Poincaré inequalities for dimension reduction and efficient sampling

Part IV: Optimal preconditionner for Langevin-type sampler

Olivier Zahm
olivier.zahm@inria.fr

INRIA – Université Grenoble-Alpes
https://team.inria.fr/airsea/en/

ETICS
October 2025

Poincaré inequalities

A probability distribution μ on $\mathbb{R}^d$ satisfies the **Poincaré inequality** if there exists $C(\mu) < \infty$ such that

$$\mathrm{Var}_\mu(f) \leq C(\mu) \int \|\nabla f\|^2 \mathrm{d}\mu$$

for any smooth function $f : \mathbb{R}^d \to \mathbb{R}$. The smallest constant $C(\mu)$ is called the **Poincaré constant**.

Poincaré inequalities are central in many fields, including:

- Global sensitivity analysis: $\boxed{\text{Sobol'} \leq C(\pi)\text{DGSM}}$
- (Nonlinear) Dimension reduction
- Control of PDE, domain decomposition,
- Concentration of measure, isoperimetric inequality,
- MCMC sampling,
- Stochastic differential equation (SDE),
- ...

The (overdamped) Langevin equation

Assume

$$\mathrm{d}\mu(x) \propto \exp(-V(x))\mathrm{d}x$$

for some smooth potential $V : \mathbb{R}^d \to \mathbb{R}$. The overdamped Langevin SDE reads

$$X_0 \sim \mu_0 \quad \Rightarrow \quad \boxed{\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t} \quad \Rightarrow \quad X_\infty \sim \mu$$

where B_t is the Brownian motion.

Proposition. For any $C \geq 0$ we have:

$$C(\mu) \leq C \qquad \Leftrightarrow \qquad \left\{\chi^2(\mu_t||\mu) \leq e^{-2t/C}\,\chi^2(\mu_0||\mu), \quad \forall \mu_0\right\}$$

where $\chi^2(\mu_0||\mu) = \mathrm{Var}_\mu(\mathrm{d}\mu_0/\mathrm{d}\mu)$

The (overdamped) Langevin equation

Assume

$$\mathrm{d}\mu(x) \propto \exp(-V(x))\mathrm{d}x$$

for some smooth potential $V : \mathbb{R}^d \to \mathbb{R}$. The overdamped Langevin SDE reads

$$X_0 \sim \mu_0 \quad \Rightarrow \quad \boxed{\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t} \quad \Rightarrow \quad X_\infty \sim \mu$$

where B_t is the Brownian motion.

Proposition. For any $C \geq 0$ we have:

$$C(\mu) \leq C \quad \Leftrightarrow \quad \left\{\chi^2(\mu_t||\mu) \leq e^{-2t/C}\, \chi^2(\mu_0||\mu), \quad \forall \mu_0\right\}$$

where $\chi^2(\mu_0||\mu) = \mathrm{Var}_\mu(\mathrm{d}\mu_0/\mathrm{d}\mu)$

Remark: LogSob inequality "Entropy $\leq C_{\mathrm{LSI}}(\mu)$ L2-norm of gradient"

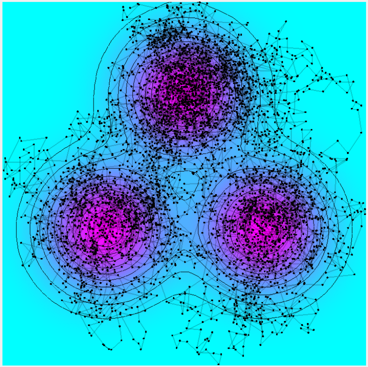
$$C_{\mathrm{LSI}}(\mu) \leq C \quad \Leftrightarrow \quad \left\{\mathrm{D}_{\mathrm{KL}}(\mu_t||\mu) \leq e^{-2t/C}\, \mathrm{D}_{\mathrm{KL}}(\mu_0||\mu), \quad \forall \mu_0\right\}$$

[Chewi et.al. 2024] "Analysis of Langevin Monte Carlo from Poincaré to log-Sobolev" FoCM

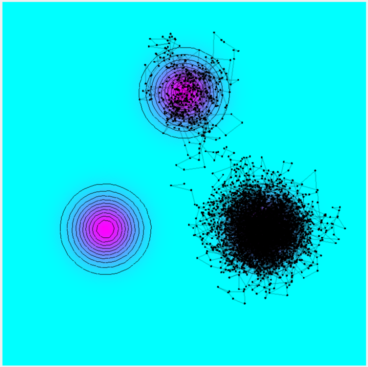
Unajusted Langevin Algorithm (ULA)

Euler-Maruyama discretization of SDE:

$$X_{n+1} = X_n - \nabla V(X_n)\Delta t + \sqrt{2\Delta t} Z_n, \qquad Z_n \sim \mathcal{N}(0, I_d)$$



Small $C(\mu)$



Large $C(\mu)$

How to control/bound $C(\mu)$?

- **Bakry Émery** theorem for log-concave distributions

$$\mathrm{Hess}(-\log \mu(x)) \succeq \rho I_d \qquad \Rightarrow \qquad C(\mu) \leq 1/\rho$$

- **Holley-Stroock** perturbation lemma

$$\alpha \mathrm{d}\nu(x) \leq \mathrm{d}\mu(x) \leq \beta \mathrm{d}\nu(x) \qquad \Rightarrow \qquad C(\mu) \leq \frac{\beta}{\alpha} C(\nu)$$

  See also [Cattiaux and Guillin 2022] for more advanced perturbation analyses.

- Combination of the two above.

