*

# Concentration and Functional Inequalities and their Relation to Markov Processes

**Part III Essay, 2019**

*

# Overview

The overall goal of this essay is to give a brief view of how functional inequalities can be used to describe the convergence of a Markov process. However, as functional inequalities are characterized only by internal geometric structure of the space (*e.g.* measure, metric and differential), so should be the construction of the Markov process. The discussions in **Chapter 1** will give a sufficient amount of detail on this construction, in a rather functional analytic language. Then in **Chapter 2**, some candidates of functional inequalities in shaping the convergence of the Markov process will be presented.

Main theoretical progresses in understanding the relation between the log-Sobolev inequality or the Sobolev inequality with Markov processes are given in **Chapter 3** and **4**. In this process, remarkable equivalences between regularity properties of the semigroup and the functional inequalities will be unveiled, and relationships between the functional inequalities will be derived.

In **Chapter 5** and **6**, applications of the log-Sobolev inequality and the Sobolev inequality will be examined. With log-Sobolev inequality, Zegarlinski's proof for convergence of spin-lattice system to a thermal equilibrium will be reproduced. For an application of Sobolev inequality, the semigroup associated to Schrödinger-type operator will be used to analyse the spectrum of the Schrödinger's equation.

# Contents

# Chapter 1 DIRICHLET FORMS AND SEMIGROUPS OF MARKOV PROCESSES

## Section 1.1 Characterizing diffusivity of a Markov process

In this essay, we are going to discuss about convergence of different Markov processes to an equilibrium measure. The aim of this chapter is to transfer the problem of convergence of Markov processes to the problem of 'regularization' properties of functions under the action of transition semigroup, which will be understood in the language of operators and functional inequalities associated to them. While the stochastic processes are described suitably well in a standard language of stochastic calculus, the language of functions and operators requires some degree of background on functional analysis and in particular, so called the Hille-Yosida theory. Although two theories are developed in parallel historically, while the language of stochastic calculus is fairly intuitive, that of functional analysis is not. So in this chapter, we devote good amount of time on how we can transfer from one to the other.

Although construction of a most general class of continuous Markov processes does not involve the 'diffusivity' of the process, it is quite natural to think of the 'diffusivity' as the key character of a physical continuous Markov process. Intuitively, if we think of a Markov process as a heat propagating across a physical solid medium, then the temperature distribution at a given time would be completely determined by the geometry and the heat diffusivity of the medium. Moreover, if we assume medium is a Riemannian manifold, then it is plausible that we can make, at least in a local level, a change of coordinate to a Euclidean space with a modification of the diffusivity of the medium to describe the same heat propagation. Also, if we are given a stopped Markov process at a given boundary, then we can just treat the process as being killed at the boundary. In this regard, we will insist to characterize the diffusivity of a system by inspecting a concrete example, and introduce the notion of semigroups and Dirichlet forms. After this, we will see how we can define a (notion of) Markov process starting from a Dirichlet form acting on an algebra of functions on the space $(X, d)$ in the next section. This approach will be helpful not only in motivating the definition of the Dirichlet form, but also instructive in how we tackle the problem of treating a Markov diffusion process.

But also note that in spite of the fact the formal setting in **Section 1.2** is more apt for developing general theories, this formality occasionally blurs the intuition and complicates the problem unnecessarily. To complement this, we give a slightly different approach with a better intuition and with less generality in this section.

Let us first think about the Markov process on real vector space, $\mathbb{R}^d$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$. We denote $\mathcal{B}$ by the Borel $\sigma$-algebra generated by the Euclidean norm on $\mathbb{R}^d$ and $L^\infty(\mathbb{R}^d)$ be the set of bounded Borel functions. We also let $m$ be the Lebesgue measure on $\mathbb{R}^d$. If a continuous adapted random process $(X_t)_{t\geq 0}$ has Markov property, then we may think of a collection of operators $(P_t)_{t\geq 0}$ acting on $L^\infty(\mathbb{R}^d)$ by

$$P_t : L^\infty(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d), \quad P_t u(x) = \mathbb{E}_x(u(X_t)) = \mathbb{E}(u(X_t)|X_0 = x)$$

given that the $u(X_t^x)$ is measurable in $x$. (Such measurability is guaranteed if the process is given a solution of a Lipschitz coefficient SDE.) Indeed, $\mathbb{E}(f(X_t))$ is uniformly bounded by $\|u\|_{L^\infty(\mathbb{R})}$ for any $u \in f \in L^\infty(\mathbb{R}^d)$. We can already see some crucial features of $P_t$ from this simple setting :

$$
\begin{array}{ll}
(i) & P_0 u = u \quad \forall u \in L^\infty(\mathbb{R}) \\
(ii) & P_t \mathbf{1}_{\mathbb{R}^d} = \mathbf{1}_{\mathbb{R}^d} \\
(iii) & P_t u \geq 0 \quad \forall u \geq 0 \\
(iv) & \|P_t\| \leq 1 \\
(v) & P_{t+s} = P_t P_s \quad \forall s, t \geq 0
\end{array} \tag{1.1}
$$

where point $(v)$ follows from Markov property of $X_t$ and appropriate regularity of the space. If we give up the measure-preserving property and instead $X_t$ to go to a cemetery state $\infty$ after some stopping times $T$, and let $u(\infty) = 0$, then we have

$$
(ii^*) \quad P_t \mathbf{1}_{\mathbb{R}^d} \leq \mathbf{1}_{\mathbb{R}^d}
$$

*i.e.* $P_t$ has sub-Markov property. We can also go in the converse direction.

**Definition**

(1) *A collection of operators $(P_t)_{t\geq 0}$ acting on $L^\infty(\mathbb{R}^d)$ is called a* ***Markov transition semigroup*** *if it satisfies* (1.1)*. If $(ii)$ is replaced by $(ii^*)$, then $(P_t)_{t\geq 0}$ is called a* ***sub-Markov transition semigroup****.*

(2) *An adapted continuous process $(X_t)_{t\geq 0}$ is called a* ***Markov process*** *with transition semigroup $(P_t)_{t\geq 0}$ if it satisfies*

$$
P_s u(x) = \mathbb{E}_x(u(X_t)|\mathcal{F}_s) \quad \forall u \in L^\infty(\mathbb{R}),\ t \leq s
$$

Let us start to think about a specific example. The most simple Markov process we can think about is the $\kappa$-Brownian motion that has transition semigroup $P_t^B$ defined by

$$
P_t u(x) = \int_{\mathbb{R}^d} \frac{1}{(2\pi\kappa t)^{d/2}} e^{-\frac{\|x-y\|}{2\kappa t}} u(y) dy
$$

Then the transition semigroup is generated by an infinitesimal generator $\mathcal{G} = -\frac{\kappa}{2}\triangle$ in the sense that

$$
\frac{d}{dt}(P_t u) = \mathcal{G} P_t u = P_t(\mathcal{G} u) \quad \forall u \in W^{2,\infty}(\mathbb{R}^d)
$$

where $W^{k,p}(U) \subset L^p(U)$ is the Sobolev space. In this sense, we may also denote $P_t = \exp(\mathcal{G}t)$. There is in fact a more general theory which guarantees the existence of such generator.

**Theorem 1.1** *Let $(X_t)_{t\geq 0}$ be a Markov process in $\mathbb{R}^d$ with transition semigroup $(P_t)_{t\geq 0}$ such that (i) $\|P_t - id\| \to 0$ as $t \to 0^+$ and (ii) $P_t f \in C_0(\mathbb{R}^d)$ for any $f \in C_0(\mathbb{R}^d)$. Then there is a sub-algebra $\mathcal{D}_P$ and an operator $\mathcal{G} : \mathcal{D}_P \to L^\infty(\mathbb{R}^d)$ such that*

$$
\frac{d}{dt} P_t u = \mathcal{G} P_t u = P_t \mathcal{G} u \quad \forall u \in \mathcal{D}_P
$$

*and $\mathcal{D}_P$ is dense in $C(\mathbb{R}^d)$ with uniform norm. Such $\mathcal{G}$ is called the* ***infinitesimal generator*** *of $P_t$ and $\mathcal{D}_P$ is called the* ***domain of*** *$(P_t)_{t\geq 0}$. This equation is called the* ***Fokker-Planck equation****, or the* ***heat equation****.*

We do not prove this theorem, but one may want to compare this statement with **Proposition 1.2**, which is more useful in general. While the notion of infinitesimal generator encodes in full generality the information about the 'diffusivity' of a Markov process given that the semigroup $(P_t)_{t\geq 0}$ satisfies conditions $(i)$ and $(ii)$ of **Theorem 1.1**, there is still an equivalently useful object called carré du champ operator that encodes the 'stress-density' of a function in the spirit of Dirichlet principle.

**Definition** *Let $\mathcal{G}$ be an infinitesimal generator of a Markov process. Then we define the **carré du champ operator** of the process as*

$$\Gamma : \mathcal{D}_P \times \mathcal{D}_P \to L^\infty(\mathbb{R}^d), \quad (f,g) \mapsto \frac{1}{2}(\mathcal{G}(fg) - f\mathcal{G}(g) - g\mathcal{G}(f))$$

*If $f = g$, we also write $\Gamma(f,f) = \Gamma(f)$. We also define the **Dirichlet form** associated to $\Gamma$ and with respect to a measure $\mu$ (such that $\mu \ll m$)*

$$\mathcal{E}(f,g) = \int_{\mathbb{R}^d} \Gamma(f,g)(x)d\mu(x)$$

*The usual choice of $\mu$ is the **invariant measure** of $P_t$, i.e. the measure $\mu \ll m$ that has density function $u \geq 0$ with $P_t u = u$ a.e. $\forall t \geq 0$ (if it exists). We also denote $\mathcal{E}(f) = \mathcal{E}(f,f)$.*

Asking of the existence of an invariant measure will also be an interesting question, but the discussion is usually case-specific, so we do not discuss about the problem in generality in this essay. Also in most of the cases, the invariant measure can be given explicitly and the invariance would follow from direct computation. For a non-trivial example, see **Chapter 5** for the Gibbs measure.

To see how a Dirichlet form relates to the problem of convergence, recall the following :

**Generalized Dirichlet Principle** *Let $U \subset \mathbb{R}^d$ be a bounded open set and let $\mathcal{G}u(x) = \sum_{i,j=1}^d \partial_i(a^{ij}(x)\partial_j u(x))$ in a weak sense where $(a^{ji})_{i,j=1}^d \subset C(\mathbb{R}^d)$ is symmetric and uniformly elliptic. Let $g$ be a continuous function in $\mathbb{R}^d$. Then the solution $u \in C(\overline{U}) \cap C^2(U)$ of the Dirichlet problem*

$$\begin{cases} Lu = 0 & \textit{in } U \\ u = g & \textit{on } \partial U \end{cases}$$

*is the (unique) minimizer of the integral form*

$$\mathcal{E}'[u] = \int_U \sum_{i,j=1}^d a^{ij}(x)\partial_i u(x)\partial_j u(x)dm(x)$$

*given the boundary condition $u = g$ on $\partial U$.*

Since $\mathcal{G}$ is a non-positive linear operator, we can use the construction in **Section 1.2** to find a (not necessarily a Markovian) semigroup $P_t$ that solves $\frac{d}{dt}P_t = \mathcal{G}P_t$ by taking exponential, $P_t = \exp(-\mathcal{G}t)$. Then we would have

$$\frac{d}{dt}\mathcal{E}'[P_t u] = -2\int_U (\mathcal{G}P_t u)^2 dm \leq 0, \quad \forall t \geq 0$$

so optimistically, application of $P_t$ on any function $u$ will give a minimiser of $\mathcal{E}'$. This principle would also be useful in our setting, but in a much more rudimentary way : the minimiser would always be a constant function.

Also, looking at the heat equation $\partial_t P_t f = \mathcal{G} P_t f$ with $\mathcal{G} : \mathcal{D}(\mathcal{G}) = W^{2,\infty}(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d)$,

$$\mathcal{G} f(x) = \sum_{i,j=1}^{d} \partial_i (a^{ij}(x) \partial_j u(x))$$

gives a different useful interpretation of $\Gamma$. If $\Gamma$ is computed using the formula $\Gamma(f,g) = \frac{1}{2}(\mathcal{G}(fg) - f\mathcal{G}(g) - g\mathcal{G}(f))$ then we have

$$\Gamma(f,g) = \sum_{i,j=1}^{d} a^{ij}(x) \partial_i f \partial_j g$$

So the coefficients of $\Gamma(f,g)$ are exactly the diffusivity of a physical medium.

### § The issue of differentiability

We will often have need to differentiate the form $\int_X g P_t f d\mu$ in time. Intuitively, the time derivative should be

$$\frac{d}{dt} \int_X g P_t f d\mu = \int_X g \mathcal{G} P_t f d\mu$$

However, the problem is that we can not say for sure that the derivative and the integral commutes or not. The way to resolve this issue is identical to what we had been doing so far : we just simply say that there is a domain $\mathcal{D}'_P$ on which we can do such operations, which is also dense in $L^p(\mu)$ for each $L^p$. Then we can develop our theory under assumption that $f \in \mathcal{D}'_P$ and extend our conclusion later on using density of $\mathcal{D}'_P$ in $L^p(\mu)$.

**Proposition 1.2** *Under assumptions of **Theorem 1.1**, there is a domain $\mathcal{D}'_P$ that is dense in each $L^p(\mu)$ $(p \in [1,\infty))$ and for each $g \in L^1(\mu)$, $f \in \mathcal{D}'_P$,*

$$\frac{d}{dt} \int_X g P_t f d\mu = \int_X g \mathcal{G} P_t f d\mu.$$

*Proof. (sketch)* Consider

$$\mathcal{D}'_P = \operatorname{span}\left\{ \int_0^\infty u(t) P_t f dt : u \in C_c^\infty(\mathbb{R}_{>0}), f \in L^\infty(\mu) \right\} \subset L^\infty(\mu).$$

Note that, by contraction property of $P_t$ (**Proposition 1.4**), $\sup_t \|P_{t+s} f - P_s f\|_\infty \to 0$ as $s \to 0$, so $\int_0^\infty u(t) P_t v dt$ converges as a Riemann integral for each $u \in C_c^\infty(\mathbb{R}_{>0})$, $f \in L^\infty(\mu)$. Also, note that

$$P_s \int_0^\infty u(t) P_t f dt = \int_0^\infty u(t) P_{t+s} f dt = \int_s^\infty u(t-s) P_t f dt$$

so any element of $\mathcal{D}'_P$ applied with $P_s$ is differentiable in time with

$$\frac{d}{ds}\left( P_s \int_0^\infty u(t) P_t f dt \right) = -\int_s^\infty u'(t-s) P_t f dt \in \mathcal{D}'_P \subset L^\infty(\mu).$$

Therefore $\frac{d}{dt}\int_X gP_tfd\mu = \int_X g\mathcal{G}P_tfd\mu$ for any $g \in L^1(\mu)$. (Actually, we need that $f \in \mathcal{D}_P$, but we have not specified what $\mathcal{D}_P$ is earlier, so let us just assume this here.)

To check $\mathcal{D}'_P$ is dense in $L^p(\mu)$ for any $p \in [1, \infty)$, just note that

$$\left\|\lambda \int_0^\infty \eta(\lambda t) P_t f dt - f\right\|_p \leq \lambda \int_0^\infty \eta(\lambda t)\|P_t f dt - f\|_p \to 0 \quad \text{as } \lambda \to \infty$$

for any $\eta \in C_c^\infty((0,\infty))$, $\int_0^\infty \eta(t)dt = 1$.

This proposition is not meant to be used as an alternative of **Theorem 1.1**, because we are assuming the differentiability of $P_tf$ here. Whenever we differentiate $\int gP_tfd\mu$, we are going to assume $f \in \mathcal{D}'_P$ even if it is not explicitly mentioned. Also, later on in **Section 1.2**, $\mathcal{D}_\mathcal{E}$ will be used in place of $\mathcal{D}'_P$.

### § Self-adjoint semigroup and reversible measure

As mentioned earlier, assuming the existence of an invariant measure is crucial in our discussion. However, this assumption is often not exhaustive. Again retracting to the intuition, a Markov process describing a heat diffusion process (without a Maxwell's demon) should have no preference over a direction. This suggests that the process should be reversible in the sense that a time-reversed process should be indistinguishable from the original process, *i.e.* the law of $(Z_{t_0-t})_{t\in[0,t_0]}$ conditioned with $Z_{t_0} = x$ should be the same as that of $(Z_t)_{t\in[0,t_0]}$ conditioned with $Z_0 = x$. In language of the semigroup, this is written as the following.

**Definition** *An invariant measure is called to be* ***reversible*** *with respect to a Markov process with transition semigroup $(P_t)_{t\geq 0}$ if*

$$\int_{\mathbb{R}^d} fP_tgd\mu = \int_{\mathbb{R}^d} gP_tfd\mu \quad \forall f, g \in L^\infty(\mu).$$

This definition is not always the most useful form in many applications. Rather, it has many equivalent forms that can be chosen depending on the setting of the problems.

**Lemma 1.3** *Let $f, g \in \mathcal{D}'_P$ and $\mu$ be a reversible measure. Then*

$$\int_{\mathbb{R}^d} f\mathcal{G}gd\mu = -\int_{\mathbb{R}^d} \Gamma(f,g)d\mu$$

*Proof.* Use integration by parts. Since $\int \mathcal{G}hd\mu = \frac{d}{dt}\int P_thd\mu|_{t=0} = \frac{d}{dt}\int hP_t\mathbf{1}d\mu|_{t=0} = 0$ for bounded $h$, we have

$$\int_X \Gamma(f,g)d\mu = \frac{1}{2}\int_X \mathcal{G}(fg) - f\mathcal{G}(g) - g\mathcal{G}(f)d\mu = -\frac{1}{2}\int_X f\mathcal{G}(g) + g\mathcal{G}(f)d\mu$$

Also, making use of reversibility, we have

$$\int_X f\mathcal{G}gd\mu = \frac{d}{dt}\int_X fP_tgd\mu = \frac{d}{dt}\int_X gP_tfd\mu = \int_X g\mathcal{G}fd\mu$$

and the desired result follows.

*(End of proof)* □

It is not hard to see that having $\int f\mathcal{G}gd\mu = \int g\mathcal{G}fd\mu$ or $\int f\mathcal{G}gd\mu = -\int \Gamma(f,g)d\mu$ is equivalent for $\mu$ being a reversible measure.

A useful fact is that if $(P_t)_{t\geq 0}$ is a Markov transition semigroup, then $\mu$ being reversible implies $\mu$ being invariant. Indeed, if $\mu$ is reversible, then

$$\int_{\mathbb{R}^d} \mathbf{1} P_t g d\mu = \int_{\mathbb{R}^d} g P_t \mathbf{1} d\mu = \int_{\mathbb{R}^d} g d\mu, \quad \forall t \geq 0$$

so $\mu$ is also invariant.

### § Contraction property of $P_t$

In relation to the convergence property of the Markov process, it would be useful to discuss about the contraction property of $P_t$ at this point. It makes a very first depiction of how $P_t$ smooths the function $f$ in time, and it would also be used as a fundamental inequality in developing theories further.

**Proposition 1.4** *(Contraction property of semigroup) Whenever $f, f^p \in \mathcal{D}_P$ and $p \in [1,\infty)$, we have*

$$|P_t f|^p \leq P_t(|f|^p) \quad \forall t \geq 0.$$

> *Proof 1.* Assume $f \geq 0$. The proof is immediate if we use the expression $P_t f(x) = \mathbb{E}_x(f(X_t))$ and Jensen's inequality :
>
> $$P_t(f^p)(x) = \mathbb{E}_x[f^p(X_t)] \geq (\mathbb{E}_x[f(X_t)])^p$$
>
> *(End of proof)* □

Instead, we can prove the contraction property without referring to the Markov process, if we are given a sufficient property on the generator or the carré du champ operator.

**Definition** *(Diffusion property) A Markov process is called a **diffusion process** if for any $\Xi \in C^\infty(\mathbb{R}^m, \mathbb{R})$ such that $\Xi(\mathbf{0}) = 0$ and $f_1, \cdots, f_m \in \mathcal{D}_P$, has $\Xi(f_1, \cdots, f_m) \in \mathcal{D}_P$ and for any $g \in \mathcal{D}_P$, has*

$$\Gamma(\Xi(f_1, \cdots, f_m), g) = \sum_{j=1}^{m} \partial_j \Xi(f_1, \cdots, f_m)\Gamma(f_j, g).$$

*$\mu$-almost everywhere, with a reversible measure $\mu$ associated to the process.*

> *Proof 2 of Proposition 1.3. (For $p \geq 2$)* (Assuming diffusion process) We may also prove the proposition without referring to the Markov process $(X_t)_{t\geq 0}$ associated to the semigroup. Without loss of generality, we may assume that $f \geq 0$. Then for non-negative $g \in \mathcal{D}'_P$, we have
>
> $$\begin{aligned}\int pf^{p-1}g\mathcal{G}fd\mu &= -p\int \Gamma(f^{p-1}g, f)d\mu = -p\int \left(f^{p-1}\Gamma(g,f) + g\Gamma(f^{p-1}, f)\right)d\mu \\ &\leq -p\int f^{p-1}\Gamma(g,f)d\mu = \int g\mathcal{G}(f^p)d\mu\end{aligned}$$

where we have used in the inequality the fact that $\Gamma(f^{p-1}, f) = (p-1)f^{p-2}\Gamma(f, f) \geq 0$. Therefore we have $pf^{p-1}\mathcal{G}f \leq \mathcal{G}(f^p)$ and so

$$\frac{d}{dt}P_t(f^p)\Big|_{t=0} = \mathcal{G}(f^p) \geq pf^{p-1}\mathcal{G}f = \frac{d}{dt}(P_tf)^p\Big|_{t=0}$$

Now let $T = \inf\{t \geq 0 : P_t(f^p) < (P_tf)^P \text{ on a non-null set}\}$. If $T < \infty$, then since $\|P_t - id\| \to 0$ as $t \to 0$, we have that $P_T(f^p) \geq (P_Tf)^p$ $\mu$-a.e., while

$$\frac{d}{dt}P_t(f^p)\Big|_{t=T} = \mathcal{G}(P_T(f^p)) \geq \mathcal{G}((P_Tf)^p) \geq p(P_Tf)^{p-1}\mathcal{G}(P_Tf) = \frac{d}{dt}(P_tf)^p\Big|_{t=T}$$

which is a contradiction. Therefore $\inf\{t \geq 0 : \exists x \text{ s.t. } P_t(f^p)(x) < (P_tf)^P(x) \text{ on a non-null set}\} = \varnothing$, *i.e.* $P_t(f^p) \geq (P_tf)^p$ for all $t \geq 0$.

*(End of proof)* □

**Corollary 1.5** *If $P_t\mathbf{1}_X = \mathbf{1}_X$, then $P_t$ acting on $L^p$ is a contraction for $p \in [1, \infty]$.*

**Remark :** These results will continue to hold in the construction in the next section for infinitesimal generator and transition semigroup under assuming diffusion property of the carré du champ operator. Once we have proved that the contraction property holds for $p \geq 2$, this can be used establish an associated process $(X_t)_{t\geq 0}$, then this in turn would prove contraction property for $p \geq 1$ case using the first proof.

## Section 1.2 Definitions and assumptions

In this section, we extend our definition working on $\mathbb{R}$ from the previous section to a Polish space $X$ to define an object called Markov triple. However, the most significance difference from the last section is not this generality, but is that it enables us to treat the problem of convergence in a more analytic fashion. In particular, although all Markov processes have a corresponding Markov triple, some Markov triple might not have a Markov process generating it, but is still useful in characterizing analytic properties of certain operators. In this concern, it is more natural to talk about a Markov triple than talking about a Markov process when working with functional inequalities.

When we talk about a diffusion process in a space $X$, it is necessary to talk about continuity of the process, and completeness of the space $X$. Also, in order to establish various results about the $\sigma$-algebra generated by the topology of $X$, it is reasonable to assume that the topology has a countable basis. In this consideration, we are always going to talk about Polish spaces $(X, d)$. That is, $(X, d)$ is a separable complete metric space. Also let $\mathcal{B}$ be the $\sigma$-algebra the Borel $\sigma$-algebra of $(X, d)$.

Let us first state with a reasonable amount of generality the definition of a Markov triple, and then present a simple example we have at back of our mind. As we are mainly focusing on the constructions and definitions, we often omit proofs.

Let $(X, d)$ be a locally compact space with sequence of compact sets $(K_n)$ such that $X = \cup_n K_n$, so in particular $(X, d)$ is a Polish space. Let $\mathcal{B}$ be the Borel $\sigma$-algebra generated by the metric and $\mu$ be a Radon probability measure on $(X, d)$. Let $\mathcal{D}_\Gamma \subset L^2(\mu)$ be a (linear-)subspace and $\Gamma : \mathcal{D}_\Gamma \times \mathcal{D}_\Gamma \to \mathcal{M} =: \{f : f \text{ is measurable}\}$ be a symmetric non-negative bilinear form. Let $\mathcal{D}_\mathcal{E} \subset \mathcal{D}_\Gamma$ be another subspace such that $\Gamma|_{\mathcal{D}_\mathcal{E}\times\mathcal{D}_\mathcal{E}}$ maps to $L^1(\mu)$ and define

$$\mathcal{E} : \mathcal{D}_\mathcal{E} \times \mathcal{D}_\mathcal{E} \to \mathbb{R}, \quad \mathcal{E}(f, g) = \int_X \Gamma(f, g)d\mu$$

We will define norm on $\mathcal{E}$ to be $\|\cdot\|_{\mathcal{E}} = \|\cdot\|_{L^2(\mu)} + \mathcal{E}(\cdot,\cdot)$, and assume that $\mathcal{D}_{\mathcal{E}}$ is complete with respect to the norms $\|\cdot\|_{\mathcal{E}}$. (If this is not the case, it is tempting to define $\mathcal{D}_{\mathcal{E}}$ as the completion, but the completion might not stay in $\mathcal{D}_{\Gamma}$. So we just instead insist that $\mathcal{D}_{\mathcal{E}} \subset \mathcal{D}_{\Gamma}$ is already complete.) We also assume $\mathcal{D}_{\mathcal{E}} \cap L^p(\mu)$ is dense in $L^p(\mu)$ for each $p \geq 1$. (The density except for $p = 2$ does not play a role, but makes proof easier later.) With this setting, we define a Markov triple.

