Ergodicity and Functional Inequalities

Xuda Ye

Peking University

May 26, 2022

Contents

Part I
Functional Inequalities

In this part we introduce the classical theory of functional inequalities in Markov processes, and demonstrate how the Poincaré inequality and the log-Sobolev inequality play the important role in the study of ergodicity. Contents in this part are based on [1, 2].

1 General Markov semigroups

We introduce the Markov semigroup, which serves as an alternative approach to characterize the Markov process. In particular, it is useful in the study of the ergodicity of the Markov process. Let $(P_t)_{t\geqslant 0}$ be the Markov semigroup corresponding to a Markov process $\{X_t\}_{t\geqslant 0}$ on E. If μ is the invariant distribution, then $(P_t)_{t\geqslant 0}$ is a contraction semigroup on $L^2(E,\mu)$.

1.1 Markov semigroup and invariant distribution

Let E be a Polish space, and $\{X_t\}_{t\geqslant 0}$ be the Borel measurable Markov process on E. The transition probablity kernel $p_t(x,\mathrm{d}y)$ thus satisfies the following conditions

- For each $x\in E$, $p_t(x,\cdot)$ is a probability measure on E;
- For each $B\in\mathcal{B}(E)$, $p_t(\cdot,B)$ is a measurable function from E to $\mathbb{R}$;
- For $0<t_1\leqslant\cdots\leqslant t_k$ and initial value $X_0=x\in E$, the distribution law of $(X_{t_1},\cdots,X_{t_k})$ is

$$p_{t_1}(x,\mathrm{d}y_1)p_{t_2-t_1}(y_1,\mathrm{d}y_2)\cdots p_{t_k-t_{k-1}}(y_{k-1},\mathrm{d}y_k). \tag{1.1}$$

For such a Markov process $\{X_t\}_{t\geqslant 0}$, define the associated Markov semigroups $(P_t)_{t\geqslant 0}$ by

$$P_tf(x):=\mathbb{E}^x[f(X_t)],\qquad t\geqslant 0,\qquad x\in E. \tag{1.2}$$

The semigroup $(P_t)_{t\geqslant 0}$ has the following properties:

1. For each $t\geqslant 0$, P_t maps bounded measurable functions to bounded measurable functions;

2. $P_0 = I$ is the indentity operator;
3. $P_t(1) = 1$, where 1 is the constant function;
4. (positivity preserving) If $f \geqslant 0$, then $P_t f \geqslant 0$;
5. (semigroup property) For $t, s \geqslant 0$, $P_{t+s} = P_t \circ P_s$.

Another useful result is given by Jensen's inequality. For any convex function $\phi : \mathbb{R} \to \mathbb{R}$,

$$P_t\big(\phi(f(x))\big) \geqslant \phi(P_t f(x)). \tag{1.3}$$

In particular, by choosing $\phi(x) = |x|^p$, one obtains

$$P_t(|f(x)|^p) \geqslant |P_t f(x)|^p. \tag{1.4}$$

The Markov semigroup $(P_t)_{t\geqslant 0}$ induces a dual semigroup $(P_t^*)_{t\geqslant 0}$ acting on probablity distributions:

$$\int_E P_t f(x)\nu(\mathrm{d}x) = \int_E f(x)(P_t^*\nu)(\mathrm{d}x). \tag{1.5}$$

Formally speaking, if ν is the distribution law of X_0, then $P_t^*\nu$ is the distribution law of X_t.

It is often convenient to use a probability density to represent the Markov semigroup $(P_t)_{t\geqslant 0}$ or the transirion kernel $p_t(x, \mathrm{d}y)$. Recall that P_t and $p_t(x, \mathrm{d}y)$ are related by

$$P_t f(x) = \int_E f(y) p_t(x, \mathrm{d}y). \tag{1.6}$$

Let m be a reference measure on E, and assume $p_t(x, \cdot)$ is absolutely continuous with respect to m, then there exists a density function $p_t(x, y)$ satisfying

$$p_t(x, \mathrm{d}y) = p_t(x, y) m(\mathrm{d}y), \qquad x, y \in E. \tag{1.7}$$

Hence the Markov semigroup can be represented as

$$(P_t f)(x) = \int_E f(y) p_t(x, y) m(\mathrm{d}y), \qquad x, y \in E. \tag{1.8}$$

In the viewpoint of $\{X_t\}_{t\geqslant 0}$, $p_t(x, y)$ is the transition probability density from x to y in time t.

