\alpha}$-LSI (resp. $\tilde{\alpha}$-PI) with

$$\tilde{\alpha} = \alpha \cdot e^{-2\operatorname{osc}(g)}$$



Strong log-concavity

Hessian of relative entropy

Relative entropy:

$$H_\nu(\rho) = \mathbb{E}_\rho \left[\log \frac{\rho}{\nu} \right]$$

Lemma: Hessian of relative entropy is a quadratic form

$$\operatorname{Hess} H_\nu(\rho) \colon \mathsf{T}_\rho \mathcal{P} \times \mathsf{T}_\rho \mathcal{P} \to \mathbb{R}$$

that sends $\phi = -\nabla \cdot (\rho \nabla u) \in \mathsf{T}_\rho \mathcal{P}$ to

$$\operatorname{Hess} H_\nu(\rho)(\phi, \phi) = \mathbb{E}_\rho \left[\|\nabla^2 u\|_{\mathsf{HS}}^2 + \langle \nabla u, (\nabla^2 f) \nabla u \rangle \right]$$

- Decomposition of relative entropy $H_\nu(\rho) = -H(\rho) + \mathbb{E}_\rho[f]$

[Villani, *Optimal Transport: Old and New*, 2008, Formula 15.7]

Convexity of relative entropy

Relative entropy:

$$H_\nu(\rho) = \mathbb{E}_\rho\left[\log \frac{\rho}{\nu}\right]$$

Theorem:

1. If ν is log-concave, then H_ν is convex
2. If ν is α-strongly log-concave, then H_ν is α-strongly convex

- In $\mathcal{P}(\mathbb{R}^n)$ with W_2 metric, H_ν inherits convexity of $f = -\log \nu$ in $\mathbb{R}^n$

Proof: Assume $\nu \propto e^{-f}$ is α-strongly log-concave, so $\nabla^2 f(x) \succeq \alpha I$

Then for any $\phi = -\nabla \cdot (\rho \nabla u) \in \mathrm{T}_\rho \mathcal{P}$

$$\begin{aligned}\mathrm{Hess}\, H_\nu(\rho)(\phi,\phi) &= \mathbb{E}_\rho\Big[\underbrace{\|\nabla^2 u\|^2_{\mathrm{HS}}}_{\geq 0} + \underbrace{\langle \nabla u, (\nabla^2 f)\nabla u\rangle}_{\geq \alpha \|\nabla u\|^2}\Big] \\ &\geq \alpha\, \mathbb{E}_\rho\left[\|\nabla u\|^2\right] \\ &= \alpha \|\phi\|^2_\rho\end{aligned}$$

Therefore,

$$\mathrm{Hess}\, H_\nu(\rho) \succeq \alpha I$$

which means H_ν is α-strongly convex. $\square$

Convergence of Langevin dynamics under SLC

Theorem: Assume $\nu \propto e^{-f}$ is α-strongly log-concave. Then the Fokker-Planck equation:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot \left(\rho_t \nabla \log \frac{\rho_t}{\nu} \right)$$

converges exponentially fast:

$$W_2(\rho_t, \nu)^2 \le e^{-2\alpha t} W_2(\rho_0, \nu)^2$$

- In fact, a *contraction*: For coevolving $\rho_t, \tilde{\rho}_t$ along the FP equation

$$W_2(\rho_t, \tilde{\rho}_t)^2 \le e^{-2\alpha t} W_2(\rho_0, \tilde{\rho}_0)^2$$

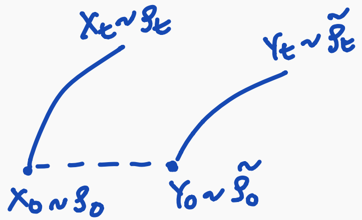
- Proof 1: This follows from ν α-SLC $\Rightarrow$ H_ν α-strongly convex. $\square$

Proof 2: (Direct proof via *coupling*)

Let $(X_0, Y_0) \sim \pi_0$ be the optimal coupling of $X_0 \sim \rho_0$ and $Y_0 \sim \tilde{\rho}_0$, so

$$W_2(\rho_0, \tilde{\rho}_0)^2 = \mathbb{E}[\|X_0 - Y_0\|^2]$$

Run two Langevin dynamics with the same Brownian motion dW_t (this is *synchronous coupling*):



$$dX_t = -\nabla f(X_t)\, dt + \sqrt{2}\, dW_t$$
$$dY_t = -\nabla f(Y_t)\, dt + \sqrt{2}\, dW_t$$