**Definition** *We will call a tuple $(X, \mu, \Gamma)$ is a **Markov triple** if $f \in \mathcal{D}_{\mathcal{E}}$ and $g \in L^2(\mu)$ are such that $|g(x) - g(y)| \leq |f(x) - f(y)|$, $|g(x)| \leq |f(x)|$ then $g \in \mathcal{D}_{\mathcal{E}}$ and $\mathcal{E}(g) \leq \mathcal{E}(f)$.*

Although the assumption does not seem very sapient, the consequence is far-reaching.

**Theorem 1.6** *Given a Markov triple $(X, \mu, \Gamma)$, there is a non-positive (might be unbounded) self-adjoint linear operator $\mathcal{G}$ such that the domain of $\sqrt{-\mathcal{G}}$ equals to $\mathcal{D}_{\mathcal{E}}$ and $(u, \mathcal{G}v) = -\Gamma(u, v)$ for any $u, v$ in the domain of $\mathcal{G}$.*

*Conversely, given a non-positive self-adjoint operator $\mathcal{G}$, there is an operator $\Gamma$ that makes $(X, \mu, \Gamma)$ a Markov triple with $\mathcal{D}_{\mathcal{E}}$ dense in $L^2(\mu)$, complete with respect to $\|\cdot\|_{\mathcal{E}}$-norm and $(u, \mathcal{G}, v) = -\Gamma(u, v)$ for any $u, v$ in the domain of $\mathcal{G}$.*

*We say $\mathcal{G}$ is the **generator** associated to the triple and denote the domain as*

$$\mathcal{D}_{\mathcal{G}} \subset \text{Domain}(\sqrt{-\mathcal{G}}) = \mathcal{D}_{\mathcal{E}}.$$

Given a generator $\mathcal{G}$, we can prove a converse statement to **Theorem 1.1** to find a semigroup $(P_t)_{t\geq 0} = \exp(\mathcal{G}t)$ generated by the generator.

**Theorem 1.7** *Let $\mathcal{G}$ be a non-positive self-adjoint linear operator and $P'_t = \exp(\mathcal{G}t)$ on $\mathcal{D}_{\mathcal{G}}$. Then $P'_t$ has an extension $P_t : L^2(\mu) \to L^2(\mu)$ that satisfies $P_0 = id$, $\|P_t\|_{L^2} \leq 1$, $P_{t+s} = P_t P_s$ for any $t, s \geq 0$, $\|P_t - id\|_{L^2} \to 0$ as $t \searrow 0$ and*

$$\mathcal{G}f = -\lim_{t\searrow 0} \frac{P_t f - f}{t} \quad \forall f \in \mathcal{D}_{\mathcal{G}}$$

*i.e. $\mathcal{G}$ is a generator of $P_t$ in the sense of **Theorem 1.1**.*

*Such $P_t$ is called the **semigroup** associated to the Markov triple.*

Even more, we also have an analogous version of **Proposition 1.2** with $\mathcal{D}'_P$ replaced by $\mathcal{D}_{\mathcal{E}}$.

**Lemma 1.8** *For $u, v \in \mathcal{D}_{\mathcal{E}}$, we have*

$$\lim_{t\to 0} \Big(\frac{u - P_t u}{t}, v\Big)_{L^2(\mu)} = \mathcal{E}(u, v)$$

Having came this far, it is not still not clear whether the $(P_t)_{t\geq 0}$ satisfies (1.1), and it is not true that any Markov triple $(X, \mu, \Gamma)$ gives rise to a Markov transition semigroup. We first present a simple example where we can verity whether the associated semigroup is a Markov transition semigroup.

**Example :** Let $(X, \mu, \Gamma)$ be a Markov triple and $\Gamma$ satisfy *diffusion property*. That is, for any $\Xi \in C^\infty(\mathbb{R}^m, \mathbb{R})$ such that $\Xi(\mathbf{0}) = 0$ and $f_1, \cdots, f_m \in \mathcal{D}_\Gamma$, has $\Xi(f_1, \cdots, f_m) \in \mathcal{D}_\Gamma$ and for any $g \in \mathcal{D}_\Gamma$, has

$$\Gamma(\Xi(f_1, \cdots, f_m), g) = \sum_{j=1}^m \partial_j \Xi(f_1, \cdots, f_m)\Gamma(f_j, g),$$

Moreover, assume that constant functions are in $\mathcal{D}_\Gamma$ with $\Gamma(c, f) = 0$ whenever $c$ is a constant function, $f \in \mathcal{D}_\Gamma$. Then first note that for $c \in \mathbb{R}$, $f \in \mathcal{D}_\mathcal{E}$, we have $\Gamma(c, f) = 0$ so

$$\frac{d}{dt}(P_t c, f)_{L^2} = \frac{d}{dt}(c, P_t f)_{L^2} = -\mathcal{E}(c, P_t f) = 0$$

and therefore $(P_t c, f)_{L^2} = (c, f)_{L^2}$ for any $f \in \mathcal{D}_\mathcal{E}$, and in particular $P_t c = c$ for any $t \geq 0$. Also *Proof 2* of **Proposition 1.6** directly applies in the present setting, *i.e.* $P_t(f^p) \geq (P_t f)^p$ whenever $f \geq 0$, $f \in L^p(\mu)$, $p \in [2, \infty]$ and $t \geq 0$. This fact together with that $P_t c = c$ indicates that whenever $u(x) \in [0, 1]$ for each $x \in X$, then $P_t(2u - 1) \geq -1$ by $L^\infty$-contractivity, and so $P_t u \geq 0$, *i.e.* $P_t$ is *positivity preserving*. Then we may define a Markov process $(Z_t)_{t\geq 0}$ satisfying $\mathbb{P}(Z_t \in A | Z_0 = x) = P_t \mathbf{1}_A(x)$ - this makes sense only because $P_t$ is an $L^\infty$-contraction and positivity-preserving now. Then conversely we can repeat the construction of a Markov transition semigroup from the last section to see that $P_t$ is exactly the Markov transition semigroup for the process $(Z_t)_{t\geq 0}$.

As seen in the example of Markov triple with diffusion property, having the $L^\infty$-contraction property and positivity preserving property of $P_t$ are the crucial condition for the Markov triple to have a sub-Markovian transition semigroup. The property that $P_t c = c$ for any constant $c$ will also be of separate interest, in which case the Markov triple has an associated Markovian transition semigroup.

**Definition** *We call that $\mathcal{E}$ is a **Dirichlet form** on the $(X, \mu, \Gamma)$ if the associated semigroup $(P_t)_{t\geq 0}$ is a sub-Markov transition semigroup. This is equivalent to $P_t$ having (i) $L^\infty$-contraction property and (ii) positivity preserving property.*

*In this setting, we call $\Gamma$ a **carré du champ operator** and call the triple $(X, \mu, \Gamma)$ a **standard Markov triple**.*

*If in addition $P_t$ is **mass-preserving** in that $P_t \mathbf{1}_X = \mathbf{1}_X$ for any $t \geq 0$ (and in particular constant functions are in $\mathcal{D}_\mathcal{E}$), then the standard Markov triple is also called to preserve mass, and this is equivalent for $P_t$ being a Markov transition semigroup.*

A standard Markov triple will be assumed to be mass-preserving if it is used without an explicit sign. Given a standard Markov triple, we get plenty of approximation results that can be used in proving functional inequalities later on.

**Lemma 1.9** *Let $(X, \mu, \Gamma)$ be a standard Markov triple. Then the following hold :*

*(i)* *$\mathcal{D}_\mathcal{E}$ forms a lattice, i.e. whenever $f, g \in \mathcal{D}_\mathcal{E}$, we have $f \vee g \in \mathcal{D}_\mathcal{E}$. Moreover, $f \wedge c \in \mathcal{D}_\mathcal{E}$ for any constant function $c \in \mathbb{R}$ and $f \in \mathcal{D}_\mathcal{E}$.*

*(ii)* *For $c \in \mathbb{R}$, $(f \wedge c) \vee (-c) \to f$ as $c \to \infty$ in $\|\cdot\|_\mathcal{E}$-norm.*

*(iii)* *$\mathcal{D}_\mathcal{E} \cap L^\infty(\mu)$ forms a vector subspace of $L^2(\mu)$, and forms an algebra. By (ii), this is dense in $\mathcal{D}_\mathcal{E}$ with $\|\cdot\|_\mathcal{E}$-norm and dense in $L^2(\mu)$ with $L^2$-norm. Also, for $f, g \in \mathcal{D}_\mathcal{E} \cap L^\infty(\mu)$, the Dirichlet form satisfies bound*

$$\mathcal{E}(f \cdot g) \leq \left( \|f\|_\infty \sqrt{\mathcal{E}(g)} + \|g\|_\infty \sqrt{\mathcal{E}(f)} \right)^2.$$

Although we have defined a standard Markov triple starting from the carré du champ operator, we actually rarely construct a Markov process in this order. In the following example, we will see a Markov triple along with the infinitesimal generator and the transition semigroup and verify that this system satisfies the required properties *a posteriori*.

**Example :** (Generalized Ornstein-Uhlenbeck process) Let $(X, d) = (\mathbb{R}^d, \|\cdot\|)$, $\mu(dx) = \frac{1}{Z}e^{-F(x)}dx$ with $Z = \int_{\mathbb{R}^d} e^{-F(x)}dx < \infty$ and $F \in C^1(\mathbb{R}^d, \mathbb{R})$. Let $\mathcal{G}f(x) = \frac{1}{2}\triangle f - \frac{1}{2}\nabla F \cdot \nabla f$, $\Gamma(f, g) = \frac{1}{2}\nabla f \cdot \nabla g$, then setting $\mathcal{D}_\Gamma = W^{1,2}_{loc}(\mathbb{R}^d)$, the local Sobolev space, would be sufficient. Indeed, since $\Gamma(f, g) = \frac{1}{2}(\mathcal{G}(fg) - f\mathcal{G}(g) - g\mathcal{G}(f))$ holds, integration by parts gives

$$\begin{aligned}\int \Gamma(f, g)d\mu &= \frac{1}{2}\int(\nabla f \cdot \nabla g)e^{-F}dx \\ &= -\frac{1}{2}\int f(\triangle g)e^{-F}dx + \frac{1}{2}\int f(\nabla g \cdot \nabla F)e^{-F}dx = -\int f\mathcal{G}g d\mu\end{aligned}$$

for any $f, g \in \mathcal{D}_\mathcal{E} = W^{1,2}(\mu)$. We may also set $\mathcal{D}_\mathcal{G} = \mathcal{D}_\Gamma = W^{1,2}_{loc}(\mathbb{R}^d)$.

The transition semigroup would be $P_t f(x) = \mathbb{E}_x[f(X_t)]$ where $X_t$ is the solution to the SDE

$$dX_t = -\frac{1}{2}\nabla F(X_t)dt + dB_t, \quad X_0 = x$$

To verify this, use Ito's formula to see that $df(X_t) = (-\frac{1}{2}\nabla F(X_t) \cdot \nabla f(X_t) + \frac{1}{2}\triangle f(X_t))dt - \frac{1}{2}\nabla F(X_t) \cdot dB_t$ for any $f \in C^1(\mathbb{R}^d, \mathbb{R})$, and therefore

$$\frac{d}{dt}\mathbb{E}_x[f(X_t)]\Big|_{t=0} = -\frac{1}{2}\nabla F(x) \cdot \nabla f(x) + \frac{1}{2}\triangle f(x) = \mathcal{G}f(x)$$

Since the SDE governing the evolution of $X_t$ is homogeneous in time and this holds for any $x \in \mathbb{R}^d$, this is sufficient to verify that

$$\frac{d}{dt}P_t f(x) = \mathcal{G}P_t f(x) \quad \forall x \in \mathbb{R}^d, \forall t \in \mathbb{R}_{\geq 0}$$

Furthermore, we we automatically have that $(X, \mu, \Gamma)$ it is a Markov triple by **Theorem 1.6** without verifying the defining property. Finally, we see by induction that diffusion property holds for the carré du champ operator and $\Gamma(c, f) = 0$ for any $c \in \mathbb{R}$, $f \in \mathcal{D}_\Gamma$. Therefore $(X, \mu, \Gamma)$ is a mass-conserving standard Markov triple.

For one to invent a Markov triple, it is often the case that one should also think of the Markov process associated to the process, but once the Markov triple, infinitesimal generator and transition semigroup are obtained, we may only think of how the operators work on the function spaces when working with functional inequalities.

# Chapter 2 NOTIONS OF CONVERGENCE AND FUNCTIONAL INEQUALITIES

Ahead of actually jumping into proving any results about convergence of diffusion Markov processes, we first have to give characterisations for the convergence. Once given the notions of convergence, we will next start to think about how the condition for the convergence can be written just in terms of the carré du champ operator (or the Dirichlet form) and the measure in the Markov triple $(X, \mu, \Gamma)$. When applied to familiar Radon probability measures in $\mathbb{R}^d$, these will often be in a form that draws relations between function spaces. In this consideration, such relations, given in inequalities, are called functional inequalities.

Making a connection between the functional inequalities and the convergence of Markov processes is often not a trivial problem. While functional inequalities are just written in terms of functions and operators acting on them, Markov processes are probabilistic objects that often do not have explicit representations. So this is where the transition semigroup $(P_t)_t$ plays a significant role - it lays a bridge between the functional inequalities and the Markov processes. Although we can not specify $X_t$, if the Markov process converges, then the equality $P_t f(x) = \mathbb{E}_x(f(X_t))$ suggests that whenever $X_t$ converges to some distribution as $t \to \infty$, then $P_t f$ would also get better regularity properties as it evolves in time.

We will be discussing about three main functional inequalities in this section. The first one is Poincaré inequality associated to the variance of a function, the second one is Sobolev inequality associated to $L^p$-norm, and the final one is log-Sobolev inequality associated to the entropy. In this section, we will see the regularity property of the semigroup when Poincaré inequality is assumed. The regularity property is of more complicated nature for the case of Sobolev and log-Sobolev inequality, so defer the discussions to **Chapter 3** and **Chapter 4**.

Throughout this chapter, we will assume $(X, \mu, \Gamma)$ is a mass-preserving standard Markov triple with infinitesimal generator $\mathcal{G}$ and transition semigroup $(P_t)_{t\geq 0}$.

## Section 2.1 Variance and $L^p$-norm

We will get introduced to notions of variance and $L^p$-norm convergence in this section. The corresponding inequality to variance is the Poincaré inequality and the corresponding one for $L^p$-norms are Sobolev inequalities. They are introduced within the same context because two inequalities look similar, but their nature of convergence are very different.

**Definition** *Let $f \in L^1(X, \mu)$. Then the* ***variance*** *of $f$ is defined as*

$$\mathrm{Var}(f) = \int_X (f - \mu(f))^2 d\mu$$

If $P_t f$ converges to a constant function, then the variance would converge to 0 as $t \to \infty$. So it is natural to look at how $\mathrm{Var}(P_t f)$ evolves in time. Let us try to differentiate $\mathrm{Var}(P_t f)$ in time. For $f \in \mathcal{D}_{\mathcal{E}}$, we have

$$\begin{aligned} \frac{d}{dt}(\mathrm{Var}(P_t f)) &= \frac{d}{dt}\Big(\mu((P_t - \mu(f))^2)\Big) = \mu\Big(2(P_t - \mu(f))\mathcal{G}P_t f\Big) \\ &= -2\mu(\Gamma(P_t f, P_t f)) = -2\mathcal{E}(P_t f) \end{aligned}$$

In the first equality, we have used the fact that

$$\frac{d}{dt}\mu(P_t f) = \mu(\mathcal{G}P_t f) = \int_X \mathbf{1}\mathcal{G}P_t f d\mu = -\int_X \Gamma(\mathbf{1}, \mathcal{G}P_t f) d\mu = \int_X (\mathcal{G}\mathbf{1}) P_t f d\mu = 0$$

So if we have some inequality relation between the variance and the Dirichlet form, then we would have an exponential decay of the variance. This motivate us to define the following.

**Definition** *A Markov triple is said to satisfy **Poincaré inequality** with constant $C_P$ if*

$$\mathrm{Var}(f) \leq C_P \mathcal{E}(f) \quad \forall f \in \mathcal{D}_{\mathcal{E}}$$

*This is denoted as $PI(C_P)$.*

As an immediate consequence, we have

$$\mathrm{Var}(P_t f) \leq e^{-2t/C_P} \mathrm{Var}(f) \quad \forall t \geq 0$$

whenever $(X, \mu, \Gamma)$ satisfies $PI(C_P)$. But in fact, we also have that whenever the variance of $P_t f$ has exponential decay, the Markov triple should satisfy a Poincaré inequality.

**Theorem 2.1** *The following are equivalent :*

*(i)* $(X, \mu, \Gamma)$ *satisfies* $PI(C_P)$.

*(ii) For every function* $f \in L^2(\mu)$,

$$\mathrm{Var}(P_t f) \leq e^{-2t/C_P} \mathrm{Var}(f) \quad \forall t \geq 0$$

*(iii) For every function* $f \in L^2(\mu)$

$$\mathrm{Var}(P_t f) \leq e^{-2t/C_P} v(f) \quad \forall t \geq 0$$

*for some constant* $v(f) > 0$ *only depending on* $f$.

*Proof.* Proving $(i) \Rightarrow (ii)$ for $f \in \mathcal{D}_{\mathcal{E}}$ is already done. To prove this for $f \in L^2(\mu)$, we use density of $\mathcal{D}_{\mathcal{E}}$ in $L^2(\mu)$. First, whenever $f_n \to f$ as $n \to \infty$ in $L^2(\mu)$, we have

$$\begin{aligned}
\mathrm{Var}(P_t f_n) - \mathrm{Var}(P_t f) &= \int_X P_t(f_n - f) \cdot P_t(f_n + f) d\mu \\
&\leq \int_X (P_t(f_n - f))^2 d\mu \int_X (P_t(f_n + f))^2 d\mu \\
&\leq \|f_n - f\|_{L^2(\mu)} \|f_n + f\|_{L^2(\mu)} \to 0 \quad \text{as } n \to \infty
\end{aligned}$$

where we have used contraction property of $P_t$ in the last inequality. Therefore the inequality

$$\mathrm{Var}(P_t f) \leq e^{-2t/C_P} \mathrm{Var}(f) \quad \forall t \geq 0$$

still holds for all $f \in L^2(\mu)$.

Implication $(ii) \Rightarrow (iii)$ is immediate.

To prove $(iii) \Rightarrow (ii)$, note that $\log(\|P_t f\|_2)$ is convex as a function of $t \geq 0$. Indeed, whenever $f \in \mathcal{D}_{\mathcal{E}}$, we have

$$\begin{aligned}
\frac{d}{dt}\Big(\log(\mu((P_t f)^2))\Big) &= \frac{2\mu(P_t f \mathcal{G} P_t f)}{\mu((P_t f)^2)} \\
\frac{d^2}{dt^2}\Big(\log(\mu((P_t f)^2))\Big) &= 4\frac{\mu((\mathcal{G} P_t f)^2)\mu((P_t f)^2) - (\mu(P_t f \mathcal{G} P_t f))^2}{\mu((P_t f)^2)^2} \geq 0
\end{aligned}$$

by Cauchy-Schwarz inequality. For general $f \in L^2(\mu)$, use density of $\mathcal{D}_{\mathcal{E}}$ in $L^2(\mu)$ and that $\|P_t f\|_2$ is continuous in $f$ with respect to $L^2(\mu)$-norm to prove that $t \mapsto \log(\|P_t f\|_2)$ is convex. Now condition $(iii)$ and $(ii)$ can be reformulated as

$$(iii) \ : \ \lambda(t) \leq \log(v(f)) - \frac{2}{C_P}t \quad \forall t \geq 0$$
$$(ii) \ : \ \lambda(t) \leq \lambda(0) - \frac{2}{C_P}t \quad \forall t \geq 0$$

with $\lambda(t) = \log(\|P_t(f - \mu(f))\|_2)$ and convention $\log(0) = -\infty$. Suppose for contradiction that $(iii)$ holds, but $(ii)$ fails, say $\lambda(s) > \lambda(0) - \frac{2}{C_P}s$ for some (fixed) $s > 0$, so there is $\epsilon > 0$ with

$$\lambda(s) = \lambda(0) - \Big(\frac{2}{C_P} - \epsilon\Big)s > \lambda(0) - \frac{2}{C_P}s$$

Then by convexity, for any $\alpha \geq 1$,

$$\frac{\lambda(\alpha s) + (\alpha - 1)\lambda(0)}{\alpha} \geq \lambda(s) = -\Big(\frac{2}{C_P} - \epsilon\Big)s + \lambda(0)$$
$$\Rightarrow \qquad \lambda(\alpha s) \geq -\Big(\frac{2}{C_P} - \epsilon\Big)\alpha s + \lambda(0)$$

but by $(iii)$, we also have that $\lambda(\alpha s) \leq \log(v(f)) - \frac{2}{C_P}\alpha s$ for any $\alpha \geq 1$, which is a contradiction.

Implication $(ii) \Rightarrow (i)$ for comes from Taylor expanding $(ii)$ about $t = 0$, *i.e.* for $f \in \mathcal{D}_{\mathcal{E}}$,

$$\mathrm{Var}(P_t f) = \mathrm{Var}(f) - 2\mathcal{E}(f)t + o(t^2) \leq \Big(1 - \frac{2t}{C_P} + o(t^2)\Big)\mathrm{Var}(f)$$

so $PI(C_P)$ holds.

*(End of proof)* $\square$

From this result we immediately see that Poincaré inequality implies $P_t f \xrightarrow{t\to\infty} \mu(f)$ in $L^2(\mu)$. Also, if given the right regularity of the probability space and initial distribution, the associated Markov process would converge in distribution.

**Example :** Let $X = \mathbb{T}^d \simeq \mathbb{R}^d/(\mathbb{Z})^d$ $(L > 0)$ with Riemannian metric inherited from that of $\mathbb{R}^d$, and consider the Markov triple $(X, \mu, \Gamma)$ associated to the Brownian motion. Then $\mu(dx) = m(dx)$ is the Lebesgue measure and $\Gamma(f) = \frac{1}{2}|\nabla f|^2$ and $\mathcal{G} = -\frac{1}{2}\triangle$. But the spectrum of Laplacian in the cube is a subset of $\{4\pi^2 k : k \in \mathbb{Z}_{\geq 0}\}$, with only constant function in the eigenspace of 0. Therefore we have the bound $\|P_t|_H\|_{2,2} \leq e^{-8\pi^2 t}$ where $H$ is the orthogonal complement of constant functions in $L^2(\mathbb{T})$ and $PI(1/4\pi^2)$ holds.

**Example :** *A generalized Ornstein-Uhlenbeck process with curvature condition* Let $X = \mathbb{R}^d$, $\mu = e^{-F}dm$ and $\Gamma(f) = \frac{1}{2}|\nabla f|^2$ be the Markov triple for Ornstein-Uhlenbeck process. Further, assume that $D^2 F \geq \frac{1}{C} I_{\mathbb{R}^d}$. Then the system satisfies $PI(C)$. This follows from Bakry-Émery criterion (**Lemma 5.3**) and implication from a tight log-Sobolev inequality to a Poincaré inequality (**Proposition 3.8**).

Hence we have invented a reasonably useful and general notion of convergence only using relatively straightforward calculus, and we already see that functional inequalities stated just in terms of the equilibrium measure and the Dirichlet form can be used to characterize the concentration behaviour of a Markov process. There are more involved theories related to Poincaré inequality, but in this essay, we are only going to use Poincaré inequality as a tool for aiding different inequalities. Rather, we are now going to move on to study log-Sobolev inequalities and Sobolev inequalities.

### § Sobolev inequality

There is a notion of convergence stated in terms of $\|f-\mu(f)\|_{L^p(\mu)}$ for $p>2$ (when $p=2$, it is just the variance). However, we usually omit $\mu(f)$ here and just use $\|f\|_{L^p(\mu)}$ because the main purpose of using $L^p$-norm when $p>2$ is to measure the regularity of the function. Statements related to the regularity property and convergence in $L^p$ norm will be characterized and proved in **Chapter 4**. In the current section, we will get introduced to the corresponding functional inequality for $L^p$-convergence.

**Definition** *A Markov triple is said to satisfy **Sobolev inequality** with constants $C_S>0$, $\delta_S\geq 1$ if*

$$\|f\|_p^2 \leq C_S\Big(\delta_S\|f\|_2^2 + \mathcal{E}(f)\Big) \quad \forall f \in \mathcal{D}_{\mathcal{E}}.$$

*This is denoted as $SI^p(C_S,\delta_S)$, and when $\delta_S C_s = 1$, we call the inequality is **tight** and denote $SI^p(C_S, 1/C_S) = SI^p(C_S)$.*

Note that $\|f\|_p^2$ on the left hand side is compensated by $\delta_S C_S\|f\|_2^2$, so using a constant function as a test function, we see that having $\delta_S \geq 1/C_S$ is necessary. From the definition of the Sobolev inequality, it is far from clear if the Sobolev inequality implies a convergence in $L^p$-norm. However, we will be seeing in **Chapter 4** that a tight Sobolev inequality implies a Poincaré inequality and that there is time $\tau>0$ such that $\|P_t\|_{2,p}\leq M$ for some $M>0$ and $t\geq\tau$, therefore

$$\|P_t f - \mu(f)\|_p \leq M\|f-\mu(f)\|_2 \leq Me^{-2t/C'}\mathrm{Var}(f), \quad t\geq\tau.$$

One would also like to compare the Sobolev inequality with a classical Sobolev inequality.