Riemannian Poincaré inequalities

Given $W : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ a field of symmetric positive matrices, let $C(\mu, W)$ be the smallest constant such that

$$\boxed{\mathsf{Var}_\mu(f) \leq C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)}$$

Riemannian Poincaré inequalities

Given $W : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ a field of symmetric positive matrices, let $C(\mu, W)$ be the smallest constant such that

$$\boxed{\mathrm{Var}_\mu(f) \le C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)}$$

This generalizes severall variants of Poincaré inequalities, including:

- **Brascamp-Lieb inequality** for strongly log-concave distributions μ

$$\mathrm{Var}_\mu(f) \le \int \nabla f(x)^\top \left(-\mathrm{Hess} \log \mu(x)\right)^{-1} \nabla f(x) \mathrm{d}\mu(x)$$

Riemannian Poincaré inequalities

Given $W : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ a field of symmetric positive matrices, let $C(\mu, W)$ be the smallest constant such that

$$\boxed{\mathrm{Var}_\mu(f) \le C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)}$$

This generalizes severall variants of Poincaré inequalities, including:

- **Brascamp-Lieb inequality** for strongly log-concave distributions μ

$$\mathrm{Var}_\mu(f) \le \int \nabla f(x)^\top \left(-\mathrm{Hess}\log \mu(x)\right)^{-1} \nabla f(x) \mathrm{d}\mu(x)$$

- **Mirror PI.** Given a convex function φ, let $C(\mu, \varphi)$ be the smallest constant such that

$$\mathrm{Var}_\mu(f) \le C(\mu, \varphi) \int \nabla f(x)^\top \left(\mathrm{Hess}\,\varphi(x)\right)^{-1} \nabla f(x) \mathrm{d}\mu(x)$$

Riemannian Poincaré inequalities

Given $W : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ a field of symmetric positive matrices, let $C(\mu, W)$ be the smallest constant such that

$$\boxed{\mathrm{Var}_\mu(f) \leq C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)}$$

This generalizes severall variants of Poincaré inequalities, including:

- **Brascamp-Lieb inequality** for strongly log-concave distributions μ

$$\mathrm{Var}_\mu(f) \leq \int \nabla f(x)^\top \left(-\mathrm{Hess}\log\mu(x)\right)^{-1} \nabla f(x) \mathrm{d}\mu(x)$$

- **Mirror PI.** Given a convex function φ, let $C(\mu, \varphi)$ be the smallest constant such that

$$\mathrm{Var}_\mu(f) \leq C(\mu, \varphi) \int \nabla f(x)^\top \left(\mathrm{Hess}\,\varphi(x)\right)^{-1} \nabla f(x) \mathrm{d}\mu(x)$$

- **Transported PI.** For any diffeomorphism $T : \mathbb{R}^d \to \mathbb{R}^d$, we have

$$\mathrm{Var}_\mu(f) \leq C(T_\sharp \mu) \int \|\nabla T(x)^{-1} \nabla f(x)\|^2 \mathrm{d}\mu(x)$$

where $T_\sharp \mu$ is the pushforward measure of μ by T.

Riemannian PI and Langevin dynamic

Proposition

Consider the preconditioned (overdamped) Langevin dynamic $X_t \sim \mu_t$

$$\mathrm{d}X_t = (\mathrm{div}(W) - W\nabla V)\mathrm{d}t + \sqrt{2W}\mathrm{d}B_t$$

Then, for any $C \geq 0$, we have

$$C(\mu, W) \leq C \qquad \Leftrightarrow \qquad \left\{\chi^2(\mu_t||\mu) \leq e^{-2t/C}\,\chi^2(\mu_0||\mu), \quad \forall \mu_0\right\}$$

(proof: Theorem 4.2.5 from [Bakry, Gentil, Ledoux (2014)])

Riemannian PI and Langevin dynamic

Proposition

Consider the preconditioned (overdamped) Langevin dynamic $X_t \sim \mu_t$

$$\mathrm{d}X_t = (\mathrm{div}(W) - W\nabla V)\mathrm{d}t + \sqrt{2W}\mathrm{d}B_t$$

Then, for any $C \geq 0$, we have

$$C(\mu, W) \leq C \qquad \Leftrightarrow \qquad \left\{\chi^2(\mu_t||\mu) \leq e^{-2t/C}\, \chi^2(\mu_0||\mu), \quad \forall \mu_0\right\}$$

(proof: Theorem 4.2.5 from [Bakry, Gentil, Ledoux (2014)])

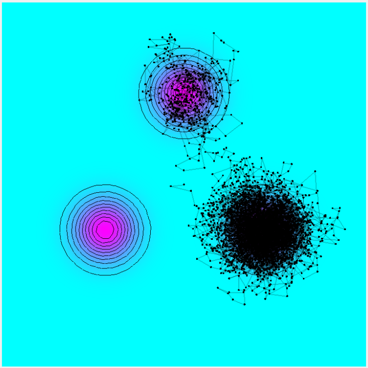
Objective of the talk: optimal preconditioning of Langeving dynamic via

$$\boxed{\min_{W:x\to W(x)\succeq 0} C(\mu, W)}$$

- Existence of solution via the notion of **moment measure**
- Numerical solution using the **finite elements method**

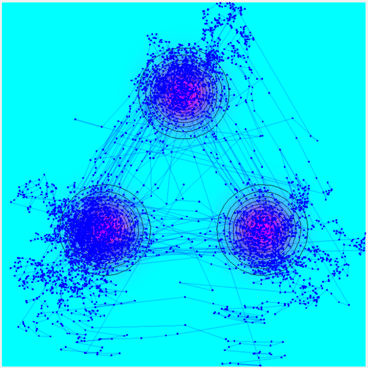
Spoiler

ULA:



$C(\mu) = 61.8$

Preconditioned ULA:



$C(\mu, W) = 1.005$

Existence of solution via moment maps

Scaling for W

$$\mathrm{Var}_\mu(f) \le C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)$$