Then $X_t \sim \rho_t$, $Y_t \sim \tilde{\rho}_t$, and by definition, $W_2(\rho_t, \tilde{\rho}_t)^2 \le \mathbb{E}[\|X_t - Y_t\|^2]$

The difference $X_t - Y_t$ follows

$$\dot{X}_t - \dot{Y}_t = \frac{d}{dt}(X_t - Y_t) = -(\nabla f(X_t) - \nabla f(Y_t))$$

Since f is α-strongly convex, ∇f is *strongly monotone*:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2$$

Then

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[\|X_t - Y_t\|^2] &= 2\mathbb{E}\left[\langle X_t - Y_t, \dot{X}_t - \dot{Y}_t \rangle\right] \\ &= -2\mathbb{E}\left[\langle X_t - Y_t, \nabla f(X_t) - \nabla f(Y_t) \rangle\right] \\ &\leq -2\alpha\, \mathbb{E}[\|X_t - Y_t\|^2] \end{aligned}$$

Integrating:

$$\mathbb{E}[\|X_t - Y_t\|^2] \leq e^{-2\alpha t}\mathbb{E}[\|X_0 - Y_0\|^2] = e^{-2\alpha t} W_2(\rho_0, \tilde{\rho}_0)^2$$

Therefore,

$$W_2(\rho_t, \tilde{\rho}_t)^2 \leq \mathbb{E}[\|X_t - Y_t\|^2] \leq e^{-2\alpha t} W_2(\rho_0, \tilde{\rho}_0)^2$$

In particular, if $\tilde{\rho}_0 = \nu$, then $\tilde{\rho}_t = \nu$ for all $t > 0$. $\square$

Strongly log-concave distributions

ν α-**SLC**:

$$\nu \propto e^{-f}, \qquad \nabla^2 f(x) \succeq \alpha I$$

- Nice properties: Unimodal, concentration, Gaussian tail
- Can be sampled efficiently: Langevin dynamics converges fast

$$dX_t = -\nabla f(X_t)\, dt + \sqrt{2}\, dW_t$$

$\Rightarrow$ Coupling technique works since ∇f is strongly monotone

- Analogous to the class of strongly convex functions
  - $*$ f strongly convex $\Leftrightarrow$ ν SLC $\Rightarrow$ H_ν strongly convex
- Log-concave sampling (of ν) $\equiv$ convex optimization (of H_ν)

Log-Sobolev inequality

Log-Sobolev inequality

Definition: A probability distribution ν on $\mathbb{R}^n$ satisfies α-**log-Sobolev inequality** (α-LSI) if for all $\rho \in \mathcal{P}(\mathbb{R}^n)$:

$$J_\nu(\rho) \geq 2\alpha \, H_\nu(\rho)$$

where

- $H_\nu(\rho) = \mathbb{E}_\nu[\log \frac{\rho}{\nu}]$ is relative entropy
- $J_\nu(\rho) = \mathbb{E}_\nu[\|\nabla \log \frac{\rho}{\nu}]\|^2$ is relative Fisher information
- This is the gradient domination condition of relative entropy:

$$J_\nu(\rho) = \|\mathrm{grad}\, H_\nu(\rho)\|_\rho^2 \geq 2\alpha \, H_\nu(\rho)$$

- Implies exponential convergence rate of Langevin dynamics

Convergence of Langevin dynamics under LSI

Theorem: Assume $\nu \propto e^{-f}$ satisfies α-log-Sobolev inequality.
Then the Fokker-Planck equation:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot \left(\rho_t \nabla \log \frac{\rho_t}{\nu} \right)$$

converges exponentially fast:

$$H_\nu(\rho_t) \le e^{-2\alpha t} H_\nu(\rho_0)$$

- Proof: $\frac{d}{dt} H_\nu(\rho_t) = -J_\nu(\rho_t) \le -2\alpha\, H_\nu(\rho_t)$. $\square$

