Poincaré Inequality

Reference: ① Ramon van Handel, Probability in High Dimension, Lecture notes, Princeton University, 2016 (RvH book)

② Solution sets to problems in Prob. in High Dim., Available online at: tylerkf.github.io/prob-reading/

③ Sinho Chewi, Log-concave sampling, Online notes, Nov. 2024

(I closely follow Chaps. 2,3 in ①)

In this and the next couple of notes, $\mu = \mu_*$ unless otherwise noted.

We have a M.P. $(\mathcal{L}, \mu)$. The Poincaré inequality relates the variance of a function (average energy) with the Dirichlet form (average of local energy). In the Gaussian case (OU Process),

$$\mathcal{E}(f,f) = \|\nabla f\|^2_{L^2(\mu)} = \int_{\mathbb{R}^d} \left(\sum_{i=1}^{d} \left| \frac{\partial}{\partial x_i} f(x) \right|^2 \right) d\mu(x)$$

$(f: \mathbb{R}^d \to \mathbb{R})$.

average

local energy, as it is the square of the "spatially local differences" (the gradient)

$$\mathrm{Var}_\mu(f) = \int (f - \mu(f))^2 \, d\mu$$

In the Gaussian case, Poincaré Inequality states that:

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

(Gaussian Poincaré Inequality)

$$= \int \|\nabla f(x)\|_2^2 \, d\mu(x).$$

The Poincaré inequality is fundamental to both concentrations of tensor products of measures (aka functions of iid random variables) as well as convergence of M.P. The concentrations connections is also known as the Efron-Stein inequality, which is discussed in the Advanced Prob. class (see notes on my webpage under the "Teaching" tab).

Here, we will discuss the M.P. convergence connection. This will, together with the Bakry-Emery criterion, allow us to derive a dimension-independent convergence rate for the Langevin Dynamics.

→ (*)

Theorem 1 (2.18 in PvH): Let $\{P_t\}$ be a reversible ergodic M.P., with generator $\mathcal{L}$, and invariant measure μ. Then, given $C \geq 0$, the following statements are equivalent. Choose any $f \in L^2(\mu)$.

(1) $\mathrm{Var}_\mu(f) \leq C \cdot \mathcal{E}(f,f)$ (Poincaré Inequality)

(2) $\| P_t f - \mu(f) \|_{L^2(\mu)} \leq e^{-t/C} \| f - \mu(f) \|_{L^2(\mu)} \quad \forall t \geq 0$

($= \mu(P_t f)$, $\because \mu$ stationary)

(3) $\mathcal{E}(P_t f, P_t f) \leq e^{-2t/C} \mathcal{E}(f,f) \quad \forall t \geq 0$

(4) Given $f \in L^2(\mu)$, $\exists$ $K(f)$ s.t.

$$\| P_t f - \mu(f) \|_{L^2(\mu)} \leq K(f) e^{-t/C}$$

(5) Given any $f \in L^2(\mu)$, $\exists$ $K(f)$ s.t.

$$\mathcal{E}(P_t f, P_t f) \leq K(f) e^{-2t/C}$$

Discussion: ⓐ For simplicity, let us center $f(\cdot)$ such that $E_{x_0 \sim M}[f(x_0)] = 0$. Then, let ($M$ stat.)

i.e., $X_t \overset{d}{=} X_0 \ \forall t$

$$h_t(x) = E\left[f(X_t) \,|\, X_0 = x\right].$$ Observe that:

$$h_0(x) = f(x), \quad \text{and} \quad h_\infty(x) = 0 \ \forall x.$$ (M-a.s.)

$$M(P_t f) = E_{x_0 \sim M}[h_t(X_0)] = E_{x_0 \sim M}\left[E\left[f(X_t) \,|\, X_0 = x\right]\right]$$

$$= E_{x_0 \sim M}[f(X_t)] = E_{x_0 \sim M}[f(X_0)] = 0.$$

$$\|f\|_{L^2(M)} = E_{x_0 \sim M}\left[f(X_0)^2\right]^{1/2}$$

$$\|h_t\|_{L^2(M)} = \|P_t f\|_{L^2(M)}$$

$$= E_{x_0 \sim M}\left[h_t(X_0)^2\right]^{1/2}$$

$\therefore \|h_\infty\|_{L^2(M)} = 0$. The question is: How fast does $\|h_t\|_{L^2(M)}$ go to 0? The Poincaré Inequality implies that it ↘ down exponentially fast in time, with rate $(2/c)$.