Example Consider the standard Brownian motion B_t in $\mathbb{R}^d$. The heat kernel $p_t(x, y)$ is given by

$$p_t(x, y) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \exp\left(- \frac{|x-y|^2}{4t} \right). \tag{1.9}$$

Now we define the invariant distribution, which plays the central role in the study of the long time behavior of a Markov process.

Definition 1.1 (invariant) *Given the Markov semigroup $(P_t)_{t\geqslant 0}$ on the Polish space E, a probability distribution μ on E is said to be invariant for $(P_t)_{t\geqslant 0}$, if for any bounded measurable function $f : E \to \mathbb{R}$ and every $t \geqslant 0$,*

$$\int_E P_t f(x)\mu(\mathrm{d}x) = \int_E f(x)\mu(\mathrm{d}x). \tag{1.10}$$

An equivalent characterization is $P_t^*\mu = \mu$ for any $t \geqslant 0$. The invariant distribution μ induces the Banach space $L^p(E,\mu)$ with the norm

$$\|f\|_{L^p(E,\mu)} = \begin{cases} \left(\mu(|f|^p)\right)^{\frac{1}{p}}, & 1 \leqslant p < \infty \\ \operatorname{ess\,sup} |f(x)|, & p = \infty \end{cases} \tag{1.11}$$

Integrating (1.4) with the invariant measure μ, one obtains

$$\int_E |P_t f(x)|^p \mu(\mathrm{d}x) \leqslant \int_E P_t(|f(x)|^p)\mu(\mathrm{d}x) = \int_E |f(x)|^p \mu(\mathrm{d}x), \tag{1.12}$$

which can be equivalently written as

$$\|P_t f\|_{L^p(E,\mu)} \leqslant \|f\|_{L^p(E,\mu)}. \tag{1.13}$$

For any $1 \leqslant p \leqslant \infty$, the Markov semigroup $(P_t)_{t\geqslant 0}$ is a contraction in $L^p(E,\mu)$.

To identify $(P_t)_{t\geqslant 0}$ as a contraction semigroup, we further require the following condition. For each $f \in L^2(E,\mu)$, $P_t f$ converges to f in $L^2(E,\mu)$ as $t \to 0$. This usually reflects the reflects the regularity properties of the associated Markov process $\{X_t\}_{t\geqslant 0}$. To this end, we claim that $(P_t)_{t\geqslant 0}$ is a contraction semigroup on the Banach space $L^2(E,\mu)$.

Remark The definition of the contraction semigroup $(P_t)_{t\geqslant 0}$ relies on the invariant measure μ. For particular Markov processes, for example, the overdamped and underdamped Langevin dynamics, the invariant distribution μ can be explicitly derived.

1.2 Infinitesimal generator and Fokker-Planck equation

Let the associated Markov semigroup $(P_t)_{t\geqslant 0}$ be a contraction semigroup on $L^2(E,\mu)$. Based on the classical Hille-Yosida theory, $(P_t)_{t\geqslant 0}$ has an infinitesimal generator $\mathcal{L}$ on $L^2(E,\mu)$. Formally,

$$\frac{\partial}{\partial t} P_t = P_t \mathcal{L} = \mathcal{L} P_t. \tag{1.14}$$

Formally, one may write the Markov semigroup $(P_t)_{t\geqslant 0}$ as $P_t = \exp(t\mathcal{L})$. Using the infinitesimal generator $\mathcal{L}$, the invariant distribution μ can be interpreted as

$$\int_E \mathcal{L} f(x)\mu(\mathrm{d}x) = 0, \quad \forall f \in L^1(E,\mu). \tag{1.15}$$