Convergence of Langevin dynamics under LSI

Theorem: Assume $\nu \propto e^{-f}$ satisfies α-log-Sobolev inequality.
Then the Fokker-Planck equation:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot \left(\rho_t \nabla \log \frac{\rho_t}{\nu} \right)$$

converges exponentially fast:

$$H_\nu(\rho_t) \leq e^{-2\alpha t} H_\nu(\rho_0)$$

- Proof: $\frac{d}{dt} H_\nu(\rho_t) = -J_\nu(\rho_t) \leq -2\alpha\, H_\nu(\rho_t)$. $\square$
- Since α-LSI $\Rightarrow$ α-TI, also implies

$$W_2(\rho_t, \nu)^2 \leq \frac{2}{\alpha} e^{-2\alpha t} H_\nu(\rho_0)$$

(but: $W_2(\rho_t, \tilde{\rho}_t)^2 \cancel{\leq} e^{-2\alpha t} W_2(\rho_0, \tilde{\rho}_0)^2$)

SLC implies LSI

Lemma: If ν is α-SLC, then ν satisfies α-LSI

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

- Proof: ν SLC $\Rightarrow$ H_ν strongly convex $\Rightarrow$ gradient dominated $=$ LSI.

$\square$

SLC implies LSI

Lemma: If ν is α-SLC, then ν satisfies α-LSI

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

- Proof: ν SLC $\Rightarrow$ H_ν strongly convex $\Rightarrow$ gradient dominated $=$ LSI. $\square$

- Proof 2: (Bakry-Emery)
  Consider Langevin dynamics from $\rho_0 = \rho$, so $\frac{d}{dt} H_\nu(\rho_t) = -J_\nu(\rho_t)$.
  Show that $\frac{d}{dt} J_\nu(\rho_t) \leq -2\alpha J_\nu(\rho_t)$, so $J_\nu(\rho_t) \leq e^{-2\alpha t} J_\nu(\rho)$.
  Integrate from $t = 0$ to ∞ to get $H_\nu(\rho) \leq \underbrace{(\int_0^\infty e^{-2\alpha t} dt)}_{= \frac{1}{2\alpha}} J_\nu(\rho)$. $\square$

Log-Sobolev inequality

α-**LSI**: For all $\rho \in \mathcal{P}(\mathbb{R}^n)$

$$H_\nu(\rho) \leq \frac{1}{2\alpha} J_\nu(\rho)$$

- Equivalent to: For all $h\colon \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}_\nu[h^2] < \infty$,

LSI :
$$\mathrm{Ent}_\nu(h^2) \leq \frac{2}{\alpha} \mathbb{E}_\nu[\|\nabla h\|^2]$$

where

$$\mathrm{Ent}_\nu(h^2) = \mathbb{E}_\nu[h^2 \log h^2] - \mathbb{E}_\nu[h^2] \log \mathbb{E}_\nu[h^2]$$

Plug in $h^2 = \frac{\rho}{\nu}$, $\mathbb{E}_\nu[h^2] = \int \nu \cdot \frac{\rho}{\nu} = \int \rho = 1$

then get $H_\nu(\rho) \leq \frac{1}{2\alpha} J_\nu(\rho)$

Log-Sobolev inequality

α-**LSI**: For all $\rho \in \mathcal{P}(\mathbb{R}^n)$

$$H_\nu(\rho) \le \frac{1}{2\alpha} J_\nu(\rho)$$

- Equivalent to: For all $h \colon \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}_\nu[h^2] < \infty$,

$$\mathrm{Ent}_\nu(h^2) \le \frac{2}{\alpha} \mathbb{E}_\nu[\|\nabla h\|^2]$$

  where

$$\mathrm{Ent}_\nu(h^2) = \mathbb{E}_\nu[h^2 \log h^2] - \mathbb{E}_\nu[h^2] \log \mathbb{E}_\nu[h^2]$$

- Gross (1975) proved in Gaussian case $\nu = \mathcal{N}(0, I)$

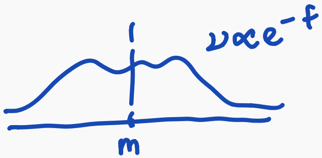
  Stam (1959) proved in equivalent formulation

$$\mathcal{P}(\rho) \cdot J(\rho) \ge n$$

  where $\mathcal{P}(\rho) = \frac{1}{2\pi e} \exp(\frac{2}{n} H(\rho))$ is the *entropy power*

Log-Sobolev inequality on $\mathbb{R}^1$

Necessary and sufficient conditions for $\nu\colon \mathbb{R} \to \mathbb{R}$ to satisfy LSI:



$$\sup_{x \geq m} \nu([x,\infty)) \left(\int_m^x \frac{dt}{\nu(t)} \right) \log \frac{1}{\nu([x,\infty))} < +\infty$$

$$\sup_{x \leq m} \nu((-\infty,x]) \left(\int_x^m \frac{dt}{\nu(t)} \right) \log \frac{1}{\nu((-\infty,x])} < +\infty$$

where $m \in \mathbb{R}$ is a *median* of ν, i.e. $\nu((-\infty, m]) = \nu([m,\infty)) = \frac{1}{2}$

- If $f(x) \sim |x|^p$ as $|x| \to \infty$, then $\nu \propto e^{-f}$ satisfies LSI $\quad\Leftrightarrow\quad p \geq 2$.

