401-3382-25L: Log-Sobolev Inequalities and Markov Semigroups

Lecture 2 – OU semigroup and Gaussian LSI

Lecturer: Yuansi Chen *Week 3-4* *Spring 2025*

Key concepts:

- Ornstein-Uhlenbeck (OU) semigroup
- Poincaré inequality
- Logarithmic Sobolev inequality
- Implication I: convergence to equilibrum
- Implication II: concentration of Lipschitz functions
- Alternative proof via local Poincaré inequality
- Alternative proof via tensorization and Central Limit Theorem.

2.1 Ornstein-Uhlenbeck (OU) semigroup

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space equipped with a filtration $\mathcal{F}_t$ carrying a standard n-dimensional Brownian motion $(B_t)_{t\geq 0}$. We consider the following stochastic differential equation (SDE)

$$dX_t = -X_t dt + \sqrt{2} dB_t. \tag{2.1}$$

We could follow the standard existence and uniqueness theorems (see e.g. Chap 3 of [Oks13]) on the solution of the above SDE. But actually, it can be solved explicitly. We have

$$d(e^t X_t) = e^t dX_t + e^t X_t dt = \sqrt{2} e^t dB_t.$$

Then

$$X_t = e^{-t} X_0 + \sqrt{2} \int_0^t e^{s-t} dB_s,$$

is a solution to Eq. (2.1). It is not hard to see that $(X_t)_{t\geq 0}$ defined above is a Markov process. And if $X_0 = x$, then $X \sim \mathcal{N}(\sqrt{\rho} x, (1-\rho)\mathbb{I}_n)$ where $\rho = e^{-2t}$.

In the following, we derive the following objects of interest for this Markov process.

1. Associated Markov semigroup
2. Invariant measure, reversibility
3. Kernel
4. Infinitesimal generator
5. Carré du champ operator
6. Fokker-Planck equation

Derive its associated Markov semigroup We have

$$\begin{aligned} P_t f(x) &= \mathbb{E}\left[f(X_t) \mid X_0 = x\right] \\ &\overset{(i)}{=} \mathbb{E}_{G\sim\mathcal{N}(0,\mathbb{I}_n)}\left[f\left(e^{-t}x + \sqrt{1-e^{-2t}}G\right)\right] \\ &= \int_{\mathbb{R}^n} f\left(e^{-t}x + \sqrt{1-e^{-2t}}y\right)\gamma_n(dy). \end{aligned}$$

For (i), note that $\sqrt{2}\int_0^t e^{s-t} dB_s$ is a Gaussian vector centered at 0 and having covariance matrix (via Itô's isometry)

$$2\int_0^t e^{2(s-t)} ds \mathbb{I}_n = (1-e^{2t})\mathbb{I}_n.$$

Here γ_n is the density of standard Gaussian $\mathcal{N}(0,\mathbb{I}_n)$.

Proposition 2.1.1 (Merhler formula). *For any test function f (such that the integrals are well-defined) and every $x \in \mathbb{R}^n$,*

$$\begin{aligned} P_t f(x) &= \int_{\mathbb{R}^n} f(\sqrt{\rho}x + \sqrt{1-\rho}y)\gamma_n(dy) \\ &= f * g_{1-\rho}(\sqrt{\rho}x), \end{aligned}$$

where $*$ denotes convolution, $\rho = e^{-2t}$, $g_{1-\rho}$ is the density of $\mathcal{N}(0,(1-\rho)\mathbb{I}_n)$.

From the Merhler formula, we deduce the semigroup property $P_{s+t} = P_s \circ P_t$ because the convolution of two Gaussian with mean 0 and covariance Σ_1 and Σ_2 results in a Gaussian with mean 0 and covariance $\Sigma_1 + \Sigma_2$.

$$\begin{aligned} P_s(P_t f)(x) &= P_s(f(\sqrt{\rho_t}\cdot) * g_{1-\rho_t}(\frac{1}{\sqrt{\rho_t}}\cdot))(x) \\ &= f(\sqrt{\rho_t}\cdot) * g_{1-\rho_t}(\frac{1}{\sqrt{\rho_t}}) * g_{1-\rho_s}(\sqrt{\rho_s}x) \\ &= f(\sqrt{\rho_t}\cdot) * g_{1/\rho_t - \rho_s}(\cdot)(\sqrt{\rho_s}x) \\ &= f * g_{1-\rho_t\rho_s}(\cdot)(\sqrt{\rho_t\rho_s}x). \end{aligned}$$

Derive its invariant measure

Proposition 2.1.2. *The standard Gaussian measure γ_n is reversible for (P_t).*

Proof. We have

$$\begin{aligned}\int f(P_t h)d\gamma_n &= \int\int f(x)h(\sqrt{\rho}x+\sqrt{1-\rho}y)\gamma_n(dy)\gamma_n(dx)\\ &\overset{(i)}{=} \int\int f(x)h(\sqrt{\rho}x-\sqrt{1-\rho}y)\gamma_n(dy)\gamma_n(dx)\\ &\overset{(ii)}{=} \int\int f(\sqrt{\rho}w+\sqrt{1-\rho}z)h(w)\gamma_n(dw)\gamma_n(dz)\\ &= \int h(P_t f)d\gamma_n.\end{aligned}$$

(i) follows from the symmetry of Gaussian around 0, (ii) uses change of variance $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{\rho}w+\sqrt{1-\rho}z \\ -\sqrt{1-\rho}w+\sqrt{\rho}z \end{bmatrix}$, with Jacobian 1. □

Hence, γ_n is also an invariant measure for (P_t).