Instead of asking about the $L^2(\mu)$ norm, i.e., $\|h_t\|_{L^2(\mu)}$, we can look at higher order norms, i.e.,

$$\|h_t\|_{L^q(\mu)} \quad , \text{for} \quad q > 2.$$

- As q increases, $\|h_t\|_{L^q(\mu)}$ will become larger, as a larger q amplifies "spikes" in the function.

- As t increases, $\|h_t\|_{L^q(\mu)}$ will become smaller, because P_t smooths the function ($h_t \to 0$ as $t \to \infty$)

This trade-off is characterized through hyper-contractivity (Gross's hyper-contractivity theorem), where under the Poincaré Inequality with constant 'c', we have

$$\|h_t\|_{L^{1+(p-1)(4t/c)}(\mu)} \leq \|f\|_p \quad , p \geq 0. \quad \rightarrow (a)$$

In comparison, the Theorem above has:

$$\|h_t\|_{L^2(\mu)} \leq e^{-t/c} \|f\|_2 \quad \rightarrow (b)$$

Thus for $p=2$, we see that we can either get a exponential decrease in value (b), or a linear increase in norm, as time progresses.

Please also see:

- Djalil Chafai, Log-Sobolev and Bakry-Emery, Blog post: djalil.chafai.net/2023/01/12/log-sobolev-and-bakry-emery.

- My notes on hyper-contractivity for the discrete Rademacher case and Gaussian r.v.s (Bonomi-Beckner Thm, Nelson's Thm, etc.).

(b) Condition (2) in Theorem 1 is:

$$(2)\quad \|P_t f - \underbrace{\mu(f)}_{=\mu(P_t f),\ \because \mu \text{ stationary}}\|_{L^2(\mu)} \leq e^{-t/c}\, \|f-\mu(f)\|_{L^2(\mu)} \quad \forall t \geq 0$$

Since the norm is w.r.t. the stat. measure μ, one

might ask how this relates to convergence to stationarity starting from any initial condition. After all, if $X_0 \sim \mu$, then $X_t \sim \mu \ \forall t \geq 0$.

We would ideally want something like: for each $x \in \mathbb{R}^d$,

$$\left| E\left[f(X_t) \mid X_0 = x\right] - E\left[f(X_\infty) \mid X_0 = x\right] \right| \leq e^{-t/c} \cdot (\cdot)$$

($E[f(X_\infty) \mid X_0 = x]$: constant function indep. of x because of mixing)

i.e., $$\left| P_t f(x) - \mu(f) \right| \leq e^{-t/c} (\cdot) \longrightarrow (c)$$

(or perhaps, sup-norm over $x \in \mathbb{R}^d$, if that is well-defined for f)

Condition (2) is indeed a softer version of this:

$$\left(\int \left(P_t f(x) - \mu(f) \right)^2 d\mu(x) \right)^{1/2} \leq e^{-t/c} (\cdot)$$

$= \| P_t f - \mu(f) \|_{L^2(\mu)}$

$\longrightarrow (d)$

(Sob-kv inequality: $\| f - Ef \|_q \leq c_q \| \nabla f \|_2$, $q > 2$)

We could get closer to (c) by using a higher norm, and develop similar results (but with a different constant). For a concrete construction, see: P. Cattiaux, A. Guillin, C. Roberto, "Poincaré inequality and the L^p convergence of semigroups". Elec. Comm. in Prob, 2010.

(Gaussian Poincaré Inequality): $\{X_t, t \geq 0\}$ is the OU process with:

$$\mathcal{L}(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle$$
$$\mu = N(0, I_d).$$

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix}, \text{ with } Z_i \sim N(0,1), \quad Z_i \perp\!\!\!\perp Z_j \ \forall i,j.$$

$$\mathrm{Var}_\mu(f) = E\left[\left(f(Z_1,\dots,Z_d) - E\left[f(Z_1,\dots,Z_d)\right]\right)^2\right]$$
$$\leq E\left[\sum_{i=1}^{d}\left(\frac{\partial f}{\partial x_i}(Z_i)\right)^2\right]$$

i.e., $\mathrm{Var}_\mu(f) \leq \mathcal{E}(f,f)$ $\quad (c=1)$.

