Fokker Planck equation, Poincare inequaltiy, and convergence of Markov Process

Jian Li

IIIS Tsinghua

ATCS 2023

Recall Fokker Planck equation

- Consider a diffusion process on $\mathbb{R}^d$ with time-independent drift and diffusion coefficients. The Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = -\sum_{j=1}^{d} \frac{\partial}{\partial x_j}(a_i(x)p) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(b_{ij}(x)p), \ t > 0, \ x \in \mathbb{R}^d,$$

$$p(x,0) = f(x), \quad x \in \mathbb{R}^d.$$

Fokker Planck equation for OU-proces: $dX_t = -\alpha X_t dt + \sqrt{2D} dW_t$

Set $a(t,x) = -\alpha x$, $b(t,x) \equiv 2D > 0$:

$$\frac{\partial p}{\partial t} = \alpha \frac{\partial (xp)}{\partial x} + D \frac{\partial^2 p}{\partial x^2}.$$

Fokker Planck equation

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

- The corresponding FP equation is:

$$\frac{\partial p}{\partial t} = \nabla \cdot (\nabla V p) + D\Delta p.$$

The stationary distribution of the above Markov process is the following Gibbs distribution:

$$p(x) = \frac{1}{Z} e^{-V(x)/D}$$

one can verify it satifies FP equation

where the normalization factor Z *is the* **partition function**

$$Z = \int_{\mathbb{R}^d} e^{-V(x)/D}\,dx.$$

A normalized version

- It is more convenient to normalize the solution of the
- Fokker-Planck equation wrt the invariant distribution

let $\rho(x)$ *be the Gibbs distribution*

Define $h(x,t)$ *through*

$$p(x,t) = h(x,t)\rho(x).$$

Then the function h *satisfies the* **backward Kolmogorov equation***:*

$$\frac{\partial h}{\partial t} = -\nabla V \cdot \nabla h + D\Delta h, \quad h(x,0) = p(x,0)\rho^{-1}(x).$$

Proof. The initial condition follows from the definition of h. We calculate the gradient and Laplacian of p:

$$\nabla p = \rho \nabla h - \rho h D^{-1} \nabla V$$

and

$$\Delta p = \rho \Delta h - 2\rho D^{-1} \nabla V \cdot \nabla h + h D^{-1} \Delta V \rho + h |\nabla V|^2 D^{-2} \rho.$$

We substitute these formulas into the FP equation to obtain

$$\rho \frac{\partial h}{\partial t} \;=\; \rho\Big(- \nabla V \cdot \nabla h + D \Delta h \Big),$$

from which the claim follows. □

The self-adjoint generator

- Consider the Hilbert space with the following inner product

$$(f, h)_\rho = \int_{\mathbb{R}^d} f h \rho(x)\, dx.$$

Proposition 3. *Assume that $V(x)$ is a smooth potential and assume that condition (7) holds. Then the operator*

$$\mathcal{L} = -\nabla V(x) \cdot \nabla + D\Delta$$

is self-adjoint in L^2_ρ. Furthermore, it is non-positive, its kernel consists of constants.

The self-adjoint generator

Proof. Let $f, \in C_0^2(\mathbb{R}^d)$. We calculate

$$
\begin{aligned}
(\mathcal{L}f, h)_\rho &= \int_{\mathbb{R}^d} (-\nabla V \cdot \nabla + D\Delta) f h \rho \, dx \\
&= \int_{\mathbb{R}^d} (\nabla V \cdot \nabla f) h \rho \, dx - D \int_{\mathbb{R}^d} \nabla f \nabla h \rho \, dx - D \int_{\mathbb{R}^d} \nabla f h \nabla \rho \, dx \\
&= -D \int_{\mathbb{R}^d} \nabla f \cdot \nabla h \rho \, dx,
\end{aligned}
$$

from which self-adjointness follows.

The self-adjoint generator

If we set $f = h$ in the above equation we get

$$(\mathcal{L}f, f)_\rho = -D\|\nabla f\|_\rho^2,$$

which shows that $\mathcal{L}$ is non-positive.