Derive its kernel Since

$$\begin{aligned}P_t f(x) &= \int_{\mathbb{R}^n} f(\sqrt{\rho}x+\sqrt{1-\rho}y)\gamma_n(dy)\\ &= \int_{\mathbb{R}^n} f(z)\frac{1}{(2\pi(1-\rho))^{\frac{n}{2}}}e^{-\frac{|z-\sqrt{\rho}x|^2}{2(1-\rho)}}dz,\end{aligned}$$

we identify the kernel (with respect to γ_n)

$$p_t(x,y)d\gamma_n(y) = \frac{1}{(1-\rho)^{\frac{n}{2}}}e^{-\frac{|y-\sqrt{\rho}x|^2}{2(1-\rho)}+\frac{|y|^2}{2}}d\gamma_n(y).$$

Derive its infinitesimal generator Since

$$P_t f(x) = \int_{\mathbb{R}^n} f(e^{-t}x+\sqrt{1-e^{-2t}}y)\gamma_n(dy)$$

Taking derivative with respect to t around 0 gives

$$\frac{d}{dt}P_t f(x) = \int_{\mathbb{R}^n}\left\langle -e^{-t}x+e^{-2t}(1-e^{-2t})^{-\frac{1}{2}}y, \nabla f(e^{-t}x+\sqrt{1-e^{-2t}}y)\right\rangle \gamma_n(dy).$$

It gives two terms. The first term gives $-x \cdot \nabla f(x)$ when $t \to 0$, and the second term gives $\Delta f(x)$ via integration by parts. Finally, the generator

$$L = \Delta - x \cdot \nabla,$$

defined for functions in C_b^2. Consequently, the associated Fokker-Planck equation is

$$\partial_t p_t = \Delta p_t - x \cdot \nabla p_t.$$

Derive carré du champ operator We have

$$\Gamma(f) = \frac{1}{2}\left[L(f^2) - 2fLf\right] = |\nabla f|^2.$$

Integration by parts formula for OU semigroup For f and g in C_b^2, we have the following integration by parts formula

$$\int_{\mathbb{R}^n} (Lf) g d\gamma_n = -\int_{\mathbb{R}^n} \langle \nabla f, \nabla g \rangle \, d\gamma_n \tag{2.2}$$

Proof. Since (P_t) has reversible measure γ_n, we have

$$\int_{\mathbb{R}^n} (P_t f) g d\gamma_n = \int_{\mathbb{R}^n} f (P_t g) d\gamma_n.$$

For $f, g \in C_b^2$ we can apply dominated convergence and differentiate at $t = 0$, and obtain

$$\int_{\mathbb{R}^n} (Lf) g d\gamma_n = \int_{\mathbb{R}^n} f (Lg) d\gamma_n.$$

Similarly, since (P_t) has reversible measure γ_n, we have

$$\int_{\mathbb{R}^n} P_t(fg) d\gamma_n = \int_{\mathbb{R}^n} fg d\gamma_n.$$

Differentiating gives

$$\int_{\mathbb{R}^n} L(fg) \gamma_n = 0.$$

And easy computation show that, in the case of OU semigroup,

$$L(fg) = (Lf)g + f(Lg) + 2\langle \nabla f, \nabla g \rangle.$$

We conclude. □

2.2 Gaussian Poincaré inequality

Theorem 2.2.1 (Gaussian Poincaré inequality)**.** *For every test function f whose gradient belongs to $\mathbb{L}^2(\gamma_n)$, we have*

$$\mathrm{Var}_{\gamma_n}(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 \, d\gamma_n.$$

ϕ-entropy Given a convex function ϕ on an interval $I \subseteq \mathbb{R}$, the ϕ-entropy of $f : \mathbb{R} \to I$ is

$$\mathbb{E}^{\phi}_{\gamma_n}(f) = \int_{\mathbb{R}^n} \phi(f) d\gamma_n - \phi\left(\int_{\mathbb{R}^n} f d\gamma_n\right).$$

Remark 1. *Taking $\phi(x) = x^2$ on $I = \mathbb{R}$ recovers variance, and $\phi(x) = x\log(x)$ on $I = \mathbb{R}_+$ recovers entropy.*

Proposition 2.2.1 (semigroup expression of ϕ-entropy)**.** *Let $\phi : I \to \mathbb{R}$ be C^2 on an interval $I \subseteq \mathbb{R}$ and $f : \mathbb{R}^n \to I$ be C^2_b. Then*

$$\mathbb{E}^{\phi}_{\gamma_n}(f) = \int_0^{\infty} \int_{\mathbb{R}^n} \phi''(P_t f) \, |\nabla P_t f|^2 \, d\gamma_n dt.$$

Proof. Let

$$\alpha(t) = \int_{\mathbb{R}^n} \phi(P_t f) d\gamma_n.$$

Note that

- At time 0,

$$\alpha(0) = \int \phi(f) d\gamma_n.$$

- At time $+\infty$,

$$\lim_{t\to+\infty} \alpha(t) \overset{(i)}{=} \int_{\mathbb{R}^n} \lim_{t\to+\infty} \phi(P_t f) d\gamma_n = \phi\left(\int_{\mathbb{R}^n} f d\gamma_n\right).$$

  (i) switches the order of limit and integral (by dominated convergence theorem, and boundedness of f).

Hence, we can write

$$\mathbb{E}^{\phi}_{\gamma_n}(f) = \alpha(0) - \alpha(+\infty) = -\int_0^{\infty} \alpha'(t)dt.$$

We have

$$\begin{aligned}
\alpha'(t) &\overset{(i)}{=} \int_{\mathbb{R}^n} \partial_t \phi(P_t f) d\gamma_n \\
&= \int_{\mathbb{R}^n} \phi'(P_t f)(LP_t f) d\gamma_n \\
&\overset{(ii)}{=} -\int_{\mathbb{R}^n} \langle \nabla \phi'(P_t f), \nabla P_t f \rangle \, d\gamma_n \\
&= -\int_{\mathbb{R}^n} \phi''(P_t f) \left|\nabla P_t f\right|^2 d\gamma_n
\end{aligned}$$

(i) switches the order of derivative and integral (by dominated convergence theorem). (ii) applies integration by parts (2.2) and the boundary terms are 0. Additionally, for $LP_t f$ to be well-defined, it suffices to have f in C_b^2. □

Lemma 1 (Gradient and OU semigroup commutation, pointwise). *If f is smooth with bounded derivative, (P_t) is OU semigroup, then*

$$\nabla P_t f(x) = e^{-t} P_t(\nabla f)(x),$$

where $P_t(\nabla f)$ is defined by extending P_t to $\mathbb{R}^n$-valued function coordinate wise.