Because $C(\mu, \alpha W) = \alpha^{-1} C(\mu, W)$, we impose the following scaling for W:

$$\boxed{\int \mathrm{trace}(W) \mathrm{d}\mu = \mathrm{trace}(\mathrm{Cov}_\mu)}$$

Scaling for W

$$\mathrm{Var}_\mu(f) \le C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)$$

Because $C(\mu, \alpha W) = \alpha^{-1} C(\mu, W)$, we impose the following scaling for W:

$$\boxed{\int \mathrm{trace}(W) \mathrm{d}\mu = \mathrm{trace}(\mathrm{Cov}_\mu)}$$

Proposition [Cui, Tong and Zahm 2024]

- For any SPD matrix field W satisfying the above scaling, we have

$$C(\mu, W) \ge 1$$

- Furthermore, if the Riemannian PI is saturated by all **affine functions** f, then

$$C(\mu, W) = 1 \qquad \text{and} \qquad \int W \mathrm{d}\mu = \mathrm{Cov}_\mu$$

Scaling for W

$$\mathrm{Var}_\mu(f) \le C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)$$

Because $C(\mu,\alpha W) = \alpha^{-1} C(\mu, W)$, we impose the following scaling for W:

$$\boxed{\int \mathrm{trace}(W) \mathrm{d}\mu = \mathrm{trace}(\mathrm{Cov}_\mu)}$$

Proposition [Cui, Tong and Zahm 2024]

- For any SPD matrix field W satisfying the above scaling, we have

$$C(\mu, W) \ge 1$$

- Furthermore, if the Riemannian PI is saturated by all **affine functions** f, then

$$C(\mu, W) = 1 \qquad \text{and} \qquad \int W \mathrm{d}\mu = \mathrm{Cov}_\mu$$

Remark: [Lelièvre et al. 2024] imposes the constrains $\alpha I_d \preceq W \preceq \beta I_d$ and

$$\int \|W\|_F^p \mathrm{d}\mu \le 1$$

$\Rightarrow$ existence of solutions via **compactness** arguments

Existence of solution via **moment measure**

We say that μ is a **moment measure** if there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that

$$\mu = T_\sharp \nu \qquad \begin{cases} T(x) = \nabla\varphi(x) \\ \mathrm{d}\nu(x) \propto \exp(-\varphi(x))\mathrm{d}x \end{cases}$$

Existence of solution via **moment measure**

We say that μ is a **moment measure** if there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that

$$\mu = T_\sharp \nu \qquad \begin{cases} T(x) = \nabla\varphi(x) \\ \mathrm{d}\nu(x) \propto \exp(-\varphi(x))\mathrm{d}x \end{cases}$$

Sufficient condition: $\mathrm{supp}(\mu)$ is **convex** and $\int x \mathrm{d}\mu = 0$, $\int \|x\|^2 \mathrm{d}\mu < \infty$
[Cordero-Erausquin and Klartag 2015]

Existence of solution via **moment measure**

We say that μ is a **moment measure** if there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that

$$\mu = T_\sharp \nu \qquad \begin{cases} T(x) = \nabla\varphi(x) \\ \mathrm{d}\nu(x) \propto \exp(-\varphi(x))\mathrm{d}x \end{cases}$$

Sufficient condition: $\mathrm{supp}(\mu)$ is **convex** and $\int x \mathrm{d}\mu = 0$, $\int \|x\|^2 \mathrm{d}\mu < \infty$
[Cordero-Erausquin and Klartag 2015]

Theorem [Cui, Tong and Zahm 2024]

Assume μ is the moment measure of a **smooth** ($\mathcal{C}^2$) strictly convex function φ. Then

$$\boxed{W(x) = \text{Hess } \varphi(\nabla\varphi^{-1}(x))}$$

is an optimal metric with $C(\mu, W) = 1$

Existence of solution via **moment measure**

We say that μ is a **moment measure** if there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that

$$\mu = T_\sharp \nu \qquad \begin{cases} T(x) = \nabla\varphi(x) \\ \mathrm{d}\nu(x) \propto \exp(-\varphi(x))\mathrm{d}x \end{cases}$$

Sufficient condition: $\mathrm{supp}(\mu)$ is **convex** and $\int x \mathrm{d}\mu = 0$, $\int \|x\|^2 \mathrm{d}\mu < \infty$
[Cordero-Erausquin and Klartag 2015]

Theorem [Cui, Tong and Zahm 2024]

Assume μ is the moment measure of a **smooth** ($\mathcal{C}^2$) strictly convex function φ. Then

$$\boxed{W(x) = \text{Hess } \varphi(\nabla\varphi^{-1}(x))}$$

is an optimal metric with $C(\mu, W) = 1$

Sufficient condition: [Berman and Berndtsson, 2013] μ is supported on a **compact** and **convex** domain, and

$$0 < \alpha \leq \frac{\mathrm{d}\mu(x)}{\mathrm{d}x} \leq \beta < \infty$$

Link with **Stein kernels** [Fathi 2019]

A **Stein kernel** for μ is a matrix field $W : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ (not necessarily positive) such that

$$\int W \nabla f \mathrm{d}\mu = \int (x - m) f \mathrm{d}\mu$$

for any smooth f. Stein kernels are central objects in **Stein's method** to analyze CLT in Wasserstein spaces [Ledoux, Nourdin and Peccati 2015].

Link with **Stein kernels** [Fathi 2019]

A **Stein kernel** for μ is a matrix field $W : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ (not necessarily positive) such that

$$\int W\nabla f \mathrm{d}\mu = \int (x-m) f \mathrm{d}\mu$$

for any smooth f. Stein kernels are central objects in **Stein's method** to analyze CLT in Wasserstein spaces [Ledoux, Nourdin and Peccati 2015].