Clearly, constants are in the null space of $\mathcal{L}$. Assume that $f \in \mathcal{N}(\mathcal{L})$. Then, from the above equation we get

$$0 = -D\|\nabla f\|_\rho^2,$$

and, consequently, f is a constant. $\square$

Dirichlet Form and Poincare inequality

Remark 1. *The expression* $(-\mathcal{L}f, f)_\rho$ *is called the* **Dirichlet form** *of the operator* $\mathcal{L}$. *In the case of a gradient flow, it takes the form*

$$(-\mathcal{L}f, f)_\rho = D\|\nabla f\|_\rho^2.$$

Proposition 4. *Assume that the potential* V *satisfies the convexity condition*

$$D^2 V \geqslant \lambda I.$$

Then the corresponding Gibbs measure satisfies the Poincaré inequality with constant λ:

$$\int_{\mathbb{R}^d} f\rho = 0 \quad \Rightarrow \quad \|\nabla f\|_\rho \geqslant \sqrt{\lambda}\|f\|_\rho. \tag{11}$$

How should we understand Poincare inequality?

Poincare inequality essentially asserts that the spectral gap of self-adjoint operator L is at least λ.

$$\int_{\mathbb{R}^d} f\rho = 0 \quad \Rightarrow \quad \|\nabla f\|_\rho \geqslant \sqrt{\lambda}\|f\|_\rho.$$

Note that the first eigenvalue of L is 0 (with eigenfunction being the constant function)

Larger spectral gap implies faster convergence (to the stationary distribution). Later.

Theorem 2. *Assume that* $p(x,0) \in L^2(e^{V/D})$. *Then the solution* $p(x,t)$ *of the Fokker-Planck equation* (6) *converges to the Gibbs distribution exponentially fast:*

$$\|p(\cdot,t) - Z^{-1}e^{-V}\|_{\rho^{-1}} \leqslant e^{-\lambda D t}\|p(\cdot,0) - Z^{-1}e^{-V}\|_{\rho^{-1}}.$$

Discrete Markov Chain

It would be instructive to consider the discrete Markov chain with uniform stationary distribution (e.g., an undirected graph with uniform degree) (Here A is the transition matrix)

Let A be a symmetric $n \times n$ matrix. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigen values of A and $v_1, v_2, \ldots, v_n$ be the corresponding eigenvectors. Then

$$\lambda_1 = \max_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x} \text{ and } \lambda_2 = \max_{x \perp v_1} \frac{x^T A x}{x^T x}.$$

$$\int_{\mathbb{R}^d} f\rho = 0 \quad \Rightarrow \quad \|\nabla f\|_\rho \geqslant \sqrt{\lambda}\|f\|_\rho.$$

f is orthognal to 1 (w.r.t. inner prod $\langle , \rangle_\rho$)

recall $(\mathcal{L}f, f)_\rho = -D\|\nabla f\|_\rho^2$

So, the spectral gap of self-adjoint operator L is at least λ

Convergence of Discrete Markov Chain

Definition (mixing time)

Let π be the stationary of the chain, and $p_x^{(t)}$ be the distribution after t steps when the initial state is x.

- $\Delta_x(t) = \|p_x^{(t)} - \pi\|_{TV}$ is the distance to stationary distribution π after t steps, started at state x.
- $\Delta(t) = \max_{x\in\Omega} \Delta_x(t)$ is the maximum distance to stationary distribution π after t steps.
- $\tau_x(\epsilon) = \min\{t \mid \Delta_x(t) \leq \epsilon\}$ is the time until the total variation distance to the stationary distribution, started at the initial state x, reaches ϵ.
- $\tau(\epsilon) = \max_{x\in\Omega} \tau_x(\epsilon)$ is the time until the total variation distance to the stationary distribution, started at the worst possible initial state, reaches ϵ.

Theorem

Let P be the transition matrix for a symmetric Markov chain on state space Ω where $|\Omega| = N$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the spectrum of P and $\lambda_{\max} = \max\{|\lambda_2|, |\lambda_N|\}$. The mixing rate of the Markov chain is

$$\tau(\epsilon) \leq \frac{\frac{1}{2}\ln N + \ln\frac{1}{2\epsilon}}{1 - \lambda_{\max}}.$$

Recall that due to Perron-Frobenius theorem, $\lambda_1 = 1$. And $\mathbf{1}P = \mathbf{1}$ since P is double stochastic, thus $v_1 = \frac{\mathbf{1}}{\|\mathbf{1}\|_2} = \left(\frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}}\right)$

Source: https://tcs.nju.edu.cn/wiki/index.php?title=%E9%9A%8F%E6%9C%BA%E7%AE%97%E6%B3%95_(Fall_2015)/Expander_Graphs_and_Mixing

Proof.