Proof. This is a consequence of Mehler's formula in Prop 2.1.1, by switching the order of derivative and integral. □

Now we are ready to prove the Gaussian Poincaré inequality in Thm 2.2.1

Proof of Thm 2.2.1. First, start with $f \in C_b^2$ and $\phi(x) = x^2$. The semigroup expression of variance in Prop 2.2.1 applies, and we have

$$\mathrm{Var}_{\gamma_n}(f) = 2\int_0^{+\infty} \int_{\mathbb{R}^n} \left|\nabla P_t f\right|^2 d\gamma_n.$$

Using the pointwise commutation property in Lem 1, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} \left|\nabla P_t f\right|^2 d\gamma_n &= \int_{\mathbb{R}^n} e^{-2t} \left|P_t(\nabla f)\right|^2 d\gamma_n \\
&\overset{(i)}{\leq} e^{-2t} \int_{\mathbb{R}^n} P_t(|\nabla f|^2) d\gamma_n \\
&\overset{(ii)}{=} e^{-2t} \int_{\mathbb{R}^n} \left|\nabla f\right|^2 d\gamma_n
\end{aligned}$$

(i) applies Cauchy-Schwarz inequality and (ii) uses the fact that γ_n is stationary. Note that $\int_0^\infty e^{-2t} = \frac{1}{2}$. We conclude by noting that C_b^2 functions are dense in $L^2(\gamma_n)$. **YC — The density argument could be done in two steps 1. C_c^∞ is dense in L^p (see e.g. Prop 8.17 in [Fol99]) 2. Because of Gaussian tail decay, we can approximate C_c^∞ via C_b^2 functions using a Mollifier.** □

Remark 2 (On the optimality of the Poincaré constant)**.** *Check on affine functions $f(x) = \langle x, a\rangle$ that the inequality is sharp. Leave as exercise.*

2.3 Gaussian logarithmic Sobolev inequality

Theorem 2.3.1 (Gaussian logarithmic Sobolev inequality)**.** *For every test function f, which is nonnegative, C_b^1, integrable, the following inequality holds*

$$Ent_{\gamma_n}(f) \le \frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} d\gamma_n.$$

Proof of Thm 2.3.1. First, we start with $f \in C_b^2$ and $f \ge \epsilon$ for some $\epsilon > 0$. Then we can apply the semigroup expression in Prop 2.2.1 for $\phi(x) = x \log x$ to obtain

$$\mathrm{Ent}_{\gamma_n}(f) = \int_0^{+\infty}\int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f} d\gamma_n dt.$$

Using the pointwise commutation property in Lem 1, we have

$$\begin{aligned} |\nabla P_t f|^2 &= e^{-2t}|P_t(\nabla f)|^2 \\ &= e^{-2t}\left|P_t\left(\frac{\nabla f}{\sqrt{f}}\sqrt{f}\right)\right|^2 \\ &\overset{(i)}{\le} e^{-2t} P_t\left(\frac{|\nabla f|^2}{f}\right) P_t(f) \end{aligned}$$

(i) uses Cauchy-Schwarz inequality. Then using stationarity, we have

$$\int_{\mathbb{R}^n} \frac{|\nabla P_t f|^2}{P_t f} d\gamma_n \le e^{-2t}\int_{\mathbb{R}^n} P_t\left(\frac{|\nabla f|^2}{f}\right) d\gamma_n \le e^{-2t}\int_{\mathbb{R}^n}\frac{|\nabla f|^2}{f} d\gamma_n$$

We conclude by noting that C_b^2 and > 0 functions are dense in the desired test function class. □

Remark 3 (On the optimality of the log-Sobolev constant)**.** *Check on functions $f(x) = e^{\langle x,a\rangle}$ that the inequality is sharp. Leave as exercise.*

Poincaré inequality can be formally derived from the log-Sobolev inequality Let h be a bounded, C^1-smooth function. Take $\epsilon > 0$ small enough such that $1 + \epsilon h$ is nonnegative. We have the Taylor expansion

$$(1+t)\log(1+t) = t + \frac{t^2}{2} + o(t^2),$$

for t around 0. Then

$$\mathrm{Ent}_{\gamma_n}(1+\epsilon h) = \frac{\epsilon^2}{2}\mathrm{Var}_{\gamma_n}(h) + o(\epsilon^2).$$

Similarly,

$$\int_{\mathbb{R}^n} \frac{|\nabla(1+\epsilon h)|^2}{1+\epsilon h} d\gamma_n = \epsilon^2 \int_{\mathbb{R}^n} |\nabla h|^2 \, d\gamma_n + o(\epsilon^2).$$

Finally, apply LSI in Thm 2.3.1 to $f = 1 + \epsilon h$, and take the limit $\epsilon \to 0$, we obtain the Poincaré inequality for h.

2.4 Implication I: convergence to equilibrium

One might ask what are the implications of PI or LSI. We discuss two main applications

- Exponential convergence to equilibrium
- Concentration of Lipschitz functions

We still focus on the Gaussian case, the general case is deferred to a later lecture.

We have seen that the OU process $X_t \to \gamma_n$ in law as t tends to $+\infty$. We shall see now that the Poincaré inequality and the logarithmic Sobolev inequalities allow to quantify this convergence.