**A classical Sobolev inequality in $\mathbb{R}^d$** *Let $U\subset\mathbb{R}^d$ be open, bounded and has $C^1$-boundary. If $1\leq q<n$, then there is a constant $C(d,q,U)>0$ such that*

$$\|f\|_{L^{q^*}(U)} \leq C(d,q,U)\|f\|_{W^{1,q}(U)}, \quad \forall f\in W^{1,q}(U)$$

*where $\frac{1}{q^*} = \frac{1}{q} - \frac{1}{n}$.*

If $U$ has a finite measure, then setting $q=2$, $p\leq q^* = \frac{2d}{d-2}$, $\Gamma(f,g) = \nabla f\cdot\nabla g$ and $\mu(dx) = m(dx)$ would give $SI^p(C(d,q,U), C(d,q,U))$. So it would be reasonable to expect that the 'dimension' of the system has to be greater or equal to 3 and exponent $p\leq\frac{2d}{d-2}$ to obtain a $SI^p(C_S,\delta_S)$.

Also note that, in this setting, the triple $(U,m,\Gamma)$ gives a Brownian motion killed at the boundary $\partial U$, so the process is only sub-Markovian. In particular, we would have $P_t f\to 0$ as $t\to\infty$ for any $f\in L^2(U)$. However, we never refer to a converging point in the definition of $SI^p(C_S,\delta_S)$ in contrast to the case of Poincaré inequality. So we may expect that the theory of Sobolev inequality is equally useful in dealing with killed processes, *i.e.* when $P_t$ is only a sub-Markovian semigroup.

## Section 2.2 Entropy

There is a seemingly very different concept of convergence of functions, called the entropy. However, when studied with sufficient care, it will be revealed that it can play a similar role to that of variance and $L^p$ norms of $p > 2$. One of the results that highlights this point is that the exponential decay of entropy implies the exponential decay of variance. One equivalent form of the exponential decay of entropy, stated in terms of the Dirichlet form is called the log-Sobolev inequality. ((The exponential decay of entropy and the equivalence will be dealt in **Chapter 3**.)

**Definition** *For a measure space $(X, \mathcal{B}, \mu)$ and a non-negative function $f \in L^1(\mu)$, we define the **entropy** of $f$ by*

$$\mathrm{Ent}_\mu(f) = \int_X f \log f d\mu - \Big(\int_X f d\mu\Big) \log\Big(\int_X f d\mu\Big) \in [0, +\infty]$$

*with convention $0 \log 0 = 0$. We often omit $\mu$ and write $\mathrm{Ent}(f) \equiv \mathrm{Ent}_\mu(f)$.*

*A Markov Triple $(X, \mu, \Gamma)$ is said to satisfy a **log-Sobolev inequality** with log-Sobolev constant $C_{lS} > 0$ and local norm $\gamma_{lS}$ if*

$$\mathrm{Ent}(f^2) \leq C_{ls}\Big(\gamma_{lS} \int_X f^2 d\mu + 2\mathcal{E}(f)\Big) \quad \forall f \in \mathcal{D}(\mathcal{E})$$

*and is denoted $LSI(C_{lS}, \gamma_{lS})$. Equivalently, whenever $f \geq 0$, $\sqrt{f} \in \mathcal{D}_\mathcal{E}$,*

$$\mathrm{Ent}(f) \leq C_{ls}\Big(\gamma_{lS} \int_X f d\mu + \frac{1}{2} \int_X \frac{\Gamma(f, f)}{f} d\mu\Big)$$

*When $\gamma_{lS} = 0$, then we say the inequality is **tight** and denote $LSI(C_{lS}, 0) = LSI(C_{lS})$.*

Again using a constant function as a test function, we see that we should expect $\gamma_{lS} \geq 0$.

For a sanity check, note that $\mathrm{Ent}(cf) = c\mathrm{Ent}(f)$ for any $c \geq 0$ a constant, and therefore both sides of $LSI(C, \gamma)$ scales with the same factor as we multiply $f$ with a constant factor. To see that $\mathrm{Ent}(f) \geq 0$ whenever $f \in L^1(\mu)$, $f \geq 0$, note that the function $F : x \mapsto x \log x$ is convex for $x \geq 0$ (with convention $0 \log 0 = 0$). Now Jensen's inequality implies

$$\int_X F(f) d\mu \geq F\Big(\int_X f d\mu\Big).$$

so $\mathrm{Ent}_\mu(f) = \mu(F(f)) - F(\mu(f)) \geq 0$.

Comparing with $SI^p(C_S, \delta_S)$, we see that $\mathrm{Ent}(f^2)$ plays the role of $\|f\|_{L^P}$, but at the same time, when $\gamma_{lS} = 0$, the role of $\mathrm{Ent}(f^2)$ is also comparable to the role of $\mathrm{Var}(f)$. So it will be not surprising to find out that there are analogues in the log-Sobolev inequality for some results on Poincaré inequality and Sobolev inequality.

### § Contraction property of semigroups and the spectral gap of infinitesimal generators

A generally-applicable strategy for proving Poincaré inequality is to find the 'spectral gap' of infinitesimal generator. In order to see how this makes sense, consider the space

$$H = \{f \in L^2(\mu) : f = \text{constant}\}^\perp = \{h \in L^2(\mu) : \mu(h) = 0\}$$

which is closed subspace of $L^2(\mu)$. Since $\frac{d}{dt}\int_X cP_t h d\mu = \int_X c\mathcal{G} h d\mu = \int h\mathcal{G} c d\mu = 0$ for any constant $c$ and $h \in H \cap D'_P$, we see that $P_t$ acts on $H$ as $P_t|_H : H \to H$. Then the exponential variance decay condition of Poincaré inequality can be stated as

$$\|P_t|_H h\|_2 \le e^{-t/C}\|h\|_2, \quad \forall h \in H.$$

Recalling the definition $P_t = e^{t\mathcal{G}}$ on domain $\mathcal{D}_\mathcal{E}$, this suggests that Poincaré inequality is equivalent to having having spectral gap of $\mathcal{G}$ at 0. That is, the spectrum of $\mathcal{G}$ excluding 0 is well-separated from 0.

One way to prove such a bound on $\|P_T|_H\|_{2,2}$ is to seek for a 'superior-contraction' property of $P_t$, *i.e.* to seek for $p < q$ and $t > 0$ with nice decay behaviour of $\|P_t\|_{p,q}$. For example, if we can prove the existence of $\tau > 0$ such that $\|P_\tau\|_{2,4} \coloneqq \sup\{\|P_\tau f\|_4 : \|f\|_2 \le 1\} \le 1$, then we would have, for any $f \in H$ with $\|f\|_2 = 1$ and $c \in \mathbb{R}$,

$$\begin{aligned}\|f + c\|_2^4 &= (\mu(f)^2 + c^2)^2 \\ &\ge \|P_\tau f + c\|_4^4 \ge c^4 + 6c^2\|P_\tau f\|_2^2 - 4|c|\|P_\tau f\|_3^3 + \|P_\tau f\|_4^4\end{aligned}$$

so $6c^2\|P_\tau f\|_2^2 + o(c^2) \le 2c^2\|f\|_2^2$ as $c \to \infty$. Therefore we have $\|P_\tau f\|_2 \le \frac{2}{3}\|f\|_2$. Because the semigroup $(P_t)_t$ is self-adjoint, we see that

$$\|P_t|_H\|_{2,2} \le \left(\frac{2}{3}\right)^{t/\tau},$$

which concludes that the system satisfies the Poincaré inequality. Studying carefully between the lines, it becomes convincing that a 'superior-contraction' property does not only makes us able to prove Poincaré inequality, but also gives a very good regularizing property of the semigroup as $t$ evolves. In later two chapters, we will see that this indeed is true for both Sobolev inequality and the log-Sobolev inequality in the name of 'Ultracontractivity' and 'Hypercontractivity' respectively.

### § A brief view on relations between the inequalities

In later two chapters, we will be studying how we can characterize the concentration properties of the semigroup under the Sobolev inequality and the log-Sobolev inequality. Apparently, the easiest way to prove a such a result is to exploit the results on the Poincaré inequality we already have seen in the current chapter. However, proofs do not go in a direct way as it seems. Instead, we will formulate various equivalent forms of the inequalities and show the relations between the equivalent forms. As a result, we will have the following :

$$SI^p(C,\delta) \Longrightarrow LSI(C,\delta)$$

$$SI^p(C) \Longrightarrow LSI(C') \Longrightarrow PI(C')$$

Although there are strong and weak sense of convergence, the functional inequalities characterizing the convergence are each of independent mathematical and practical interest, so one should not think of these relations as superior or inferior relations.

# Chapter 3 LOG-SOBOLEV INEQUALITIES

In this chapter, we are going to look at some equivalent formulations of log-Sobolev inequality. We will mainly focus on the proving that exponential decay of entropy and hypercontractivity are equivalent to a log-Sobolev inequality. Exponential decay of entropy gives an analogous result for exponential decay of variance for Poincaré inequality, and hypercontractivity shows a regularisation property of $P_t$ in evolution of time.

We will let $(X, \mu, \Gamma)$ be a mass-preserving standard Markov triple and $\Gamma$ have diffusion property. In particular, for any choice $s \geq 1$, the diffusion property indicates that the log-Sobolev inequality $LSI(C, \gamma)$ is equivalent to having

$$\mathrm{Ent}(f^s) \leq C\Big(\frac{s^2}{2}\int_X f^{s-2}\Gamma(f,f)d\mu + \gamma\int_X f^s ds\Big)$$

for each $f \in \mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$. Alternatively, we can just replace the notion of $LSI(C, \gamma)$ by this family of inequalities for all $s \geq 1$.

## Section 3.1 Basic properties of entropy

Let $f$ be a non-negative function. Then we have defined the entropy of $f$ by

$$\mathrm{Ent}_\mu(f) = \mu(f \log f) - \mu(f)\log(\mu(f))$$

with convention $0 \log 0 = 0$. Because of the singularity of logarithmic function at 0, dealing with limits and derivatives of functions $f$ that are not strictly separated away from 0 requires some technical tricks. Also, entropy is not a metric, so we can not use naive additive properties as for the case of $L^p$ or $L^2$-norm. So we introduce a couple of tricks to deal with such problems here, and establish some results on analytic properties of entropy function.

**Proposition 3.1** *Let $f, g : X \to \mathbb{R}$ be measurable functions such that $f \geq 0$ is integrable and $\int_X e^g d\mu < \infty$. Then*

$$\int_X fg d\mu \leq \mathrm{Ent}(f) + \mu(f)\log(\mu(e^g))$$

*In particular, we have*

$$\mathrm{Ent}(f) = \sup\Big\{\int_X fg d\mu \ : \ \int_X e^g d\mu \leq 1\Big\}$$

*Proof.* Since the whole inequality is linear in $f$, we may rescale $f$ so that $\mu(f) = 1$, and define a new probability measure $\nu(A) = \mu(f\mathbf{1}_A)$. If the entropy is infinite, the result is trivial. If not, then by Jensen's inequality,

$$\int_X f(g - \log(f))d\mu = \nu(\log(e^g/f)) \leq \log(\nu(e^g/f)) = \log(\int_X e^g d\mu)$$

and the inequality follows.

For the later part, it is again sufficient to establish the relation for $f$ with $\mu(f) = 1$. We readily have from above inequality that $\int_X fg d\mu \leq \mathrm{Ent}(f)$ whenever $\int_X e^g d\mu \leq 1$. Also, by choosing $g = \log(f)$, we have $\int_X e^g = 1$ and $\mathrm{Ent}(f) = \int_X fg d\mu$, hence we have the result.

*(End of proof)* $\square$

### § Computing entropy using approximation

Since positive functions strictly separated away from 0 behaves well under taking logarithm, we will often approximate any non-negative function $f$ by functions of form $f + c$ and approximate $\mathrm{Ent}(f)$ with $\mathrm{Ent}(f + c)$. For this approximation to work properly, we will need $\mathrm{Ent}(f + c) \to \mathrm{Ent}(f)$ as $c \searrow 0$.

**Lemma 3.2** *Let $f : X \to \mathbb{R}_{\geq 0}$ be integrable. Then for each $c > 0$, $\mathrm{Ent}(f + c) \leq \mathrm{Ent}(f)$. Also, whenever $(c_n)_n$ is a sequence of functions such that $c_n(x) \searrow 0$ as $n \to \infty$ uniformly and $\mathrm{Ent}(f + c_n) < \infty$ for each $n$, then*

$$\mathrm{Ent}(f + c_n) \to \mathrm{Ent}(f) \quad \text{as } n \to \infty$$

*Proof.* Recall, $\mathrm{Ent}(f) = \sup\{\int fg d\mu : \int_X e^g d\mu \leq 1\}$. For each $\epsilon > 0$, let $g_\epsilon$ be such that $\int_X e^{g_\epsilon} \leq 1$ and $\int (f + c) g_\epsilon d\mu \geq \mathrm{Ent}(f + c) - \epsilon$. Then

$$\mathrm{Ent}(f) \geq \int_X f g_\epsilon d\mu \geq \mathrm{Ent}(f + c) - \epsilon - c \int_X g_\epsilon d\mu,$$

while by Jensen's inequality,

$$\int_X g_\epsilon d\mu \leq \log \int_X e^{g_\epsilon} d\mu \leq 0$$

so $\mathrm{Ent}(f) \geq \mathrm{Ent}(f + c) - \epsilon$. This is true for any $\epsilon > 0$, so $\mathrm{Ent}(f) \geq \mathrm{Ent}(f + c)$.

Too see the later part, let

$$\begin{aligned} &A_k = \Big\{x : f(x) \leq \frac{1}{e} - \frac{1}{k}\Big\}, \quad B_k = \Big\{x : f(x) \geq \frac{1}{e} + \frac{1}{k}\Big\}, \\ &C_k = X \backslash (A_k \cup B_k). \end{aligned}$$

Then by monotone convergence theorem, as $n \to \infty$,

$$\begin{aligned} \int_{A_k} (f + c_n) \log(f + c_n) d\mu &\to \int_{A_k} f \log(f) d\mu \\ \int_{B_k} (f + c_n) \log(f + c_n) d\mu &\to \int_{B_k} f \log(f) d\mu. \end{aligned}$$

Also, $\int_{C_k} (f + c_n) \log(f + c_n) d\mu \xrightarrow{n\to\infty} \int_{C_k} f \log f$ as $n \to 0$ by dominated convergence theorem,

*(End of proof)* □

We have a seemingly a similar lemma, which would be of different use. The lemma is due to {R2}.

**Lemma 3.3** *(Rothaus) For any $c \in \mathbb{R}$, $f^2 \in L^1(\mu)$, we have*

$$\mathrm{Ent}(f^2) \leq \mathrm{Ent}((f - c)^2) + 2\|f - c\|_2^2$$

*Proof.* We prove an equivalent statement saying that $\mathrm{Ent}((f-c)^2) \leq \mathrm{Ent}(f^2)+2\|f\|_2^2$. To obtain such a bound, we start from the discrete case in order to get around the singularity of logarithmic function at 0. Then the result would naturally extend to the general case. For $a_1, \cdots, a_n \in \mathbb{R}$, we want to prove

$$\begin{aligned}\Lambda(t) = &-\frac{1}{n}\sum_i (ta_i+c)^2 \log(ta_i+c)^2 + \frac{1}{n}\sum_i (ta_i+c)^2 \log\Big(\frac{1}{n}\sum_i (ta_i+c)^2\Big) \\ &+ \frac{1}{n}\sum_i t^2a_i^2 \log(t^2a_i^2) - \frac{1}{n}\sum_i t^2a_i^2 \log\Big(\frac{1}{n}\sum_i t^2a_i^2\Big) + \frac{2}{n}\sum_i t^2a_i^2\end{aligned}$$

has $\Lambda(1) \geq 1$. We utilize the convexity of $\Lambda$. Indeed,

$$\begin{aligned}\Lambda'(t) = &-\frac{2}{n}\sum a_i(ta_i+c)(2\log(ta_i+c)+1) + 2(1+\log\Big(\frac{1}{n}\sum(ta_i+c)^2\Big))\frac{1}{n}\sum_i a_i(ta_i+c) \\ &+ \frac{1}{n}\sum\Big(2ta_i\log(t^2a_i^2)+2ta_i^2\Big) - 2(1+\log\Big(\frac{1}{n}\sum(ta_i)^2\Big))\frac{1}{n}\sum_i a_i(ta_i) + \frac{4}{n}\sum ta_i^2\end{aligned}$$

and

$$\begin{aligned}\frac{\Lambda''(t)}{2} = &-\frac{1}{n}\sum a_i^2\log(ta_i+c)^2 + \frac{1}{n}\sum a_i^2 \log\Big(\frac{1}{n}\sum(ta_i+c)^2\Big) \\ &+ \frac{1}{n}\sum x_i^2\log(a_i^2) - \frac{1}{n}\sum a_i^2\log\Big(\frac{1}{n}\sum a_i^2\Big) + 2\Big(\frac{1}{n}\sum a_i(ta_i)\Big)^2\Big/\Big(\frac{1}{n}\sum(ta_i+c)^2\Big).\end{aligned}$$

With normalisation $\frac{1}{n}\sum_i(ta_i+c)^2 = 1$, **Proposition 3.1** gives that

$$\frac{1}{n}\sum a_i^2 \geq \frac{1}{n}\sum a_i^2\log(ta_i+c)^2$$

and therefore $\Lambda''(t) \geq 0$. The inequality does not depend on the normalisation, so we have $\Lambda''(t) \geq 0$ for $t \in [0,1]$. Also we have that $\Lambda(0) = \Lambda'(0) = 0$, and therefore we have the desired conclusion.

*(End of proof)* □

## § de Brujin's identity

As for proving exponential decay in variance, it is natural to look at the derivative of entropy $\mathrm{Ent}(P_tf)$ to derive exponential decay of entropy under $LSI(A,C)$. So we derive an expression for the time derivative of $\mathrm{Ent}(P_tf)$.

**Lemma 3.4** *(de Brujin's identity) We have*

$$\frac{d}{dt}\mathrm{Ent}(P_tf) = -\int_X \frac{\Gamma(P_tf)}{P_tf}d\mu$$

*whenever $f(x) \in [m,M] \subset (0,\infty)$. We call $\int_X \frac{\Gamma(f)}{f} =: I_\mu(f) =: I(f)$ the **Fisher information** of $f$.*

*Proof.* With assumption that $f$ is bounded in $[m,M] \subset (0,\infty)$, $P_tf$ is also bounded in $[m,M]$ for all $t \geq 0$ and

$$\begin{aligned}\frac{d}{dt}\mathrm{Ent}(P_tf) &= \frac{d}{dt}\Big(\int_X P_tf\log(P_tf)d\mu\Big) = \int_X \log(P_tf)\mathcal{G}P_tf + \mathcal{G}P_tfd\mu \\ &= \int_X \log(P_tf)\mathcal{G}fd\mu = -\int_X \Gamma(\log(P_tf),P_tf)d\mu \\ &= -\int_X \frac{\Gamma(P_tf)}{P_tf}d\mu\end{aligned}$$

*(End of proof)* □

This result can be extended to non-negative functions $f$ with some extra integrability conditions using approximation techniques, but we do not need those results in the present setting.

## Section 3.2 Exponential decay of entropy under log-Sobolev inequality

By the de Brujin's identity, the exponential decay in entropy follows immediately from a tight log-Sobolev inequality. But like the case of variance, the converse also holds.

**Theorem 3.5** *(Entropy decay) For a Markov triple $(X, \mu, \Gamma)$, having $LSI(C)$ is equivalent to having*

$$\mathrm{Ent}(P_t f) \leq e^{-2t/C}\mathrm{Ent}(f) \quad \forall t \geq 0$$

*for each $f \geq 0$, $f \in L^1(\mu)$.*

*Proof.* First assume $LSI(C)$. Since the inequality is trivial when $f$ has infinite entropy, assume it has finite entropy. Also, assume $f$ is bounded above by $M$ and $f \geq 0$, recalling that $(f \wedge M) \vee 0 \in \mathcal{D}_{\mathcal{E}}$ whenever $f \in \mathcal{D}_{\mathcal{E}}$. By an equivalent formulation of $LSI(C)$ inequality, for any $c > 0$,

$$\mathrm{Ent}(P_t(f+c)) \leq \frac{C}{2}\int_X \frac{\Gamma(P_t(f+c))}{P_t(f+c)}d\mu = \frac{C}{2}I(P_t f + c) = -\frac{C}{2}\frac{d}{dt}\mathrm{Ent}(P_t(f+c))$$

so Grönwall's inequality gives $\mathrm{Ent}(P_t(f+c)) \leq e^{-2t/C}\mathrm{Ent}(f+c)$. Letting $c \searrow 0$, we have $\mathrm{Ent}(P_t f) \leq e^{-2t/C}\mathrm{Ent}(f)$.

If $f$ is not bounded above, let $f^M(x) = f(x) \wedge M$. Then $\mathrm{Ent}(f^M) \to \mathrm{Ent}(f)$ as $M \to \infty$ by monotone convergence. By monotonicity of $P_t$, we see that

$$P_t(f^M)(x) \leq P_t(f^{M+1})(x) \leq P_t f(x) \quad \forall x.$$

Also, $P_t f^M \to P_t f$ as $M \to \infty$ in $L^1(\mu)$, so there is a sequence $(M_k)_k \subset \mathbb{N}$ such that $P_t f^{M_k} \nearrow P_t f$ $\mu$-a.e. as $k \to \infty$. Therefore, monotone convergence implies that $\mathrm{Ent}(P_t f^{M_k}) \to \mathrm{Ent}(P_t f)$ as $k \to \infty$. So we conclude $\mathrm{Ent}(P_t f) \leq e^{-2t/C}\mathrm{Ent}(f)$ for any $f \geq 0$.

Conversely, if we assume the exponential decay, then Taylor's expansion about $t = 0$ gives, for $f(x) \in [m, M] \subset (0, \infty)$,

$$(\mathrm{Ent}(f) - I(f)t + o(t)) \leq \Big(1 - \frac{2t}{C} + o(t)\Big)\mathrm{Ent}(f)$$

so $\mathrm{Ent}(f) \leq \frac{C}{2}I(f)$. For $f$ just bounded above, we have

$$\mathrm{Ent}(f) = \lim_{c\searrow 0}\mathrm{Ent}(f+c) \leq \frac{C}{2}\liminf_{c\searrow 0}\int_X \frac{\Gamma(f,f)}{f+c}d\mu = \frac{C}{2}\int_X \frac{\Gamma(f,f)}{f}d\mu$$

by monotone convergence theorem, with convention $0/0 = 0$. Hence, $\mathrm{Ent}(f^2) \leq 2C\mathcal{E}(f)$ holds for any $f \in L^\infty(\mu)$, and in particular in $\mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$. If $f \in \mathcal{D}_{\mathcal{G}}$, then we may find $(f_n)_n \subset \mathcal{D}_{\mathcal{E}}$ such that $f_n \to f$ in $\|\cdot\|_{\mathcal{E}}$ and $\mu$-a.e., so Fatou's lemma gives

$$\mathrm{Ent}(f) \leq \liminf_n \mathrm{Ent}(f_n) \leq \liminf_n 2C\mathcal{E}(f_n) = 2C\mathcal{E}(f).$$

*(End of proof)* □

## Section 3.3 Hypercontractivity

Complementing the role of entropy decay inequality, there is a result called hypercontractivity that characterizes how regularity of $P_tf$ evolves in $t$ as $t \to \infty$. Unlike entropy decay, hypercontractivity does not require tight log-Sobolev inequality.

### § Motivation

Recall our argument from **Section 2.2** that existence of $\tau > 0$ with $\|P_{\tau(\epsilon)}\|_{2,4} \leq 1$ is sufficient for proving a Poincaré inequality. In a reasonably generalized form, we would like to see what we can infer from $\|P_\tau\|_{p,q} \leq M$ for some $\tau > 0$, $1 < p < q$.

First fix $\epsilon$ and rescale $p$ to 2 and $M$ to $1+\epsilon$ : application of Riesz-Thorin interpolation theorem with

$$\|P_\tau\|_{p,q} \leq 1+\epsilon, \quad \|P_\tau\|_{1,1} \leq 1$$

gives $\|P_t\|_{\overline{p},\overline{q}} \leq 1+\epsilon$ for $\overline{p} < \overline{q}$ sufficiently close to 1. Using interpolation once more if necessary, we may assume that $\overline{p} \leq 2$, and consecutively using interpolation with bound $\|P_t\|_{\infty,\infty} \leq 1$, we may assume $\overline{p} \leq 2 < \overline{q}$.

