Poincaré Constant estimation

Loucas Pillaud-Vivien

Joint work with Alessandro Rudi and Francis Bach
In collaboration with G.Stolz and T.Lelièvre CERMICS/INRIA

January 14, 2020

Outline

A historical perspective

Work of H.Poincaré: PDE for mathematical physics Fourier eigenvalue problem for the heat equation:

$$\begin{aligned} \Delta U + kU &= 0 \qquad \Omega \subset \mathbb{R}^3, \quad \Omega \text{ bounded domain} \\ \frac{\partial U}{\partial n} &= 0 \qquad \partial\Omega \end{aligned}$$

A historical perspective

Spectral problem: for $j \geqslant 1$,

$$\Delta u_j + k_j u_j = 0$$
$$\frac{\partial u_j}{\partial n} = 0,$$

and he showed that k_j admitted another characterization:

$$k_j = \frac{\int u_j(-\Delta u_j)dx}{\int u_j^2 dx} = \frac{\int \|\nabla u_j\|^2 dx}{\int u_j^2 dx},$$

$k_2 \geqslant \kappa(\Omega) > 0$ and $k_j \longrightarrow +\infty$.

Sur les Equations aux Dérivées Partielles de la Physique Mathématique.

PAR H. POINCARÉ.

Quand on envisage les divers problèmes de Calcul Intégral qui se posent naturellement lorsqu'on veut approfondir les parties les plus différentes de la Physique, il est impossible de n'être pas frappé des analogies que tous ces problèmes présentent entre eux. Qu'il s'agisse de l'électricité statique ou dynamique, de la propagation de la chaleur, de l'optique, de l'élasticité, de l'hydrodynamique, on est toujours conduit à des équations différentielles de même famille et les conditions aux limites, quoique différentes, ne sont pas pourtant sans offrir quelques resemblances. Nous ne citerons ici que quelques exemples.

J'imagine d'abord que l'on se propose de trouver la température finale d'un corps solide conducteur, homogène et isotrope, lorsque les divers points de la surface de ce corps sont maintenus artificiellement à des températures données.

Ce problème traduit dans le langage analytique s'énonce comme il suit:

Trouver une fonction V qui dans une portion de l'espace satisfasse à l'équation de Laplace,

$$\Delta V = \frac{d^2 V}{dx^2} + \frac{d^2 V}{dy^2} + \frac{d^2 V}{dz^2} = 0,$$

Rappelons d'abord la définition de k_2; k_2 est le minimum du rapport

$$\frac{\int\left[\left(\frac{dV}{dx}\right)^2+\left(\frac{dV}{dy}\right)^2+\left(\frac{dV}{dz}\right)^2\right]d\tau}{\int V^2 d\tau}$$

quand la fonction V et assujettie à la condition :

$$\int V d\tau = 0. \qquad (10)$$

and after come calculations...

Par conséquent *pour un solide convexe quelconque on a* :

$$k_2 > \frac{6K_0 W}{\pi\lambda^5},$$

K_0 désignant une constante numérique, W le volume du corps, et λ la plus grande distance de deux points de la surface du corps.

Poincaré inequality for bounded open convex set in $\mathbb{R}^n$

Theorem (H.Poincaré 1890)

For Ω open bounded convex set of $\mathbb{R}^d$, f smooth from $\bar{\Omega}$ to $\mathbb{R}$ such that, $\int_\Omega f dx = 0$,

$$\int_\Omega f^2 dx \leqslant \mathcal{P} \int_\Omega \|\nabla f\|^2 dx,$$

where $\mathcal{P} = \mathcal{P}(\Omega) \leqslant K_d \, \mathrm{Diam}(\Omega)^2 < +\infty$.

(Has been optimized in the 60's $\longrightarrow \mathcal{P}(\Omega) = \mathrm{Diam}(\Omega)^2/\pi^2$.)

Poincaré inequalities: definition in modern language

Definition (Poincaré inequality)

$\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies a Poincaré Inequality with constant $\mathcal{P}$ if

$$\mathrm{Var}_\mu(f) \leqslant \mathcal{P}_\mu \int \|\nabla f\|^2 d\mu,$$

for all (bounded) $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ of class $\mathcal{C}^1$.

Recall that :

- $\mathrm{Var}_\mu(f) = \int f^2 d\mu - \left(\int f d\mu\right)^2 = \int \left(f - \int f d\mu\right)^2 d\mu$
- $\int \|\nabla f\|^2 d\mu = \mathcal{E}(f)$ is the *Dirichlet Energy*.