[Bobkov & Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, Journal of Functional Analysis, 1999]

Necessary or sufficient conditions for LSI

Theorem:

1. If ν_1 and ν_2 satisfy α-LSI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ satisfies α-LSI on $\mathbb{R}^{n_1+n_2}$

Necessary or sufficient conditions for LSI

Theorem:

1. If ν_1 and ν_2 satisfy α-LSI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ satisfies α-LSI on $\mathbb{R}^{n_1+n_2}$
2. If $\nu \propto e^{-f}$ where $\nabla^2 f(x) \succeq \alpha I$, then ν satisfies α-LSI

Necessary or sufficient conditions for LSI

Theorem:

1. If ν_1 and ν_2 satisfy α-LSI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ satisfies α-LSI on $\mathbb{R}^{n_1+n_2}$
2. If $\nu \propto e^{-f}$ where $\nabla^2 f(x) \succeq \alpha I$, then ν satisfies α-LSI
3. Let $\tilde{\nu} = \nu e^{-g}$ where ν satisfies α-LSI and g is bounded. Then $\tilde{\nu}$ satisfies $\tilde{\alpha}$-LSI where $\tilde{\alpha} = \alpha e^{-2\mathrm{osc}(g)}$, $\mathrm{osc}(g) = \sup g - \inf g$

Necessary or sufficient conditions for LSI

Theorem:

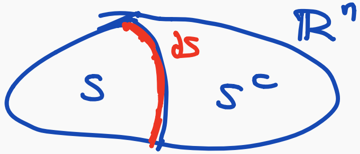
1. If ν_1 and ν_2 satisfy α-LSI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ satisfies α-LSI on $\mathbb{R}^{n_1+n_2}$
2. If $\nu \propto e^{-f}$ where $\nabla^2 f(x) \succeq \alpha I$, then ν satisfies α-LSI
3. Let $\tilde{\nu} = \nu e^{-g}$ where ν satisfies α-LSI and g is bounded. Then $\tilde{\nu}$ satisfies $\tilde{\alpha}$-LSI where $\tilde{\alpha} = \alpha e^{-2\mathrm{osc}(g)}$, $\mathrm{osc}(g) = \sup g - \inf g$
4. If ν satisfies LSI, then $\int_{\mathbb{R}^n} e^{c\|x\|^2} \nu(x) dx < +\infty$ for some $c > 0$

[Villani, *Topics in Optimal Transportation*, AMS, 2003, Theorem 9.9]

Gaussian isoperimetry

We say ν satisfies **Gaussian isoperimetry** if for all $S \subset \mathbb{R}^n$

$$\nu(\partial S) \geq \psi \cdot \mathcal{G}(\nu(S))$$



where $\mathcal{G} = \gamma_1 \circ \Gamma^{-1}$ is the Gaussian isoperimetry function

- $\gamma_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is Gaussian density $\mathcal{N}(0,1)$ on $\mathbb{R}$
- $\Gamma(x) = \int_{-\infty}^{x} \gamma_1(x) dx$ is Gaussian CDF
- $\mathcal{G}(x) \sim x\sqrt{2\log(1/x)}$ as $x \to 0$

Log-isoperimetry

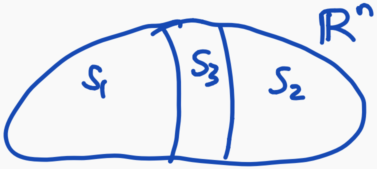
Gaussian isoperimetry: For all $S \subset \mathbb{R}^n$

$$\nu(\partial S) \geq \psi \cdot \mathcal{G}(\nu(S))$$

Isoperimetry

$$\phi = \inf_{\substack{S \subset \mathbb{R}^n \\ \nu(S) \le \frac{1}{2}}} \frac{\nu(\partial S)}{\nu(S)}$$