Proof: (We'll prove the 1-d case, and avoid vector notation. Similar proof for the d-dimensional case.

From notes (5. Markov Processes), we have:

$$P_t f(x) = E_{Z \sim N}\left[f\left(e^{-t} x + \sqrt{1 - e^{-2t}} \cdot Z \right) \right]$$

$$\therefore \quad \frac{d}{dx} P_t f(x) = e^{-t} E\left[f'\left(e^{-t} x + \sqrt{1 - e^{-2t}}\, Z \right) \right]$$

$$= e^{-t} P_t f'(x).$$

$$\therefore \quad \mathcal{E}(P_t f, P_t f) = \left\| (P_t f)' \right\|^2_{L^2(\mu)}$$

$$= e^{-2t} \left\| P_t f' \right\|^2_{L^2(\mu)}$$

$$\leq e^{-2t} \left\| f' \right\|^2_{L^2(\mu)} = e^{-2t} \mathcal{E}(f, f).$$

$\therefore$ From main theorem (3) $\Rightarrow$ (1)

$\Rightarrow$ (2).

$$\therefore \quad \mathrm{Var}_\mu(f) \leq \mathcal{E}(f, f)$$

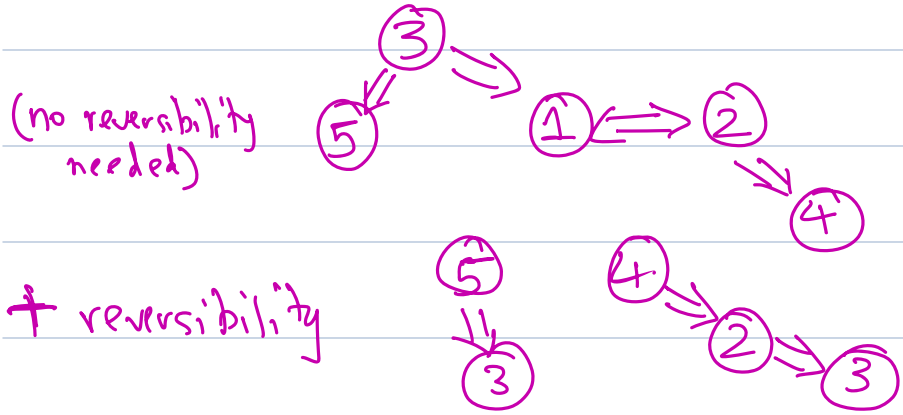
$$\left\| P_t f - \mu f \right\|_{L^2(\mu)} \leq e^{-t} \left\| f - \mu(f) \right\|_{L^2(\mu)} \quad \forall t \geq 0$$

i.e., the OU process converges exponentially fast

to its stationary dist. Further, the calculation for higher d shows that $C=1$, and does not depend on dimension d. This implies that the OU process converges exponentially fast to its stat. dist. independent of dimension. Note that this is not very surprising, because each dimension is essentially an independent 1-d OU process, with no coupling across dimensions.

However, we will later show that a similar property holds for the Langevin Dynamics for any strongly log-concave distribution.

Proof of Poincaré Inequality Thm:



① $\mathrm{Var}_\mu(f) \le C \cdot \mathcal{E}(f,f)$

② $\|P_t f - \mu f\|_{L^2(\mu)} \le e^{-t/C} \|f - \mu f\|_{L^2(\mu)}$

③ $\mathcal{E}(P_t f, P_t f) \le e^{-2t/C} \mathcal{E}(f,f)$

④ Given $f \in L^2(\mu)$, $\exists\ K(f)$ s.t.
$$\|P_t f - \mu(f)\|_{L^2(\mu)} \le K(f) e^{-t/C}$$

⑤ Given any $f \in L^2(\mu)$, $\exists\ K(f)$ s.t.
$$\mathcal{E}(P_t f, P_t f) \le K(f) e^{-2t/C}$$

$\mu(f) = E_{Z\sim\mu}[f(Z)]$

Lemma 1 (2.28 in RvH book)

$$\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\,\mathcal{E}(P_t f, P_t f)$$

Proof: $\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = \frac{d}{dt}\left(\mu\left((P_t f)^2\right) - \underbrace{\mu(P_t f)^2}_{=\mu(f)}\right)$

$\mathrm{Var}_\mu(g) = E[g(Z)^2] - E[g(Z)]^2$, $Z\sim\mu$

$$= \frac{d}{dt}\mu\left((P_t f)^2\right) = \mu\left(2P_t f\,\frac{d}{dt}(P_t f)\right)$$

$$= \mu\left(2P_t f\cdot \mathcal{L}P_t f\right) = -2\,\mathcal{E}(P_t f, P_t f)$$

∎

Recall (from Markov Processes notes) that $\frac{d}{dt}\mathrm{Var}(P_t f) \le 0$.