Spectral interpretation: $\mathcal{E}(f) = \int \nabla f \cdot \nabla f d\mu = \int f(-\mathcal{L} f) d\mu$

$\longrightarrow 1/\mathcal{P} = \lambda_2$, first non-trivial eigenvalue of $\mathcal{L}$.

Poincaré inequalities : Basic examples 1

Poincaré-Wirtinger: f : $\bar{\Omega} = [0,1] \longrightarrow \mathbb{R}$, smooth and periodic : $f(0) = f(1)$ and $\int_0^1 f dx = 0$, then

$$\int_0^1 f^2 dx \leqslant \frac{1}{4\pi^2} \int_0^1 f'^2 dx,$$

Proof: $f(x) = \sum_{m \geqslant 1} a_m \cos(2\pi mx) + b_m \sin(2\pi mx)$, $a_0 = \int_0^1 f dx = 0$.

$$\frac{1}{2\pi} \int_0^1 f^2 dx = \sum_{m \geqslant 1} a_m^2 + b_m^2$$

$$\frac{1}{2\pi} \int_0^1 f'^2 dx = \sum_{m \geqslant 1} 4\pi^2 m(a_m^2 + b_m^2)$$

Poincaré inequalities : Basic examples 2

Poincaré-Wirtinger spherical version: σ uniform measure on the sphere of dimension d. f : $\mathbb{S}^d \longrightarrow \mathbb{R}$, smooth such that $\int_{\mathbb{S}^d} f d\sigma = 0$, then

$$\int_{\mathbb{S}^d} f^2 d\sigma \leqslant \frac{1}{d} \int_{\mathbb{S}^d} \|\nabla f\|^2 d\sigma,$$

proof by expansion in spherical harmonics.

Poincaré for gaussian: μ gaussian probability measure: $d\mu(x) : \frac{1}{(2\pi)^{d/2}} e^{-\|x\|^2/2} dx$ f : $\mathbb{R}^d \longrightarrow \mathbb{R}$, smooth such that $\int_{\mathbb{R}^n} f d\mu = 0$, then

$$\int_{\mathbb{R}^d} f^2 d\mu \leqslant \int_{\mathbb{R}^d} \|\nabla f\|^2 d\mu,$$

proof by expansion in Hermite polynomials.

Ok, fine. But what are the applications of such inequalities?

Bounding the variance

Poincaré's inequalities are a powerful tool to **bound the variance**.
Assume that X a random vector distributed accorded to μ s.t.:

1. μ satisfies a Poincaré inequality $\mathrm{Var}(f(X)) \leqslant \mathcal{P}\mathbb{E}\|\nabla f(X)\|^2$
2. f is L-lipschitz

Then,

$$\mathrm{Var}(f(X)) \leqslant \mathcal{P}L.$$

Applications:

- X standard Gaussian random vector, f Lipschitz, then: $\mathrm{Var}(f(X)) \leqslant 1$
- The variance of the largest singular value, $\sigma(A)$, of a random matrix A taking values in $[0, 1]$ is bounded by 1.

Convergence to equilibrium for Diffusions

Let us consider the overdamped Langevin diffusion in $\mathbb{R}^d$:

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

- **Stationnary measure**: $d\mu(x) = e^{-V(x)}dx$.
- **Semi-group**: $P_t(f)(x) = \mathbb{E}[f(X_t)|X_0 = x] \longrightarrow$ "law of X_t".
- **Infinitesimal generator**: $\mathcal{L}\phi = \Delta\phi - \nabla V \cdot \nabla\phi$.

We can verify that the law of X_t follows the dynamics:

$$\frac{d}{dt}P_t(f) = \mathcal{L}P_t(f).$$

Convergence to equilibrium for Diffusions

Theorem (Poincaré implies convergence to equilibrium)

With the notations above, the following propositions are equivalent:

- μ *satisfies a Poincaré Inequality with constant* $\mathcal{P}$
- *For all* f *smooth,* $\mathrm{Var}_\mu(P_t(f)) \leqslant e^{-2t/\mathcal{P}} \mathrm{Var}_\mu(f)$ *for all* $t \geqslant 0$.