$\Leftrightarrow$ **Log-isoperimetry**: For $\nu(S) \leq \frac{1}{2}$



$$\nu(\partial S) \geq \psi \cdot \nu(S) \sqrt{\log \frac{1}{\nu(S)}}$$

Log-isoperimetry

$$\psi = \inf_{\substack{S \subset \mathbb{R}^n \\ \nu(S) \le \frac{1}{2}}} \frac{\nu(\partial S)}{\nu(S) \cdot \sqrt{\log \frac{1}{\nu(S)}}}$$

$\Leftrightarrow$ For all partition $\mathbb{R}^n = S_1 \cup S_2 \cup S_3$

$$\nu(S_3) \geq \psi \cdot d(S_1, S_2) \cdot \min\{\nu(S_1), \nu(S_2)\} \cdot \sqrt{\log \frac{1}{\min\{\nu(S_1), \nu(S_2)\}}}$$

where $d(S_1, S_2) = \min\{\|x - y\|_2 : x \in S_1, y \in S_2\}$

LSI $\Rightarrow$ log-isoperimetry

Theorem: If ν satisfies α-LSI:

$$J_\nu(\rho) \geq 2\alpha \, H_\nu(\rho)$$

then it satisfies log-isoperimetry:

$$\nu(\partial S) \geq C\sqrt{\alpha} \cdot \mathcal{G}(\nu(S))$$

for some universal constant C

[Ledoux, *A simple analytic proof of an inequality by P. Buser*, AMS, 1994]

Properties of H_ν in terms of ν

Relative entropy on $\mathcal{P}(\mathbb{R}^n)$

$$H_\nu(\rho) = \mathbb{E}_\rho\left[\log \frac{\rho}{\nu}\right]$$

Strong convexity:

$$\mathrm{Hess}\, H_\nu(\rho) \succeq \alpha I$$

Gradient dominated:

$$\|\mathrm{grad}\, H_\nu(\rho)\|_\rho^2 \geq 2\alpha\, H_\nu(\rho)$$

Quadratic growth:

$$H_\nu(\rho) \geq \frac{\alpha}{2} d(\rho, \nu)^2$$

Target distribution on $\mathbb{R}^n$

$$\nu \propto e^{-f}$$

Strong log-concavity:

$$\nabla^2 f(x) \succeq \alpha I$$

Log-Sobolev inequality:

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

Talagrand inequality:

$$H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$

Talagrand inequality

Talagrand inequality

Definition: ν satisfies α-**Talagrand inequality** (α-TI) if for all $\rho \in \mathcal{P}(\mathbb{R}^n)$:

$$H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$

- Also known as Transport Inequality
- This is the quadratic growth condition of relative entropy:

$$H_\nu(\rho) \geq \frac{\alpha}{2} d(\rho, \nu)^2$$

- First shown by Talagrand (1996) for Gaussian $\nu = \mathcal{N}(0, I)$

[Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. 6, 1996]

LSI implies TI

Lemma: If ν satisfies α-LSI:

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

then it also satisfies α-TI:

$$H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$

- Proof: H_ν gradient dominated $=$ LSI $\Rightarrow$ quadratic growth $=$ TI. $\square$
- If ν is log-concave, then the converse also holds (for some α')

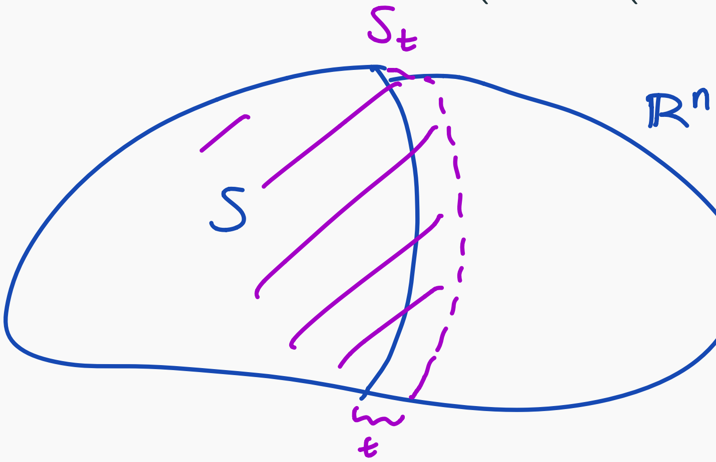
[Otto & Villani, *Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality*, Journal of Functional Analysis, 2000]