Next, using the fact that $\|P_\tau\|_{\infty,\infty} \leq 1$ for any $t \geq 0$, one may extend the bound for $\|P_\tau\|_{2,\pi}$ for larger values of $\pi > \overline{q}$ by increasing $t$. Indeed, using bounds

$$\|P_\tau\|_{2,\overline{q}} \leq 1+\epsilon, \quad \|P_\tau\|_{\infty,\infty} \leq 1$$

and Riesz-Thorin interpolation theorem, we obtain the bound $\|P_\tau\|_{2\theta,\overline{q}\theta} \leq (1+\epsilon)^{1/\theta}$ for any $\theta \geq 1$. Choosing $\theta = (\overline{q}/2)^k$ for $k \in \mathbb{N}$, we have $\|P_\tau\|_{2(\overline{q}/2)^k,\overline{q}(\overline{q}/2)^k} \leq (1+\epsilon)^{(2/\overline{q})^k}$ and therefore by induction,

$$\|P_{(k-1)\tau}\|_{2,\overline{q}(\overline{q}/2)^k} \leq (1+\epsilon)^{1+(2/\overline{q})+\cdots+(2/\overline{q})^{k-1}} \leq (1+\epsilon)^{\frac{\overline{q}}{\overline{q}-2}}, \quad \forall k \in \mathbb{N}.$$

So for any $\epsilon > 0$, there is $\tau > 0$ such that $\|P_t\|_{2,4} \leq 1+\epsilon$ for any $t \geq \tau$. So we may think of having $\|P_\tau\|_{p,q} \leq M$ for any $\tau > 0$, $p < q$ to be the most strongest result we desire for the moment. This is precisely what we prove to be the case for the log-Sobolev inequality.

### § Gross' hypercontractivity

Although we know we should make estimation of $\|P_t\|_{q,p}$ for some exponents $p, q > 1$, it is not easy to make the right choice of these exponents. Naively differentiating $\|P_tf\|_p$ in $t$ with fixed $q > 1$ would not work, because it would not take finite value in general if $f \in L^q(\mu)$ with $\|f\|_p = \infty$. So the idea is to make the exponent $p$ to have dependence on time with $p(t=0) = q$. After we obtain an expression for the derivative $\|P_t\|_{q,p(t)}$ for general $p(t)$, the exact form of the function $p(t)$ would become clear.

We first investigate the differentiability of $\|P_tf\|_{p(t)}$ in $t$.

**Lemma 3.6** *Let $q > 1$ and $p(t) \in C^1(\mathbb{R})$, $p(0) = q$. Then with $f \in \mathcal{D}_\mathcal{E}$, $f \in [m, M] \subset (0,\infty)$, we have $\|P_tf\|_{p(t)}$ differentiable in $t$ at $t = 0$ and*

$$\frac{d}{dt}\|P_tf\|_{p(t)}\Big|_{t=0} = \Big(\frac{p'(0)}{q^2}\mathrm{Ent}(f^q) + \int f^{q-1}\mathcal{G}f\,d\mu\Big)\|f\|_q^{1-q}$$

*Proof.* It is enough to prove this for $f \in [m,1]$ by rescaling. Then $|f^p(x) - f^p(y)| \leq |f(x) - f(y)|$ for each $x, y \in X$ and $|f^p(x)| \leq |f(x)|$, so by definition of being a Markov triple, we have $f^p \in \mathcal{D}_\mathcal{E} \cap L^\infty(\mu)$. The rest follows from basic calculus.

*(End of proof)* $\square$

In order to make use of log-Sobolev inequality, we invoke that the log-Sobolev inequality can be written as

$$\mathrm{Ent}(f^p) \leq C\Big(\frac{p^2}{2}\int_X f^{p-2}\Gamma(f,f)d\mu + \gamma\int_X f^p d\mu\Big) = C\Big(-\frac{p^2}{2(p-1)}\int_X f^{p-1}\mathcal{G}f d\mu + \gamma\int_X f^p d\mu\Big).$$

If we put this in the expression for $\frac{d}{dt}\|P_t f\|_{p(t)}$ and choose $p(t)$ so that $p'(t) = 2(p(t)-1)/C$ then the $\int_X f^{p-1}\mathcal{G}f d\mu$ term will vanish and obtain a bound of $\frac{d}{dt}\|P_t f\|_{p(t)}$ just in terms of $\|P_t f\|_{p(t)}$. This is done formally in the proof of the next theorem.

**Theorem 3.7** *(Hypercontractivity Theorem, Gross) For fixed $C > 0$, $\gamma \geq 0$, the following are equivalent :*

*(i) The Markov triple $(X, \mu, \Gamma)$ satisfies $LSI(C, \gamma)$.*

*(ii) $\|P_t\|_{q,p(t)} \leq e^{M(t,q)}$ where $p(t) = 1 + (q-1)e^{2t/C}$, $M(t,q) = \gamma C\big(\frac{1}{q} - \frac{1}{p(t)}\big)$, for any $q \in (1,\infty)$ and $t \geq 0$.*

*Proof.* $(i) \Rightarrow (ii)$ : Consider $f \in \mathcal{D}_\mathcal{E}$, $f \in [m, M] \subset (0,\infty)$. Then by the previous lemma,

$$\begin{aligned}\frac{d}{dt}\big\|P_t f\big\|_{p(t)} &= \Big(\frac{p'(t)}{p(t)^2}\mathrm{Ent}((P_t f)^{p(t)}) + \int (P_t f)^{p(t)-1}\mathcal{G}(P_t f)d\mu\Big)\big\|P_t f\big\|_{p(t)}^{1-p(t)} \\ &= \Big(\frac{2(p-1)}{Cp^2}\mathrm{Ent}((P_t f)^p) - (p-1)\int (P_t f)^{p-2}\Gamma(P_t f)d\mu\Big)\big\|P_t f\big\|_p^{1-p}.\end{aligned}$$

Also, by putting $g^{s/2}$ in $LSI(C,\gamma)$, we have (given sufficient integrability condition of $g$, *e.g.* $g \in \mathcal{D}_\mathcal{E} \cap L^\infty(\mu)$)

$$\mathrm{Ent}(g^s) \leq C\Big(\frac{s^2}{2}\int_X g^{s-2}\Gamma(g)d\mu + \gamma\int_X g^s d\mu\Big) \quad \forall s \geq 1$$

so putting these together,

$$\frac{d}{dt}\big\|P_t f\big\|_{p(t)} \leq \frac{2(p-1)}{p^2}\gamma\big\|P_t f\big\|_p$$

Noting that $M'(t,q) = \frac{2(p(t)-1)}{p(t)^2}\gamma$, Grönwall's inequality gives

$$\|P_t f\|_{p(t)} \leq \|f\|_q e^{M(t,q)}$$

An easy density argument allow us to extend this to $f \in \mathcal{D}_\mathcal{E}$, $f \geq 0$. Now since $\mathcal{D}_\mathcal{E}$ is dense in $L^q(\mu)$, this concludes the proof of forward implication.

$(ii) \Rightarrow (i)$ : This direction is proved using Taylor expansion as usual. That is, for $f \in \mathcal{D}_\mathcal{E}$, $f \in [m, M] \subset (0, \infty)$, as $t \to 0$,

$$\begin{aligned}\|P_t f\|_{p(t)} &= \|f\|_q + \Big(\frac{2(q-1)}{Cq^2}\mathrm{Ent}(f^q) - (q-1)\int_X f^{q-2}\Gamma(f)d\mu\Big)\|f\|_q^{1-q} + o(t) \\ &\leq \Big(1 + 2\gamma\frac{q-1}{q^2} + o(t)\Big)\|f\|_q\end{aligned}$$

In particular, if $q=2$, and if we just assume $f \in [0,M]$, we have

$$\mathrm{Ent}((f+c)^2) \le C\Big(\mathcal{E}(f,f) + \gamma \int_X (f+c)^2 d\mu\Big)$$

and putting $c \searrow 0$ gives $LSI(C,\gamma)$ for $f \in \mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$, $f \ge 0$. To see the inequality holds for general $f \in \mathcal{D}_{\mathcal{E}}$, $f \ge 0$, just observe that $f^M = f \wedge M \to f$ as $M \to \infty$ and recalling the fact that $f^M \to f$ in $\|\cdot\|_{\mathcal{E}}$ as $M \to \infty$, Fatou's lemma gives

$$\mathrm{Ent}(f^2) \le \liminf_{n\to\infty} \mathrm{Ent}(f_n^2) \le C \liminf_{n\to\infty} \Big(\mathcal{E}(f_n,f_n) + \gamma\|f_n\|_2\Big) = C\Big(\mathcal{E}(f,f) + \gamma\|f\|_2\Big)$$

*(End of proof)* □

## Section 3.4 Convergence properties of $LSI(C,\gamma)$

### § Implication of Poincaré inequality

Although we have proved many interesting results from log-Sobolev inequality, we have not seen yet that log-Sobolev inequality implies convergence of the Markov process $(M_t)_{t\ge 0}$ associated to the Markov triple $(X,\mu,\Gamma)$. We show this via proving that $LSI(C)$ implies $PI(C)$.

**Proposition 3.8** *$LSI(C)$ implies $PI(C)$.*

There are multiple ways of showing this.

*Proof 1.* (Using exponential decay) Let $u \in \mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$ and $f$ be given by $f = 1 + \epsilon u$, for a (small enough) parameter $\epsilon \in \mathbb{R}$. Then $\mathcal{E}(f) = \epsilon^2 \mathcal{E}(u)$ so $f \in \mathcal{D}(\mathcal{E})$. Also,

$$\begin{aligned}\mathrm{Ent}(f^2) &= 2\int_X (1 + 2\epsilon u + \epsilon^2 u^2)(\epsilon u - \frac{1}{2}\epsilon^2 u^2 + o(\epsilon^2)) d\mu \\ &\quad - (1 + \epsilon\mu(u^2))(2\epsilon\mu(u) + \epsilon^2\mu(u^2) - 2\epsilon^2\mu(u)^2 + o(\epsilon^2)) \\ &= 2\epsilon^2\mu(u^2) + o(\epsilon^2)\end{aligned}$$

so exponential decay of entropy implies

$$2\epsilon^2\mu(u^2) + o(\epsilon^2) \le e^{-\frac{2t}{C}}\mu((P_t u)^2) + o(\epsilon^2)$$

hence $\mathrm{Var}(u) \le e^{-\frac{2}{C}t}\mathrm{Var}(u)$ holds for bounded $u$. Using density of $\mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$ inside $L^2(\mu)$ gives

$$\mathrm{Var}(u) \le e^{-\frac{2t}{C}}\mathrm{Var}(P_t u) \quad \forall u \in L^2(\mu)$$

which is equivalent to having $PI(C)$.

*Proof 2.* Let $u \in \mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$ and $f$ be given by $f = 1 + \epsilon u$, for $\epsilon \in \mathbb{R}$. Then by Taylor expansion and $LSI(C)$,

$$\mathrm{Ent}(f^2) = 2\epsilon^2\mu(u)^2 + o(\epsilon^2) \le 2C\epsilon^2\mathcal{E}(u)$$

so we have $PI(C)$ for $u \in \mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$. Again use density argument to conclude.

*Proof 3.* (Using hypercontractivity) (proof for having $PI(C')$ for some $C' > 0$) Assuming $LSI(C)$, using hypercontractivity, there is some $t_0 > 0$ such that $\|P_{t_0}\|_{2,4} \le 1$. Hence according to the procedure described in **Section 1.2**, we have a Poincaré inequality.

*(End of proof)* □

## § Exponential integrability and heat kernel bound

So far, we had been seeing how $P_t$ acts on functions on a local level, how regularity of a function improves on action of $P_t$. However, we can also establish results for how tail distribution of $P_t f$ behaves under $LSI(C)$. That is, we can prove results on bounds on $P_T 1_A(x)$ as $\|x\| \to \infty$ for some bounded set $A$.

In order to work out such a bound, it would be best to compute the bound on heat kernel of $P_t$, based on the observation that $A \mapsto P_t \mathbf{1}_A(x)$ is a probability measure. The definition of a heat kernel should be obvious, but let us make the notion concrete first.

**Definition** *Let $(X, \mathcal{F})$ be a measurable space and $Q : X \times \mathcal{F} \to [0,1]$ be such that (i) $X \to [0,1]$, $x \mapsto Q(x, A)$ is measurable for each $A \in \mathcal{F}$, (ii) $\mathcal{F} \to [0,1]$, $A \mapsto Q(x, A)$ is a probability measure for each $x \in X$. Then $Q$ is called a* ***(Markovian) transition kernel****, and $Q$ acts on measurable function $f$ (such that $f \in L^1(Q(\cdot, dy))$ for each $x$) by*

$$Qf(x) = \int_X f(y)Q(x, dy)$$

*If a background measure $\mu$ is given, a function $q(x,y)$ is called a* ***density kernel*** *of $Q$ (with respect to $\mu$) if*

$$Q(x, dy) = q(x,y)d\mu(y)$$

Apparently, each $P_t$ is a transition kernel, but the existence of a density kernel $p_t(x,y)$ with sufficient level of regularity depends on $P_t$. However, let us assume here that a bound on $p_t(x,y)$ exists for all $t \geq 0$, say $p_t(x,y) \leq A(t)$ (In next chapter, we will see that this holds true when the system satisfies a Sobolev inequality). In order to measure decay rate of $p_t(x,y)$, we may define a metric on $X$ just in terms of diffusivity of the system.

**Definition** *For $f \in \mathcal{D}_\Gamma$, we define $\|f\|_{Lip} = \|\Gamma(f)\|_\infty^{1/2}$, and we say $f$ is* ***1-Lipschitz*** *if $\|f\|_{Lip} \leq 1$. Also, define the* $\Gamma$***-metric*** *$d_\Gamma : X \times X \to \mathbb{R}_{\geq 0}$ by*

$$\begin{aligned} d_\Gamma(x,y) &= \operatorname{esssup}\Big\{|f(x) - f(y)| \,:\, \|f\|_{Lip} \leq 1,\ x, y \in X\Big\} \\ &= \sup\Big\{s \,:\, \|f\|_{Lip} \leq 1,\ |f(x) - f(y)| \leq s \text{ for } \mu \otimes \mu\text{-a.e. } x, y \in X\Big\} \end{aligned}$$

It is not very difficult to see that this actually defines a metric on $X$. Now using $P_t \circ P_t = 2P_t$ and that $d_\Gamma(x,y)^2 \leq (d_\Gamma(x,z) + d_\Gamma(z,y))^2 \leq 2d_\Gamma(x,z)^2 + 2d_\Gamma(y,z)^2$, we have

$$\begin{aligned} p_{2t}(x,y) &= \int_X p_t(x,z)p_t(z,y)d\mu(z) \\ &\leq e^{-d_\Gamma(x,y)^2/c} \int_X p_t(x,z)e^{2d_\Gamma(x,z)^2/c} p_t(z,y)e^{2d_\Gamma(z,y)^2/c} d\mu(z) \\ &\leq e^{-d_\Gamma(x,y)^2/c} \int_X p_t(x,z)^2 e^{2d_\Gamma(x,z)^2/c} d\mu(z) \end{aligned}$$

for any constant $c > 0$, where the last inequality follows from Cauchy-Schwarz inequality. Since we have assumed $p_t(x,y)$ is bounded, if we can prove

$$\int_X p_t(x,z)e^{4d_\Gamma(x,z)^2/c} d\mu(z) < \infty$$

then we will obtain $p_t(x,y) \leq Ke^{-d_\Gamma(x,y)^2/c}$, for each fixed time $t > 0$, with $K$ and $c$ depending on time. Indeed, this holds, in a more general setting (noting that $d_\Gamma(\cdot, z)$ is also a 1-Lipschitz function).

**Lemma 3.9** *(Exponential integrability) If $LSI(C)$ holds, then whenever $F$ is 1-Lipschitz,*

$$\int_X e^{2\lambda F} d\mu \leq e^{2\lambda\mu(F)+2\lambda^2 C}$$

*and integrating this along distribution $e^{-\frac{2\lambda^2}{\sigma^2}}$, one obtains $\int_X e^{\frac{1}{2}\sigma^2 F^2} d\mu < \infty$ whenever $C\sigma^2 < 1$.*

*Proof.* First consider the case where $F$ is bounded, say $|F| \leq M$. We want to apply $LSI(C)$ on $\exp(\lambda F)$, so we first have to prove that $\exp(\lambda F) \in \mathcal{D}_{\mathcal{E}}$. By **Lemma 1.9** ***(iii)***, we have $\mathcal{E}(F^2) \leq (2M\sqrt{\mathcal{E}(F)})^2 = 4M^2\mathcal{E}(F)$. Using induction, we obtain $\mathcal{E}(F^n) \leq 2^n M^{2(n-1)}\mathcal{E}(F)$. Therefore, if we define the partial sum $S_n = \sum_{j=0}^n \frac{1}{j!}\lambda^j f^j$, we have

$$\begin{aligned}
\mathcal{E}(S_n) = \mathcal{E}\Big(\sum_{j=0}^n \frac{\lambda^j}{j!}F^j\Big) &= \sum_{i,j=0}^n \frac{1}{i!j!}\mathcal{E}(f^i, f^j)\lambda^{i+j} \\
&\leq \sum_{i,j=0}^n \frac{1}{i!j!}(\sqrt{2}M\lambda)^{i+j}\frac{\mathcal{E}(F)}{M^2} \\
&= \frac{\mathcal{E}(F)}{M^2}\Big(\sum_{i=0}^n \frac{1}{i!}(\sqrt{2}M\lambda)^{i+j}\Big)^2 \\
&\leq \frac{\mathcal{E}(f)}{M^2}e^{2\sqrt{2}M\lambda}
\end{aligned}$$

So $LSI(C)$ applies to $f = e^{\lambda F}$. Setting $H(\lambda) = \mu(e^{2\lambda F})$,

$$\begin{aligned}
\mathrm{Ent}(e^{2\lambda F}) &= 2\lambda\int_X Fe^{2\lambda F}d\mu - \mu(e^{2\lambda F})\log(\mu(e^{2\lambda F})) = \lambda H'(\lambda) - H(\lambda)\log H(\lambda) \\
\mathcal{E}(e^{\lambda F}) &= \lambda^2\int_X e^{\lambda F}\Gamma(F)d\mu \leq \lambda^2 H(\lambda)
\end{aligned}$$

where the fact $\Gamma(F) \leq 1$ had been exploited in the second line. So $LSI(C)$ gives,

$$\lambda H'(\lambda) - H\log H \leq 2C\lambda^2 H$$

and

$$\frac{d}{d\lambda}\Big(\frac{1}{\lambda}\log(H(\lambda))\Big) = \frac{1}{\lambda}\frac{H'(\lambda)}{H\lambda)} - \frac{1}{\lambda^2}\log H(\lambda) \leq 2C$$

so we have $\int_X e^{2\lambda F}d\mu \leq \exp\Big(2\lambda\mu(F) + 2C\lambda^2\Big)$

To prove for general $F \in \mathcal{D}_{\mathcal{E}}$, let $F^M = (F \wedge M) \vee (-M)$ for $M > 0$, then $F^M$ satisfies the inequality for each $M$. Then Fatou's lemma gives

$$\int_X e^{2\lambda F}d\mu \leq \liminf_{M\to\infty} e^{2\lambda\mu(F^M)+2\lambda^2 C} = e^{2\lambda\mu(F)+2\lambda^2 C}$$

*(End of proof)* □

Exponential integrability has one particularly interesting implication, related to concentration phenomenon of random variables.

**Corollary 3.10** *Under $LSI(C)$, for any Lipschitz function $F$,*

$$\mu(|F - \mu(F)| \geq \epsilon) \leq 2e^{-\epsilon^2/(2C\|F\|_{Lip}^2)}$$

*Proof.* This is a direct consequence of Tchebyshev's inequality, *i.e.* after scaling $F$ so that $\|F\|_{Lip} = 1$,

$$\mu(F \geq \mu(F) + \epsilon) = \mu(e^{\lambda F} \geq e^{\lambda(\mu(F)+\epsilon)^2}) \leq e^{-\lambda(\mu(F)-\epsilon)^2}\mu(e^{\lambda F}) = e^{-\epsilon^2/2C}$$

The same principle applies for $\mu(F \leq \mu(F) - \epsilon)$, and we have the result.

*(End of proof)* $\square$

# Chapter 4 SOBOLEV INEQUALITIES

We have defined a Sobolev inequality $SI^p(C,\delta)$ by

$$\|f\|_p^2 \le C(\delta\|f\|_2^2 + \mathcal{E}(f)) \quad \forall f \in \mathcal{D}_{\mathcal{E}}$$

As is widely known, the classical Sobolev inequality has fruitful applications in the theory of partial differential equations. A most immediate and important result of a classical Sobolev inequality is that it implies that $W^{1,2}(\mathbb{R}^d)$ can be embedded in $L^{\frac{2d}{d-2}}(\mathbb{R}^d)$. However, as for our previous results, there are even stronger consequences of the Sobolev inequality if we are dealing with probability measures $\mu$. In particular, we will see in the next section that the Sobolev space $W^{1,2}(U,\mu)$ for bounded open $U$ can be compactly embedded in $L^2(U,\mu)$, which can be considered as an analogue of Rellich-Kondrachov theorem. However, in order to see this, we need some preliminary works in consideration on Sobolev inequality.

In this chapter, we let $(X,\mu,\Gamma)$ be a standard Markov triple and $(P_t)_{t\ge 0}$ be the semigroup associated to it, but we do not in general assume that it is mass-preserving. Also, instead of assuming that $\mathbf{1}_X \in \mathcal{D}_{\mathcal{E}}$, we will assume the existence of a sequence $(\psi_m)_{m\in\mathbb{N}} \subset \mathcal{D}_{\mathcal{E}}$ such that $0 \le \psi_m \le \psi_{m+1} \le 1$, $\psi_m \xrightarrow{L^2,\,a.e.} \mathbf{1}_X$ and $\mathcal{E}(\psi_m) \to 0$ as $m \to \infty$.

## Section 4.1 Bound on kernel density

### § A model example

Let $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d \subset \mathbb{R}^d$ and $dm$ be the Lebesgue measure on $\mathbb{T}^d$. In view of the general theory of Sobolev inequalities, compact embedding of $W^{1,2}(\mathbb{T}^d,dm)$ in $L^2(\mathbb{T}^d,dm)$ is related to the compactness of the semigroup $(P_t)_{t\ge 0}$ associated to the heat kernel. In order to see this, observe that if $P_t$ is a compact operator, then whenever $(u_k)_{k\in\mathbb{N}}$ is bounded by $K$ in $W^{1,2}(\mathbb{T}^d,dm)$, each $u_k$ can be approximated as $P_t u_k$, and $P_t u_k$ has a $L^2$-converging subsequence, say $P_t u_{k_l^{(t)}}$. Having this in mind, we may write

$$\|u_k - u\|_2 \le \|u_k - P_t u_k\|_2 + \|P_t(u-u_k)\|_2 + \|P_t u - u\|_2.$$

and attempt to make each term converge to 0 as $k \to \infty$, upto a subsequence. Notice that, for any $g \in \mathcal{D}_{\mathcal{E}}$,

$$\begin{aligned}
\frac{d}{dt}\|P_t g - g\|_2 &= \frac{d}{dt}\int_{\mathbb{T}^d}(P_t g - g)(P_t g - g)d\mu = 2\int_{\mathbb{T}^d}(P_t g - g)\mathcal{G}P_t g dm \\
&= -2\mathcal{E}(P_t g) - 2\int_{\mathbb{T}^d} P_{t/2}g\mathcal{G}P_{t/2}g dm = -2\mathcal{E}(P_t g) + 2\mathcal{E}(P_{t/2}g) \\
&\le 2\mathcal{E}(P_{t/2}g).
\end{aligned}$$

Also,

$$\frac{d}{ds}\mathcal{E}(P_s g) = -\int_{\mathbb{T}^d}(\mathcal{G}P_t g)^2 - P_t g(\mathcal{G}^2 P_t g)dm = -2\int_{\mathbb{T}^d}(\mathcal{G}P_t f)^2 dm \le 0$$

so we obtain $\mathcal{E}(P_s g) \le \mathcal{E}(g)$ whenever $s \ge 0$. Therefore

$$\|P_t g - g\|_2 \le 2t\mathcal{E}(g) \quad \forall g \in \mathcal{D}_{\mathcal{E}}.$$

So we have, for each $k \in \mathbb{N}$,

$$\|P_t u_k - u_k\|_2 \le 2t\mathcal{E}(u_k) \le 2tK.$$

Now by induction, define collection of nested subsequence $(u_{k_l^{1/(n+1)}})_l \subset (u_{k_l^{(1/n)}})_l \subset \cdots \subset (u_k)_k$ where $P_{1/n}u_{k^{(1/n)}l} \to P_{1/n}u$ in $L^2(\mathbb{T}^d, dm)$ and pointwise $m$-a.e., for each $n \in \mathbb{N}$ and $\|P_{1/(n+1)}u_{k_l^{(1/(n+1))}} - P_{1/n}u\|_2 \leq \|P_{1/n}u_{k_l^{(1/n)}} - P_{1/n}u\|_2$. Let $u_{k_l} = u_{k_l^{(1/l)}}$. Then by Fatou's lemma,

$$\|P_t u - u\|_2 \leq \liminf_{l\to\infty} 2t\mathcal{E}(u_{k_l^{(t)}}) \leq 2tK$$

hence we have

$$\|u_{k_l} - u\|_2 \leq \|u_{k_l} - P_{1/l}u_{k_l}\|_2 + \|P_{1/l}(u - u_{k_l})\|_2 + \|P_{1/l}u - u\|_2 \xrightarrow{l\to\infty} 0.$$

Therefore $W^{1,2}(\mathbb{T}^d, dm)$ can be compactly embedded in $L^2(\mathbb{T}^d, dm)$.

To see that the heat semigroup is compact, just note that the transition kernel have bound is given by

$$\sup_{x,y\in\mathbb{T}^d} p_t(x,y) = \sup_{x,y\in\mathbb{T}^d} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{1}{\sqrt{(2\pi t)^{d/2}}} e^{-\frac{\|y-x-\boldsymbol{k}\|^2}{2t}} < \infty$$

and therefore

$$\int_{\mathbb{T}^d\times\mathbb{T}^d} p_t^2(x,y)dm(x,y) < \infty$$

*i.e.* $P_t$ is a Hilbert-Schmidt operator.