Proof: Integration by part formula (μ is reversible),

$$-\int f(\mathcal{L}g)\, d\mu = \int \nabla f \cdot \nabla g \, d\mu = -\int (\mathcal{L}f) g \, d\mu, \quad \text{hence,}$$

$$\begin{aligned}\frac{d}{dt}\mathrm{Var}_\mu(P_t(f)) = \frac{d}{dt}\int (P_t(f))^2 d\mu &= 2\int P_t(f)(\mathcal{L}P_t(f)) d\mu \\ &= -2\int \|\nabla P_t(f)\|^2 d\mu \\ &\leqslant -2/\mathcal{P}\ \mathrm{Var}_\mu(P_t(f))\end{aligned}$$

Application to the Ornstein-Uhlenbeck process

The diffusion of the **Ornstein-Uhlenbeck process** follows the SDE in $\mathbb{R}^d$:

$$dX_t = -X_t dt + \sqrt{2} dB_t,$$

Denote $\mathcal{L}$ the operator $\mathcal{L}\phi = \Delta\phi - x \cdot \nabla\phi$, then

1. For $d\mu(x) = \frac{1}{(2\pi)^{d/2}} e^{-\|x\|^2/2} dx$, $\mathcal{L}$ is self adjoint in L^2_μ
2. μ stationnary measure of O-U process
3. μ verifies Poincaré inequality with constant 1.
4. for all f smooth, for all $t \geqslant 0$.

$$\mathrm{Var}_\mu(P_t(f)) \leqslant e^{-2t}\mathrm{Var}_\mu(f).$$

Poincaré implies Concentration of measure

Definition (Concentration of measure)

One says that μ satisfies the concentration of measure property if for any set A such that $\mu(A) \geqslant 1/2$, we have:

$$\mu(A_r) \geqslant 1 - \exp(-r^2/2),$$

with $A_r = \{x \in \mathbb{R}^d \mid \mathrm{dist}(x, A) \leqslant r\}$.

Theorem (Poincaré implies Concentration of measure)

Assume μ satisfies a Poincaré Inequality with constant $\mathcal{P}$, then for any set A such that $\mu(A) \geqslant 1/2$, we have:

$$\mu(A_r) \geqslant 1 - \exp\left(-\frac{r}{2\sqrt{\mathcal{P}}}\right),$$

Statistical estimation of the Poincaré Constant

Setting of the problem

Let X a random vector of $\mathbb{R}^d$ be distributed according to the probability measure μ that satisfies the following Poincaré :

$$\mathrm{Var}_\mu\left(f(X)\right) \leqslant \mathcal{P}_\mu \, \mathbb{E}_\mu\left[\, \|\nabla f(X)\|^2 \right]$$

Goal: given $(x_1, \ldots, x_n)$ n i.i.d samples of the probability measure of $d\mu$, our goal is to estimate $\mathcal{P}_\mu$.

Two steps approach:

1. Construct a estimator $\hat{\mathcal{P}}_\mu^n$
2. Prove its statistical consistency : $\hat{\mathcal{P}}_\mu^n \underset{n\to\infty}{\longrightarrow} \mathcal{P}_\mu$

Reformulation of the problem in a RKHS

Let $\mathcal{F}$ be a dense RKHS in $\mathcal{C}^1$ associated with kernel K, then:

1. $\mathcal{F} = \overline{\mathrm{span}}\{K(\cdot, x),\ x \in \mathbb{R}^d\}$, and in particular $y \longrightarrow K(y, x) \in \mathcal{F}$ that we will note K_x.
2. **Reproducing property**: $\forall f \in \mathcal{F}$ and $\forall x \in \mathbb{R}^d$, $f(x) = \langle f, K(\cdot, x)\rangle_{\mathcal{F}}$. In other words, functions evaluations are equal to dot products with canonical elements of the RKHS.
3. **Derivation** corresponds to the following: $\forall f \in \mathcal{F}$ and $\forall x \in \mathbb{R}^d$, $\partial_j f(x) = \langle f, \partial_j K(\cdot, x)\rangle_{\mathcal{F}}$.

Example: gaussian kernel: $K_\sigma(x, y) = \exp\left(\frac{\|x-y\|^2}{2\sigma^2}\right)$.