$\therefore$ **Corollary 2** (2.29 in RvH) $\quad \mathcal{E}(g,g) \ge 0 \quad \forall g.$

Pf: Let $g = P_t f$ for some t, f. Then

$$\mathcal{E}(g,g) = -\frac{1}{2}\frac{d}{dt}\mathrm{Var}_\mu(P_t f) \ge 0$$

∎

Corollary 3 (2.30 in RvH) (α, μ) ergodic. Then $\forall f \in \mathcal{F}$,

$$\mathrm{Var}_\mu(f) = 2\int_0^\infty \mathcal{E}(P_t f, P_t f)\, dt$$

Pf: Ergodicity $\iff P_t f \xrightarrow{t\to\infty} \mu(f)$

$\Downarrow$

$\mathrm{Var}_\mu(P_t f) \xrightarrow{t\to\infty} \mathrm{Var}_\mu(\mu(f)) = 0$

(constant vector: $\mu(f) = E_\mu[f(Z)]$)

$$\therefore \mathrm{Var}_\mu(f) = \mathrm{Var}_\mu(P_0 f) - \underbrace{\mathrm{Var}_\mu(\mu(f))}_{=0}$$

$$= \mathrm{Var}_\mu(P_0 f) - \lim_{t\to\infty} \mathrm{Var}_\mu(P_t f)$$

Fundamental Thm of Calculus

$$= -\int_0^\infty \frac{d}{dt}\mathrm{Var}_\mu(P_t f)\, dt$$

Lemma 1

$$= 2\int_0^\infty \mathcal{E}(P_t f, P_t f)\, dt$$

∎

Proof of Main Theorem (2.18 in RvH)

① $\mathrm{Var}_\mu(f) \le c \cdot \mathcal{E}(f,f)$

② $\|P_t f - \mu f\|_{L^2(\mu)} \le e^{-t/c} \|f - \mu f\|_{L^2(\mu)}$

③ $\mathcal{E}(P_t f, P_t f) \le e^{-2t/c} \mathcal{E}(f,f)$

④ Given $f \in L^2(\mu)$, $\exists\ K(f)$ s.t.
$\|P_t f - \mu(f)\|_{L^2(\mu)} \le K(f) e^{-t/c}$

⑤ Given any $f \in L^2(\mu)$, $\exists\ K(f)$ s.t.
$\mathcal{E}(P_t f, P_t f) \le K(f) e^{-2t/c}$

② ⇒ ④ immediate by defn

③ ⇒ ⑤ "

---

③ ⇒ ①: ③ $\mathcal{E}(P_t f, P_t f) \le e^{-2t/c} \mathcal{E}(f,f)$ ⇒ ① $\mathrm{Var}_\mu(f) \le c \cdot \mathcal{E}(f,f)$

$$\mathrm{Var}_\mu(f) \overset{\text{Corr. 3}}{=} 2\int_0^\infty \mathcal{E}(P_t f, P_t f)\, dt$$

$$\overset{③}{\le} 2\mathcal{E}(f,f) \int_0^\infty e^{-t/c}\, dt = c\,\mathcal{E}(f,f)$$

---

① ⇒ ② ① $\mathrm{Var}_\mu(f) \le c \cdot \mathcal{E}(f,f)$ ⇒ ② $\|P_t f - \mu f\|_{L^2(\mu)} \le e^{-t/c} \|f - \mu f\|_{L^2(\mu)}$

Lemma 1 ⇒

$$\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\mathcal{E}(P_t f, P_t f)$$
$$\overset{①}{\le} -\frac{2}{c}\mathrm{Var}_\mu(P_t f)$$

Lemma 1:
$\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\mathcal{E}(P_t f, P_t f)$

(∵ From ①, $-\frac{1}{c}\mathrm{Var}_\mu(g) \ge -\mathcal{E}(g,g)$)

$$\therefore\ \mathrm{Var}_\mu(P_t f) \le e^{-2t/c} \cdot \mathrm{Var}_\mu(P_0 f)$$
$$= e^{-2t/c}\mathrm{Var}_\mu(f).$$