### § Ultracontractivity

The general strategy to show that $\mathcal{D}_{\mathcal{E}}$ compactly embeds in $L^2(\mu)$ given $S^p(C,\delta)$ is exactly the same. We will first show that $P_t$ is a Hilbert-Schmidt operator or by showing $\|p_t(\cdot,\cdot)\|_\infty < \infty$ or $\|p_t(\cdot,\cdot)\|_2 < \infty$, *i.e.* $\|P_t\|_{1,\infty} < \infty$ or $\|P_t\|_{1,2} < \infty$ for each $t > 0$. These two results are in fact equivalent.

**Proposition 4.1** *Let $(P_t)_{t\geq 0}$ be a (self-adjoint) transition semigroup. Then there is a non-negative, non-increasing function $a : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ such that $\|P_t\|_{1,\infty} \leq a(t)$ for all $t > 0$ if and only if there is a non-increasing function $a' : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ with $\|P_t\|_{1,2} \leq a'(t)$ for all $t > 0$.*

*Proof.* First assume that $\|P_t\|_{1,2} \leq a(t)$. Then by duality,

$$\begin{aligned}
\|P_t\|_{1,2} &= \sup\{\|P_t f\|_2 : \|f\|_1 \leq 1\} \\
&= \sup\{\int_X gP_t f d\mu : \|f\|_1 \leq 1,\ \|g\|_2 \leq 1\} \\
&= \sup\{\int_X fP_t g d\mu : \|g\|_2 \leq 2,\ \|f\|_1 \leq 1\} \\
&= \sup\{\|P_t g\|_\infty : \|g\|_2 \leq 1\} = \|P_t\|_{2,\infty}
\end{aligned}$$

so $\|P_t\|_{1,\infty} \leq \|P_{t/2}\|_{1,2} \cdot \|P_{t/2}\|_{2,\infty} \leq a(t/2)^2$.

Conversely, assume $\|P_t\|_{1,\infty} \leq a(t)$. Then $P_t$ admits a density kernel $p_t(x,y)$ of $P_t$, so we should have $\|p_t(\cdot,\cdot)\|_\infty \leq 1$, so the implication is trivial. Even if not mentioning the density kernel, we may use Riesz-Thorin interpolation theorem with bounds

$$\|P_t\|_{1,1} \leq 1, \quad \|P_t\|_{1,\infty} \leq a(t)$$

to obtain bound $\|P_t\|_{1,q} \leq a(t)^{1-\frac{1}{q}}$ for each $p \in [1, \infty]$, and in particular we have

$$\|P_t\|_{1,2} \leq \sqrt{a(t)}.$$

*(End of proof)* $\square$

This motivates us to define the ultracontractive semigroup.

**Definition** *A transition semigroup $(P_t)_{t\geq 0}$ is said to be* ***ultracontractive*** *if there is a non-negative, non-increasing function $a : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ such that*

$$\|P_t\|_{1,\infty} \leq a(t), \quad \forall t > 0$$

*Or equivalently, there is a non-increasing function $a : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ with*

$$\|P_t\|_{1,2} \leq a(t), \quad \forall t > 0$$

Remarkably, it turns out a semigroup being ultracontractive is equivalent with the Markov triple satisfying $SI^p(C, \delta)$ given any $\delta \geq 0$ and $C > 0$.

**Theorem 4.2** *Suppose $p > 2$ and $p = \frac{2n}{n-2}$. Then the following are equivalent.*

*(i) The semigroup is polynomial-class ultracontractive with $\|P_t\|_{1,2} \leq C''/t^{n/4}$ holds for $t \in (0, 1)$, some $C'' > 0$.*

*(ii) $S^p(C, \delta)$ holds for some $C > 0$, $\delta \geq 0$.*

We will prove the proof later.

## Section 4.2 Nash's approach to ultracontractivity

The implication from Sobolev inequality to ultracontractivity was first proved by Varopoulos ({V}) by observing how Hardy-Littlewood estimate of $P_t f$ evolves under time $t$ under Sobolev inequality, in order to control the regularity of $f$ with $P_t$ for $t$ in a neighbourhood of 0. Although this gives a direct proof of the equivalence, it involves too much tedious technical details. So instead, we will first investigate how the problem was first formulated by John Nash ({N}, {CKS}). The equivalent condition to ultracontractivity that Nash originally stated in his paper is called the Nash inequality. We will first show that a polynomial-exponential ultracontractivity is equivalent to a Nash inequality, and then prove that the Nash inequality is in fact also equivalent to a Sobolev inequality.

**Definition** *A Markov triple is said to satisfy a* ***Nash inequality*** *with constants $C > 0$, $\delta \geq 0$ and exponent $\nu > 0$ if*

$$\|f\|_2^{2+4/\nu} \leq C(\mathcal{E}(f) + \delta\|f\|_2^2)\|f\|_1^{4/\nu}, \quad \forall f \in \mathcal{D}_{\mathcal{E}}.$$

*We will denote this as $NI^\nu(C, \delta)$. The inequality is called to be* ***tight*** *if $C\delta = 1$ and is denoted $NI^\nu(C)$ in this setting.*

**Theorem 4.3** *Fix $\nu \in (0,\infty)$ and $\delta \in [0,\infty)$. Then the following are equivalent :*

*(i) $NI^\nu(C,\delta)$ holds for some $C > 0$.*

*(ii) $\|P_t\|_{1,\infty} \leq Be^{\delta t}/t^{\nu/2}$ for some $B > 0$.*

*Proof.* Using density argument, proving the equivalence for non-negative functions $f \in \mathcal{D}_{\mathcal{E}}$ is sufficient. This is because $\|P_t f\|_\infty \leq \|P_t|f|\|_\infty$ by contraction property and $\mathcal{E}(|f|) \leq \mathcal{E}(f)$. Also, we may only consider $f$ such that $\|f\|_1 = 1$ by rescaling. Note that sub-Markov property gives $\|P_t f\|_1 \leq 1$ for each $t \geq 0$.

Assume $(i)$ holds. Then by $(i)$,

$$\begin{aligned} -\frac{d}{dt}e^{-2\delta t}\|P_t f\|_2^2 &= 2e^{-2\delta t}(\mathcal{E}(P_t f) + \delta\|P_t f\|_2^2) \geq \frac{2}{C}\Big(e^{-2\delta t}\|P_t f\|_2^2\Big)^{1+2/\nu} \\ \Rightarrow \quad \frac{d}{dt}\Big(e^{-2\delta t}\|P_t f\|_2^2\Big)^{-2/\nu} &= -\frac{2}{\nu}\Big(e^{-2\delta t}\|P_t f\|_2^2\Big)^{-1-2/\nu}\frac{d}{dt}\Big(e^{-2\delta t}\|P_t f\|_2^2\Big) \geq \frac{4}{\nu C} \end{aligned}$$

so

$$e^{-2\delta t}\|P_t f\|_2^2 \leq \Big(\frac{4t}{\nu C}\Big)^{-\nu/2}$$

and therefore $\|P_t\|_{1,2} \leq Ce^{\delta t}t^{\nu/4}$. Then the bound on $\|P_t\|_{1,\infty}$ follows from **Proposition 4.1**.

Next Assume $(ii)$ holds. Let $u(t,x) = e^{-\delta t}P_t f(x)$. Then $(ii)$ implies that

$$\int_X fu(t,\cdot)d\mu \leq B/t^{\nu/2}$$

and

$$\begin{aligned} &\frac{\partial}{\partial t}u(t,\cdot) = -\delta u(t,\cdot) - \mathcal{G}u(t,\cdot) \\ \Leftrightarrow \quad &u(t,\cdot) = f - \int_0^t (\mathcal{G} + \delta \cdot id)u(s,\cdot)ds. \end{aligned}$$

Putting these together,

$$\begin{aligned} B/t^{\nu/2} &\geq \int_X fu(t,\cdot)d\mu = \|f\|_2^2 - \int_0^t\int_X f(x)(\mathcal{G} + \delta\cdot id)u(s,x)d\mu(x)ds \\ &\geq \|f\|_2^2 - t(\mathcal{E}(f) + \delta\|f\|_2^2) \end{aligned}$$

or equivalently,

$$\|f\|_2^2 \leq t\delta\|f\|_2^2 + \frac{B}{t^{\nu/2}} + t\mathcal{E}(f).$$

To optimize the inequality over $t > 0$, choose $t = \Big(\frac{B\nu/2}{\delta\|f\|_2^2 + \mathcal{E}(f)}\Big)^{1/(1+\nu/2)}$, then we have the result.

*(End of proof)* □

## Section 4.3 Equivalence of Sobolev inequality and Nash inequality

Proving that Sobolev is equivalent to Sobolev inequality is more tricky.

**Theorem 4.4** *Let $\nu > 2$ and $p = \frac{2\nu}{\nu-2} > 2$. The following are equivalent.*

*(i) $SI^p(C,\delta)$ holds for some $C > 0$ and $\delta \geq 0$.*

*(ii) $NI^\nu(C',\delta')$ holds for some $C' > 0$ and $\delta' \geq 0$.*

*Moreover, $SI^p(C)$ if and only if $NI^\nu(C')$ holds.*

*Proof.* Assume $f \in \mathcal{D}_\mathcal{E}$ is non-negative. Again, we will prove the inequalities are equivalent for only such functions and use density argument to conclude.

Assume $(i)$ holds. Rescale to set $\|f\|_2 = 1$ and assume that $f$ is bounded below by $m > 0$. Define $\lambda(s) = \log(\|f\|_{1/s})$ for $s \in (0,1]$. Then

$$\frac{d}{ds}\lambda(s) = \frac{d}{ds} s\log(\|f\|_{1/s}^{1/s}) = \log(\|f\|_{1/s}^{1/s}) - \frac{1}{s}\frac{\int f^{1/s}\log f d\mu}{\|f\|_{1/s}^{1/s}} = -\frac{1}{\|f\|_{1/s}^{1/s}}\mathrm{Ent}(f^{1/s})$$

$$\frac{d^2}{ds^2}\lambda(s) = \frac{1}{s^3\|f\|_{1/s}^{2/s}}\Big(\mu((\log f)^2 f^{1/s})\mu(f^{1/s}) - (\mu(f^{1/s}\log f))^2\Big) \geq 0$$

where the last inequality follows from Cauchy-Schwarz inequality. Hence $\lambda$ is a convex function. Hence

$$\frac{\lambda(1/2)-\lambda(1/p)}{\frac{1}{2}-\frac{1}{p}} = \nu(\lambda(1/2)-\lambda(1/p)) \leq \frac{\lambda(1)-\lambda(1/2)}{1-\frac{1}{2}} = 2(\lambda(1)-\lambda(1/2)),$$

and we have

$$\begin{aligned}\|f\|_2^{\nu+2} &= e^{(\nu+4)\lambda(\|f\|_2)} \\ &\leq e^{2\log\|f\|_1}e^{\nu\log(\|f\|_p)} \\ &= \|f\|_1^2(\|f\|_p)^\nu\end{aligned}$$

If $f$ does not have lower-bound $m > 0$, then use dominated convergence on the inequality $\|f+c\|_2^{\nu+2} \leq \|f+c\|_1^2\|f+c\|_p^\nu$ on taking limit $c \searrow 0$ to obtain $\|f\|_2^{\nu+2} \leq \|f\|_1^2\|f\|_p^\nu$. Then using $SI^p(C,\delta)$ gives

$$\|f\|_2^{\nu+2} \leq \|f\|_1^2(C\delta\|f\|_2^2 + C\mathcal{E}(f))^{\nu/2}$$

the $NI^\nu(C,\delta)$.

For the converse direction, assume $NI^\nu(C,\delta)$. To prove a Sobolev inequality, a combinatorial trick called 'slicing' is used. The idea is to approximate $f$ with level functions $h$ by slicing $f$ according to the levels of $f$ in such a way that does not lose too much information about the norms of $f$. Ideally, any slicing with sufficiently narrow slicing would work, but in order to make addition of norms of level functions easy, we make choice

$$U_k = \{x : f(x) \geq 2^k\}, \quad f_k = ((f-2^k)\wedge 2^k)\vee 0, \quad \forall k \in \mathbb{Z}$$

so that $\sum_{k\in\mathbb{Z}} f_k = f$ and $\mathbf{1}_{U_k}2^k \geq f_k \geq \mathbf{1}_{U_{k+1}}2^k$. We will approximate $f$ piecewise on $U_k$ in terms of $\mathbf{1}_{U_k}$ and $\mathbf{1}_{U_{k+1}}$. We immediately have that

$$\|f_k\|_2^2 \in [2^{2k}\mu(U_{k+1}), 2^{2k}\mu(U_k)].$$

Then application of Nash inequality on $f_k$ gives

$$\|f_k\|_2^{2+4/\nu} \leq C(\mathcal{E}(f_k) + \delta\|f_k\|_2^2)\|f_k\|_1^{4/\nu}.$$

Setting $L_k = C(\mathcal{E}(f_k) + \delta 2^{2k}\mu(U_k))$ and $H_k = 2^{pk}\mu(U_k)$, this inequality with the previous bound on $\|f_k\|_2^2$ gives

$$H_{k+1} \leq 2^p (L_k)^{\nu/(\nu+2)} H_k^{4/(\nu+2)}, \quad \forall k \in \mathbb{Z}.$$

and therefore upon summing over $k \in \mathbb{Z}$ and using Hölder inequality, we have

$$\begin{aligned} \sum_{k\in\mathbb{Z}} H_{k+1} &\leq 2^p \sum_{k\in\mathbb{Z}} (L_k)^{\nu/(\nu+2)} H_k^{4/(\nu+2)} \\ &\leq 2^p \Big(\sum_{k\in\mathbb{Z}} L_k\Big)^{\nu/(\nu+2)} \Big(\sum_{k\in\mathbb{Z}} \alpha_k^2\Big)^{2/(\nu+2)} \\ &\leq 2^p \Big(\sum_{k\in\mathbb{Z}} L_k\Big)^{\nu/(\nu+2)} \Big(\sum_{k\in\mathbb{Z}} \alpha_k\Big)^{4/(\nu+2)} \\ \Rightarrow \quad \Big(\sum_{k\in\mathbb{Z}} H_k\Big)^{1-\frac{4}{\nu+2}} &\leq 2^p \Big(\sum_{k\in\mathbb{Z}} L_k\Big)^{\nu/(\nu+2)}. \end{aligned}$$

Also, using the bound $f \leq \sum_{k\in\mathbb{Z}} 2^{k+1}\mathbf{1}_{U_k\setminus U_{k-1}}$, we have $\|f\|_p^p \leq \sum_{k\in\mathbb{Z}} 2^{p(k+1)}\mu(U_k\setminus U_{k-1}) \leq 2^p \sum_k H_k$, and putting this in the inequality above gives

$$\begin{aligned} \Big(\frac{1}{2}\|f\|_p\Big)^{\frac{2\nu}{\nu-2}\cdot\frac{\nu-2}{\nu+2}} &\leq 2^p \Big(\sum_{k\in\mathbb{Z}} L_k\Big)^{\nu/(\nu+2)} \\ \Leftrightarrow \quad \|f\|_p^2 &\leq C'(\delta'\|f_k\|_2^2 + \sum_k \mathcal{E}(f_k)) \end{aligned}$$

for some $C' > 0$ and $\delta' \geq 0$. Since $f_k$'s have disjoint support, we have $\sum_{k=-N}^{N} \mathcal{E}(f_k) = \mathcal{E}((f_k \wedge 2^N) \vee 2^{-N})$, and if we let $g_N = (f_k \wedge 2^N) \vee 2^{-N}$, then by definition of being a Markov triple, $|g_N(x) - g_N(y)| \leq |f(x) - f(y)|$, $|g_N(x)| \leq |f(x)|$ for all $x, y \in X$ gives $\mathcal{E}(g_N) \leq \mathcal{E}(f)$. This holds for all $N$, so we have $\sum_k \mathcal{E}(f_k) \leq \mathcal{E}(f)$, which concludes the proof.

*(End of proof)* □

## Section 4.4 Relation with Different inequalities

**Remark :** When we were deriving in the proof of **Theorem 4.4** $(ii) \Rightarrow (i)$, we have used the convexity of the function $\lambda(s) = \log(\|f\|_{1/s})$. Again, assume $\|f\|_2 = 1$. Exploiting the convexity further, we have

$$\frac{\lambda(1/2) - \lambda(1/p)}{\frac{1}{2} - \frac{1}{p}} = \nu(\lambda(1/2) - \lambda(1/p)) \leq \lambda'(1/2) = -\frac{\mathrm{Ent}(f^2)}{\|f\|_2^2}$$

so a Sobolev inequality $SI^\nu(C, \delta)$ implies

$$\mathrm{Ent}(f^2) \leq \nu \log(\|f\|_p) \leq \frac{\nu}{2}\log(C(\delta\|f\|_2^2 + \mathcal{E}(f))).$$

Therefore, whenever $\|f\|_2 = 1$, we have

$$\mathrm{Ent}(f^2) \leq \frac{\nu}{2} C(\delta + \mathcal{E}(f)),$$

a log-Sobolev inequality. So we have the following.

**Corollary 4.5** *A Sobolev inequality $SI^p(C,\delta)$ with $C > 0$ and $\delta \geq 0$ implies a log-Sobolev inequality $LSI(C',\gamma')$.*

To appreciate the full power of the implication, it will be better to see a more general statement that any reasonable degree of ultracontractivity is sufficient for a log-Sobolev inequality.

**Theorem 4.6** *Let $P_t$ be ultracontractive with $\|P_t\|_{1,2} \leq e^{b(t)}$ for all $t > 0$, where $b(t)$ is a monotonically decreasing and continuous. Then for all $\tau > 0$*

$$\mathrm{Ent}(f^2) \leq 2\tau\mathcal{E}(f) + 2b(\tau)\|f\|_2^2, \quad \forall f \in \mathcal{D}_{\mathcal{E}}.$$

*for $f \in \mathcal{D}_{\mathcal{E}}$ and integrable.*

*Proof.* *(sketch)* By rescaling, assume $\|f\|_2 = 1$, $f \in \mathcal{D}'_P$, and also $f$ is bounded by $[m, M] \subset \mathbb{R}_{\geq 0}$. We have not defined how to extend the semigroup $P_t$ for $t \in S =: \{t \in \mathbb{C} : 0 \leq \mathrm{Re}(t) \leq \tau\}$, but assume we have, and also assume that the family of operators $(P_z)_{z\in S}$ satisfies the assumptions of Stein interpolation theorem. Then we may apply Stein interpolation theorem with bounds $\|P_0\|_{2,2} \leq 1$, $\|P_\tau\|_{2,\infty} = \|P_\tau\|_{1,2} \leq e^{b(\tau)}$ to obtain the bound

$$\|P_s\|_{2,r(s)} \leq e^{\frac{s}{\tau}b(\tau)}$$

where $r(s) = \frac{2\tau}{\tau - s}$. In particular, we have $\|P_s f\|_{r(s)}^{r(s)} \leq \exp(s \cdot r(s)b(\tau)/\tau)$ and therefore

$$\frac{d}{ds}\int_X (P_s f)^{r(s)} d\mu\bigg|_{s=0} = 2\mathcal{E}(f) + \frac{1}{2}\int_X f^2 \log(f^2) d\mu \leq \frac{d}{ds} e^{sr(s)b(\tau)/\tau}\bigg|_{s=0} = 2M(\tau)/\tau$$

and therefore rearrangement gives the result.

For general $f \in \mathcal{D}_{\mathcal{E}}$, use usual density argument.

Nevertheless, a tight Sobolev inequality also implies a Poincaré inequality, given that $\Gamma$ is mass-preserving.

**Proposition 4.7** *For a mass-preserving standard Markov triple, a tight Sobolev inequality $SI(C)$ implies a Poincaré inequality $PI(C/(p-2))$.*

*Proof.* Consider $g \in \mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$ and $\epsilon > 0$ so that $\|\epsilon g\|_\infty < 1$. Then

$$\begin{aligned}
&\|1 + \epsilon g\|_p^2 - \|1 + \epsilon g\|_2^2 \\
&= 1 + 2\epsilon\mu(g) + (p-1)\epsilon^2\mu(g^2) + \frac{1}{p}\Big(\frac{2}{p} - 1\Big)\epsilon^2 p^2 \mu(g)^2 - 1 - 2\epsilon\mu(g) - \epsilon^2\mu(g^2) + o(\epsilon^3) \\
&= (p-2)\epsilon^2(\mu(g^2) - \mu(g)^2) \leq \liminf_{m\to\infty} C\mathcal{E}(\psi_m + \epsilon g) = \epsilon^2 C\mathcal{E}(g)
\end{aligned}$$

as $\epsilon \to 0$. Therefore

$$(p-2)\mathrm{Var}(g) \leq C\mathcal{E}(g)$$

as desired. For general $g \in \mathcal{D}_{\mathcal{E}}$, use density of $\mathcal{D}_{\mathcal{E}} \cap L^\infty(\mu)$ in $\mathcal{D}_{\mathcal{E}}$.

*(End of proof)* □

Putting these results together, we can make a tight log-Sobolev inequality from a tight Sobolev inequality. This makes use of the following lemma.

**Lemma 4.8** *Suppose we have $LSI(C, \gamma)$ and $PI(C')$. Then we also have $LSI(C'')$ for some $C'' > 0$*

*Proof.* First, apply $LSI(C, \gamma)$ to the function $(f - \mu(f))$, then

$$\mathrm{Ent}((f - \mu(f))^2) \leq C\mathcal{E}(f) + C\delta\|f - \mu(f)\|_2^2 \leq C\mathcal{E}(f) + C'C\delta\mathcal{E}(f) = C(1 + \gamma C')\mathcal{E}(f)$$

Hence if we can obtain some bound of $\mathrm{Ent}(f^2)$ in terms of $\mathrm{Ent}((f - \mu(f))^2)$, then we can prove a log-Sobolev inequality. Indeed, from **Lemma 3.3**, we have

$$\mathrm{Ent}(f^2) \leq \mathrm{Ent}((f - \mu(f))^2) + 2\mathrm{Var}(f) \leq (C + \gamma CC' + 2C')\mathcal{E}(f)$$

so we have the conclusion.

*(End of proof)* □

**Corollary 4.9** *A tight Sobolev inequality $SI(C)$ implies a tight log-Sobolev inequality $LSI(C')$.*

# Chapter 5 CONCENTRATION BEHAVIOUR OF ISING MODEL

In this chapter, we see one application of log-Sobolev inequality.

### § A general introduction to applications of log-Sobolev inequalities

There are many interesting applications of the log-Sobolev inequality. True power of the log-Sobolev inequality comes when dealing with a high-dimensional system. In a reductionist's point of view, this is a consequence of the fat that the log-Sobolev inequality is preserved under independent product. That is, if $(X_j, \mu_j, \Gamma_j)$ are Markov triples satisfying $LSI(C, \gamma)$, and $(X, \nu, \Gamma) = (\prod_{j=1}^N X_j, \otimes_{j=1}^N \mu, \otimes_{j=1}^N \Gamma_j)$, then the new Markov triple $(X, \nu, \Gamma')$ also satisfies $LSI(C, \gamma)$. This can be justified using a simple argument involving hypercontractivity : using induction, it is sufficient to prove this just for the case $N = 2$. Letting $q, p(t)$ and $M(t, q)$ be as in the statement of Gross' hypercontractivity theorem and $P_t^1$, $P_t^2$ two transition semigroups, we have $\|P_t^1 \otimes P_t^2\|_{q,p(t)} \le \exp(M(t, q))$, so $LSI(C, \gamma)$ holds on the new product system.

However, this result has very limited use, since the system $(X, \mu, \Gamma)$ of interest usually has the measures $(\mu_i)_{i=1}^N$ dependent to each other. For example, if we think of $(X, \mu, \Gamma)$ as the configuration space of $N$ particles and $\mu_j$ as the distribution of a particles labelled $j \in \{1, \cdots, N\}$, then the independence of measures $\mu_j$ can be understood as saying that the particles have no interaction. This is rarely of interest. Instead, we usually think of systems of weak interactions, where the log-Sobolev inequality can be conserved under the limit $N \to \infty$, while the dependences between $\mu_j$'s are still present. (More sanely, we make reasonable hypothesis on interactions that makes us able to preserve the log-Sobolev inequality, and decide to the call the system with the hypothesis to have weak interactions.)