Theorem 2.4.1. *Let (P_t) be the OU semigroup. For every $f \in L^2(\gamma_n)$, we have*

$$\mathrm{Var}_{\gamma_n}(P_t f) \le e^{-2t}\,\mathrm{Var}_{\gamma_n}(f).$$

For every nonnegative integrable f, we have

$$Ent_{\gamma_n}(P_t f) \le e^{-2t}\, Ent_{\gamma_n}(f).$$

Proof of Thm 2.4.1. We always start with test function f with enough regularity. Let $\alpha(t) = \mathrm{Var}_{\gamma_n}(P_t f)$. By stationarity, we have

$$\alpha(t) = \int_{\mathbb{R}^n} (P_t f)^2 d\gamma_n - \left(\int_{\mathbb{R}^n} f d\gamma_n\right)^2.$$

Using the semigroup expression in Prop 2.2.1, we have

$$\alpha'(t) = -2 \int_{\mathbb{R}^n} |\nabla P_t f|^2 \, d\gamma_n.$$

Applying Gaussian Poincaré inequality in Thm 2.2.1, we obtain

$$\alpha' t(t) \le -2\alpha(t), \quad \forall t \ge 0.$$

By Gronwall's inequality, we obtain

$$\alpha(t) \le e^{-2t}\alpha(0),$$

which is exactly the first inequality.

For the second inequality, let $\beta(t) = \mathrm{Ent}_{\gamma_n}(P_t f)$. Taking derivative, using the semigroup expression and apply LSI, we obtain

$$\beta'(t) \le -2\beta(t).$$

Using Gronwall's inequality, we obtain

$$\beta(t) \le e^{-2}\beta(0).$$

□

As we might generalize this convergence to other measures. Note that the decay takes the form

- $e^{-\frac{2}{C_{PI}}t}$ for the variance
- $e^{-\frac{4}{C_{LSI}}t}$ for the entropy

with $C_{PI} = 1$ and $C_{LSI} = 2$ for γ_n.

Remark 4. *Note that we can reformulate the above result in terms of convergence of measure as follows*

$$\chi^2(\mu P_t \,||\, \gamma_n) \le e^{-2t}\chi^2(\mu \,||\, \gamma_n)$$
$$KL(\mu P_t \,||\, \gamma_n) \le e^{-2t} KL(\mu \,||\, \gamma_n).$$

Here $\chi^2(\mu \,||\, \nu)$ *denotes the chi-square divergence defined by*

$$\chi^2(\mu \,||\, \nu) := \mathrm{Var}_\nu(f), \text{ where } f := \frac{d\mu}{d\nu}.$$

Here $KL(\,\|\,)$ denotes the Kullback-Leibler divergence defined by

$$KL(\mu \parallel \nu) := Ent_\nu(f) = \int f \log f d\nu, \text{ where } f := \frac{d\mu}{d\nu}.$$

It is customary to set $\chi^2(\mu \,||\, \nu) = +\infty$ and $KL(\mu \parallel \nu) = +\infty$ if μ is not absolutely continuous with respect to ν, or when $f \notin \mathbb{L}^2(\nu)$ or $f \log f \notin \mathbb{L}^1(\nu)$ respectively.

Given the definitions, we show that $\frac{d(\mu P_t)}{d\gamma_n} = P_t \frac{d\mu}{d\gamma_n}$ in the $L^2(\gamma_n)$ sense, by taking inner product with any test function g

$$\begin{aligned}
\int \frac{d(\mu P_t)}{d\gamma_n} g d\gamma_n &= \int g d(\mu P_t) \\
&\overset{(i)}{=} \int P_t g d\mu \\
&= \int P_t g \frac{d\mu}{d\gamma_n} d\gamma_n \\
&\overset{(ii)}{=} \int g P_t \left(\frac{d\mu}{d\gamma_n}\right) d\gamma_n
\end{aligned}$$

(i) is the definition of P_t operating on measures. (ii) follows from the reversibility of γ_n under P_t.

2.5 Implication II: concentration of Lipschitz functions

Lipschitz We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz* if

$$\|f\|_{\text{Lip}} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} < \infty.$$

A Lipschitz function is always continuous, and a theorem due to Hans Rademacher states that it is differentiable almost everywhere (see Section 3.5 of [Fol99]).

Theorem 2.5.1 (Gaussian concentration of Lipschitz functions)**.** *For any $n \geq 1$ and $f : \mathbb{R}^n \to \mathbb{R}$ Lipschitz, and any real $r > 0$, we have*

$$\mathbb{P}_{Z \sim \gamma_n}(f(Z) \geq \gamma_n(f) + r) \leq \exp\left(-\frac{r^2}{2\,\|f\|_{Lip}^2}\right),$$

where $\gamma_n(f) = \mathbb{E}_{X \sim \gamma_n}[f(X)]$.

Note that the RHS does not depend on n. Using the result for f and $-f$, with union bound, we get two-sided bound as

$$\mathbb{P}_{Z\sim\gamma_n}(|f(Z)-\gamma_n(f)|\geq r)\leq 2\exp\left(-\frac{r^2}{2\left\|f\right\|_{\mathrm{Lip}}^2}\right).$$

This means that under γ_n, f concentrates around its mean, with a width that depends on the Lipschitz constant.

To prove Theorem 2.5.1, we need the following lemma.

Lemma 2 (Sub-gaussian bound on Laplace transform of Lipschitz functions)**.** *For any $n\geq 1$, and any Lipschitz function $f:\mathbb{R}^n\to\mathbb{R}$, and any $\theta\in\mathbb{R}$,*

$$L(\theta):=\int\exp(\theta f)d\gamma_n\leq\exp\left(\frac{\theta^2}{2}\left\|f\right\|_{Lip}^2+\theta\gamma_n(f)\right).$$

In other words,

$$\int\exp\left(\theta\left(f-\gamma_n(f)\right)\right)d\gamma_n\leq\exp\left(\frac{\theta^2}{2}\left\|f\right\|_{Lip}^2\right).$$

Proof of Lem. 2. First, we observe that for any $\theta\in\mathbb{R}$, we have $e^{\theta f}\in L^1(\gamma_n)$ since

$$\int e^{\theta f}d\gamma_n\leq e^{|\theta||f(0)|}\int e^{|\theta|\|f\|_{\mathrm{Lip}}|x|}d\gamma_n(x)<\infty.$$

Similarly, we can show that $f\in L^1(\gamma_n)$.