TI implies concentration

Lemma: Suppose ν satisfies α-TI. Then for any $S \subset \mathbb{R}^n$, the neighborhood $S_t = S + tB_2^n = \{x \in \mathbb{R}^n \mid d(x, S) \leq t\}$ has

$$\nu(S_t) \geq 1 - \exp\left(-\frac{\alpha}{2}\left(t - \sqrt{\frac{2}{\alpha}\log\frac{1}{\nu(S)}}\right)^2\right) \sim 1 - e^{-\frac{\alpha}{2}t^2}$$

for large t:



TI implies concentration

Lemma: Suppose ν satisfies α-TI. Then for any $S \subset \mathbb{R}^n$, the neighborhood $S_t = S + tB_2^n = \{x \in \mathbb{R}^n \mid d(x, S) \leq t\}$ has

$$\nu(S_t) \geq 1 - \exp\left(-\frac{\alpha}{2}\left(t - \sqrt{\frac{2}{\alpha}\log\frac{1}{\nu(S)}}\right)^2 \right)$$

Proof:

$$\begin{aligned} t \leq W_2(\nu|_S, \nu|_{S_t^c}) &\leq W_2(\nu|_S, \nu) + W_2(\nu|_{S_t^c}, \nu) \\ &\leq \sqrt{\frac{2}{\alpha}H_\nu(\nu|_S)} + \sqrt{\frac{2}{\alpha}H_\nu(\nu|_{S_t^c})} \\ &= \sqrt{\frac{2}{\alpha}\log\frac{1}{\nu(S)}} + \sqrt{\frac{2}{\alpha}\log\frac{1}{\nu(S_t^c)}} \end{aligned}$$

$\square$

[Villani, *Topics in Optimal Transportation*, AMS, 2003, §9.3.2]

Properties of ν

1. **Strongly log-concave:** $\Rightarrow H_\nu$ strongly convex

$$-\nabla^2 \log \nu(x) \succeq \alpha I$$

2. **Log-Sobolev inequality:** $\Leftrightarrow H_\nu$ gradient dominated

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

3. **Talagrand inequality:** $\Leftrightarrow H_\nu$ quadratic growth

$$H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$

4. **Poincaré inequality:**

$$\mathbb{E}_\nu[\|\nabla h\|^2] \geq \alpha\, \mathrm{Var}_\nu(h)$$

Poincaré inequality

Poincaré inequality

Definition: ν satisfies α-**Poincaré inequality** (α-PI) if for all smooth $\phi \colon \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{Var}_\nu(\phi) \leq \frac{1}{\alpha} \mathbb{E}_\nu[\|\nabla \phi\|^2]$$

where $\mathrm{Var}_\nu(\phi) = \mathbb{E}_\nu[\phi^2] - \mathbb{E}_\nu[\phi]^2$ is the variance of ϕ

Poincaré inequality

Definition: ν satisfies α-**Poincaré inequality** (α-PI) if for all smooth $\phi\colon \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{Var}_\nu(\phi) \le \frac{1}{\alpha}\,\mathbb{E}_\nu[\|\nabla\phi\|^2]$$

where $\mathrm{Var}_\nu(\phi) = \mathbb{E}_\nu[\phi^2] - \mathbb{E}_\nu[\phi]^2$ is the variance of ϕ

- If $\phi(x) = \langle x, u\rangle$, then $u^\top \mathrm{Cov}_\nu(X) u \le \frac{1}{\alpha}\|u\|^2$, so

  $\nabla\phi(x) = u$

$$\mathrm{Cov}_\nu(X) \preceq \frac{1}{\alpha} I$$

PI $\Rightarrow$ Isoperimetry

α-**PI**:

$$\mathrm{Var}_\nu(\phi) \le \frac{1}{\alpha}\,\mathbb{E}_\nu[\|\nabla\phi\|^2]$$

Theorem: If ν satisfies α-PI, then it satisfies isoperimetry: For all $S \subset \mathbb{R}^n$

$$\nu(\partial S) \ge 2\sqrt{\alpha} \cdot \min\{\nu(S), 1-\nu(S)\}$$

[Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, 1970]

[Ledoux, *A simple analytic proof of an inequality by P. Buser*, AMS, 1994]

LSI $\Rightarrow$ Poincaré inequality

Theorem: If ν satisfies α-LSI:

$$J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

then it also satisfies α-PI:

$$\mathrm{Var}_\nu(\phi) \leq \frac{1}{\alpha}\, \mathbb{E}_\nu[\|\nabla \phi\|^2]$$