## Section 5.1 The Gibbs measure

Gibbs measure models the distribution of spin lattice configurations. Think of a lattice $\Lambda = \mathbb{Z}^d$ with (non-quantum) magnetic spins placed at each lattice point, with spins parallel to a single direction, say $\hat{e}$, and having real-valued magnitude denoted by $\omega(i) \in \mathbb{R}_{(i)} = \mathbb{R}$ for $i \in \Lambda$ and each configuration $\omega \in \Omega = \bigotimes_{i\in\Lambda} \mathbb{R}_{(i)}$ where $\Omega$ can be thought to be the set of all configurations of the spins. An external potential is present, so each spin at $i \in \Lambda$ has self-energy given by $V(\omega(i))$ and interaction between spins at site $i, j \in \Lambda$ makes potential $-J_{ij}\omega(i)\omega(j)$. Assume that there is no extra contribution to the potential of the system, so the total potential for each configuration $\omega \in \Omega$ would be

$$E(\omega) = -\sum_{i,j\in\Lambda} J_{ij}\omega(i)\omega(j) + \sum_{i\in\Lambda} V(\omega(i))$$

To define a measure on the configuration space $\Omega$, we first define a $\sigma$-algebra on the configuration space $\Omega$. Let

$$\Lambda^{(n)} := [-n, n]^d \cap \mathbb{Z}^d, \quad \mathcal{F}_{\Lambda^{(n)}} := \bigotimes_{i\in\Lambda^{(n)}} \mathcal{B}(\mathbb{R}_{(i)}), \quad \mathcal{F}^{(n)} := \mathcal{F}_{\Lambda^{(n)}} \otimes \{\varnothing, \mathbb{R}^{\Lambda\setminus\Lambda^{(n)}}\}.$$

Then define $\mathcal{F} = \sigma(\cup_n \mathcal{F}^{(n)} : n \in \mathbb{N})$, a $\sigma$-algebra on $\Omega$. For each subset $A \subset \Lambda$, let

$$P_A : \Omega \to \bigotimes_{i\in\Lambda} \mathbb{R}_{(i)} =: \Omega_A$$

be the projection map and $\mathcal{F}_A$ be the push-forward of $\mathcal{F}$ along $P_A$ and $\mathcal{F}^A = \mathcal{F}_A \otimes \{\varnothing, \mathbb{R}^{\Lambda\setminus A}\}$. Having these, let $\rho$ be the Lebesgue measure on $\mathbb{R}$, $(\rho_j)_{j\in\Lambda}$ be independent copies of $\rho$ at lattice point and define $\rho_{\Lambda^{(n)}} = \otimes_{i\in\Lambda^{(n)}}\rho_i$. $\rho_\Lambda$ would be the canonical extension of $\rho_{\Lambda^{(n)}}$ to the whole $\Lambda$, *i.e.* it is the unique measure with $(P_{\Lambda^{(n)}})_{\#}\rho_\Lambda = \rho_{\Lambda^{(n)}}$. Also let $(P_A)_{\#}\rho_\Lambda = \rho_A$.

Note that the $\sigma$-algebra is generated by the topology induced by the metric

$$d(\omega,\overline{\omega}) = \sum_{n=1}^{\infty} 2^{-n}\|\omega(i)-\overline{\omega}\|_{L^\infty(\Lambda^{(n)};\mathbb{R})}$$

hence $\Omega$ can be considered a Polish space. Although this fact will not be quoted explicitly henceforth, this makes us able to apply disintegration of measure on any measure on $(\Omega,\mathcal{F})$.

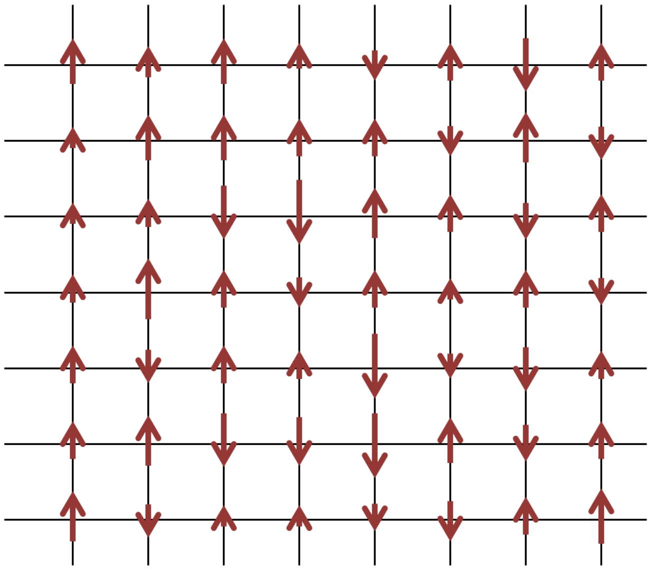

Figure 1: An example of spin configuration on $([0,7]\times[0,6])\cap\mathbb{Z}^2$.

We will define the Gibbs measure at temperature $\beta^{-1}$ to be the thermal equilibrium measure of distribution of $\{s(i) : i\in\Lambda\}$ surrounded by environment with temperature $\beta^{-1}$. According to arguments of Maxwell, Boltzmann and later further investigated by Gibbs, each configuration of a non-quantum finite system constitutes of probability proportional to $\exp(-\beta E(\omega))$ where $E(\omega)$ is the energy of configuration $\omega$, so called the Boltzmann statistics. However, in our case, since the system is infinite, the integral $\int_\Omega \exp(-\beta E(\omega))\rho(d\omega)$ may not converge. So we have to use a non-direct method to define a Gibbs measure on the system. As we have done for defining the $\sigma$-algebra and $\rho$ on $\Omega$, we will first define partial Gibbs measure for finite subsets of $\Omega$ and extend it to the whole $\Omega$ :

- Let $F(\Lambda)$ be the set of finite subsets of $\Lambda$. For $A\in F(\Lambda)$ and $S\in\Omega_{A^c}$, we define the conditional Gibbs measure on $\Omega_A$ given $S$ by

$$\pi_{A;S}(d\omega_A) \equiv \mu_A(S,d\omega_A) = Z_{A;S}^{-1}\exp(-\beta E_{A;S}(\omega_A))\rho_A(d\omega_A)$$

  where $Z_{A;S} = \int_{\Omega_A}\exp(-\beta E_{A;S}(\omega_A))\rho_A(d\omega_A)$ and

$$E_{A;S}(\omega_A) = \sum_{i\in A} V(\omega_A(i)) - \sum_{i,j\in A} J_{ij}\omega_A(i)\omega_A(j) - \sum_{i\in A, j\in A^c}\omega_A(i)S(j)$$

  From here on, we just assume $\beta=1$, as scaling $V$ and $J$ accordingly always give the same result.

- Define partial Gibbs measure on $\Omega$ given $S$ by

$$\mu_{A;S}(d\omega_A, d\omega_{A^c}) = \pi_{A;S}(d\omega_A)\otimes\delta_S(d\omega_{A^c}),\quad \omega_A\in\Omega_A,\ \omega_{A^c}\in\Omega_{A^c}$$

Now the most natural way to define an extension of the partial measures will be to let $\mu(d\omega|P_{A^c}(\omega) = S) = \mu_{A;S}(d\omega)$. However, this can not be made sense when $\mu(P_{A^c}\omega = S) = 0$, so we rather choose to invent a version of the partial measure that does not refer to a specific configuration $S$, but just on a set in $\mathcal{F}_{\Lambda\setminus A}$. For $U \in \mathcal{F}_{\Lambda\setminus A}$, $W \in \mathcal{F}_A$, define

$$\mu_A \mathbf{1}_W(P_A(\omega))\mathbf{1}_U(P_{A^c}(\omega)) = \int_{\Omega_A} \mathbf{1}_W(P_A(\varpi))\mathbf{1}_U(P_{A^c}(\varpi))\mu_{A;P_{A^c}(\omega)}(d\varpi).$$

Using monotone class-type argument, this has extension to bounded measurable functions, say

$$\mu_A f(\omega) = \int_{\Omega_A} f(\varpi)\mu_{A;P_{A^c}(\omega)}(d\varpi), \quad \forall f \in L^\infty(\Omega, \mathcal{F}, \rho).$$

- A probability measure $\mu(=\mathbb{P})$ is called a **Gibbs measure** if it satisfies $\mu(f|\mathcal{F}_{A^c}) = \mu_A f$ for any bounded measurable function, *i.e.* an $L^\infty$-random variable on $\Omega$.

If a Gibbs measure exists and is unique, it can be viewed as a canonical extension of the Boltzmann distribution on finite systems with the universal property that, whenever $\tilde{\mathbb{P}}$ is another probability measure (probably on a larger probability space) with such restriction properties, then $\tilde{\mathbb{P}} \xrightarrow{|\mathcal{F}_{A^c}} \mu_A$ factors by

$$\tilde{\mathbb{P}} \xrightarrow{\quad|\mathcal{F}_\Lambda\quad} \mu \xrightarrow{\quad|\mathcal{F}_{A^c}\quad} \mu_A.$$

Although we have made sense of what the equilibrium measure $\mu$ should be, we are still far away from what we actually intend to prove. In particular, there are three major questions that are crucial to our survey :

*1.* The existence of the Gibbs measure has to be proved.

*2.* The stochastic process that governs the evolution of spin configuration has to be constructed.

*3.* If *1, 2* are solved, then the convergence of the process to the Gibbs measure should be proved. Provided the convergence, the Gibbs measure would be unique.

Question 1 and 2 are not that hard to prove given a right condition on the potential functions $J$ and $V$, but they are not quite relevant to studying the convergence of the process, so we are only going to see a brief introduction to the key ideas and notations that will be needed later on. For answering Question 3, we will establish a log-Sobolev inequality on the system, given the so called 'weak-interactions' assumption.

## Section 5.2 Existence of Gibbs measure and the Glauber-Langevin process

As is hinted from the construction of a Gibbs measure, for a Gibbs measure to exist for an infinite lattice, it is necessary for the total energy of the system to be bounded under a reasonable scaling. The easiest way to make this work is to make strong bounds on $J$ and $V$.

**Hypothesis** *We assume the following assumptions to hold true in the rest of the chapter :*

$(H1)$ *The coupling constant $J(i)$ is non-zero on finite number of points and $J(i) = J(-i)$.*
$(H2)$ *The external potentail $V$ is a non-constant polynomial bounded below.*

*In particular, $V$ is polynomial of even degree.*

In a physical point of view, if the external field does not make contribution on the potential of the system, then the potential would have multiple ground states, from which we may expect multiple Gibbs measures. Such a setting makes the problem extraordinarily complicated, so in this regard, we always restrict $V$ to be non-constant in $(H2)$. Also, in a similar spirit, we will often fix $V$ and make statements when $J$ is sufficiently small compared to $V$, *i.e.* we will find $\sigma > 0$ such that certain statements hold whenever $\sum_{i\in\Lambda} |J(i)| \leq \sigma$. This is called the **weak-interaction** assumption on the system.

Now under the given hypothesis, the following holds :

**Proposition 5.1** *With $V$ fixed, whenever $\sum_{i\in\Lambda} |J(i)| \leq \sigma = \inf V''(x)$ or $V$ has degree $\geq 4$, then there is at least one Gibbs measure that is supported on the set*

$$\Omega' := \{\omega \in \Omega : |\omega(i)| \text{ has polynomial growth as } |i| \to \infty\}.$$

*Proof. (sketch)* If we can show tightness of the sequence of measures $(\mu_{\Lambda^{(n)};\mathbf{0}})_{n\in\mathbb{N}} \subset \mathcal{P}(\Omega')$, then Prohorov's theorem indicates that there is a limit point $\mu$. Let us just check that $\mu$ has the desired properties.

First we note the following : let $A \subset B(\subset \Lambda)$ are two finite sets and $f$ a bounded measurable function, then we have $\mu_B\mu_A f = \mu_A f$. To see this, first assume that $f$ is a simple function of form $f(u_A, v_{B\setminus A}, w_{B^c}) = \mathbf{1}_U(u_A)\mathbf{1}_V(v_{B\setminus A})\mathbf{1}_W(w_{B^c})$ for $U \in \Omega_A$, $V \in \Omega_{B\setminus A}$ and $W \in \Omega_{B^c}$. Then

$$\begin{aligned}
\mu_B\mu_A f(w_A, w_{B\setminus A}, w_{B^c}) &= \int_{\Omega_B}\int_{\Omega_A} \mathbf{1}_U(u_A)\mathbf{1}_V(v_{B\setminus A})\mathbf{1}_W(w_{B^c})\pi_{A;(v_{B\setminus A}, w_{B^c})}(du_A)\pi_{B;(w_{B^c})}(dv_B) \\
&= \int_{\Omega_{B\setminus A}}\int_{\Omega_A} \mathbf{1}_U(u_A)\mathbf{1}_V(v_{B\setminus A})\mathbf{1}_W(w_{B^c})\pi_{A;w_{B^c}}(du_A, dv_{B\setminus A}) \\
&= \mu_B f(w_A, w_{B\setminus A}, w_{B^c})
\end{aligned}$$

by integrating over $dv_A$ in the second line (where $x_C = P_C(x)$ for $x \in \Omega$, $C \subset \Lambda$). So monotone class argument gives $\mu_B\mu_A f = \mu_B f$.

Now $A \subset \Lambda$ be any bounded set and $f$ be bounded continuous function that $f(\omega)$ only depends on $P_A(\omega)$. There is $N$ such that $A \subset \Lambda^{(n)}$ for each $n \geq N$. Then we have $\mu_{\Lambda^{(n)}}\mu_A f = \mu_{\Lambda^{(n)}} f$ for each $n \geq N$, and therefore the weak convergence $\mu_{\Lambda^{(n)}} \to \mu$ gives

$$\mu_{\Lambda^{(n)}} f = \mu_{\Lambda^{(n)}}\mu_A f \to \mu\,\mu_A f = \mu f$$

so $\mu$ is a Gibbs measure.

Note that the statement $\mu_B\mu_A = \mu_B$ whenever $A \subset B$ is in accordance with the computation

$$\mu_B\mu_A = \mu(\mu(\cdot|\mathcal{F}_{A^c})|\mathcal{F}_{B^c}) = \mu(\cdot|\mathcal{F}_{B^c}) = \mu_B$$

So it says nothing more than that the measures $(\mu_A)_{A\in F(\Lambda)}$ are related to each other by conditional expectations.

Next, as already notified, we investigate the stochastic process associated to the measure system, so called **Glauber-Langevin diffusion process**. We will assume that the associated process is a diffusion process that can be viewed as a type of generalized Ornstein-Uhlenbeck process. Although we do not justify this physically, this becomes plausible if we invoke that a generalized Ornstein-Uhlenbeck process $dY_t = -\frac{1}{2}\nabla F(Y_t)dt + dB_t$ has equilibrium measure $\frac{1}{Z}e^{-F}dm(x)$, which is exactly the form of a Boltzmann distribution on a finite lattice system. In order to obtain $Z^{-1}\exp(-\beta U_{\Lambda^{(n)}}(x,S))$ as a equilibrium measure for $n \in \mathbb{N}$ and $S \in \Omega_{A^c}$, $x \in \Omega_A$, we define a conditioned Glauber-Langevin process on $\Lambda^{(n)}$ to be

$$X_t^{(\Lambda^{(n)},S,x)} = x + B_t^{(\Lambda^{(n)})} - \int_0^t \nabla E_{A;S}(X_u^{(\Lambda^{(n)},S,x)})du$$

where $B_t^{(\Lambda^{(n)})}$ is a vector of independent Brownian motions on lattice points of $A$. Then one can show that the sequence $X^{(\Lambda^{(n)},S,x_n)}$ converges as $n \to \infty$, whenever $x_n$ is given as the projection of a configuration $\omega \in \Omega'$ on $\Lambda^{(n)}$ and $S$ is a projection of some configuration in $\Omega'$ ton $(\Lambda^{(n)})^c$. So we have the following.

**Proposition 5.2** *For any $x \in \Omega'$, there is a continuous process $(X_t^x)_{t\geq 0}$ taking place in $\Omega'$ and $X_0^x = x$, which is unique up to a.s. modification. Furthermore, any Gibbs measure is reversible for the associated semigroup.*

Now call the associated semigroup to such a process $(P_t)_{t\geq 0}$. Note that now the carré du champ operator is $\Gamma(f) = \sum_i |D_i f|^2$ and Dirichlet form is specified as $\mu(\sum_i |D_i f|^2)$, with appropriate domains.

**Remark :** Once we know that the sequence of measures $(\mu_{\Lambda^{(n)};\mathbf{0}})_{n\in\mathbb{N}}$ is tight, this implies that any Gibbs measure should be obtained as a weak limit of some subsequence of such measures. Also, the Dirichlet form is given as a limit of finite-dimensional Dirichlet form, so if we can verify a log-Sobolev inequality that is independent of choice of $n$, then this implies that log-Sobolev inequality also holds for the Gibbs measure.

## Section 5.3 Log-Sobolev inequality on the Glauber-Langevin process

Let $f$ be a bounded measurable function on $(\Omega, \mathcal{F}, \rho)$ and has bounded derivative in the $i^{th}$ coordinate. Let the set of such functions be $\mathcal{D}_0 \subset \mathcal{D}_{\mathcal{E}}$.

We first show a preliminary result on log-Sobolev inequality of each spin in the lattice with the surrounding environment fixed. The convergence rate turns out to have a bound uniform over the surrounding configurations. We will prove this result following the tradition of using the celebrated Bakry-Émery criterion.

**Lemma 5.3** *(Bakry-Émery criterion) Let $F \in C^2(\mathbb{R}^d)$, with $\nabla^2 F \geq \frac{1}{C} I$, $\mu(dx) = e^{-F(x)}m(dx)$ and $\Gamma(f) = |\nabla f|^2$. Then $(\mathbb{R}^d, \mu, \Gamma)$ satisfies a tight log-Sobolev inequality with constant $C$.*

*Proof.* Let $P_t$ be the semigroup associated to the process $dX_t = -\frac{1}{2}\nabla F(X_t)dt + dB_t$, and $\mathcal{G}f = \frac{1}{2}\triangle f - \frac{1}{2}\nabla F \cdot \nabla f$ be the generator of the semigroup. Letting $\Gamma_2(f) = \frac{1}{2}(\mathcal{G}\Gamma(f) - 2\Gamma(\mathcal{G}f, f))$, after a lengthy computation, we see that $D^2 F \geq \frac{1}{C} I$ implies $\Gamma_2(f) \geq \frac{1}{C}\Gamma(f)$ for any $f \in \mathcal{D}_{\mathcal{E}}$. Then for $f \in C_b^1(\mathbb{R}^d)$,

$$\begin{aligned}
\frac{d}{ds} P_{t-s}|\nabla(P_s f)|^2 &= -P_{t-s}\mathcal{G}|\nabla(P_s f)|^2 + 2P_{t-s}\nabla(P_s \mathcal{G} f) \cdot \nabla(P_s f) \\
&= P_{t-s}(-\mathcal{G}|\nabla P_s f|^2 + 2\Gamma(P_s f, \mathcal{G} P_s f)) \\
&= -2P_{t-s}\Gamma_2(\nabla P_s f) \\
&\leq -\frac{2}{C} P_{t-s}|\nabla(P_s f)|^2.
\end{aligned}$$

So we have $|\nabla(P_t f)|^2 \le e^{-2t/C} P_t |\nabla f|^2$. So if $g \in C_b^1(\mathbb{R}^d)$ is bounded below by $m > 0$, then

$$\begin{aligned}
\mathcal{E}(P_t g, \log P_t g) &= -\mu((\log P_t g)\mathcal{G} P_t g) = -\mu(P_t \log P_t g \mathcal{G} g) = \mu(\Gamma(P_t \log P_t g, g)) \\
&\le \mu(\Gamma(g)^{1/2}\Gamma(P_t \log P_t g)^{1/2}) \\
&\le \mu(g^{-1}\Gamma(g))^{1/2}\mu(g\Gamma(P_t \log P_t g))^{1/2} \\
&\le \mathcal{E}(\log g, g)^{1/2} e^{-t/C}\mu(g P_t \Gamma(\log P_t g))^{1/2} \\
&= e^{-t/C}\mathcal{E}(\log g, g)^{1/2}\mu(P_t g \Gamma(\log(P_t f)))^{1/2} \\
&= e^{-t/C}\mathcal{E}(\log g, g)^{1/2}\mathcal{E}(\log P_t g, P_t g)^{1/2}
\end{aligned}$$

using the inequality obtained above, chain rule and Cauchy-Schwarz inequality multiple times. So we have

$$-\frac{d}{dt}\mathrm{Ent}(P_t g) = \mathcal{E}(\log P_t g, P_t g) \le e^{-2t/C}\mathcal{E}(\log g, g)$$

Integrating both sides with $t$, we obtain $\mathrm{Ent}(g) - \mathrm{Ent}(P_t g) \le \frac{C}{2}(1 - e^{-2t/C})\mathcal{E}(g, \log g)$, so upon taking $t \to \infty$, we obtain $\mathrm{Ent}(g) \le \frac{C}{2}\mathcal{E}(g, \log g) + \limsup_{t\to} \mathrm{Ent}(P_t g)$.

We next show $\mathrm{Ent}(P_t g) \to 0$ as $t \to \infty$. For brevity, denote $\psi : x \mapsto x \log x$, and $\|\psi\|_{Lip} = K$. Then

$$\begin{aligned}
\left| \int \psi(P_t g) - \psi(P_t g) d\mu \right| &\le K \int |P_t g - \mu(g)| d\mu = K \int \left| \int (P_t g(x) - P_t g(y))\mu(dy) \right| \mu(dx) \\
&\le K \iint |P_t g(x) - P_t g(y)| \mu \otimes \mu(dx, dy)
\end{aligned}$$

In order to estimate $P_t g(x) - P_t g(y)$, consider two processes $M_t^x$ and $M_t^y$ starting at $x$ and $y$ respectively and having generator $\mathcal{G}$. Then Itô's formula gives

$$d|M_t^x - M_t^y|^2 = -2(M_t^x - M_t^y)(\nabla(F)(M_t^x) - \nabla(F)(M_t^y))dt \le -\frac{2}{C}|M_t^x - M_t^y|^2 dt$$

so Grönwall's lemma gives $\mathbb{E}(|M_t^x - M_t^y|)^2 \le e^{-2t/C}||x - y|^2$. This implies

$$|P_t g(x) - P_t g(y)| = |\mathbb{E}(g(M_t^x) - g(M_t^y))| \le \|\nabla g\|_\infty |x - y| e^{-t/C} \xrightarrow{t\to\infty} 0$$

and therefore we also have $\mu(\psi(P_t g)) - \psi(\mu(P_t g)) \to 0$ as $t \to \infty$.

Therefore we have $\mathrm{Ent}(g) \le \frac{C}{2}\mathcal{E}(g, \log g) = \int_{\mathbb{R}^d} \Gamma(g, g)/g d\mu$ for $g$ bounded below by $m > 0$. For general non-negative integrable $g$, just use a routine argument,

$$\mathrm{Ent}(g) = \lim_{c\searrow 0} \mathrm{Ent}(g + c) \le \liminf_{c\searrow 0} 2C \int_{\mathbb{R}^d} \frac{\Gamma(g + c, g + c)}{g + c} d\mu = 2C \int_{\mathbb{R}^d} \frac{\Gamma(g, g)}{g} d\mu$$

*(End of proof)* □

In order to use this criterion to prove a log-Sobolev inequality with potential $V$, is is desirable to extract out a convex smooth part from $V$. This can be achieved via the following procedure.

**Lemma 5.4** *Let $V : \mathbb{R} \to \mathbb{R}$ be a non-constant polynomial bounded below. Then there is a decomposition $V = V_1 + V_2$ where $V_1$ has a compact support and $V_2 \in C^2(\mathbb{R})$ has $\inf_x V''(x) > 0$.*

*Proof.* Let $u$ be a smooth function supported and strictly positive on the set $\{x : V''(x) \leq 0\}$ and for $d > 1$, let

$$v_d(x) = \begin{cases} \int_{-d-1}^{x} dz \int_{-d-1}^{z} dy \Big( -2u(y) + u(y+d) + u(y-d) \Big) & \text{if } |x| \leq d+1 \\ 0 & \text{if otherwise} \end{cases}$$

Then for $c$ and $d$ sufficiently large, $V(x) - cv_d(x)$ has second derivative strictly positive, and $cv_d(x)$ has compact support.

*(End of proof)* □

**Lemma 5.5** *Suppose a Markov triple $(X, \mu, \Gamma)$ satisfies $LSI(C)$ and $\nu$ is a measure on $X$ such that $\alpha^{-1}\nu \leq \mu \leq \alpha\nu$ for some $\alpha \geq 1$. Then $(X, \nu, \Gamma)$ satisfies $LSI(\alpha^2 C)$.*

*Proof.* Note that, for $c(x) = x \log x$, we have $c(x) - c(y) - c'(y)(x - y) \geq 0$ by convexity, so choosing $x = \mu(f)$ gives $c(\mu(f)) - c(y) - c'(y)(\mu(f) - y) \geq 0$ and therefore

$$\mu(c(f)) - c(\mu(f)) = \mathrm{Ent}_\mu(f) \leq \mu(c(f) - c(y) - c'(y)(f - y))$$

Also, if we choose $y = \mu(f)$ then this becomes an equality, so we have

$$\begin{aligned} \mathrm{Ent}_\mu(f) &= \inf_{y \in \mathbb{R}} \mu(c(f) - c(y) - c'(y)(f - y)) \\ \text{and} \quad \mathrm{Ent}_\nu(f) &= \inf_{y \in \mathbb{R}} \nu(c(f) - c(y) - c'(y)(f - y)) \end{aligned}$$

Therefore we have $\mathrm{Ent}_\nu(f) \leq \alpha \mathrm{Ent}_\mu(f)$. We also trivially has $\mathcal{E}_\mu(f) \leq \alpha \mathcal{E}_\nu(f)$, so $LSI(\alpha^2 C)$ holds for $(X, \nu, \Gamma)$.

*(End of proof)* □

With the extraction procedure described in the preceding two lemmas, we have the log-Sobolev inequality on each spin without much difficulty.

**Proposition 5.6** *(Pointwise log-Sobolev inequality) The measure $\mu_{i,\omega'} \equiv \mu_{\{i\},\omega'}$ satisfies a tight log-Sobolev inequality uniform over the choice $\omega' \in \Omega_{\{i\}^c}$, i.e. there is $C_p$ with*

$$\mu_{i,\omega'}(f^2 \log(f^2)) - \mu_{i,\omega'}(f^2) \log(\mu_{i,\omega'}(f^2)) \leq C_p \mu_{i,\omega}(|\nabla_i f|^2), \quad \forall f \in \mathcal{D}_0$$

*Proof.* Using **Lemma 5.4**, we have decomposition $V(x) = V_1(x) + V_2(x)$ where $V_2 \in C_c^\infty(\mathbb{R})$ and $\inf V_1'' \geq C_0 > 0$. Then the potential can be written as

$$\begin{aligned} F(x) \coloneqq E_{i,\omega'}(\omega_i) &= V(\omega_i(i)) - \sum_{j \neq i} J(j - i)\omega'(j)\omega_i(i) \\ &= V_1(x) + V_2(x) - Jx \end{aligned}$$

where $x = \omega_i(i)$ and $J = \sum_{j \neq i} J(j - i)\omega'(j)$. Then by the Bakry-Émery criterion(**Lemma 5.3**), the measure $\mu_1(dx) = e^{-V_1(x)+Jx} m(dx)$ satisfies a $LSI(C_0)$. Now since $V_2$ has compact support, so $\mu(dx) = e^{-V(x)+Jx} m(dx)$ only differs upto $\alpha = e^{\|V_2\|_{L^\infty}} < \infty$ from $\mu_1$. Now **Lemma 5.5** applies, and concludes that $\mu$ satisfies $LSI(\alpha^2 C_0)$.