Proposition [Cui, Tong and Zahm 2024]

Any optimal metric W with $C(\mu, W) = 1$ is a **positive Stein Kernel**. As a consequence, the preconditioned Langevin equation writes

$$\mathrm{d}X_t = \underbrace{(\mathrm{div}(W) - W\nabla V)}_{\Downarrow \atop -(X_t - m)}\mathrm{d}t + \sqrt{2W}\mathrm{d}B_t$$

$\Rightarrow$ Gradient-free Langevin sampler!

Link with the central limit theorem (CLT)

For **independent** copies $X_1, \dots, X_N \sim \mu$, with $N \gg 1$ we have

$$\frac{1}{N}\sum_{i=1}^{N} X_i \quad \approx \quad \mathcal{N}\left(\mathbb{E}[X], \frac{1}{N}\,\mathrm{Cov}_\mu\right)$$

Link with the central limit theorem (CLT)

For **independent** copies $X_1, \ldots, X_N \sim \mu$, with $N \gg 1$ we have

$$\frac{1}{N}\sum_{i=1}^{N} X_i \quad \approx \quad \mathcal{N}\left(\mathbb{E}[X], \frac{1}{N}\,\mathrm{Cov}_\mu\right)$$

For $\mathbb{E}[X] = 0$ and $\mathrm{Cov}_\mu = I_d$, the CLT states that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i \;\sim\; \mu_N \qquad \xrightarrow[N\to\infty]{d} \qquad \gamma = \mathcal{N}\left(0, I_d\right)$$

Stein's methods provide **quantitive estimate** of the above convergence, see *e.g.* **[Ross 2011: Fundamentals of Stein's method]**

Link with the central limit theorem (CLT)

For **independent** copies $X_1, \ldots, X_N \sim \mu$, with $N \gg 1$ we have

$$\frac{1}{N}\sum_{i=1}^{N} X_i \quad \approx \quad \mathcal{N}\left(\mathbb{E}[X], \frac{1}{N}\,\mathrm{Cov}_\mu\right)$$

For $\mathbb{E}[X] = 0$ and $\mathrm{Cov}_\mu = I_d$, the CLT states that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i \;\sim\; \mu_N \qquad \xrightarrow[N\to\infty]{d} \qquad \gamma = \mathcal{N}(0, I_d)$$

Stein's methods provide **quantitive estimate** of the above convergence, see *e.g.* **[Ross 2011: Fundamentals of Stein's method]**

Proposition

For $p \geq 2$ and W being an optimal Riemannian metric, we have

$$\mathcal{W}_p(\mu_N, \gamma) \leq \frac{C(p) d^{1-2/p}}{\sqrt{N}} \left(\int \|W - I_d\|_p^p \mathrm{d}\mu \right)^{1/p}$$

where $\mathcal{W}_p(\cdot, \cdot)$ is the p-Wasserstein distance.

Analytical solutions

For $d = 1$ we have

$$\boxed{W(x) = \frac{1}{\mu(x)} \int_{t=x}^{\infty} (t - \mathbb{E}[X]) \mathrm{d}\mu(t)}$$

[Saumard 2019] "Weighted Poincaré inequalities, concentration inequalities and tail [...]", Bernoulli.
[Song et al. 2019] "Derivative-based new upper bound of Sobol' sensitivity measure", RESS.
[Ernst et al. 2020] "First-order covariance inequalities via Stein's method", Bernoulli.
[Germain Swan 2023] "A note on one-dim Poincaré inequalities by Stein-type integration", Bernoulli.
[Heredia Joulin Roustant 2025] "On one-dim weighted Poincaré inequalities for GSA", J. Math. Anal.

For product measures $\mu(x) = \mu_1(x_1) \dots \mu_d(x_d)$ we have

$$\boxed{W(x) = \begin{pmatrix} W_1(x_1) & & 0 \\ & \ddots & \\ 0 & & W_d(x_d) \end{pmatrix}}$$

- Gaussian $\mu(x) \propto \prod_{i=1}^{d} \exp(-\frac{|x_i|^2}{2})$

$$W(x) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

- Uniform $\mu(x) \propto 1_{[0,1]^d}(x)$

$$W(x) = \begin{pmatrix} \frac{x_1(1-x_1)}{2} & & 0 \\ & \ddots & \\ 0 & & \frac{x_d(1-x_d)}{2} \end{pmatrix}$$

- Laplace $\mu(x) \propto \prod_{i=1}^{d} \exp(-|x_i|)$

$$W(x) = \begin{pmatrix} 1+|x_1| & & 0 \\ & \ddots & \\ 0 & & 1+|x_d| \end{pmatrix}$$

- Cauchy $\mu(x) \propto \prod_{i=1}^{d} (1+|x_i|^2)^{-\beta}$, $\beta > 1$

$$W(x) = \begin{pmatrix} \frac{1+x_1^2}{2(\beta-1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1+x_d^2}{2(\beta-1)} \end{pmatrix}$$

Analytical solutions

In the homogenized limit ($d \geq 1$) [Lelièvre et al. 2024]

$$\boxed{W(x) = \frac{1}{\mu(x)} \; M}$$

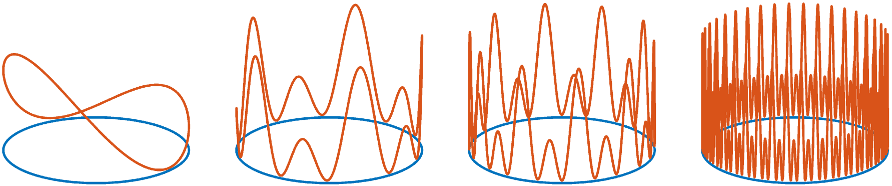
solves

$$\min_{\substack{W : x \to W(x) \succeq 0 \\ \int W \mathrm{d}\mu = M}} \; \lim_{k \to \infty} C(\mu_{\#,k}, W_{\#,k})$$

where

$$\mu_{\#,k}(x) = \mu(kx)$$
$$W_{\#,k}(x) = W(kx)$$

are the "k-periodized" μ and W defined on the torus $\mathbb{T}^d = (\mathbb{R} \backslash \mathbb{Z})^d$.