Second, let's start with f bounded and C^∞. For any $\theta>0$, applying the Gaussian LSI to $e^{\theta f}$, gives

$$\mathbb{E}\left[\theta f e^{\theta f}\right]-\mathbb{E}[e^{\theta f}]\log\mathbb{E}[e^{\theta f}]\leq\frac{1}{2}\mathbb{E}\theta^2 e^{\theta f}\left|\nabla f\right|^2\leq\frac{\left\|f\right\|_{\mathrm{Lip}}^2}{2}\mathbb{E}\theta^2 e^{\theta f}.$$

Using the notation $L(\theta)=e^{\theta f}$, we have

$$\theta L'(\theta)-L(\theta)\log L(\theta)\leq\frac{\left\|f\right\|_{\mathrm{Lip}}^2\theta^2}{2}e^{\theta f}.$$

Define $K(\theta):=\frac{\log L(\theta)}{\theta}$. Then the above gives

$$K'(\theta)\leq\frac{\left\|f\right\|_{\mathrm{Lip}}^2}{2}.$$

Since $L(0)=1$ and $L'(0)=\gamma_n(f)$, and $K(0)=(\log L)'(0)=L'(0)/L(0)=\gamma_n(f)$. We obtain that $K(\theta)\leq\gamma_n(f)+\theta\frac{\|f\|_{\mathrm{Lip}}^2}{2}$. That is what we wanted. For $\theta<0$ it suffices to apply the above to $-f$.

Third, we show that any Lipschitz function can be approximated by bounded and C^∞ functions, and the inequality remains valid under limits. We outline the proof sketch of this approximation argument. Let f be Lispschitz. We define a sequence of functions as follows, for $k \geq 1$ and $\epsilon > 0$,

$$f_{k,\epsilon} := \max(-k, \min(k, f)) * \rho_\epsilon$$

where $*$ denotes convolution and $\rho_\epsilon \in C^\infty(\mathbb{R}^n, \mathbb{R})$, is a mollifier, satisfies

$$\operatorname{supp}(\rho_\epsilon) \subseteq \{x \in \mathbb{R}^n : |x| \leq \epsilon\}, \rho_\epsilon \geq 0, \text{ and } \int \rho_\epsilon(x)dx = 1.$$

For example, we can take $\rho_\epsilon(x) = \epsilon^{-n}\rho(\epsilon^{-1}|x|)$ with $\rho(x) := c^{-1}e^{-\frac{1}{1-x^2}}\mathbf{1}_{|x|<1}, c := \int \rho(x)dx$. Now

- $\|f_{k,\epsilon}\|_{\text{Lip}} \leq \|f\|_{\text{Lip}}$ since

$$|f_{k,\epsilon}(x) - f_{k,\epsilon}(y)| \leq \int |f_k(x-z) - f_k(y-z)| \rho_\epsilon(z)dz \leq \|f\|_{\text{Lip}} |x-y|,$$

  where $f_k := \max_{-k,\min(k,f)}$.

- $f_{k,\epsilon}$ tends to f pointwise, because if $|x-y| \leq \epsilon$, then

$$|f_k(y) - f(x)| \leq |f_k(y) - f_k(x)| + |f_k(x) - f(x)| \leq \epsilon \|f\|_{\text{Lip}} + |f(x)| \mathbf{1}_{|f(x)|\geq k} \underset{k\to\infty,\epsilon\to 0}{\to} 0.$$

  and

$$|f_{k,\epsilon}(x) - f(x)| \leq \int_{|x-y|\leq\epsilon} |f_k(y) - f(x)| \rho_\epsilon(x-y)dy \leq \epsilon \|f\|_{\text{Lip}} + |f(x)| \mathbf{1}_{|f(x)|\geq k} \underset{k\to\infty,\epsilon\to 0}{\to} 0.$$

- By dominated convergence, $f_{k,\epsilon}$ tends to f in $L^1(\gamma_n)$.

- With pointwise convergence, we apply Fatou's lemma to obtain the desired inequality

$$\begin{aligned}
\int e^{\theta f} d\gamma_n = \int \lim_{k\to\infty,\epsilon\to 0} e^{\theta f_{k,\epsilon}} d\gamma_n &\leq \liminf_{k\to\infty,\epsilon\to 0} \int e^{\theta f_{k,\epsilon}} d\gamma_n \\
&\leq \liminf_{k\to\infty,\epsilon\to 0} \exp\left(\frac{\theta^2}{2} \|f_{k,\epsilon}\|_{\text{Lip}}^2 + \theta \int f_{k,\epsilon} d\gamma_n\right) \\
&\leq \exp\left(\frac{\theta^2}{2} \|f\|_{\text{Lip}}^2 + \theta \int f d\gamma_n\right).
\end{aligned}$$

□

Now we are ready to prove Theorem 2.5.1.

Proof of Theorem 2.5.1. It suffices to prove for $\|f\|_{\text{Lip}} = 1$ and $\gamma_n(f) = 0$ by scaling and translation. For $r > 0$ and $\theta > 0$, we have

$$\begin{aligned}\mathbb{P}_{Z\sim\gamma_n}(f(Z) \geq r) &= \mathbb{P}_{Z\sim\gamma_n}(e^{\theta f(Z)} \geq e^{\theta r}) \\ &\overset{(i)}{\leq} \frac{\mathbb{E}[e^{\theta f(Z)}]}{e^{\theta r}} \\ &\overset{(ii)}{\leq} \frac{e^{\theta^2/2}}{e^{\theta r}} \\ &\overset{(iii)}{\leq} e^{-\frac{1}{2}r^2}.\end{aligned}$$

(i) applies Markov's inequality. (ii) applies Lemma 2. (iii) picks the best $\theta = \frac{1}{2r^2}$ which makes the bound the smallest. □