In (1.3), take the limit $t \to 0$. For any convex function $\phi : \mathbb{R} \to \mathbb{R}$, one obtains

$$\mathcal{L}(\phi(f))(x) \geqslant \phi'(f(x))(\mathcal{L}f)(x). \tag{1.16}$$

If for some positive function $f \geqslant 0$, one defines the quantity

$$\Lambda(t) := \int_E \phi(P_t f)\mathrm{d}\mu, \tag{1.17}$$

then

$$\Lambda'(t) = \int_E \phi'(P_t f)\mathcal{L}P_t f \mathrm{d}\mu \leqslant \int_E \mathcal{L}\phi(P_t f)\mathrm{d}\mu = 0, \tag{1.18}$$

hence $\Lambda(t)$ is decreasing in time.

The infinitesimal generator $\mathcal{L}$ also depicts the time evolution of the distribution law. For the Markov process $\{X_t\}_{t\geqslant 0}$, assume the initial distribution is $\nu_0 \in \mathcal{P}(E)$, and the distribution law of X_t is $\nu_t = P_t^* \nu_0 \in \mathcal{P}(E)$. For any measurable and bounded function f on E, the evolution of $\nu_t(f) := \int_E f \mathrm{d}\nu_t$ is described via the Liouville equation,

$$\frac{\partial}{\partial t}\nu_t(f) = \frac{\partial}{\partial t}\nu_0(P_t f) = \nu_0(P_t \mathcal{L} f) = \nu_t(\mathcal{L} f), \tag{1.19}$$

or equivalently,

$$\frac{\partial}{\partial t}\int_E f(x)\nu_t(\mathrm{d}x) = \int_E \mathcal{L}f(x)\nu_t(\mathrm{d}x). \tag{1.20}$$

To describe the dynamics of the distribution law in a more closed form, introduce the Fokker-Planck equation. Given the reference measure, suppose the density of ν_t is $\rho(x,t)$, i.e.,

$$\nu_t(\mathrm{d}x) = \rho(x,t)m(\mathrm{d}x). \tag{1.21}$$

When $E = \mathbb{R}^d$ is the whole space, m is usually the Lebesgue measure. The Liouville equation (1.20) thus becomes

$$\frac{\partial}{\partial t}\int_E f(x)\rho(x,t)m(\mathrm{d}x) = \int_E \mathcal{L}f(x)\rho(x,t)m(\mathrm{d}x). \tag{1.22}$$

If $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$ in $L^2(E,m)$, then $\rho(x,t)$ satisfies the following PDE:

$$\frac{\partial \rho}{\partial t}(x,t) = \mathcal{L}^*\rho(x,t). \tag{1.23}$$

The Fokker-Planck equation (1.23) describes how the density function $\rho(x,t)$ evolves with time. The operator $\mathcal{L}^*$ can be seen as the infinitesimal generator of the dual semigroup P_t^*. The density function of the invariant distribution can be solved from the PDE $\mathcal{L}^*\rho = 0$.

1.3 Example: overdamped Langevin dynamics

The overdamped Langevin dynamics is the one of the most important stochastic dynamics in modern physics. Let the state space $E = \mathbb{R}^d$, and the reference measure m be the Lebesgue measure. Given the drift $b : \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times l}$, consider the overdamped Langevin dynamics given by the SDE

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t. \tag{1.24}$$

Define the diffusion matrix $a(x) = \sigma\sigma^{\mathrm{T}}(x)$, then the infinitesimal generator is

$$\mathcal{L}f(x) = b(x)\cdot\nabla f(x) + \frac{1}{2}a(x) : \nabla^2 f(x), \tag{1.25}$$

and the corresponding Fokker-Planck operator is

$$\mathcal{L}^*\rho = -\nabla\cdot(b(x)\rho(x)) + \frac{1}{2}\nabla\cdot(a(x)\nabla\rho(x)). \tag{1.26}$$

In other words, the carré du champ operator captures the second-order part of the generator.