$k = 1$ $\quad$ $k = 5$ $\quad$ $k = 10$ $\quad$ $k = 40$

Numerical solution for $d \geq 2$

Convex analysis

$$\mathrm{Var}_\mu(f) \leq C(\mu, W) \int \nabla f^\top W \nabla f \mathrm{d}\mu$$

The problem is equivalent to **maximizing**

$$\boxed{\frac{1}{C(\mu, W)} = \inf_{\substack{f \in H^1(\mu, W) \\ \int f \mathrm{d}\mu = 0}} \frac{\int \nabla f^\top W \nabla f \mathrm{d}\mu}{\int f^2 \mathrm{d}\mu}}$$

where

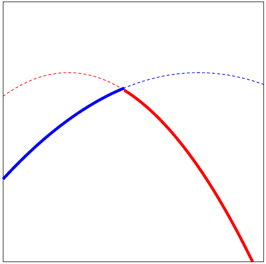
$$H^1(\mu, W) = \left\{ f : \mathbb{R}^d \to \mathbb{R} : \|f\|^2_{H^1(\mu, W)} = \int f^2 \mathrm{d}\mu + \int \nabla f^\top W \nabla f \mathrm{d}\mu < \infty \right\}.$$

Proposition [Cui, Tong and Zahm 2024]

The function $W \mapsto C(\mu, W)^{-1}$ is **concave**, therefore it admits a superdifferential $\partial_W (C(\mu, W)^{-1})$.

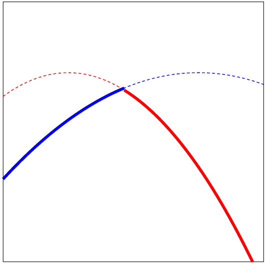
Superdifferential of concave function

A concave function,

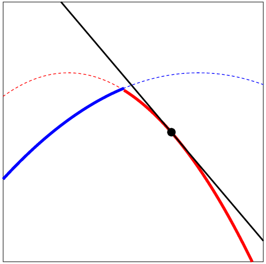


Superdifferential of concave function

A concave function,

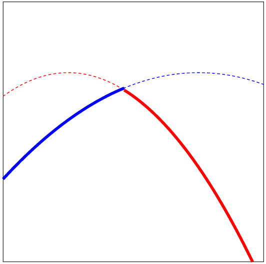


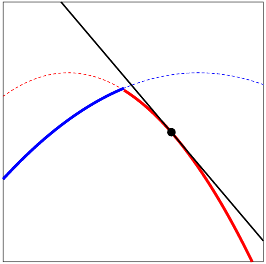
its gradient,

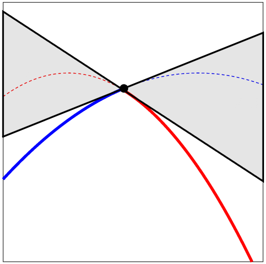


Superdifferential of concave function

A concave function, its gradient, and its superdifferential.







Spectral gap of a diffusion operator $\mathcal{L}_\mu^W$

Let $\mathcal{L}_\mu^W$ be the linear operator on $H^1(\mu, W)$ defined by

$$\boxed{\int f \mathcal{L}_\mu^W g \mathrm{d}\mu = -\int \nabla f^\top W \nabla g \mathrm{d}\mu}$$

and consider its **eigendecomposition** $-\mathcal{L}_\mu^W u_i^W = \lambda_i^W u_i^W$ with

$$0 = \lambda_0^W \leq \lambda_1^W \leq \lambda_2^W \leq \dots$$

Then $\boxed{C(\mu, W) = 1/\lambda_1^W}$ and

$$\partial_W (C(\mu, W)^{-1}) = \overline{\mathrm{conv}} \left\{ \frac{(\nabla u_i^W)(\nabla u_i^W)^\top}{\int (u_i^W)^2 \mathrm{d}\mu} : i \text{ such that } \lambda_i^W = \lambda_1^W \right\}$$

Spectral gap of a diffusion operator $\mathcal{L}_\mu^W$

Let $\mathcal{L}_\mu^W$ be the linear operator on $H^1(\mu, W)$ defined by

$$\boxed{\int f \mathcal{L}_\mu^W g \mathrm{d}\mu = -\int \nabla f^\top W \nabla g \mathrm{d}\mu}$$

and consider its **eigendecomposition** $-\mathcal{L}_\mu^W u_i^W = \lambda_i^W u_i^W$ with

$$0 = \lambda_0^W \leq \lambda_1^W \leq \lambda_2^W \leq \dots$$

Then $\boxed{C(\mu, W) = 1/\lambda_1^W}$ and

$$\partial_W (C(\mu, W)^{-1}) = \overline{\mathrm{conv}} \left\{ \frac{(\nabla u_i^W)(\nabla u_i^W)^\top}{\int (u_i^W)^2 \mathrm{d}\mu} : i \text{ such that } \lambda_i^W = \lambda_1^W \right\}$$

Takeaway message

- Evaluating the superdifferential $\partial_W (C(\mu, W)^{-1})$ requires **computing the smallest eigenvectors** of $\mathcal{L}_\mu^W$
- For the **optimal** W, we have

$$1 = \lambda_1^W = \ldots = \lambda_d^W \leq \lambda_{d+1}^W \leq \dots$$

- ... and the eigenvectors $u_1^W, \ldots, u_d^W$ are **affine functions**.