Example 1 (Application to empirical means)**.** *For any integer $n \geq 1$ and $N \geq 1$, if $X_1, \ldots, X_N$ are independent and identically distributed random variables with law γ_n, then for any Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, any real $r \geq 0$,*

$$\mathbb{P}\left(\left|\frac{\sum_{j=1}^N f(X_j)}{N} - \mathbb{E}[f(X_1)]\right| \geq r\right) \leq 2\exp\left(-\frac{Nr^2}{2\|f\|_{Lip}^2}\right).$$

Proof. Consider the function $(\mathbb{R}^n)^N \to \mathbb{R}$,

$$F(x) = \frac{1}{N}\left(\sum_{j=1}^N f(x_j)\right).$$

Show that it is Lipschitz with

$$\|F\|_{\text{Lip}} \leq \frac{\|f\|_{\text{Lip}}}{\sqrt{N}}.$$

□

Remark 5. *An examination of the proofs above reveals that the sub-Gaussian concentration inequalities are still valid when the Gaussian measure γ_n is replaced by any probability measure on $\mathbb{R}^n$ which satisfies a logarithmic Sobolev inequality. This kind of proof is known as Herbst argument [GR98].*

Remark 6 (Poincaré inequality and exponential tail)**.** *The Herbst argument allows to show that if a probability measure satisfies to a Poincaré inequality then this implies sub-exponential concentration around the mean for Lipschitz functions.*

2.6 Alternative proof via local Poincaré inequality

We have used semigroup expression of ϕ-entropy from time 0 to ∞, to prove Gaussian Poincaré inequality and LSI. We will see that a version of Poincaré inequality and LSI that holds locally

Alternative proof of Thm 2.2.1, 2.3.1. Fix $t \geq 0$ and $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to I$, for $s \in [0, t]$, define

$$\beta(s) := P_s(\phi(P_{t-s}f))(x)$$

where $\phi(z) = z^2, I = \mathbb{R}$ or $\phi(z) = z \log z, I = \mathbb{R}_+$.

We have

$$\beta(t) - \beta(0) = P_t(\phi(f)) - \phi(P_t f).$$

Taking derivative, with respect to s, and setting $g = P_{t-s}f$, gives

$$\beta'(s) = P_s \left[L\phi(g) - Lg\phi'(g) \right].$$

Recall that for OU semigroup $L = \Delta - x \cdot \nabla$, then

$$L\phi(g) - Lg\phi'(g) = \phi''(g) \left| \nabla g \right|^2.$$

Thus

$$\begin{aligned}
\beta'(s) &= P_s \left[\phi''(P_{t-s}f) \left| \nabla P_{t-s}f \right| \right] \\
&\overset{(i)}{\leq} e^{-2(t-s)} P_s(\phi''(P_{t-s}f) P_{t-s}(\left| \nabla f \right|^2)) \\
&\overset{(ii)}{\leq} e^{-2(t-s)} P_s(P_{t-s}(\phi''(f) \left| \nabla f \right|^2)) \\
&= e^{-2(t-s)} P_t(\phi''(f) \left| \nabla f \right|^2)
\end{aligned}$$

(i) follows from Mehler's formula which gives the sub-commutation $|\nabla P_{t-s}f| \leq e^{-(t-s)} P_{t-s}(|\nabla f|)$, and (ii) follows from Jensen's inequality for the convex function $(u, v) \mapsto \phi''(u)v^2$.

Integrating on $[0, t]$, we obtain

$$\beta(t) - \beta(0) \leq \frac{1 - e^{-2t}}{2} P_t \left(\phi''(f) \left| \nabla f \right|^2 \right).$$

This gives

$$P_t(\phi(f))(x) - \phi(P_t(f)(x)) \leq \frac{1 - e^{-2t}}{2} P_t \left(\phi''(f) \left| \nabla f \right|^2 \right).$$

Sending t to ∞, we obtain the Gaussian PI and LSI.

Why local Poincaré inequality or local LSI? $P_t(\cdot)(x)$ can be interpreted as a measure local at x as follows

$$P_t(f)(x) = \mathbb{E}[f(X_t) \mid X_0 = x] = \mathbb{E}_{P_t(\cdot)(x)}[f]$$

where $P_t(\cdot)(x) = \mathcal{N}(xe^{-t}, 1 - e^{-2t})$. So at any fixed time t, the above gives a Poincaré and a log-Sobolev inequality with constants $(1 - e^{-2t})$ and $(1 - e^{-2t})/2$ for $P_t(\cdot)(x)$. $\square$

2.7 Alternative proof via tensorization and Central Limit Theorem

We study the tensorization property of the variance and the entropy. It allows to provide a proof of the Poincaré and of the log-Sobolev inequalities for the Gaussian measure by using the Central Limit Theorem, starting from elementary inequalities on the two-point space.

The Gaussian measure appears as a limiting distribution in the asymptotic analysis of product spaces, due to the central limit phenomenon. The simplest product space is the discrete cube $\{0, 1\}^n$ equipped with the product Bernoulli probability measure $\mu_n = \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right)^{\otimes n}$, which is the uniform probability measure. This model is called "the two-point-space" when $n = 1$.

Given $X \sim \mu_n$, since each coordinate is *i.i.d.* with mean $\frac{1}{2}$ and variance $\sigma^2 = \frac{1}{4}$, the Central Limit Theorem states that

$$\frac{1}{\sqrt{n\sigma^2}} \sum_{j=1}^{n} \left(X_j - \frac{1}{2}\right) \xrightarrow{d}_{n\to\infty} \gamma_1.$$

In other words, for any continuous and bounded $f : \mathbb{R} \to \mathbb{R}$

$$\int_{\{0,1\}^n} g_n d\mu_n \underset{n\to\infty}{\to} \int f d\gamma_1,$$

where $g_n(x) := f\left(\frac{2}{\sqrt{n}} \sum_{j=1}^{n} (x_j - \frac{1}{2})\right)$.