Operators of the problem

Let us define the following positive semi-definite operators:

- The operators from $\mathcal{F}$ to $\mathcal{F}$, Σ and $\hat{\Sigma}$ respectively the covariance and the empirical covariance operators,

$$\Sigma = \mathbb{E}\left[K_x \otimes K_x\right] = \int_{\mathbb{R}^d} K_x \otimes K_x d\mu(x),$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} K_{x_i} \otimes K_{x_i}$$

- The operators from $\mathcal{F}$ to $\mathcal{F}$, Δ and $\hat{\Delta}$,

$$\Delta = \mathbb{E}\left[\nabla K_x \otimes \nabla K_x\right] = \int_{\mathbb{R}^d} \nabla K_x \otimes \nabla K_x d\mu(x),$$

$$\hat{\Delta} = \frac{1}{n} \sum_{i=1}^{n} \nabla K_{x_i} \otimes \nabla K_{x_i}$$

Spectral characterization of the Poincaré constant

Back to Poincaré's work:

Lemma (Spectral characterization of the Poincaré constant)

Let $\mathcal{P}_\mu$ be the Poincaré constant of $\mathcal{F} \subset H^1(\mathbb{R}^d, d\mu)$, then $\mathcal{P}_\mu$ is the maximum of the following Rayleigh ratio:

$$\mathcal{P}_\mu = \sup_{f \in \mathcal{F}} \frac{\langle f, Cf \rangle}{\langle f, \Delta f \rangle} = \left\| \Delta^{-1/2} C \Delta^{-1/2} \right\|,$$

where $\| \cdot \|$ is the operator norm, $C = \Sigma - m \otimes m$ and $m = \int_{\mathbb{R}^d} K_x d\mu(x) \in \mathcal{F}$.

Spectral characterization of the estimator of Poincaré constant

Definition (of the estimator of Poincaré constant)

$$\hat{\mathcal{P}}_\mu = \mathcal{P}_{\hat{\mu}_n} = \sup_{f \in \mathcal{F}} \frac{\langle f, \hat{C} f \rangle}{\langle f, (\hat{\Delta} + \lambda I) f \rangle} = \left\| \hat{\Delta}_\lambda^{-1/2} \hat{C} \hat{\Delta}_\lambda^{-1/2} \right\|,$$

where $\hat{C} = \hat{\Sigma} - \hat{m} \otimes \hat{m}$, $\hat{m} = \frac{1}{n} \sum_{i=1}^n K_{x_i}$ and $\hat{\Delta}_\lambda = \hat{\Delta} + \lambda I$ is a regularized empirical version of the operator Δ.

Two remarks :

1. We **need to regularize** because the kernel of $\hat{\Delta}$ is no longer strictly included in the kernel of $\hat{\Sigma}$.
2. We only need the sup over a **finite dimensionnal space**: $\mathrm{Im}(\hat{C}) \oplus \mathrm{Im}(\hat{\Delta})$. $f = \sum_{i=1}^n \alpha_i K_{x_i} + \beta_i \cdot \nabla K_{x_i}$.

Statisical consistency: Bias - Variance decomposition

Let us denote $\mathcal{P}_\mu^\lambda$ the Poincaré constant of the regularized problem $\mathcal{P}_\mu^\lambda = \sup_{f\in\mathcal{F}} \frac{\langle f, Cf \rangle}{\langle f, (\Delta+\lambda I) f\rangle} = \left\| \Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2} \right\|$, then

$$|\hat{\mathcal{P}}_\mu - \mathcal{P}_\mu| \leqslant \underbrace{|\mathcal{P}_\mu^\lambda - \mathcal{P}_\mu|}_{\text{Bias}} + \underbrace{|\hat{\mathcal{P}}_\mu - \mathcal{P}_\mu^\lambda|}_{\text{Variance}}$$

- **Variance:** $\|\hat{\Delta}_\lambda^{-1/2} \hat{C} \hat{\Delta}_\lambda^{-1/2}\| \underset{n\to\infty}{\longrightarrow} \|\Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2}\|$. Relies strongly on concentration of operators $\hat{C}$ to C and $\hat{\Delta}$ to Δ.
- **Bias:** $\|\Delta_\lambda^{-1/2} C \Delta_\lambda^{-1/2}\| \underset{\lambda\to 0}{\longrightarrow} \|\Delta^{-1/2} C \Delta^{-1/2}\|$. Need one assumption to make it converge.

Statisical consistency: Bias - Variance decomposition

- **Variance Bound proof.** Relies strongly on concentration of operators $\hat{C}$ to C and $\hat{\Delta}$ to Δ.

