# Poincaré and Log-Sobolev Inequalities

Xuda Ye

October 9, 2024

In this note, we show that for a diffusion process in $\mathbb{R}^d$, how the functional inequalities are related to exponential convergence of the diffusion process. In particular, the *log-Sobolev inequality* implies the exponential decay of the KL divergence, while the *Poincaré inequality* implies the exponential decay of the $\chi^2$-divergence. The content of this paper is based on [1].

## 1 Problem setup

Let $V(x)$ be the potential function in $\mathbb{R}^d$, and we consider the diffusion process $(X_t)_{t\geqslant 0}$ defined by

$$\mathrm{d}X_t = -\nabla V(X_t) + \sqrt{2}\mathrm{d}B_t, \tag{1}$$

where $(B_t)_{t\geqslant 0}$ is the Brownian motion in $\mathbb{R}^d$. Under mild conditions on the potential function $V(x)$, the unique invariant distribution is

$$\pi(x) = \frac{1}{Z}e^{-V(x)}, \qquad Z = \int_{\mathbb{R}^d} e^{-V(x)}\mathrm{d}x.$$

The Fokker–Planck equation corresponding to the diffusion process (1) is

$$\partial_t \mu_t = \nabla \cdot (\nabla V(x)\mu_t) + \Delta \mu_t, \tag{2}$$

where $\mu_t(x)$ is the density of the distribution law of $X_t$.

To measure the difference of a distribution $\mu(x)$ from the target distribution, introduce two the KL divergence and the $\chi^2$-divergence by

$$\begin{aligned}
\mathsf{KL}(\mu||\pi) &:= \mathbb{E}_\pi\left[\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\log\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right] = \int_{\mathbb{R}^d}\mu(x)\log\frac{\mu(x)}{\pi(x)}\mathrm{d}x, \\
\chi^2(\mu||\pi) &:= \mathbb{E}_\pi\left[\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\Big)^2\right] = \int_{\mathbb{R}^d}\frac{\big(\mu(x)-\pi(x)\big)^2}{\pi(x)}\mathrm{d}x = \int_{\mathbb{R}^d}\frac{\mu^2(x)}{\pi(x)}\mathrm{d}x - 1.
\end{aligned}$$

Using the inequality

$$\mu(x)\log\frac{\mu(x)}{\pi(x)} \leqslant \mu(x)\left(\frac{\mu(x)}{\pi(x)}-1\right) = \frac{\mu^2(x)}{\pi(x)} - \mu(x),$$

we immediately obtain

$$\mathsf{KL}(\mu||\pi) = \int_{\mathbb{R}^d} \mu(x)\log\frac{\mu(x)}{\pi(x)}\mathrm{d}x \leqslant \int_{\mathbb{R}^d}\frac{\mu^2(x)}{\pi(x)} - \int_{\mathbb{R}^d}\mu(x)\mathrm{d}x = \chi^2(\mu||\pi).$$

Hence $\chi^2$-divergence is stronger than the KL divergence.

## 2 Diffusion process as Wasserstein-2 gradient flow

An important property of the diffusion process (1) is that its Fokker–Planck equation (2) can be equivalently interpreted as the gradient flow of the KL divergence in the Wasserstein-2 metric. Recall that $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$ is a complete metric space of probability distributions, where $\mathcal{P}_2(\mathbb{R}^d)$ contains the distribution $\mu$ with finite second-order moments, namely,

$$\int_{\mathbb{R}^d} |x|^2\mu(x)\mathrm{d}x < +\infty.$$

Furthermore, $\mathcal{W}_2$ is the Wasserstein-2 distance of the probaility distributions in $\mathcal{P}_2(\mathbb{R}^d)$:

$$\mathcal{W}_2(\mu,\nu) := \left(\inf_{\gamma\in\Gamma(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\gamma(\mathrm{d}x\mathrm{d}y)\right)^{\frac{1}{2}},$$

where $\Gamma(\mu,\nu)$ is the set of the joint distributions on $\mathbb{R}^d\times\mathbb{R}^d$ whose marginal distributions in the first and second components are $\mu$ and $\nu$, respectively.

For a general functional $F(\mu)$ in $\mathcal{P}_2(\mathbb{R}^d)$, the Wasserstein-2 gradient flow of the distributions $(\mu_t)_{t\geqslant 0}$ is defined by

$$\partial_t\mu_t = \nabla\cdot\left(\mu_t\nabla\frac{\partial F}{\partial\mu}(\mu_t)\right), \tag{3}$$

which is an anolgue of the continuity equation. When $F(\mu)$ is the KL divergence, we have

$$\frac{\partial F}{\partial \mu} = 1 + \log\frac{\mu(x)}{\pi(x)} = \log\mu(x) + V(x) + \text{const.} \Longrightarrow \nabla\frac{\partial F}{\partial\mu} = \frac{\nabla\mu(x)}{\mu(x)} + \nabla V(x).$$

Then we find that (3) is exactly the same with the Fokker–Planck equation (2). Note that (2) and (3) can also be written as

$$\partial_t\mu_t = \nabla\cdot\left(\mu_t\nabla\log\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right),$$

which avoids using the potential $V(x)$ explicitly. These findings above imply the decay of the $\mathsf{KL}(\mu||\pi)$ can be used to quantify the convergence rate of the diffusion process.