The overdamped Langevin dynamics (1.24) does not have a closed form of the invariant distribution. However, if $b(x)$ is in the gradient form and σ is constant, for example,

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \tag{1.27}$$

The invariant distribution is explicitly the Boltzmann distribution

$$\mu(\mathrm{d}x) = \frac{1}{Z}\exp(-V(x))\mathrm{d}x. \tag{1.28}$$

2 Symmetric Markov semigroups

The symmetric Markov semigroup allows one to study the long time behavior of the Markov process. If the invariant distribution μ is reversible, then the infinitesimal generator $\mathcal{L}$ is self-adjoint in $L^2(E,\mu)$. The spectral gap of $\mathcal{L}$ reveals the convergence rate to the equilibrium. The Markov triple (E,μ,Γ) is sufficient to build a symmetric Markov semigroup.

2.1 Symmetric semigroup

Definition 2.1 (reversibility) *The Markov semigroup $(P_t)_{t\geqslant 0}$ is said to be symmetric with respect to μ (or μ is reversible), if for any $f,g\in L^2(E,\mu)$ and $t\geqslant 0$,*

$$\int_E f(x)(P_t g)(x)\mu(\mathrm{d}x) = \int_E g(x)(P_t f)(x)\mu(\mathrm{d}x). \tag{2.1}$$

In terms of the generator $\mathcal{L}$, the reversibility can be expressed as for any $f,g\in L^2(E,\mu)$,

$$\int_E f(x)(\mathcal{L}g)(x)\mu(\mathrm{d}x) = \int_E g(x)(\mathcal{L}f)(x)\mu(\mathrm{d}x). \tag{2.2}$$

In terms of the transition kernel $p_t(x,\mathrm{d}y)$, the reversibility can also be written as

$$p_t(x,\mathrm{d}y)\mu(\mathrm{d}x) = p_t(y,\mathrm{d}x)\mu(\mathrm{d}y), \tag{2.3}$$

which is known as the detailed balance condition.

Now we introduce the carré du champ operator, which will be useful in functional inequalities.

Definition 2.2 *Given the Markov semigroup $(P_t)_{t\geqslant 0}$ and the infinitesimal generator $\mathcal{L}$, the carré du champ Γ operator is defined by*

$$\Gamma(f,g) = \frac{1}{2}\big(\mathcal{L}(fg) - f\mathcal{L}g - g\mathcal{L}f\big). \tag{2.4}$$

The Dirichlet form is defined by

$$\mathcal{E}(f,g) = \int_E \Gamma(f,g)\mathrm{d}\mu. \tag{2.5}$$

It is clear that for symmetric Markov semigroups, $\Gamma,\mathcal{E}$ are symmetric in f,g and

$$\mathcal{E}(f,g) = \int_E \Gamma(f,g)\mathrm{d}\mu = -\int_E f\mathcal{L}g\mathrm{d}\mu = -\int_E g\mathcal{L}f\mathrm{d}\mu. \tag{2.6}$$

Starting from a symmetric operator Γ, one may determine generator $\mathcal{L}$ from (2.6) such that the resulting Markov semigroup is symmetric. Therefore, it is equivalent to use a Γ operator or a generator $\mathcal{L}$ to describe a symmetric semigroup. In the following (E,μ,Γ) is called a Markov triple.