Numerical solution

- Ensure the **positivity** and the **scaling** of W via

$$W(x) = V(x)^2 \frac{\mathrm{trace}(\mathrm{Cov}_\mu)}{\int \mathrm{trace}(V^2)\mathrm{d}\mu}$$

where $V(x)$ is a field of symmetric matrices (best parametrization amongst the ones we tried...).

Numerical solution

- Ensure the **positivity** and the **scaling** of W via

$$W(x) = V(x)^2 \frac{\mathrm{trace}(\mathrm{Cov}_\mu)}{\int \mathrm{trace}(V^2)\mathrm{d}\mu}$$

  where $V(x)$ is a field of symmetric matrices (best parametrization amongst the ones we tried...).

- Use the **finite element** method to compute

$$J(V) := \frac{1}{C(\mu, W)} = \inf_{\substack{f \in H^1(\mu, W) \\ \int f \mathrm{d}\mu = 0}} \frac{\int \nabla f^\top W \nabla f \mathrm{d}\mu}{\int f^2 \mathrm{d}\mu}$$

  and its gradients $\nabla J(V)$, which can be expressed using u_1^W.

  - f: piecewise affine function (finite element space)
  - V: piecewise constant matrix field

Numerical solution

- Ensure the **positivity** and the **scaling** of W via

$$W(x) = V(x)^2 \frac{\mathrm{trace}(\mathrm{Cov}_\mu)}{\int \mathrm{trace}(V^2)\mathrm{d}\mu}$$

  where $V(x)$ is a field of symmetric matrices (best parametrization amongst the ones we tried...).

- Use the **finite element** method to compute

$$J(V) := \frac{1}{C(\mu, W)} = \inf_{\substack{f \in H^1(\mu, W) \\ \int f \mathrm{d}\mu = 0}} \frac{\int \nabla f^\top W \nabla f \mathrm{d}\mu}{\int f^2 \mathrm{d}\mu}$$

  and its gradients $\nabla J(V)$, which can be expressed using u_1^W.

  - f: piecewise affine function (finite element space)
  - V: piecewise constant matrix field

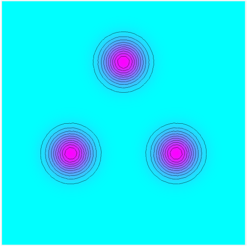
- **Gradient ascent**, constant step size $\rho > 0$

$$V^{(k+1/2)} = V^{(k)} + \rho \nabla J(V^{(k)})$$

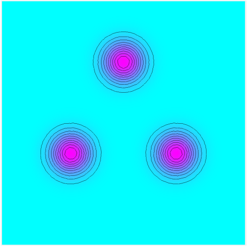
$$V^{(k+1)} = \frac{V^{(k+1/2)}}{(\int \| V^{(k+1/2)} \|_F^2 \mathrm{d}\mu)^{1/2}}$$

  Variants: Adaptive step size, **Nesterov's Accelerated Gradient (NAG)**,

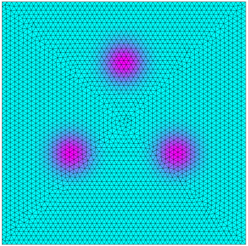
μ



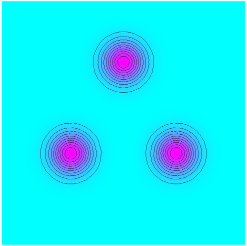
μ



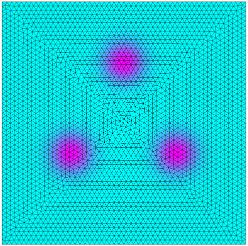
FE discretization



μ



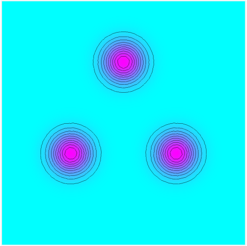
FE discretization



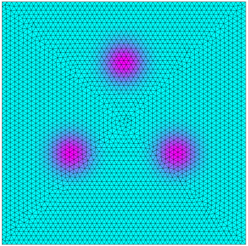
Assemble the diffusion operator

$$\mathcal{L}_\mu^W \equiv \text{''}\begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & & \\ & & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix}\text{''}$$

μ



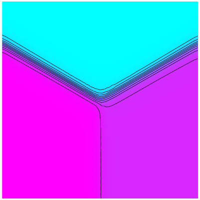
FE discretization



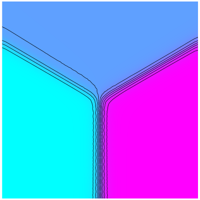
Assemble the diffusion operator

$$\mathcal{L}_\mu^W \equiv \text{''}\begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & & \\ & & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix}\text{''}$$

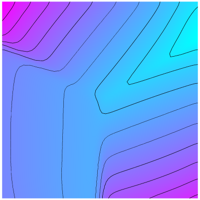
First 4 modes of $\mathcal{L}_\mu^W$ with $W(x) \propto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$