As analysed above, if P is symmetric, it has orthonormal eigenvectors $v_1, \ldots, v_N$ such that any distribution q over Ω can be expressed as

$$q = \sum_{i=1}^{N} c_i v_i = \pi + \sum_{i=2}^{N} c_i v_i$$

with $c_i = q^T v_i$, and

$$qP^t = \pi + \sum_{i=2}^{N} c_i \lambda_i^t v_i.$$

When q is a distribution, i.e., q is a nonnegative vector and $\|q\|_1 = 1$, it holds that $c_1 = q^T v_1 = \frac{1}{\sqrt{N}}$ and $c_1 v_1 = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right) = \pi$, thus

$$q = \sum_{i=1}^{N} c_i v_i = \pi + \sum_{i=2}^{N} c_i v_i,$$

$$qP^t = \pi P^t + \sum_{i=2}^{N} c_i v_i P^t = \pi + \sum_{i=2}^{N} c_i \lambda_i^t v_i.$$

Thus,

$$\begin{aligned}
\|qP^t - \pi\|_1 &= \left\| \sum_{i=2}^{N} c_i \lambda_i^t v_i \right\|_1 \\
&\leq \sqrt{N} \left\| \sum_{i=2}^{N} c_i \lambda_i^t v_i \right\|_2 \qquad \text{(Cauchy-Schwarz)} \\
&= \sqrt{N} \sqrt{\sum_{i=2}^{N} c_i^2 \lambda_i^{2t}} \\
&\leq \sqrt{N} \lambda_{\max}^t \sqrt{\sum_{i=2}^{N} c_i^2} \\
&= \sqrt{N} \lambda_{\max}^t \|q\|_2 \\
&\leq \sqrt{N} \lambda_{\max}^t.
\end{aligned}$$

The last inequality is due to a universal relation $\|q\|_2 \leq \|q\|_1$ and the fact that q is a distribution.

Then for any $x \in \Omega$, denoted by $\mathbf{1}_x$ the indicator vector for x such that $\mathbf{1}_x(x) = 1$ and $\mathbf{1}_x(y) = 0$ for $y \neq x$, we have

$$\begin{aligned} \Delta_x(t) = & \quad \mathbf{1}_x P^t - \pi \quad {}_{TV} = \frac{1}{2} \quad \mathbf{1}_x P^t - \pi \quad {}_{1} \\ & \leq \frac{\sqrt{N}}{2} \lambda_{\max}^t \leq \frac{\sqrt{N}}{2} \mathrm{e}^{-t(1-\lambda_{\max})}. \end{aligned}$$

Therefore, we have

$$\tau_x(\epsilon) = \min\{t \mid \Delta_x(t) \leq \epsilon\} \leq \frac{\frac{1}{2} \ln N + \ln \frac{1}{2\epsilon}}{1 - \lambda_{\max}}$$

for any $x \in \Omega$, thus the bound holds for $\tau(\epsilon) = \max_x \tau_x(\epsilon)$.

□

Poincare inequality implies exponential convergence

Theorem 2. *Assume that* $p(x,0) \in L^2(e^{V/D})$. *Then the solution* $p(x,t)$ *of the Fokker-Planck equation* (6) *converges to the Gibbs distribution exponentially fast:*

$$\|p(\cdot,t) - Z^{-1}e^{-V}\|_{\rho^{-1}} \leqslant e^{-\lambda D t}\|p(\cdot,0) - Z^{-1}e^{-V}\|_{\rho^{-1}}.$$

Proof.

$$\begin{aligned} -\frac{d}{dt}\|(h-1)\|_\rho^2 &= -2\left(\frac{\partial h}{\partial t}, h-1\right)_\rho = -2\left(\mathcal{L}h, h-1\right)_\rho \\ &= \left(-\mathcal{L}(h-1), h-1\right)_\rho = 2D\|\nabla(h-1)\|_\rho \\ &\geqslant 2D\lambda\|h-1\|_\rho^2. \end{aligned}$$

$$(\mathcal{L}f, f)_\rho = -D\|\nabla f\|_\rho^2$$

$$\|\nabla f\|_\rho \geqslant \sqrt{\lambda}\|f\|_\rho$$

Our assumption on $p(\cdot,0)$ implies that $h(\cdot,0) \in L^2_\rho$. Consequently, the above calculation shows that

$$\|h(\cdot,t) - 1\|_\rho \leqslant e^{-\lambda D t}\|H(\cdot,0) - 1\|_\rho.$$

□