- Can obtain via linearization: $\rho = (1 + \eta\phi)\nu$ as $\eta \to 0$
- Also α-TI $\Rightarrow$ α-PI via linearization

[Rothaus, *Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities*, J. Funct. Anal., 1985]

Necessary or sufficient conditions for PI

Theorem:

1. If ν_1 and ν_2 satisfy α-PI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ also satisfies α-PI on $\mathbb{R}^{n_1+n_2}$

Necessary or sufficient conditions for PI

Theorem:

1. If ν_1 and ν_2 satisfy α-PI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ also satisfies α-PI on $\mathbb{R}^{n_1+n_2}$
2. If $\nu \propto e^{-f}$ where $\nabla^2 f(x) \succeq \alpha I$, then ν satisfies α-PI

Necessary or sufficient conditions for PI

Theorem:

1. If ν_1 and ν_2 satisfy α-PI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ also satisfies α-PI on $\mathbb{R}^{n_1+n_2}$
2. If $\nu \propto e^{-f}$ where $\nabla^2 f(x) \succeq \alpha I$, then ν satisfies α-PI
3. Let $\tilde{\nu} = \nu e^{-g}$ where ν satisfies α-PI and g is bounded. Then $\tilde{\nu}$ satisfies $\tilde{\alpha}$-PI where $\tilde{\alpha} = \alpha e^{-2\mathrm{osc}(g)}$, $\mathrm{osc}(g) = \sup g - \inf g$

Necessary or sufficient conditions for PI

Theorem:

1. If ν_1 and ν_2 satisfy α-PI on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, then $\nu_1 \otimes \nu_2$ also satisfies α-PI on $\mathbb{R}^{n_1+n_2}$
2. If $\nu \propto e^{-f}$ where $\nabla^2 f(x) \succeq \alpha I$, then ν satisfies α-PI
3. Let $\tilde{\nu} = \nu e^{-g}$ where ν satisfies α-PI and g is bounded. Then $\tilde{\nu}$ satisfies $\tilde{\alpha}$-PI where $\tilde{\alpha} = \alpha e^{-2\mathrm{osc}(g)}$, $\mathrm{osc}(g) = \sup g - \inf g$
4. If ν satisfies PI, then $\int_{\mathbb{R}^n} e^{c\|x\|}\nu(x)dx < +\infty$ for some $c > 0$

- E.g., exponential distribution $\nu(x) \propto e^{-c\|x\|}$ satisfies PI but not LSI

Poincaré inequality as spectral gap

L^2 space

Let $L^2(\nu) = \{\phi\colon \mathbb{R}^n \to \mathbb{R} \mid \int_{\mathbb{R}^n} \phi(x)^2 \, d\nu(x) < \infty\}$

Define inner product and norm

$$\langle g, h \rangle_\nu := \mathbb{E}_\nu[gh] = \int_{\mathbb{R}^n} g(x)h(x)d\nu(x)$$
$$\|h\|_\nu^2 := \mathbb{E}_\nu[h^2] = \int_{\mathbb{R}^n} h(x)^2 d\nu(x)$$

- χ^2-**divergence** of ρ with respect to ν with density $h = \dfrac{\rho}{\nu}$ is

$$\chi^2_\nu(\rho) = \int_{\mathbb{R}^n} \nu(x) \left(\frac{\rho(x)}{\nu(x)} - 1 \right)^2 dx = \|h - 1\|_\nu^2$$

Laplacian

Let $\nu \propto e^{-f}$ be a probability distribution on $\mathbb{R}^n$

Define **Laplacian** operator $L\colon L^2(\nu) \to L^2(\nu)$ by

$$L = -\Delta + \nabla f \cdot \nabla$$

$$(Lh)(x) = -\Delta h(x) + \langle \nabla f(x), \nabla h(x) \rangle$$

- Characterizes integration by parts in $L^2(\nu)$:

$$\langle Lg, h \rangle_\nu = \int_{\mathbb{R}^n} (Lg)h \, d\nu = \int_{\mathbb{R}^n} \langle \nabla g, \nabla h \rangle \, d\nu = \int_{\mathbb{R}^n} g(Lh) \, d\nu = \langle g, Lh \rangle_\nu$$