*(End of proof)* □

We next prove a bound that will be useful later on.

**Proposition 5.7** *For $V$ fixed, there is $\sigma > 0$ such that whenever $\sum_{j\in\Lambda} |J(j)| \leq \sigma$, the measure $\mu_{i,\omega'} \equiv \mu_{\{i\},\omega'}$ satisfies the following : there are constants $(Q_{ij})_{i,j}$ such that $Q_{ij} = Q_{ji}$, $Q_{ii} = 0$, $\sum_{i\neq j} Q_{ij} \leq 1/8$ for each $j$ and*

$$|D_j(\mu_i f^2)^{1/2}| \leq (\mu_i(D_j f)^2)^{1/2} + Q_{ji}(\mu_i(D_i f)^2)^{1/2}$$

*Proof.* If $i = j$, then $D_j(\mu_i f^2) = 0$, so assume $i \neq j$. By explicit computation, we obtain

$$\begin{aligned} D_j(\mu_i g)(\omega) &= \frac{\partial}{\partial \omega(j)} \int_{\mathbb{R}} g(\overline{\omega}, \omega_{\{i\}^c}) Z^{-1}_{\overline{\omega},\omega_{\{i\}^c}} e^{-V(\omega(i)) + \sum_{k\neq i} J(i-k)\overline{\omega}(i)\omega(k)} \rho(d\overline{\omega}(i)) \\ &= \mu_i(D_j g)(\omega) - J(i,j)\mu_i(\overline{\omega}(\omega)\mu_i(g)(\omega) J(i-j)\mu_i(\overline{\omega}\cdot g)(\omega) \end{aligned}$$

so we have

$$|D_i(\mu_i f^2)| = \frac{1}{2}\cdot\frac{|D_j(\mu_i f^2)|}{|\mu_i f^2|^{1/2}} \leq \frac{|\mu_i(f D_j f)| + |J(i-j)|\cdot|\mu_i(\overline{\omega} f^2) - \mu_i(\overline{\omega}\mu_i(f^2)|}{|\mu_i f^2|^{1/2}}$$

Terms on the numerator can be bounded using Cauchy-Schwarz inequality by

$$\begin{aligned} \mu_i(f D_j f) &\leq \Big(\mu_i(f^2)\mu_i((D_j f)^2)\Big)^{1/2} \\ \mu_i(\overline{\omega} f^2) - \mu_i(\overline{\omega})\mu_i(f^2) &= \frac{1}{2}\Big[\mu_i'\mu_i\Big((\overline{\omega}-\overline{\omega}')(f^2(\omega) - f^2(\omega'))\Big)\Big]^{1/2} \\ &\leq \frac{1}{2}\Big[\mu_i'\mu_i\Big((\overline{\omega}-\overline{\omega}')^2(f(\omega)-f(\omega'))^2\Big)\mu_i'\mu_i((f(\omega)+f(\omega')^2)\Big]^{1/2} \\ &= \Big[\mu_i'\mu_i\Big((\overline{\omega}-\overline{\omega}')^2(f(\omega)-f(\omega'))^2\Big)\mu_i(f^2)\Big]^{1/2} \end{aligned}$$

where $\mu_i'$ means that we are applying $\mu_i$ over the variable $(\overline{\omega}', \omega')$. Therefore

$$|D_i(\mu_i f^2)| \leq (\mu_i(D_j f)^2)^{1/2} + |J(i-j)|\Big[\mu_i'\mu_i\Big((\overline{\omega}-\overline{\omega}')^2(f(\omega)-f(\omega'))^2\Big)\Big]^{1/2}$$

We now seek to bound the last term. However, it is not easy to make a bound on this function in this form. Rather, we make change of coordinates by writing $\overline{\omega}(i) - \overline{\omega}'(i) = s$ and $\overline{\omega}(i) + \overline{\omega}'(i) = t$ so that the last term is

$$\begin{aligned} &\mu_i\Big((\overline{\omega}-\overline{\omega}')^2(f(\omega)-f(\omega'))^2\Big)(\omega) \\ &= \frac{2}{(Z_{i;\omega_{\{i\}^c}})^2}\int_{\mathbb{R}\times\mathbb{R}} s^2\Big(f\Big(\frac{s+t}{2}, \omega_{\{i\}^c}\Big) - f\Big(\frac{t-s}{2}, \omega_{\{i\}^c}\Big)\Big)^2 e^{-E_{i,\omega_{\{i\}^c}}(\frac{s+t}{2}) - E_{i,\omega_{\{i\}^c}}(\frac{t-s}{2})}\rho(dt)\rho(ds). \end{aligned}$$

and

$$-E_{i,\omega_{\{i\}^c}}\Big(\frac{s+t}{2}\Big) - E_{i,\omega_{\{i\}^c}}\Big(\frac{t-s}{2}\Big) = -V\Big(\frac{s+t}{2}\Big) - V\Big(\frac{t-s}{2}\Big) + \sum_{j\neq i} J(i-j)\omega(j)t.$$

If we integrate over the variable $s$ first, then we can temporarily forget about the coupling terms, which makes the computation significantly easier. Noting that $\deg(V(\frac{s+t}{2}) + V(\frac{t-s}{2})) = \deg(V) = 2m$, we just need to make estimate for integrals of form $\int_{\mathbb{R}} s^2 g^2(s)\exp(-\tilde{V}(s))\rho(ds)$ in terms of $g$.

**Claim.** if $\int g(s)e^{-\tilde{V}(s)}ds = 0$, then there is a constant $\tilde{C} > 0$ such that

$$\int_{\mathbb{R}} s^2 g^2(s)\exp(-\tilde{V}(s))\rho(ds) \leq \tilde{C}\int_{\mathbb{R}} |g'(s)|^2 e^{-\tilde{V}(s)}\rho(ds)$$

for $g(s)$ with sufficient decay rate as $s \to \infty$.

: Integrating by parts, we have

$$\int g'(s)\cdot g'(s)e^{-\tilde V(s)}\rho(ds) = \int_{\mathbb{R}} g(s)(-g''(s)+\tilde V'(s)g'(s))e^{-\tilde V(s)}\rho(ds)$$

$$-\int (ge^{-\tilde V/2})'(ge^{-\tilde V/2})'\rho(ds) = \int (ge^{-\tilde V/2})(ge^{-\tilde V/2})''\rho(ds)$$

$$= \int g\Big(g'' - \tilde V' g' - \frac{g}{2}\tilde V'' + \frac{(\tilde V')^2}{4}g\Big)e^{-\tilde V}\rho(ds) \le 0$$

so putting these together, we have

$$\int g'(s)\cdot g'(s)e^{-\tilde V(s)}\rho(ds) \ge \int_{\mathbb{R}} g^2\Big(-\frac{1}{2}\tilde V'' + \frac{1}{4}(\tilde V')^2\Big)\rho(ds)$$

If $\deg(V)=2$, then $\tilde V(s) = as^2 + at^2 - b$ for some $a>0$, $b\in\mathbb{R}$, so $\frac{1}{4}(\tilde V')^2 - \frac{1}{2}\tilde V'' = a^2s^2 - a$. If $\deg(V)\ge 4$, then $\deg(\tilde V^2) > \deg(\tilde V'')$ and so we have bound $\frac{1}{4}(\tilde V')^2 - \frac{1}{2}\tilde V'' \ge a(s\vee 1)^{2m-2} - b \ge a(s\vee 1)^2 - b$. Therefore

$$\int g'(s)\cdot g'(s)e^{-\tilde V(s)}\rho(ds) \ge a\int_{|s|\ge 1} s^2g^2(s)e^{-\tilde V(s)}\rho(ds) - b\int g^2e^{-\tilde V}d\rho$$

and it follows that

$$\int g'(s)\cdot g'(s)e^{-\tilde V(s)}\rho(ds) + (b+1)\int g^2e^{-\tilde V}d\rho \ge a\int s^2g^2(s)e^{-\tilde V(s)}\rho(ds)$$

Now by previous discussions, it is apparent that a new diffusion process with generator $-\frac{1}{2}\triangle + \frac{1}{2}\nabla\tilde V\cdot\nabla$ satisfies a log-Sobolev inequality, hence a Poincaré inequality, which tells us that $\int g^2e^{-\tilde V}d\rho$ has a bound linear in $\int |g'|^2e^{-\tilde V}d\rho$.

Having the claim, just observe that $f(\frac{s+t}{2},\omega_{\{i\}^c}) - f(\frac{t-s}{2},\omega_{\{i\}^c})$ is an odd function and hence satisfies the condition of the claim, so application of the claim with the previous inequality proves the existence of $(Q_{ji})_{i\neq j}$ upon choosing $Q_{ji} = 2\tilde C|J(i-j)|$. Now, if we choose $\sum_{i\neq j}|J(i-j)|$ small so that $2\tilde C\sum_{i\neq j}|J(i-j)|\le 1/2$, then we have the desired bounds for $(Q_{ij})_{i\neq j}$.

*(End of proof)* □

We can now prove the log-Sobolev inequality for the system of weak interactions using **Proposition 5.6** and **Proposition 5.7**.

**Theorem 5.8** *For $V$ fixed, there is $\sigma>0$ such that whenever $\sum_{j\in\Lambda}|J(j)|\le\sigma$, any measure $\mu_{A,\omega'}$ satisfies a tight log-Sobolev inequality independent of $A\in F(\Lambda)$ and $\omega'\in\Omega_{A^c}$, i.e. there is $C$ independent of $A$ and $\omega'$ such that*

$$\mu_A(f^2\log(f^2)) - \mu_A(f^2)\log(\mu_A(f^2)) \le C\sum_{i\in A}\mu_A(D_if)^2, \quad \forall f\in\mathcal{D}_0.$$

*Proof.* Without loss of generality, let $f\ge 0$. By relabelling the points in $\Lambda$ if necessary, we may write $A=\{1,2,\cdots,|A|\}$. We will fix $z\in\Omega_{A^c}$, and omit the label $z$ whenever it does not ambiguity. Using **Proposition 5.6** and **Proposition 5.7**, we obtain $C_p>0$, and set $\sum_j|J(j)|$ small enough so we get $(Q_{ij})_{i,j}$ with the given properties.

Recalling the property that $\mu_A\mu_B = \mu_B$ whenever $B \subset A$, and the pointwise logarithmic Sobolev inequality, we find that

$$\begin{aligned}\mu_A(f^2\log f^2) &= \mu_A(\mu_1(f^2\log f^2)) \\ &\leq \mu_A\Big(C_p(D_1 f)^2 + \mu_1(f^2)\log(\mu_1 f^2)\Big) \\ &= C_p\mu_A(D_1 f)^2 + \mu_A(f^2\log(\mu_1 f^2))\end{aligned}$$

and letting $\varphi_1 = (\mu_1 f^2)^{1/2}$, we can apply this computation again to obtain

$$\mu_A(f^2\log f^2) \leq C_p\mu_A((D_1 f)^2 + (D_2\varphi_1)^2) + \mu_A(\varphi_1^2\log(\mu_2\varphi_1^2))$$

If we define $\varphi_l = (\mu_l\mu_{l-1}\cdots\mu_1 f^2)^{1/2} =: (H_l f^2)^{1/2}$ for $l = 1, \cdots, |A|$ and $\varphi_0 = f$, then induction will give

$$\mu_A(f^2\log f^2) \leq C_p\mu_A\Big(\sum_{l=1}^{N}(D_l\varphi_{l-1})^2\Big) + \mu_A(\varphi_N^2\log(\varphi_N^2))$$

The inequality can be applied on the function $\varphi_N$ again, then another induction process on applying $H_{|A|} =: H$ on $f$ will give log-Sobolev inequality on $\mu_A$, depending on the following two observations.

**Claim 1.** Let $\psi_k = (H^k f^2)^{1/2}$ where $H = \mu_{|A|}\mu_{|A|-1}\cdots\mu_1$. Then $\mu_A(\psi_k^2\log(\psi_k^2))(\omega) \to \mu_A(f^2\log(f^2))(\omega)$ uniformly as $k \to \infty$.

: To see this, let $L : x \mapsto x\log x$ and $\|L\|_{Lip} = K$. Noting that $\mu_A(\psi_k^2) = \mu_A f^2$ for any $k > 0$, we have

$$\mu_A(L(\psi_k^2) - L(f^2))(\omega) \leq K\mu_A\mu_A'(|\psi_k^2(x) - (\psi_k')^2(x')|)(\omega)$$

where $\mu_A'$ takes $\mu_A$ over the variable $x'$. If the Lipschitz semi-norm of $\psi_k$ converges to 0 as $k \to 0$, then this would prove the convergence of the functions $\mu_A(L(\psi_k^2))$. But having each $\psi_k$ differentiable component-wise, convergence of the semi-norm $[\psi_k] = \sup_{i=1,\cdots,|A|}\|D_i\psi_k(\omega)\|_{L^\infty(\Omega)}$ to 0 will be sufficient.

Recalling the definition of $(Q_{ij})_{i,j}$, we have

$$|D_j(\mu_i g^2)^{1/2}|(\omega) \leq |D_j g|(\omega) + Q_{ji}|D_i g|(\omega)$$

for any $g \in \mathcal{D}_0$, $g \geq 0$. For brevity, we will denote this as $D_j\mu_i \leq D_j + Q_{ji}D_i$. Then $D_N\mu_{N-1} \leq D_N + Q_{N,N-1}D_{N-1}$. Using induction, we make a bound on $D_N\mu_{N-1}\cdots\mu_{N-k}$ : define $\Pi_{N,N-k} = \{N = i_0 \geq i_1 \geq \cdots \geq i_M = N-k\}$ and

$$B_{N,N-k} = \sum_{(i_0,\cdots,i_M)\in\Pi_{N,N-k}} Q_{i_0,i_1}\cdots Q_{i_{M-1},i_M}$$

and assume $D_{N'}\mu_{N'-1}\cdots\mu_{N'-l} \leq \sum_{i=0}^{l} B_{N',N'-i}D_{N-i}$ for each $l = 1, \cdots, k-1$ and any $N$. Then

$$D_N\mu_{N-1}\cdots\mu_{N-k} \leq D_N\mu_{N-2}\cdots\mu_{N-l} + Q_{N,N-1}D_{N-1}\mu_{N-2}\cdots\mu_{N-l} = \sum_{i=0}^{l} B_{N,N-i}D_{N-i}$$

noting that we can apply the induction hypothesis on $D_N\mu_{N-2}\cdots\mu_{N-l}$ if just relabelling the points. With this estimate, we can also write

$$\begin{aligned}D_j\mu_N\cdots\mu_1 &\leq D_j\mu_N\cdots\mu_1 + Q_{j,N}\sum_{k=0}^{N-1} B_{N,N-k}D_{N-k} \\ &\leq \sum_{m=0}^{j-1}\sum_{k=0}^{N-m-1} B_{N-m,N-m-k}Q_{j,N-m}D_{N-k-m}\end{aligned}$$

Setting $N = |A|$, we then have

$$\|D_j(Hf^2)^{1/2}\|_\infty \leq \sum_{m=0}^{j-1} \sum_{k=0}^{N-m-1} |B_{N-m,N-m-k} Q_{j,N-m}|[f]$$

but $\sum_{m=0}^{j-1} \sum_{k=0}^{N-m-1} |B_{N-m,N-m-k} Q_{j,N-m}|$ does not have multiple occurrence of each combination of $Q_{j,i}$, and therefore it is strictly smaller than $(1/4) + (1/4)^2 + \cdots + (1/4)^{|A|} \leq \frac{3}{4}$. So $[Hf] \leq \frac{3}{4}[f]$. (We could certainly make much better bound.)

**Claim 2.** We have

$$\sum_{k=0}^{M} \sum_{l=1}^{N} \mu_A\Big((D_l(H_{l-1}H^k f^2)^{1/2})^2\Big) \leq \frac{9}{4} \sum_{i \in A} \mu_A((D_i f)^2) \quad \forall M \in \mathbb{N}$$

: The similar combinatorics applies, just noting that $\sum_{k=1}^{N} D_l(H_k f^2)^{1/2}$ written in terms of $B_{i,j}$ and $D_k$ only has limited number multiple occurrence of each combination of $Q_{j,i}$. So

$$\sum_{l=0}^{M} \sum_{k=1}^{N} \mu_A\Big((D_l(H^l H_k f^2)^{1/2})^2\Big) \leq \frac{3}{4} \cdot \Big(\sum_{i=0}^{\infty} \Big(\frac{1}{4}\Big)^i\Big)^2 \sum_{j \in A} \mu_A((D_j f)^2) \leq 9 \sum_{i \in A} \mu_A((D_i f)^2)$$

From the two claims, we conclude that

$$\mu_A(f^2 \log f^2) \leq 9C_p \sum_{i \in A} \mu_A(D_i f)^2 + \mu_A(f^2)\log(\mu_A f^2)$$

so $\mu_A$ satisfies $LSI(9C_p/2)$.

*(End of proof)* □

**Corollary 5.9** *Under weak interaction assumption as in the statement of the last theorem, a Gibbs measure $\mu$ satisfies a log-Sobolev inequality. In particular, if the probability space is standard and the initial distribution of Glauber-Langevin process has $L^\infty$-bounded distribution with respect to the Gibbs measure, then the process converges to Gibbs measure in distribution.*

**Remark :** Setting the background measure $\rho$ to be a Lebesgue measure seem unnatural for most people with correct physical intuition. However, once we have seen how we have to tackle the problem, minor modifications to the problem will not make much disruption.

- If we think that each spin on each lattice point $i \in \Lambda$ is composed of non-interacting multiple spins, then the distribution of the spin of each lattice point will make Gaussian distribution, we we can take into account by adding $\frac{x^2}{2\sigma^2}$ on the potential term $V$.
- We can also prove exactly the same result when we $\rho$ is replaced by a Borel measure that is supported on a finite subset of $\mathbb{R}$, *e.g.* if $\rho(ds) = \frac{1}{2}(\delta_{-1/2}(ds) + \delta_{1/2}(ds))$, then setting the background measures as $\rho$ gives the Ising model with 1/2 spin particles. There are two main points to be verified : first is to check the pointwise log-Sobolev inequality and the second is to check the existence of the constant $(Q_{ij})$ with the properties as in **Proposition 5.7**, which are even easier to verify.

The methodology suggested here becomes essentially unhelpful if we try to transfer the problem to settings where spin values do not take values in$\mathbb{R}$. However, different authors worked on proving the log-Sobolev inequalities in many different settings. See {SZ}, for example, for the proof of log-Sobolev inequality when spin takes values in a compact smooth Riemannian manifold.

# Chapter 6 BOUNDED EIGENSTATES OF SCHRÖDINGER EQUATION ON RIEMANNIAN MANIFOLDS

Just like classical Sobolev inequality is used to investigate the properties of partial differential equations defined on a Euclidean space, the general notion of Sobolev inequalities can also be used to study certain behaviours of partial differential equations on Riemannian manifolds. Among those, we see in this chapter an application to the Schrödinger equation. This application will not be entirely probabilistic in virtue, but it will illuminate the crucial role of ultracontractivity in applications of Sobolev inequalities. But before starting to prove anything about it, it would be worth sparing a section just to get introduction about diffusion processes on a Riemannian manifold, and minimize the notational discordance in the main discussion of **Section 6.2**.

## Section 6.1 Dirichlet form on Riemannian manifolds

Let $M$ be a $d$-dimensional connected Riemannian manifold with Riemannian metric $(g_{ij}(x))_{x\in M}$ and tangent space $TM$. To define a diffusion process in the Riemannian metric analogous to that of $\mathbb{R}^d$, we need to fix notions of derivatives.

First, fix a dual basis of $T_x^*M$ (the cotangent bundle at point $x \in M$) as $\partial_1, \cdots, \partial_d$. A vector field $\xi : C^\infty(M) \to C^\infty(M)$ is written as

$$\xi = \sum_{i=1}^{n} \xi_i(x)\frac{\partial}{\partial x_i} = \sum_{i=1}^{n} \xi_i(x)\partial_i$$

in local coordinates. Then we also have, by definition of a Riemannian metric,

$$g(\xi,\eta) = \sum_{ij} g_{ij}(x)\xi_i(x)\eta_j(x)$$

for two vector fields $\xi$ and $\eta$. We define the gradient of $u \in C^\infty(M)$ to be the vector field

$$\nabla u = \sum_{i=1}^{d} g^{ij}(\partial_j u)\partial_i$$

and the divergence of a vector field $\xi$ by

$$\nabla \cdot \xi = g^{-1/2}\partial_j(g^{1/2}\xi_j)$$

where $g(x) = \det(g_{ij}(x))$. With the measure on $M$ defined in local level with

$$d\mu_M(x) = g^{1/2}(x)dx_1 \cdots dx_n$$

the definition of $\nabla\cdot$ and $\nabla$ gives $\int u\nabla \cdot \xi d\mu_M = -\int \nabla u \cdot \xi d\mu_M$, so we can safely do integration by parts. Finally, the Laplacian is defined as usual,

$$\triangle u = \nabla \cdot (\nabla u) = g^{-1/2}\partial_i(g^{1/2}g^{ij}\partial_j u)$$

with Einstein's summation convention.

If we in addition assume that $M$ is a Polish space countable at infinity and has finite measure, then the diffusion process on the Riemannian manifold can be defined by the operator $\Gamma(u,v) = g(\nabla u, \nabla v) =: \nabla u \cdot \nabla v$ for $f, g \in C^\infty(M)$. Then on a local level, we have

$$\Gamma(u,v) = g^{ij}\partial_i u\partial_j v$$

so $\Gamma$ indeed satisfies the *diffusion property*, using Leibniz rule for the map $\partial_i$. Hence the form

$$\mathcal{E}(u,v) = \int_M \Gamma(u,v) d\mu_M$$

defined on the completion of $C_c^\infty(M)$ in the norm $\|\cdot\|_{\mathcal{E}}$ is a Dirichlet form. By the additional assumption that $M$ is a Polish space, the Dirichlet form induces a continuous Markov process that is an analogue of a Brownian motion.

In the case $\mu(M) = \infty$, then we need little more care in constructing the Brownian motion. However, with assumption that $M$ is countable at infinity and each compact set has finite measure, the extension should not be too different from the compact case with only minor technical complications in concern of stopping times of the process.

**Example 1 :** A generalized Ornstein-Uhlenbeck process can be defined using the same setting, with just the measure on $M$ modified by $d\mu = d\mu = \frac{1}{Z} e^{-F(x)} d\mu$ with $\mu(\exp(-F)) < \infty$ and $Z$ the normalization factor. Then the same carré du champ operator has diffusion property and the new Dirichlet form $\mathcal{E}(u,v) = \int_M \Gamma(u,v) d\mu$ gives a continuous Markov process regardless of whether $M$ is compact or not.

**Example 2 :** However, this is not the only way to define a Dirichlet form on a Riemannian manifold. One particularly interesting example that does not give a Markov process, but just a sub-Markov process arises by letting

$$\mathcal{G}f(x) = \triangle f(x) - V(x)$$

for some strictly positive function $V$ and the measure on $M$ such that $V^{-1} d\mu_M$ is integrable. Then with the new measure defined by $d\mu = V^{-1} d\mu_M$, the forms

$$\Gamma(u,v) = V^{-1} \nabla u \cdot \nabla v + u^2, \quad \mathcal{E}(u,v) = \mu(\Gamma(u,v))$$

is non-negative. If we also assume in addition that the form $(u,v) \mapsto (\nabla u, \nabla v)_{L^2(\mu)}$ with domain $H_0^1(M, \mu_M)$ defines a Markov triple with measure $\mu_M$, then it follows that $(M, \mu)$ with $\mathcal{E}$ also satisfies the assumption for being a Markov triple. Also since $V$ is assumed to be positive, $\mathcal{G}$ is a negative operator, so $P_t = \exp(\mathcal{G}t)$ should be $L^\infty$-contraction, and positivity preserving. Hence $(M, \mu, \Gamma)$ is a standard Markov triple.

Such a generator $\mathcal{G}$ describes the Schrödinger operator with a deep potential well. This can be compared with the situation described in the next section, where the number of bounded eigenstates can be estimated just under the assumption of having a shallow potential well.

## Section 6.2 Bounded eigenstates for Schrödinger equation

Let $M$ be a non-compact Riemannian manifold with finite measure that is countable at infinity and is a Polish space in its metric. By normalizing, just assume that $\mu \equiv \mu_M$ is a probability measures on $M$. We will assume in this section that a Sobolev inequality $SI^p(C)$ holds in the triple $(M, \mu, \Gamma)$ where $\Gamma(f) = \nabla f \cdot \nabla f = g(\nabla f, \nabla f)$ and $p = 2\nu/(\nu - 2)$ for some $\nu > 2$.