Theorem 2.7.1 (Tensorization). *Let $(E_1, \mathcal{A}_1, \nu_1), \ldots, (E_n, \mathcal{A}_n, \nu_n)$ be probability spaces. Let $\nu = \nu_1 \otimes \cdots \otimes \nu_n$ be the product probability measure on $(E_1 \times \cdots \times E_n, \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n)$. Let $\phi : I \to \mathbb{R}$ be convex and such that $(u, v) \mapsto \phi''(u)v^2$ is convex. Then for any $f : E_1 \times \cdots \times E_n \to \mathbb{R}$ such that $\phi(f) \in L^1(\nu_1 \otimes \cdots \otimes \nu_n)$,*

$$\mathbb{E}^{\phi}_{\nu}(f) \leq \sum_{i=1}^{n} \mathbb{E}_{\nu} \mathbb{E}^{\phi}_{\nu_i}(f),$$

where the subscript μ_i indicates the integration over i-th variable only while conditioning on the other variables.

Note that for $\phi(u) = u^2$ and $I = \mathbb{R}$, we obtain the tensorization of variance

$$\mathrm{Var}_{\nu_1\otimes\cdots\otimes\nu_n}(f) \leq \mathbb{E}_{\nu_1\otimes\cdots\otimes\nu_n}\left[\sum_{i=1}^n \mathrm{Var}_{\nu_i}(f)\right].$$

For $\phi(u) = u\log u$ and $I = [0,\infty)$, we obtain the tensorization of entropy

$$\mathrm{Ent}_{\nu_1\otimes\cdots\otimes\nu_n}(f) \leq \mathbb{E}_{\nu_1\otimes\cdots\otimes\nu_n}\left[\sum_{i=1}^n \mathrm{Ent}_{\nu_i}(f)\right].$$

Proof. By induction on n, we only have to consider the case $n = 2$. After rearranging the terms, it suffices to prove that

$$\mathbb{E}^{\phi}_{\mu_2}\left[\mathbb{E}_{\mu_1} f\right] \leq \mathbb{E}_{\mu_1}\left[\mathbb{E}^{\phi}_{\mu_2}(f)\right].$$

In the case of variance, this follows from the Cauchy-Schwarz inequality. Namely,

$$\mathrm{Var}_{\mu_2}(\mathbb{E}_{\mu_1} f) = \mathbb{E}_{\mu_2}\left[(\mathbb{E}_{\mu_1} f - \mathbb{E}_{\mu_2}\mathbb{E}_{\mu_1} f)^2\right] \leq \mathbb{E}_{\mu_2}\mathbb{E}_{\mu_1}((f - \mathbb{E}_{\mu_2} f)^2) = \mathbb{E}_{\mu_1}\mathrm{Var}_{\mu_2} f.$$

The general case has to use the convexity of ϕ and convexity of $\phi''(u)v^2$. First, we show that the ϕ-entropy functional

$$f \mapsto \mathbb{E}^{\phi}_{\mu}(f)$$

is convex as soon as $F^{\phi} : (u,v) \mapsto \phi''(u)v^2$ is convex. Since convexity is a unidimensional property, it suffices to show that $\alpha(t) := \mathbb{E}^{\phi}_{\mu}(tf + (1-t)g)$ is convex. We have

$$\begin{aligned}
\alpha'(t) &= \mathbb{E}_{\mu}\left[\phi'(tf + (1-t)g)(f-g)\right] - \phi'(\mathbb{E}_{\mu}(tf + (1-t)g))\mathbb{E}_{\mu}(f-g)\\
\alpha''(t) &= \mathbb{E}_{\mu}\left[\phi''(tf + (1-t)g)(f-g)^2\right] - \phi''(\mathbb{E}_{\mu}(tf + (1-t)g))\left[\mathbb{E}_{\mu}(f-g)\right]^2 \geq 0,
\end{aligned}$$

where the last step follows from Jensen's inequality via the convexity of F^{ϕ}. From the convexity of α, we have

$$\alpha(1) \geq \alpha(0) + \alpha'(0)(1-0),$$

which gives

$$\mathbb{E}^{\phi}_{\mu}(f) \geq \mathbb{E}^{\phi}_{\mu}(g) + \mathbb{E}_{\mu}\left[(\phi'(g) - \phi'(\mathbb{E}_{\mu} g))(f-g)\right],$$

for any g. Thus, observing that the equality is achieved at f.

$$\mathbb{E}^{\phi}_{\mu}(f) = \sup_{g:\phi(g)\in L^1(\mu)} \mathbb{E}^{\phi}_{\mu}(g) + \mathbb{E}_{\mu}\left[(\phi'(g) - \phi'(\mathbb{E}_{\mu} g))(f-g)\right].$$

Using the above variational formula for μ_2, we obtain

$$\begin{aligned}\mathbb{E}^{\phi}_{\mu_2}\left[\mathbb{E}_{\mu_1}(f)\right] &= \sup_g \left\{\mathbb{E}^{\phi}_{\mu_2}(g) + \mathbb{E}_{\mu_2}\left[(\phi'(g) - \phi'(\mathbb{E}_{\mu_2} g))(\mathbb{E}_{\mu_1} f - g)\right]\right\} \\ &= \sup_g \mathbb{E}_{\mu_1}\left\{\mathbb{E}^{\phi}_{\mu_2}(g) + \mathbb{E}_{\mu_2}\left[(\phi'(g) - \phi'(\mathbb{E}_{\mu_2} g))(f - g)\right]\right\} \\ &\le \mathbb{E}_{\mu_1} \sup_g \left\{\mathbb{E}^{\phi}_{\mu_2}(g) + \mathbb{E}_{\mu_2}\left[(\phi'(g) - \phi'(\mathbb{E}_{\mu_2} g))(f - g)\right]\right\} \\ &= \mathbb{E}_{\mu_1}(\mathbb{E}^{\phi}_{\mu_2}(f)).\end{aligned}$$

□

Now we proceed to see how tensorization gives an alternative proof of Gaussian Poincaré inequality

Alternative proof of Gaussian Poincaré in Thm 2.2.1. First, we prove the Poincaré inequality for γ_1. Let $\mu_n = \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right)^{\otimes n}$ be the uniform distribution on the cube $\{0,1\}^n$. For any $g : \{0,1\} \to \mathbb{R}$, we have

$$\mathrm{Var}_{\mu_1}(g) = \frac{g(0)^2 + g(1)^2}{2} - \left(\frac{g(0) + g(1)}{2}\right)^2 = \frac{(g(1) - g(0))^2}{4}.$$

Using the tensorization of variance, we obtain for $n \ge 1$ and $g : \{0,1\}^n \to \mathbb{R}$,

$$\mathrm{Var}_{\mu_n}(g) \le \frac{1}{4}\mathbb{E}_{\mu_n}\left[\sum_{i=1}^{n}(D_i g)^2\right] \tag{2.3}$$

where $(D_i g)^2(x) = (g(x + e_i) - g(x))^2$ where $(e_1, \ldots, e_n)$ is canonical basis of $\mathbb{R}^n$.