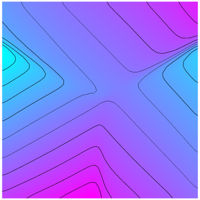
$\lambda_1 = 0.0162$



$\lambda_2 = 0.0162$

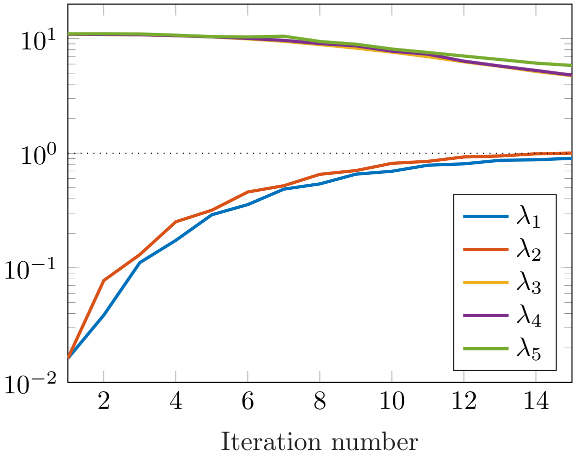


$\lambda_3 = 10.9513$

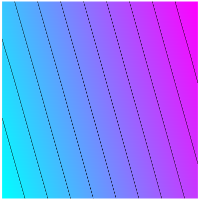


$\lambda_4 = 10.9517$

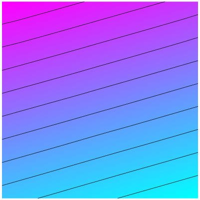
Gradient ascent (constant step size)



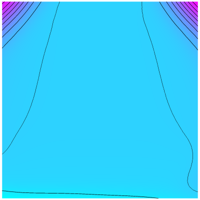
Eigenvectors of $\mathcal{L}_\mu^W$ (after 50 iterations using NAG)



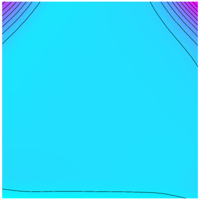
$\lambda_1 = 0.9989$



$\lambda_2 = 1.0009$



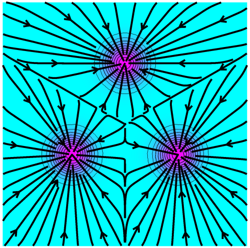
$\lambda_3 = 3.1042$



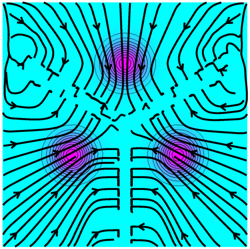
$\lambda_4 = 3.1081$

Drift and **diffusion** of Langevin dynamic

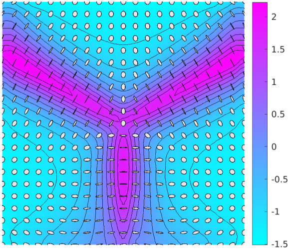
$$\mathrm{d}X_t = (\mathrm{div}(W) - W\nabla V)\mathrm{d}t + \sqrt{2W}\mathrm{d}B_t$$



∇V



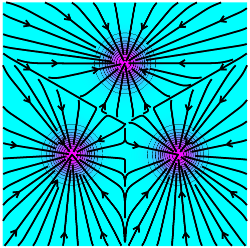
$(\mathrm{div}(W) - W\nabla V)$



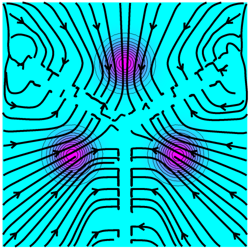
log trace$(\sqrt{2W})$

Drift and **diffusion** of Langevin dynamic

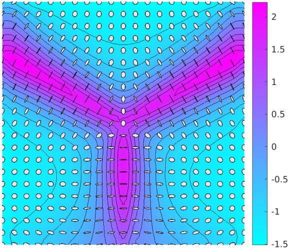
$$\mathrm{d}X_t = (\mathrm{div}(W) - W\nabla V)\mathrm{d}t + \sqrt{2W}\mathrm{d}B_t$$