## 3 Decay rate of KL and $\chi^2$-divergences

When the distribution $(\mu_t)_{t\geqslant 0}$ evolves according to the Fokker–Planck equation (2) or the gradient flow (3), we aim to find the sufficient conditions of the exponential decay of $\mathsf{KL}(\mu_t||\pi)$ or $\chi^2(\mu_t||\pi)$. Direct calculation shows that

$$\partial_t \mathsf{KL}(\mu_t||\pi) = -\int_{\mathbb{R}^d} \left\|\nabla \log \frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right\|^2 \mathrm{d}\pi = -4\int_{\mathbb{R}^d} \left\|\nabla \sqrt{\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}}\right\|^2 \mathrm{d}\pi, \tag{4}$$

where the RHS is known as the Fisher information. According to (4), a sufficient condition for the exponential decay of $\mathsf{KL}(\mu_t||\pi)$ is

$$\mathrm{Ent}(g^2) := \int_{\mathbb{R}^d} g^2 \log(g^2)\mathrm{d}\pi - \left(\int_{\mathbb{R}^d} g^2 \mathrm{d}\pi\right) \log\left(\int_{\mathbb{R}^d} g^2 \mathrm{d}\pi\right) \leqslant 2C_{\mathrm{LS}} \int_{\mathbb{R}^d} \|\nabla g\|^2 \mathrm{d}\pi, \tag{5}$$

which is known as the log-Sobolev inequality, where $C_{\mathrm{LS}}$ is the log-Sobolev constant. By choosing $g = \sqrt{\mathrm{d}\mu_t/\mathrm{d}\pi}$ in (5), we find that (5) becomes

$$\mathsf{KL}(\mu_t||\pi) = \int_{\mathbb{R}^d} \mu_t(x) \log \frac{\mu_t(x)}{\pi(x)} \mathrm{d}x \leqslant 2C_{\mathrm{LS}} \int_{\mathbb{R}^d} \left\|\nabla \sqrt{\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}}\right\|^2 \mathrm{d}\pi,$$

and thus from (4) we obtain

$$\mathsf{KL}(\mu_t||\pi) \leqslant e^{-2t/C_{\mathrm{LS}}} \mathsf{KL}(\mu_0||\pi).$$

Similar results can be established for the $\chi^2$-divergence. Using the integration by parts, it is easy to compute

$$\partial_t \chi^2(\mu_t||\pi) = -2\int_{\mathbb{R}^d} \left\|\nabla \left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right)\right\|^2 \mathrm{d}\pi, \tag{6}$$

hence a sufficient condition for the exponential decay of $\chi^2(\mu_t||\pi)$ is

$$\mathrm{Var}_\pi(g) := \int_{\mathbb{R}^d} g^2 \mathrm{d}\pi - \left(\int_{\mathbb{R}^d} g \mathrm{d}\pi\right)^2 \leqslant C_{\mathrm{P}} \int_{\mathbb{R}^d} \|\nabla g\|^2 \mathrm{d}\pi, \tag{7}$$

which is known as the Poincaré inequality, where $C_{\mathrm{P}}$ is the Poincaré constant. By choosing $g = \mathrm{d}\mu_t/\mathrm{d}\pi$ in (7), we obtain

$$\chi^2(\mu_t||\pi) = \mathrm{Var}_\pi\left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right) \leqslant C_P \int_{\mathbb{R}^d} \left\|\nabla \frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right\|^2 \mathrm{d}\pi,$$

and thus from (6) we obtain

$$\chi^2(\mu_t||\pi) \leqslant e^{-2t/C_{\mathrm{P}}} \chi^2(\mu_0||\pi).$$

Now comes an interesting question. It is known that the log-Sobolev inequality is stronger than the Poincaré inequality [2], namely, $C_{\mathrm{LS}} \geqslant C_{\mathrm{P}}$. As a consequence, the convergence derived from the Poincaré inequality is always faster than the log-Sobolev inequality. Also, since $\chi^2$-divergence is stronger than the KL divergence, our arguments imply that the Poincaré inequality yields both a faster convergence rate and a stronger convergence compared to the log-Sobolev inequality. Does it mean that it is more beneficial to employ the Poincaré inequality?

Not always. One point is that LS requires a weaker requirement on the initial distribution (weaker than $L^2$). Also, the uniform-in-$N$ ergodicity of the interacting particle system can be derived using the log-Sobolev inequality rather than the Poincaré inequality [3]. The generalized $\Gamma$ calculus is also compatible with the log-Sobolev inequality [4].

## References


[1] Chen Lu. *Upper and Lower Bounds for Sampling.* PhD thesis, Massachusetts Institute of Technology, 2023.

[2] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[3] Arnaud Guillin, Wei Liu, Liming Wu, and Chaoen Zhang. Uniform poincaré and logarithmic sobolev inequalities for mean field particle systems. *The Annals of Applied Probability*, 32(3):1590–1614, 2022.

[4] Pierre Monmarché. Generalized $\Gamma$ calculus and application to interacting particles on a graph. *Potential Analysis*, 50(3):439–466, 2019.