- $L \succeq 0$:

$$\langle Lh, h \rangle_\nu = \|\nabla h\|_\nu^2 \geq 0$$

Poincaré inequality as spectral gap of Laplacian

Laplacian:

$$L = -\Delta + \nabla f \cdot \nabla \quad \succeq 0$$

- Smallest eigenvalue is 0:

$$L\mathbb{1} = -\underbrace{\Delta 1}_{=0} + \nabla f \cdot \underbrace{\nabla 1}_{=0} = 0$$

- **Spectral gap**:

$$\lambda(L) = \inf_{h:\ \langle h,1\rangle_\nu = 0} \frac{\langle h, Lh\rangle_\nu}{\|h\|_\nu^2} = \inf_{h:\ \langle h,1\rangle_\nu = 0} \frac{\|\nabla h\|_\nu^2}{\mathrm{Var}_\nu(h)}$$

- Therefore, ν satisfies α-Poincaré inequality $\Leftrightarrow$ $\lambda(L) \geq \alpha$

$$\forall h: \quad \mathrm{Var}_\nu(h) \leq \frac{1}{\alpha} \mathbb{E}_\nu[\|\nabla h\|^2]$$

Laplacian controls evolution of relative density

Lemma: Suppose X_t follows the Langevin dynamics:

$$dX_t = -\nabla f(X_t)\, dt + \sqrt{2}\, dW_t$$

so its density $X_t \sim \rho_t$ follows the Fokker-Planck equation:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t$$

Then the density $h_t = \dfrac{\rho_t}{\nu}$ with respect to $\nu \propto e^{-f}$ follows:

$$\frac{\partial h_t}{\partial t} = -\langle \nabla f, \nabla h_t \rangle + \Delta h_t = -L h_t$$

- This is the *backward Kolmogorov equation*

Convergence rate in χ^2-divergence

Theorem: Assume $\nu \propto e^{-f}$ satisfies α-Poincaré inequality. Then the Langevin dynamics

$$dX_t = -\nabla f(X_t)\, dt + \sqrt{2}\, dW_t$$

converges exponentially fast in χ^2-divergence:

$$\chi^2_\nu(\rho_t) \le e^{-2\alpha t} \chi^2_\nu(\rho_0)$$

Proof: Let $h_t = \frac{\rho_t}{\nu}$, so $\dot{h}_t = \frac{\partial h_t}{\partial t} \overset{*}{=} -Lh_t$. Then

$$
\begin{aligned}
\frac{d}{dt}\chi^2_\nu(\rho_t) &= \frac{d}{dt}\mathbb{E}_\nu[\|h_t - 1\|^2] \\
&= 2\mathbb{E}_\nu\left[\langle h_t - 1, \dot{h}_t\rangle\right] \\
&\overset{*}{=} -2\mathbb{E}_\nu\left[\langle h_t - 1, Lh_t\rangle\right] \\
&= -2\mathbb{E}_\nu\left[\langle h_t, Lh_t\rangle\right] \\
&= -2\mathbb{E}_\nu\left[\|\nabla h_t\|^2\right] \\
&\le -2\alpha\,\mathrm{Var}_\nu(h_t) \\
&= -2\alpha\,\chi^2_\nu(\rho_t)
\end{aligned}
$$

integration by parts

by Poincaré inequality

Integrating gives the result. $\square$

PI $\Rightarrow$ Transportation inequality

Analogous to LSI $\Rightarrow$ TI

Theorem: If ν satisfies α-Poincaré inequality, then it also satisfies the χ^2-transportation inequality:

$$\chi^2_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$

- [Ding, *A note on quadratic transportation and divergence inequality*, Statist. Probab. Lett., 2015]
- [Liu, *The Poincaré inequality and quadratic transportation-variance inequalities*, Electron. J. Probab., 2020]

Properties of ν

1. **Strongly log-concave:**

$$\Downarrow \qquad -\nabla^2 \log \nu(x) \succeq \alpha I$$

2. **Log-Sobolev inequality:**

$$\Downarrow \qquad J_\nu(\rho) \geq 2\alpha\, H_\nu(\rho)$$

3. **Talagrand inequality:**

$$\Downarrow \qquad H_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$

4. **Poincaré inequality:**

$$\Downarrow \qquad \mathbb{E}_\nu[\|\nabla h\|^2] \geq \alpha\, \mathrm{Var}_\nu(h)$$

5. **χ^2-Transportation inequality:**

$$\chi^2_\nu(\rho) \geq \frac{\alpha}{2} W_2(\rho, \nu)^2$$