(solve linear ODE)

$$\therefore \|P_t f - \mu(f)\|^2_{L^2(\mu)} = \mathrm{Var}_\mu(P_t f) \le e^{-2t/c}\,\mathrm{Var}_\mu(f)$$
$$= e^{-2t/c}\,\|f - \mu(f)\|^2_{L^2(\mu)}$$

---

(2) ⇒ (1) (2) $\|P_t f - \mu f\|_{L^2(\mu)} \le e^{-t/c}\,\|f - \mu f\|_{L^2(\mu)}$ ⇒ (1) $\mathrm{Var}_\mu(f) \le c \cdot \mathcal{E}(f,f)$

($\|P_t f - \mu f\|_{L^2(\mu)} = \mathrm{Var}_\mu(P_t f)$, $\|f - \mu f\|_{L^2(\mu)} = \mathrm{Var}_\mu(f)$)

> Lemma 1:
> $\frac{d}{dt}\mathrm{Var}_\mu(P_t f) = -2\mathcal{E}(P_t f, P_t f)$

$$2\mathcal{E}(f,f) = \lim_{t\to 0} 2\mathcal{E}(P_t f, P_t f)$$

Lemma 1

$$= -\lim_{t\to 0} \frac{d}{dt}\mathrm{Var}_\mu(P_t f)$$

$$= \lim_{t\to 0} \left(\lim_{h\to 0} \frac{-\mathrm{Var}_\mu(P_{t+h} f) + \mathrm{Var}_\mu(P_t f)}{h} \right)$$

$$= \lim_{h\to 0} \frac{\mathrm{Var}_\mu(f) - \mathrm{Var}_\mu(P_h f)}{h}$$

(1)

$$\ge \lim_{h\to 0} \mathrm{Var}_\mu(f)\left(\frac{1 - e^{-h/c}}{h}\right) = \frac{2}{c}\mathrm{Var}_\mu(f)$$

$$\therefore \quad (5) \Leftarrow (3) \Rightarrow (1) \Leftrightarrow (2) \Rightarrow (4)$$

The remaining proofs require reversibility.

Lemma 4 (2.32 in RvH) Suppose $\{P_t\}$ reversible.

Then, $\ln \mathrm{Var}_\mu(P_t f)$ and $\ln \mathcal{E}(P_t f, P_t f)$ are both convex in t.

Pf: Reversibility $\Rightarrow$ $\langle P_t f, g\rangle_\mu = \langle f, P_t g\rangle_\mu$

$\langle \alpha f, g\rangle_\mu = \langle f, \alpha g\rangle_\mu$

Recall: defn $\frac{d}{dt} P_t f = \alpha P_t f = P_t \alpha f$

$$\frac{d}{dt}\mathcal{E}(P_t f, P_t f) = -\frac{d}{dt}\langle P_t f, \alpha P_t f\rangle_\mu$$

$$= -\Big(\langle \alpha P_t f, \alpha P_t f\rangle_\mu + \underbrace{\langle P_t f, \alpha^2 P_t f\rangle_\mu}_{= \langle \alpha P_t f, \alpha P_t f\rangle_\mu \text{ ∵ reversibility}}\Big)$$

$$= -2\,\|\alpha P_t f\|^2_{L^2(\mu)}.$$

Using this, we can show:

$$\frac{d^2}{dt^2} \ln \mathrm{Var}_\mu(P_t f) = 4 \frac{\|\alpha P_t f\|^2_{L^2(\mu)}}{\mathrm{Var}_\mu(P_t f)} - 4 \frac{\mathcal{E}(P_t f, P_t f)}{(\mathrm{Var}_\mu(P_t f))^2}$$

$$= \frac{4}{\mathrm{Var}_\mu(P_t f)} \left(\mathrm{Var}_\mu(P_t f) \, \|\alpha P_t f\|^2_{L^2(\mu)} - \langle P_t f, \alpha P_t f \rangle^2_\mu \right)$$

Also, $P_t 1 = 1$, $\alpha 1 = \frac{d}{dt} P_t 1 = 0$.