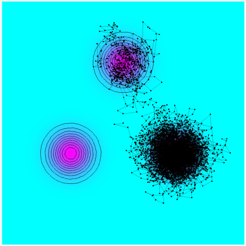
∇V



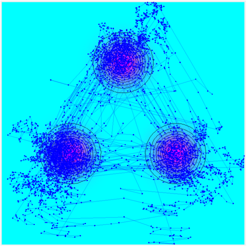
$(\mathrm{div}(W) - W\nabla V)$



log trace$(\sqrt{2W})$

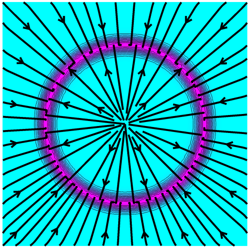


ULA

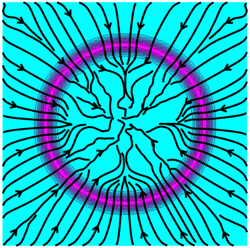


Preconditioned ULA

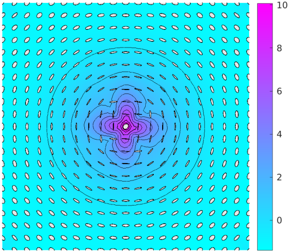
Another example: the ring



∇V

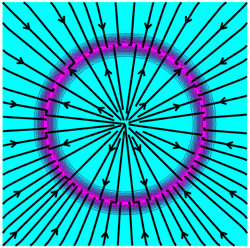


$\mathrm{div}(W) - W \nabla V$

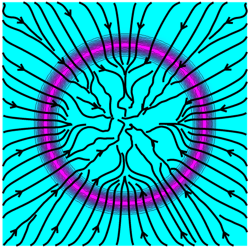


$\log \mathrm{trace}(\sqrt{2W})$

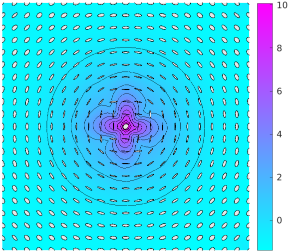
Another example: the ring



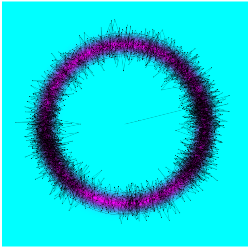
∇V



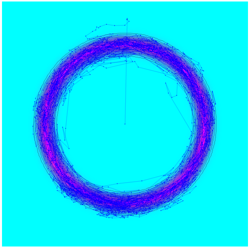
$\mathrm{div}(W) - W\nabla V$



$\log \mathrm{trace}(\sqrt{2W})$

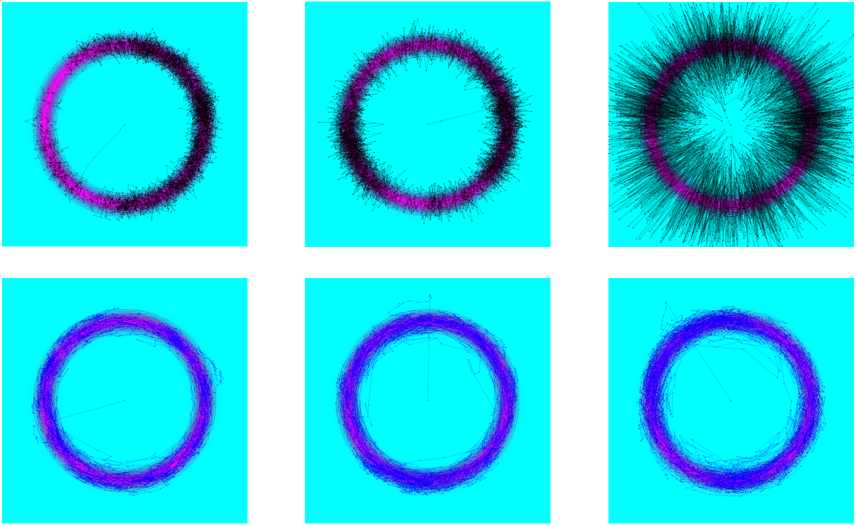


ULA



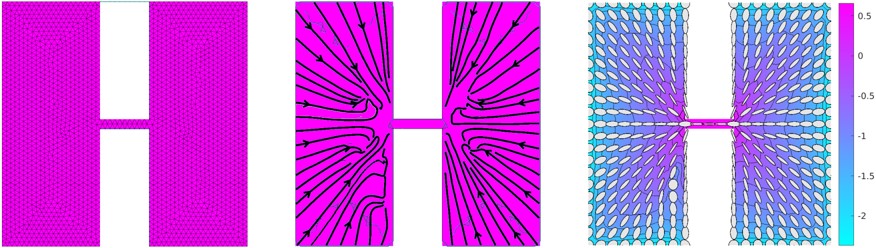
Preconditioned ULA

Impact of Δt on ULA (**top**) and precond. ULA (**bottom**)



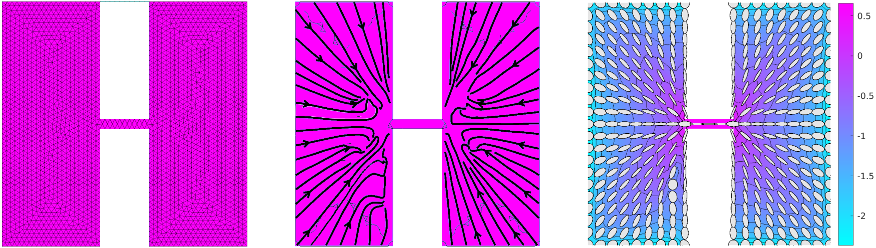
$\Delta t = 0.005$ $\qquad$ $\Delta t = 0.01$ $\qquad$ $\Delta t = 0.015$

Final example: **uniform measure** on a H-shaped domain



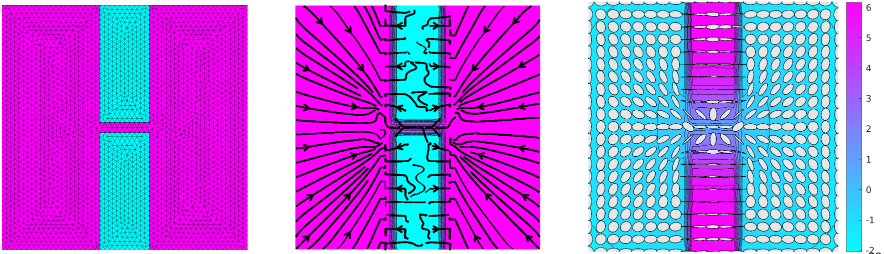
But $\boxed{C(\mu, W) = 1.09 \neq 1}$ ...

Final example: **uniform measure** on a H-shaped domain



But $\boxed{C(\mu, W) = 1.09 \neq 1}$ ...

By **convexifying** the support of $\mu \leftarrow \mu + 10^{-8} 1_{\Omega}$, we get $\boxed{C(\mu, W) = 1.001}$!



Conclusion

$$\boxed{\mathrm{Var}_\mu(f) \leq C(\mu, W) \int \nabla f(x)^\top W(x) \nabla f(x) \mathrm{d}\mu(x)}$$

Optimal Riemannian Poincaré inequalities:

- Preconditioning Langevin dynamic (space-dependent & anisotropic Δt)
- Existence of optimal W via moment maps
- Convex (but not smooth) optimization problem

Perspectives

- Uniqueness of W?
- Poincaré $\rightarrow$ Logarithmic Sobolev
- Choice of Δt for precond-ULA? for precond-MALA?
- For $d > 2$: abandon the finite elements...
- Learn W from samples (online learning during ULA/MCMC)

References

[Cui, Tong, Zahm 2024] **Optimal Riemannian metric for Poincaré inequalities and how to ideally precondition Langevin dymanics**, preprint.

[Lelièvre, Pavliotis, Robin, Santet and Stoltz 2024] **Optimizing the diffusion of overdamped Langevin dynamics**, preprint.