2.2 Spectral decomposition and ergodicity

When μ is the reversible measure, the generator $\mathcal{L}$ is self-adjoint in $L^2(E,\mu)$. Therefore, the spectral decomposition of $\mathcal{L}$ is available. Since $(P_t)_{t\geqslant 0}$ is a contraction semigroup in $L^2(E,\mu)$, all the eigenvalues of $\mathcal{L}$ are nonpositive. Suppose

$$-\mathcal{L}e_k = \lambda_k e_k, \qquad k \geqslant 0, \tag{2.7}$$

where $\{e_k\}_{k=0}^{\infty}$ is the orthonormal basis of $L^2(E,\mu)$ and $\lambda_k \geqslant 0$ is the eigenvalue. $\mathcal{L}$ has a trivial eigenpair

$$\lambda_0 = 0, \qquad e_0(x) \equiv 0. \tag{2.8}$$

The spectral gap is defined as $\lambda_1 > 0$, the difference between the two eigenvalues of $\mathcal{L}$. The spectral gap characterizes the convergence rate to the equilibrium. Formally, given the smooth measurable function f on E, define

$$u(x,t) = P_t f(x) = \mathbb{E}^x[f(X_t)], \tag{2.9}$$

then $u(x,t)$ satisfies the PDE

$$\frac{\partial u}{\partial t} = \mathcal{L}u. \tag{2.10}$$

Assume the initial value $u(x,0) = f(x)$ can be decomposed on $L^2(E,\mu)$:

$$u_0(x) = \sum_{k\geqslant 0} c_k e_k(x), \tag{2.11}$$

then c_0 is the inner product of $f(x)$ and the constant function 1,

$$c_0 = \int_E f(x)\mu(\mathrm{d}x) = \mu(f). \tag{2.12}$$

The general solution $u(x,t)$ can be expressed as

$$u(x,t) = \mu(f) + \sum_{k\geqslant 1} c_k e_k(x) e^{-\lambda_k t}. \tag{2.13}$$

As $t\to\infty$, $u(x,t)$ converges to $\mu(f)$ exponentially in the variance sense:

$$\int_E |(P_t f)(x) - \mu(f)|^2 \mu(\mathrm{d}x) \leqslant e^{-2\lambda_1 t} \int_E |f(x) - \mu(f)|^2 \mu(\mathrm{d}x), \tag{2.14}$$

In variance sense, the distribution law of X_t converges to μ exponentially, and the convergence rate is exactly $\lambda_1 > 0$. To study the long time behavior of the Markov process, the spectral gap of the generator $\mathcal{L}$ in $L^2(E,\mu)$ is the crucial quantity.

Note that the Dirichlet form $\mathcal{E}(\cdot,\cdot)$ is also closely related to the spectrum of the generator $\mathcal{L}$. Recall that $(e_k,\lambda_k)_{k\geqslant 0}$ is the orthonormal basis of $-\mathcal{L}$, then

- $\mathcal{E}(e_k,e_k) = \lambda_k$, $k\geqslant 0$;
- $\mathcal{E}(e_k,e_l) = 0$, $k\neq l$.

Therefore, $\mathcal{E}(f,g)$ can be seen as the weighted inner product. If the coefficients of f,g are f_k,g_k,

$$\mathcal{E}(f,g) = \sum_{k\geqslant 0} \lambda_k f_k g_k. \tag{2.15}$$

2.3 Curvature-dimension condition

To describe the curvature-dimension condition, we introduce the Γ_2 operator by

$$\Gamma_2(f,g) = \frac{1}{2}\big(\mathcal{L}\Gamma(f,g) - \Gamma(f,\mathcal{L}g) - \Gamma(\mathcal{L}f,g)\big). \tag{2.16}$$

Take $\phi(x) = |x|^2$ in (1.16), one obtains

$$\mathcal{L}(f^2) \geqslant 2f(\mathcal{L}f) \Longrightarrow \Gamma(f,f) \geqslant 0. \tag{2.17}$$

We simply write $\Gamma(f) = \Gamma(f,f)$ and $\Gamma_2(f) = \Gamma_2(f,f)$, then

$$\int_E \Gamma(f)\mathrm{d}\mu = -\int_E f\mathcal{L}f\mathrm{d}\mu, \qquad \int_E \Gamma_2(f)\mathrm{d}\mu = \int_E (\mathcal{L}f)^2\mathrm{d}\mu. \tag{2.18}$$

The curvature-dimension condition is an important property to derive the functional inequalities.