$\Rightarrow \langle \underset{\sim}{\mu f}, \alpha P_t f \rangle_\mu = \langle \alpha(\mu f), P_t f \rangle_\mu = 0$ (= constant vector)

$\therefore \langle P_t f, \alpha P_t f \rangle^2_\mu = \langle P_t f - \mu f, \alpha P_t f \rangle^2_\mu$

$\overset{\text{Cauchy-Schwartz Ineq.}}{\leq} \mathrm{Var}_\mu(P_t f) \, \|\alpha P_t f\|^2_{L^2(\mu)}$

$\therefore \frac{d^2}{dt^2} \ln \mathrm{Var}_\mu(P_t f) \geq 0 \Rightarrow \ln \mathrm{Var}_\mu(P_t f)$ is convex w.r.t. t.

Similar proof to show convexity of $\ln \mathcal{E}(P_t f, P_t f)$ w.r.t. t.

□

Lemma 5: (2.33 in RvH book) Suppose $g: \mathbb{R}_+ \to \mathbb{R}$ is convex and $g(t) \le k - \beta t \ \forall\, t \ge 0$, and for some $k \in \mathbb{R}$.
Then, we have $g(t) \le g(0) - \beta t \quad \forall\, t \ge 0$.

Pf: (see RvH book) ∎

We now prove ④ ⇒ ② ⇒ ③, and ⑤ ⇒ ③.

① $\mathrm{Var}_\mu(f) \le c \cdot \mathcal{E}(f,f)$

② $\|P_t f - \mu f\|_{L^2(\mu)} \le e^{-t/c} \|f - \mu f\|_{L^2(\mu)}$

③ $\mathcal{E}(P_t f, P_t f) \le e^{-2t/c} \mathcal{E}(f,f)$

④ Given $f \in L^2(\mu)$, $\exists\ K(f)$ s.t. $\|P_t f - \mu(f)\|_{L^2(\mu)} \le K(f) e^{-t/c}$

⑤ Given any $f \in L^2(\mu)$, $\exists\ K(f)$ s.t. $\mathcal{E}(P_t f, P_t f) \le K(f) e^{-2t/c}$

Proof:

② ⇒ ③. ② $\|P_t f - \mu f\|_{L^2(\mu)} \le e^{-t/c} \|f - \mu f\|_{L^2(\mu)}$ ⇒ ③ $\mathcal{E}(P_t f, P_t f) \le e^{-2t/c} \mathcal{E}(f,f)$

From Lemma 2, $\dfrac{d^2}{dt^2} \ln \mathrm{Var}_\mu(P_t f) \ge 0$

$\therefore$ $\dfrac{d}{dt} \ln \mathrm{Var}_\mu(P_t f)$ is increasing w.r.t t.

Also, $\frac{d}{dt} \ln \mathrm{Var}_\mu(P_t f) = -2 \frac{\mathcal{E}(P_t f, P_t f)}{\mathrm{Var}_\mu(P_t f)}$

$\therefore \quad -2 \frac{\mathcal{E}(P_t f, P_t f)}{\mathrm{Var}_\mu(P_t f)} \geq -2 \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\mu(f)}$

$\therefore \quad \frac{\mathcal{E}(P_t f, P_t f)}{\mathcal{E}(f,f)} \leq \frac{\mathrm{Var}_\mu(P_t f)}{\mathrm{Var}_\mu(f)} = \frac{\|P_t f - \mu(f)\|^2_{L^2(\mu)}}{\|f - \mu(f)\|^2_{L^2(\mu)}}$

---

④ $\Rightarrow$ ② ④ Given $f \in L^2(\mu)$, $\exists$ $K(f)$ s.t. $\|P_t f - \mu(f)\|_{L^2(\mu)} \leq K(f) e^{-t/c}$ $\Rightarrow$ ② $\|P_t f - \mu f\|_{L^2(\mu)} \leq e^{-t/c} \|f - \mu f\|_{L^2(\mu)}$

Taking $\ln(\cdot)$, ② leads to: $\ln \|P_t f - \mu(f)\|^2_{L^2(\mu)} \leq \ln K(f) - \frac{2}{c} t.$

Now, Lemma 2 ($\ln(\cdot)$ convex) + Lemma 3 ($g(\cdot)$ convex with $g(\cdot) \leq k - \beta t$)

$\Rightarrow \ln \|P_t f - \mu(f)\|_2^2 \leq \ln \|P_0 f - \mu(f)\|_2^2 - \frac{t}{c}$

---

⑤ $\Rightarrow$ ③ is analogous as above.