Now for any $f : \mathbb{R} \to \mathbb{R}$ and compactly supported, and set

$$g(x) = f(s_n(x)) \text{ with } s_n(x) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(x_i - \frac{1}{2}\right).$$

Using Taylor expansion, we have for $i = 1, \ldots, n$ and $x \in \{0,1\}^n$,

$$(D_i g)^2(x) = \left[\frac{2}{\sqrt{n}} f'(s_n(x)) + o\left(\frac{1}{\sqrt{n}}\right)\right]^2 = \frac{4}{n} f'^2(s_n(x)) + o\left(\frac{1}{n}\right),$$

where the o is uniform in x since f is C^2 and compactly supported. Applying Central Limit Theorem, we obtain

$$\mathrm{Var}_{\gamma_1}(f) = \lim_{n\to\infty} \mathrm{Var}_{\mu_n}(g) \le \lim_{n\to\infty} \mathbb{E}_{\mu_n}[f'^2(s_n)] = \mathbb{E}_{\gamma_1}(f'^2).$$

The middle inequality follows from Eq. (2.3).

Second, to pass from γ_1 to $\gamma_n = \gamma_1^{\otimes n}$, we can sue the tensorization of variance again to obtain

$$\begin{aligned}\mathrm{Var}_{\gamma_n}(f) &\leq \mathbb{E}_{\gamma_n}\left[\sum_{i=1}^n \mathrm{Var}_{\gamma_i}(f)\right] \\ &\leq \mathbb{E}_{\gamma_n}\left[\sum_{i=1}^n \mathbb{E}_{\gamma_i}(\partial_i f)^2\right] \\ &= \mathbb{E}_{\gamma_n}|\nabla f|^2 .\end{aligned}$$

Finally, we can enlarge the class of test function form C^2 and compactly supported to $L^2(\gamma_n)$, by approximation arguments and Fatou's lemma.

□

For the log-Sobolev inequality, we can proceed exactly as we did above for the Poincaré inequality. The starting point is the two-point space inequality: for any $g : \{0,1\} \to \mathbb{R}$,

$$\mathrm{Ent}_{\mu_1}(g^2) \leq \frac{(g(1)-g(0))^2}{2}.$$

Let $a := g(0)$ and $b := g(1)$, then the above is equivalent to

$$\frac{a^2\log(a^2)+b^2\log(b^2)}{2} - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \leq \frac{(a-b)^2}{2}.$$

By homogegeneity of the inequality, we may assume $b = 1$, this leads to

$$\frac{a^2\log a^2}{2} - \frac{a^2+1}{2}\log\frac{a^2+1}{2} \leq \frac{(a-1)^2}{2}.$$

This is just one dimensional inequality, which can be proved by studying its derivatives (for $a > 0$, the second derivative of LHS - RHS should be less than 0, so the first derivative should have a single root at 1).

The strategy of using tensorization and the Central Limit Theorem to prove log-Soobolv inequality for γ_n is from [Gro75].

2.8 Exercises

Exercise 1. *On the sharpness of Poincaré and LSI constants.*

- *Show that the Gaussian Poincaré inequality is sharp for* $f(x) = \langle x, a \rangle$.
- *Show that the Gaussian log-Sobolev inequality is sharp for* $f(x) = e^{\langle x, a \rangle}$.

Exercise 2. *Complete the proof in Example 1, which applies concentration of Lipschitz functions to empirical means.*

Exercise 3. 1. *Show that*

$$\frac{a^2 \log a^2}{2} - \frac{a^2+1}{2} \log \frac{a^2+1}{2} \leq \frac{(a-1)^2}{2},$$

for $a \in \mathbb{R}$.

2. *Complete the proof for Gaussian Log-Sobolev inequality via tensorization.*

Exercise 4 (Poincaré ineqaulity for 1-D exponential)**.** *Consider the exponential measure on* $\mathbb{R}_+$*:*

$$d\mu(x) = \lambda e^{-\lambda x} dx, x > 0$$

1. *Show that the Poincaré inequality holds*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\lambda^2} \int |f'|^2 \, d\mu.$$

2. *Compare the constant to the Gaussian case, discuss when and why it is better or worse.*

Exercise 5 (Gaussian Poincaré via Stein's method)**.** *Stein's method provides a way to study functional inequalities in the Gaussian setting.*

1. *(Stein's identity) Show that for a function* $f \in C_b^1$,

$$\mathbb{E}_{\gamma_1}[X f(X)] = \mathbb{E}[f'(X)].$$

2. *For a mean zero function* f, *consider the ordinary differential equation for a function* g

$$g'(x) - x g(x) = f(x).$$

When does the solution g *exists?*

3. *Using the above equation, prove Gaussian Poincaré inequality for* γ_1.

Bibliography

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications.* John Wiley & Sons, 1999.

[GR98] Leonard Gross and Oscar Rothaus. Herbst inequalities for supercontractive semigroups. *Journal of Mathematics of Kyoto University*, 38(2):295–318, 1998.

[Gro75] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[Oks13] Bernt Oksendal. *Stochastic differential equations: an introduction with applications.* Springer Science & Business Media, 2013.