A time-independent Schrödinger operator on $M$ is defined as

$$-\mathcal{G}f(x) = -\triangle f(x) - V(x)f(x)$$

form some function $V(x)$. Because of its pre-dominant occurrence in physics, predicting the behaviour of the Scrödinger operator is often of large interest. However, the full spectrum of $\mathcal{G}$ is computable only a limited number of cases, so there have been numerous attempts in developing an approximation scheme for computing the spectrum of the Schrd̈inger operator, both in theoretical and empirical levels. In this section, we will be showing that under the assumption $V^+ = V \vee 0 \in L^\infty(\mu) \cap L^{\nu/2}(\mu)$, we can make an estimate the number of negative eigenvalues by considering the quantity

$$n_{\mathcal{G}}(\lambda) = \text{number of eigenvalues of } -\mathcal{G} \text{ below } \lambda \text{ counted with multiplicity}$$

and setting $\lambda = 0$, where each eigenstate is considered to be in the Sobolev space $H_0^1(M)$, the completion of $C_c^\infty(M)$ in the norm $\|f\|_2 + \|\nabla f\|_2$. But keep mind that it is not even clear at this point whether the spectrum of $\mathcal{G}$ is discrete or not - this will be verified later, so just assume that we can count eigenvalues at the moment.

In an interpretation of classical dynamics, the positive part of $V$ makes a potential well that an eigenstate of $\mathcal{G}$ dwells in, so assuming $V$ is bounded above gives a good bound with $n_{\mathcal{G}}(\lambda)$. Also, if $V$ is not bounded below, then we can cut off the potential by $V_c = V \vee c$, then the eigenvalues of $-\triangle - V_c$ would also be cut down as $c$ increases. This can be made rigorous in the following sense.

**Proposition 6.1** *Define the **Rayleigh ratio** of operator $\mathcal{G}$ by*

$$R(f) = (f, \mathcal{G}f)_{L^2(\mu)}/\|f\|_{L^2(\mu)} = -\frac{\int |\nabla f|^2 - V(x)|f(x)|^2 d\mu(x)}{\int |f(x)|^2 d\mu(x)}.$$

*Then*

$$n_{\mathcal{G}}(\lambda) = \sup\{\dim(F) : F \subset C_c^\infty(M) \textit{ and if } f \in F, \textit{ then } R(f) > -\lambda\}$$

*Proof.* Surely, $C_c^\infty \subset H_0^1(M)$, so the inequality

$$n_{\mathcal{G}}(\lambda) \geq \sup\{\dim(F) : F \subset C_c^\infty(M) \text{ and if } f \in F, \text{ then } R(f) > -\lambda\}$$

is direct. To see the converse inequality, let

$$\overline{n}_{\mathcal{G}}(\mu) = \text{number of eigenvalues of } -\mathcal{G} \text{ below or equal to } \mu \text{ counted with multiplicity}$$

then $n_{\mathcal{G}}(\lambda) = \sup_{\mu<\lambda} \overline{n}_{\mathcal{G}}(\mu)$. Also, if approximating $H_0^1(M)$ functions with $C_0^\infty(M)$ functions, then we get

$$\overline{n}_{\mathcal{G}}(\mu) \leq \sup\{\dim(F) : F \subset C_c^\infty(M) \text{ and if } f \in F, \text{ then } R(f) > -\lambda\}$$

for any $\mu < \lambda$, so we get the desired equality.

*(End of proof)* □

If we let $R_c(f)$ be the Rayleigh ratio of $-\triangle - V \vee c$, then we always have $R_c(f) \geq R(f)$, and therefore $n_{\mathcal{G}}(\lambda) \leq n_{-\triangle - V\vee c}(\lambda)$. So it will be enough to make upper bound on $n_{-\triangle-V\vee c}$ in order to get an upper bound for $n_{\mathcal{G}}$. Hence, we will just be assuming that $V$ is strictly positive from now on.

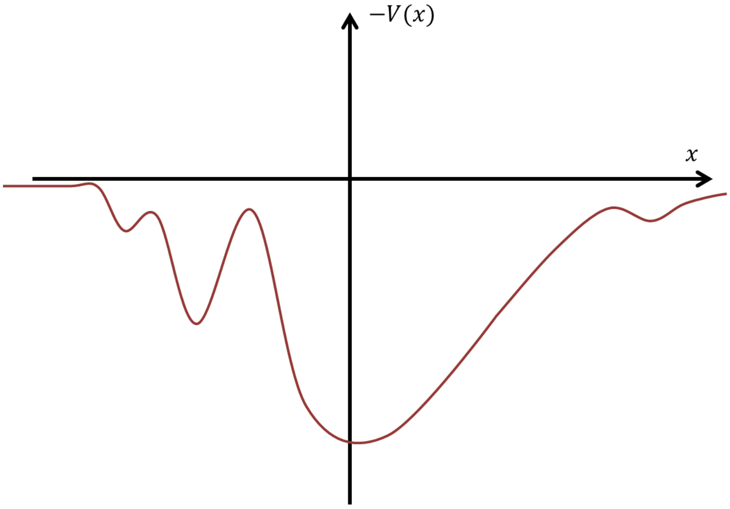


Figure 2: A strictly negative potential well.

## § Reformulation of the problem

In order to exploit the Sobolev inequality, we would like to make use of the ultracontractivity of the semigroup induced by $\Gamma$ in measure $\mu$. However, there are two major problems in doing this. First, making estimate by controlling the convergence of the integrals in this setting is not so easy (that will be made clear soon). Secondly and more fundamentally, $\mathcal{G}$ is not non-positive, so the form $\mathcal{E}(f) = -(f, \mathcal{G}f)_{L^2(\mu)}$ is not non-negative. Instead, we use a clever trick by making assumption $V \in L^{\nu/2}(\mu)$ and use the measure $d\nu = V d\mu$ to control integrability using Hölder inequalities.

In the measure system $(M, \nu)$, the form $\mathcal{E} : f \mapsto \|\nabla f\|_{L^2(\mu)}$ is written as

$$\int_M g(\nabla f, \nabla f) d\mu = \int_M f \triangle f d\mu = \int_M f(\triangle f) V^{-1} d\nu$$

but we also have

$$\int_M |f|^2 d\nu = \int_M |f|^2 V d\mu \leq \|f\|^2_{2\nu/(\nu-2)} \|V\|_{\nu/2} \leq C \|\nabla f\|_2^2 \|V\|_{\nu/2}$$

where the first inequality follows from Hölder inequality and the second inequality follows from $SI^p(C)$. Therefore an element $f \in H_0^1(M, \mu)$ has bound

$$\|f\|_{\mathcal{E}} = \|f\|_{L^2(\nu)} + \mathcal{E}(f) \leq \|\nabla f\|_{L^2(\mu)} (1 + C\|V\|_{\nu/2})$$

so $H_0^1(M, \mu)$ is automatically complete in the norm $\|\cdot\|_{\mathcal{E}}$. So we can take $H_0^1(M, \mu)$ to be the domain of $\mathcal{E}$. Also, we see from usual diffusion process on $(M, \mu)$ that $H_0^1(M, \mu)$ satisfies conditions for $(M, \nu, \Gamma)$ being a Markov triple and $\Gamma(f) = V^{-1/2} |\nabla f|^2$ satisfies a diffusion property, so $(M, \nu, \Gamma)$ is a standard Markov triple. Hence if we let $\mathcal{L}$ and $Q_t$ to be the generator and semigroup associated to the triple respectively, then $Q_t$ would be a sub-Markovian transition semigroup. Finally, the Sobolev inequality for $(M, \mu)$ gives

$$\|f\|_{L^p(\nu)} \leq C \|V\|_\infty^{2/p} \mathcal{E}(f)$$

whenever $f \in \mathcal{D}_{\mathcal{E}}$, so $SI^p(C\|V\|_\infty^{2/p})$ holds for $(M, \nu, \Gamma)$. Thus ultracontractivity indicates that $Q_t : L^2(\nu) \to L^2(\nu)$ is a compact operator for any $t > 0$. In particular, $Q_t$ has a discrete non-positive spectrum and we may write

$$Q_t = \sum_{j\in\mathbb{N}} e^{-\lambda_j t} P_{E_{\lambda_j}}$$

where $0 \le \lambda_1 < \lambda_2 < \cdots$ are negatives of eigenvalues of $\mathcal{L}$, $E_{\lambda_j}$ the associated eigenspace that are finite-dimensional if $\lambda_j \neq 0$, and $P_{E_{\lambda_j}}$ is the orthogonal projection onto $E_{\lambda_j}$. This is useful in counting the number of eigenvalues because

$$n_{\mathcal{L}}(\lambda^{-1}) = \sum_{j:\lambda_j\lambda<1} 1 \le \sum_j e^{2(1-\lambda_j\lambda)t} = e^{2t}\sum_j e^{-2\lambda_j\lambda t}.$$

Also, the following lemma gives a connection between $n_{\mathcal{L}}(\lambda^{-1})$ and $n_{\mathcal{G}}(0)$.

**Lemma 6.2** *We have $n_{\mathcal{G}}(0) = n_{\mathcal{L}}(1)$.*

*Proof.* We again make use of the Rayleigh ratio and the equality

$$n_{\mathcal{G}}(\lambda) = \sup\{\dim(F) : F \subset C_c^\infty(M) \text{ and if } f \in F, \text{ then } R(f) > -\lambda\}$$

If we also let $\tilde{R}(f)$ the Rayleigh ratio for the generator $\mathcal{L}$, *i.e.* $R_L(f) = \mathcal{E}(f)/\|f\|_{L^2(\nu)}$, then having $R(f) = (\mathcal{E}(f) - \|f\|_{L^2(\nu)})/\|f\|_{L^2(\mu)} > 0$ is equivalent to having $\tilde{R}(f) > -1$, so

$$\begin{aligned} n_{\mathcal{L}}(1) &= \sup\{\dim(F) : F \subset C_c^\infty(M) \text{ and if } f \in F, \text{ then } \tilde{R}(f) > -1\} \\ &= \sup\{\dim(F) : F \subset C_c^\infty(M) \text{ and if } f \in F, \text{ then } R(f) > 0\} = n_{\mathcal{G}}(0) \end{aligned}$$

invoking the definition of $R(f)$. (Note that this method does not work other than the case $n_{\mathcal{G}}(0)$ and $n_{\mathcal{L}}(1)$ - this is a very special case.)

*(End of proof)* □

So we are just left with estimating the quantity $\sum_i e^{-2\lambda_i\lambda t}$.

### § Estimate on $\sum_i e^{-2\lambda_i\lambda t}$

To use more analytical (and more brute-force) methods, we write the transition semigroup in terms of heat kernel, so

$$Q_t(x, dy) = q_t(x, y)d\nu(y)$$

By ultracontractivity, a bounded kernel $q_t \in L^\infty(\nu) \otimes L^\infty(\nu)$ exists. Then we can write

$$\int q_t(x, y)^2 d\nu(x)d\nu(y) = \sum_i e^{-2\lambda_i t}.$$

(Using the notation $q_t(x, y) = \sum_j e^{-\lambda_j t}\boldsymbol{e}_j(x)\boldsymbol{e}_j(y)$ where $\boldsymbol{e}_j$'s are the eigenvectors, this identity is immediate.)

**Lemma 6.3** *For each $t > 0$, $q_t(\cdot, x)$ is in $\mathcal{D}_{\mathcal{E}} = H_0^1(M, \mu)$ for a.e. $x \in M$.*

*Proof.* By self-duality of the Hilbert space $(\mathcal{D}_{\mathcal{E}}, \|\cdot\|_{\mathcal{E}})$,

$$\begin{aligned}\|\nabla q_t(\cdot,x)\|_{\mathcal{E}}^2 &= \sup\Big\{\int q_t(y,x)\xi(y)V(y)d\mu(y) + \int \nabla_y q_t(y,x)\cdot\nabla_y\xi(y)d\mu(y) : \|\xi\|_{\mathcal{E}} \le 1, \xi \in C_c^\infty(M)\Big\} \\ &\le \|V\|_\infty \|Q_t\|_{1,\infty} + \Big\{\int \nabla_y q_t(y,x)\cdot\nabla_y\xi(y)d\mu(y) : \|\xi\|_{\mathcal{E}} \le 1, \xi \in C_c^\infty(M)\Big\}.\end{aligned}$$

But since $\int \nabla f\cdot\nabla g d\mu = -\int V^{-1}(\triangle f)g d\nu$, we have $\mathcal{L} = V^{-1}\triangle$ and

$$\begin{aligned}\int \nabla_y q_t(y,x)\cdot\nabla_y\xi(y)d\mu(y) &= -\int \triangle_y q_t(y,x)\xi(y)d\mu(y) \\ &= -\int \mathcal{L}q_t(y,x)\xi(y)V(y)d\mu(y) \\ &\le \|\mathcal{L}q_t(\cdot,x)\|_{L^2(\nu)} < \infty\end{aligned}$$

*(End of proof)* □

**Lemma 6.4** *Let $Q(t) = \int q_t(x,y)^2 d\nu(x)d\nu(y)$. Then we have estimate $Q(t) \le (\|V\|_{\nu/2}C\nu/4t)^{\nu/2}$ where $C$ was the constant for the Sobolev inequality.*

*Proof.* We take the advantage of working in the measure system $d\nu$ by using Hölder inequality :

$$\begin{aligned}Q(t) &= \int d\nu(x)\Big(\int q_t(x,y)^2 V(y)d\mu(y)\Big) \\ &= \int d\nu(x)\Big(\int q_t(x,y)^{2\nu/(\nu+2)}(q_t(x,y)V(y)^{(\nu+2)/4})^{4/(\nu+2)}d\mu(y)\Big) \\ &\le \int d\nu(x)\Big(\int q_t(x,y)^p dy\Big)^{(\nu-2)/(\nu+2)}\Big(\int q_t(x,y)V(y)^{(\nu-2)/4}V(y)\Big)^{4/(\nu+2)} \\ &= \int d\nu(x)\Big(\int q_t(y,x)^p dy\Big)^{(\nu-2)/(\nu+2)}\Big(Q_t(V^{(\nu-2)/4})(x)\Big)^{4/(\nu+2)}\end{aligned}$$

By the previous lemma, we have $\mathcal{E}(\nabla q_t(\cdot,x)) < \infty$, so Sobolev inequality implies

$$\Big(q_t(x,y)^p d\mu(y)\Big)^{1/p} \le C\int |\nabla_y q_t(y,x)|^2 d\mu(y) = -\int q_t(x,y)\partial_t q_t(x,y)V(y)dy$$

where the last equality also follows from the computation $\int \nabla_y q_t(y,x)\cdot\nabla_y\xi(y)d\mu(y) = -\mathcal{L}Q_t\xi(x) = -\frac{d}{dt}Q_t\xi(x)$. Plugging this in the estimate of $Q(t)$ above,

$$\begin{aligned}Q(t) &\le \int V(x)d\mu(x)\Big(-C\int(\partial_t q_t(x,y))q_t(x,y)V(y)d\mu(y)\Big)^{\nu/(\nu+2)}\Big(Q_t(V^{(\nu-2)/4})\Big)^{4/(\nu+2)} \\ &\le \Big(C\int\int(\partial_t q_t(x,y))(q_t(x,y))V(x)V(y)d\mu(x)d\mu(y)\Big)^{\nu/(\nu+2)}\Big(K_t(V^{(\nu-2)/4})^2 V d\mu\Big)^{2/(\nu+2)} \\ &\le \Big(-\frac{1}{2}\partial_t Q(t)\Big)\Big(\int V(x)^{(\nu-2)/2}V(x)d\mu(x)\Big)^{2/(\nu+2)} \\ &= \|V\|_{\nu/2}^{\nu/(\nu+2)}\Big(-\frac{C}{2}\frac{d}{dt}Q(t)\Big)^{\nu/(\nu+2)}\end{aligned}$$

So in particular,

$$\frac{d}{dt}(Q(t))^{-2/\nu} \ge \frac{4}{\|V\|_{\nu/2}C\nu}$$

and therefore $Q(t) \le (C\|V\|_{\nu/2}\nu/4t)^{\nu/2}$.

*(End of proof)* □

Putting altogether, we conclude with the following.

**Theorem 6.5** *Let $M$ be a non-compact, Riemannian manifold satisfying $SI^p(C)$ ($p = 2\nu/(\nu - 2)$) with Riemannian measure $\mu$ and $\mu(M) < \infty$, but countable at infinity. Let $V$ be a potential such that $V \vee 0 \in L^{\nu/2}(\mu) \cap L^\infty(\mu)$, then the number of negative eigenvalues of the operator $\mathcal{G} = -\triangle - V$, counting multiplicity, is less or equal to $(\|V \vee 0\|_{\nu/2} eC)^{\nu/2}$.*

*Proof.* Fix $c > 0$ and apply the earlier result with potential $V \vee c$. The estimate $n_{\mathcal{L}}(\lambda^{-1}) \leq \|V \vee c\|_{\nu/2} e^{2t} Q(\lambda t) \leq e^{2t}(C\nu/4\lambda t)^{\nu/2}$ holds for any $t > 0$, so we make optimisation over $t$ :

$$\frac{d}{dt} e^{4t/\nu}/t = \frac{4}{\nu} e^{4t/\nu}/t - e^{4t/\nu}/t^2 = 0 \quad \Leftrightarrow \quad t = \nu/4$$

so plugging in $t = \nu/4$, we obtain

$$n_{\mathcal{G}}(0) = n_{\mathcal{L}}(1) \leq e^{\nu/2}(C)^{\nu/2} \|V \vee c\|_{\nu/2}^{\nu/2}$$

This holds for any positive $c > 0$, so we have the desired conclusion.

*(End of proof)* □

**Remark :** Requiring $M$ to be non-compact and be of finite measure at the same time seems to be a unreasonable assumption to deal with if we care about any physical systems. However, one can show that all results holds the same even if Riemannian manifold does note have finite measure, but only is $\sigma$-finite. The cost to pay would be subtle modifications in proving ultracontractivity from a Sobolev inequality in **Chapter 4**, for which we only proved results under assumption that $\mu$ in the triple $(X, \mu, \Gamma)$ is a probability measure. In particular, we see that a one-particle quantum-mechanical system with negative, bounded, integrable potential energy always have finite number of bounded states in a Euclidean space - by the classical Sobolev inequality in $\mathbb{R}^d$.

# LIST OF FREQUENTLY USED NOTATIONS

- $\coloneqq$ - if denoting $A \coloneqq \varphi[x]$ or $\varphi[x] \eqqcolon A$, then it means $A$ is defined by formula $\varphi$, possibly with variable $x$.
- $\mathbf{1}_A$ - this is the characteristic function of set $A$.
- $(X, \mu, \Gamma)$ - this is the Markov triple of interest. In most cases, $X$ is a Polish space, $\mu$ a Borel measure on $X$ and $\Gamma$ is a operator acting on a domain $\mathcal{D}_\Gamma$, called the carré du champ operator.
- $P_t, \mathcal{G}$ - if given the Markov triple, $P_t$ is the corresponding transition semigroup and $\mathcal{G}$ is the generator of the semigroup. Conversely, if $P_t$, $\mathcal{G}$ are given *apriori* to $\Gamma$, then $\Gamma$ is derived from $\mathcal{G}$ or $P_t$.
- $\Gamma$, $\mathcal{E}, \mathcal{D}_\mathcal{E}$ - they are the carré du champ operator, the Dirichlet form, and the Dirichlet domain respectively.
- $PI(C)$, $LSI(C, \gamma)$, $SI^p(C, \delta)$, $NI^\nu(C, \delta)$ - they are the Poincaré inequality, the log-Sobolev inequality, the Sobolev inequality and the Nash inequality respectively. For details, see **Chapter 2**, **Chapter 3** and **Chapter 4**.
- $\mu(A)$, $\mu(f)$ - if $\mu$ is a Borel measure and $A$ is a Borel set, then $\mu(A)$ is the measure of $B$ in $\mu$. For a Borel-measurable function $f$, $\mu(f) = \int f d\mu$ is the integral of $f$ with respect to $\mu$.
- $dm(x)$ - is always the Lebesgue measure, given the Euclidean space $\mathbb{R}^d$.
- $\|f\|_p$, $\|f\|_{L^p}$ - are the $L^p$-norms of a measurable function $f$, *i.e.* $\|f\|_p = \|f\|_{L^p} = \mu(f^p)^{1/p}$. Also, if the measure or the $\sigma$-algebra, $\mathcal{F}$, on the system need to be specified, we denote the $L^p$-norm as $\|f\|_{L^p(\mu,\mathcal{F})}$.
- $\|T\|_{a,b}$ - is the operator norm of a linear operator $T : L^a(\mu) \to L^b(\mu)$, *i.e.* $\|T\|_{a,b} = \sup\{\|Tf\|_b : \|f\|_a \leq 1\}$.
- Push-forward of $\sigma$-algebra and measure - let $f : (X, \mathcal{F}, \mu) \to Y$ be a any function. Then we define the push-forward of $\sigma$-algebra of $\mathcal{F}$ along $f$ to be the $\sigma$-algebra generated by $\{B \subset Y : f^{-1}(B) \in \mathcal{F}\}$. If $f : (X, \mathcal{F}, \mu) \to (Y, \mathcal{G})$ is a measurable function, then the push-forward of measure $\mu$ along $f$ is defined as $f_\# \mu(B) = \mu(f^{-1}(B))$ for each $B \in \mathcal{G}$.
- $\vee$, $\wedge$ - are not logical operators! We define $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$ when $a, b \in \mathbb{R}$ and $f \vee g$, $f \wedge g$ accordingly for functions $f, g$.
- $C_c^k(X)$ - this is the set of functions of compact support and $k^{th}$ continuous derivative.

## NOTES AND REFERENCES

Dominant proportion of notations and proofs throughout this essay are dues to {BGL} and {H}. In particular, notations for Dirichlet form, carré du champ operator, transition semigroups and generators are all commonly used in both references.

The construction of standard Markov triple in **Chapter 1** entirely lies on {FYT} and {Y}. For the missing proofs in constructing and establishing key properties of Dirichlet form, see {FYT}, and for explanation on relationship between semigroups and generators, see {Y}.

The proofs and definitions in **Chapter 2** also largely refer to {BGL} and {H} although the point on relation between contraction property and spectral gap is borrowed from {DS}.

In **Chapter 3**, the proof of exponential decay of entropy is dues to {BGL} and the proof of hypercontractivity theorem is based on the original paper of Gross, {G}.

In **Chapter 4**, the proof of equivalence between Nash inequality and ultracontractivity is from {CKS}, and the equivalence between Nash inequality and Sobolev inequality is from {BGL}. Then **Theorem 4.6** is proved using the method given in {D}.

In **Chapter 5**, the whole construction of Gibbs measure and convergence of Glauber-Langevin process to Gibbs measure are provided from {R1}. However, the technical points for the convergence are originally due {SZ}, {Z} and several different papers in the series of program for establishing log-Sobolev inequality on spin lattice systems. Some notations are also following the original texts.

In **Chapter 6**, the construction of a Dirichlet form on Riemannian manifolds follows that of {D}. The proofs in the **Section 6.2** are due to {LY} and {S}.

If not cited in this list of specific references, most of the results and construction can be found similarly on {BGL}, although the details between the lines differ.

### List of References


{AMTU} Anton Arnold , Peter Markowich, Giuseppe Toscani, Andreas Unterreiter, *“On Convex Sobolev Ineualities and the Rate of Convergence to Equilibrium for Fokker-Planck Type Equations"*, Commun. in PDEs, 26(1&2), 43-100 (2001)

{BGL} Dominique Bakry, Ivan Gentil, Michel Ledoux, *“Analysis and Geometry of Markov Diffusion Operators"*, Springer (2014)

{CKS} Eric Carlen, Seiichiro Kusuoka, Daniel W. Strook, *“Upper Bounds for Symmetric Markov Transition Functions"*, Annales de l'I. H. P., section B, tome 23, nº S2, p.245-287 (1987)

{D} E. Brian Davies, *“Heat kernels and spectral theory"*, CUP, (1989)

{DS} Jean-Dominique Deuschel, Daniel W. Strook, *“Large Deviations"*, Academic Press (1989)

{FYT} Masatoshi Fukushima, Yoichi Oshima, Masayoshi Takeda, *“Dirichlet Forms and Symmetric Markov Processes"*, 6th ed., De Gruyter (2001)

{G} Leonard Gross, *"Logarithmic Sobolev Inequalities"*, Amer. J. Math. 97, p.1061-1083 (1975)

{H} Ramon van Handel, *“Probability on high dimensions"*, APC 550 Lecture Notes, Princeton University (2016)

{LY} Peter Li, Shing-Tung Yaue, *“On the Schrödinger Equation and the Eigenvalue Problem"*, Commun. Math. Phys. 88, 309-318 (1983)

{N} John Nash, *“Continuity of Solutions of Parabolic and Elliptic Equations"*, Amer. Jour. of Maths. vol. 80, no 4. p.931-954 (1958)

{R1} Gilles Royer, *"An Initiation to Logarithmic Sobolev Inequalities"*, AMS (2007)

{R2} Oscar S. Rothaus, *"Diffusion on Compact Riemannian Manifolds and Logarithmic Sobolev inequalities"*, Journal of Functional Analysis 42, 102-109 (1981)

{S} Laurent Saloff-Coste, *"Aspects of Sobolev-Type Inequalities"*, CUP (2002)

{SZ} Daniel W. Strook, Boguslaw Zegarlinski, *"The Logarithmic Sobolev Inequality for Continuous Spin Systems on a Lattice"*, Journal of Functional Analysis 104, 299-326 (1992)

{V} N. Varopoulos, *"Hardy-Littlewood Theory for Semigroups"*, Jour. of Func. Analysis, 63, p.240-260 (1985)

{Y} Kösaku Yosida, *"Functional Analysis"*, Springer (1995)

{Z} Boguslaw Zegarlinski, *"The Strong Decay to Equilibrium for the Stochastic Dynamics of Unbounded Spin Systems on a Lattice"*, Commun. Math. Phys. 175, p.401-432 (1996)