Lemma (Control of the variance)

Let $\delta \in (0,1)$, $\lambda \gtrsim \lambda_0/n$, we have the following inequality with probability $1-\delta$:

$$\left|\hat{\mathcal{P}}_\mu - \mathcal{P}_\mu^\lambda\right| \lesssim \frac{\mathcal{P}_\mu^\lambda}{\sqrt{\lambda n}} \log(1/\delta).$$

- **Bias Bound proof.** Let us assume that $C \preccurlyeq \kappa\Delta^2$ (slightly stronger that Poincaré Inequality), then

$$|\mathcal{P}_\mu^\lambda - \mathcal{P}_\mu| \leqslant \kappa\lambda.$$

Statisical consistency

$$|\hat{\mathcal{P}}_\mu - \mathcal{P}_\mu| \leqslant \underbrace{|\mathcal{P}_\mu^\lambda - \mathcal{P}_\mu|}_{\text{Bias}} + \underbrace{|\hat{\mathcal{P}}_\mu - \mathcal{P}_\mu^\lambda|}_{\text{Variance}}$$
$$\lesssim \kappa\lambda + \frac{\mathcal{P}_\mu^\lambda}{\sqrt{\lambda n}} \log(1/\delta).$$

Theorem (Statistical consistency of the estimator)

For $\lambda = 1/n^{1/3}$ and fix $\delta \in (0,1)$, then $\hat{\mathcal{P}}_\mu$ is a consistent estimator of $\mathcal{P}_\mu$ and we have with probability $1-\delta$:

$$|\hat{\mathcal{P}}_\mu - \mathcal{P}_\mu| \lesssim \frac{\mathcal{P}_\mu}{n^{1/3}} \log(1/\delta)$$

Ok... but why do we want to estimate such constants?

Statistical estimation of the Poincaré Constant

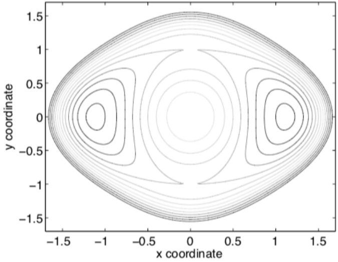
Back to the initial problem of Molecular Dynamics

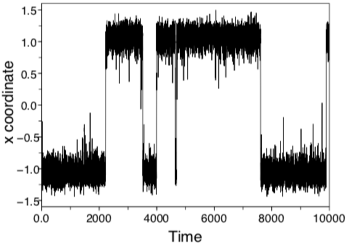
Take $X_t \in \mathbb{R}^d$ a random vector describing a molecule. Its dynamics is described by the overdamped Langevin diffusion

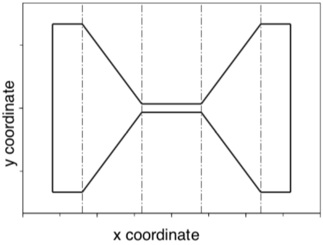
$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dB_t.$$

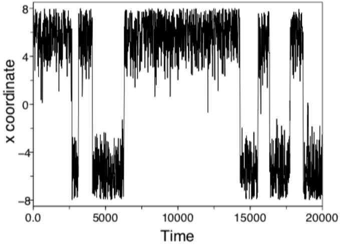
- **Goal:** Sample $(X_t)_{t\geqslant 0}$
- **Invariant measure:** $\mu(dx) = Z^{-1}\exp(-\beta V(x))dx$
- **Problem:** **Metastability**

Metastability









Escaping for metastability: learning a reaction coordinate

Goal: Find the path $\xi : \mathbb{R}^d \longrightarrow \mathbb{R}^r$ (with $r \ll d$) of the metastability, ξ is called the reaction coordinate.

" the metastability of the process is along ξ

$\Longleftrightarrow$

the measures $\mu(\cdot|\xi(x) = z)$ satisfies a Poincaré inequality with a little Poincaré constant "

Escaping for metastability: learning a reaction coordinate

" the metastability of the process is along ξ

$\Longleftrightarrow$

the measures $\mu(\cdot|\xi(x) = z)$ satisfies a Poincaré inequality with a little Poincaré constant "

Program:

1. given $x_1, \ldots, x_n$ i.i.d according to μ, estimate $\mu(\cdot|\xi(x) = z)$
2. using Part II estimate the Poincaré constant of $\mu(\cdot|\xi(x) = z)$: $\hat{\mathcal{P}}_{\mu(\cdot|\xi(x)=z)}(\xi)$
3. optimize according to ξ to get $\hat{\xi}^* = \operatorname{argmin} \hat{\mathcal{P}}_{\mu(\cdot|\xi(x)=z)}(\xi)$

Conclusion

Still a lot to do:

- Discuss with CERMICS to have more relevant hypothesis for the bias
- Subsampling techniques?
- Do some simulations!
- Technically speaking, related to kernel ICA and CCA, may be worth exploring