Definition 2.3 (curvature) *A diffusion operator $\mathcal{L}$ is said to satisfy the curvature-dimension condition $CD(\varepsilon,\infty)$, for $\varepsilon \in \mathbb{R}$, if for every function $f : E \to \mathbb{R}$,*

$$\Gamma_2(f) \geqslant \varepsilon\,\Gamma(f). \tag{2.19}$$

In Sections 3 and 4, we shall use the curvature-dimension conditions to derive the Poincaré inequality and the log-Sobolev inequality, respectively.

2.4 Example: overdamped Langevin dynamics

Consider the overdamped Langevin dynamics (1.24). We shall show that the reversibility of invariant distribution μ can be derived from a simple relation. Recalling the drift force $b : \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion matrix $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, define the flux operator $J(\rho)$ by

$$J(\rho) = b(x)\rho(x) - \frac{1}{2}a(x)\nabla\rho(x), \tag{2.20}$$

then $\mathcal{L}^*\rho = -\nabla\cdot J(\rho)$, and the Fokker-Planck equation can be written in the conservation law

$$\frac{\partial\rho}{\partial t} + \nabla\cdot J(\rho) = 0. \tag{2.21}$$

Clearly, $\rho(x)$ is the invariant measure iff $\nabla\cdot J(\rho) = 0$. In the following, we show that $\rho(x)$ is reversibile if a stronger condition $J(\rho) = 0$ is satisfied.

Lemma 2.1 (reversibility) *If $\rho(x)$ satisfies $J(\rho) = 0$, then $\rho(x)$ is reversibile.*

Proof Let $\mu(\mathrm{d}x) = \rho(x)\mathrm{d}x$ be the invariant distribution. To show μ is reversible, note that

$$\begin{aligned}
\int_{\mathbb{R}^d} f(x)(\mathcal{L}g)(x)\mu(\mathrm{d}x) &= \int_{\mathbb{R}^d} f(x)(\mathcal{L}g)(x)\rho(x)\mathrm{d}x \\
&= \int_{\mathbb{R}^d} g\mathcal{L}^*(\rho f)\mathrm{d}x \\
&= -\int_{\mathbb{R}^d} g\nabla\cdot J(\rho f) \\
&= \int_{\mathbb{R}^d} \nabla g\cdot J(\rho f).
\end{aligned}$$

Here,

$$J(\rho f) = b\rho f - \frac{1}{2}a\nabla(\rho f) = J(\rho)f - \frac{1}{2}\rho a\nabla f = -\frac{1}{2}\rho a\nabla f. \tag{2.22}$$

Hence

$$\int_{\mathbb{R}^d} f(x)(\mathcal{L}g)(x)\mu(\mathrm{d}x) = -\frac{1}{2}\int_{\mathbb{R}^d} \rho(\nabla f)^{\mathrm{T}} a\nabla g \mathrm{d}x = -\frac{1}{2}\int_{\mathbb{R}^d} a^{ij}(x)\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}\mu(\mathrm{d}x), \tag{2.23}$$

which is symmetric in the functions f, g. Therefore, $\rho(x)$ is reversible. □

When the dynamics is driven by the gradient,

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

the invariant distribution is $\mu(\mathrm{d}x) \propto e^{-V(x)}\mathrm{d}x$, and the Γ, Γ_2 operators are

$$\Gamma(f,g) = \nabla f \cdot \nabla g, \quad \Gamma_2(f,g) = \nabla^2 f : \nabla^2 g + \nabla^2 V(\nabla f, \nabla g). \tag{2.24}$$

If the potential function is strongly convex, i.e., $\nabla^2 V \geqslant \kappa I$, then the curvature-dimension condition $CD(\kappa,\infty)$ holds. In most cases below, we deal with the overdamped Langevin dynamics.

3 Poincaré inequality

The Poincaré inequality provides a simple approach to characterize the spectral gap of the generator. The good thing is that, the Poincaré inequality itself does not require the spectral knowledge explicitly, thus can be proved via other approaches. In particular, the curvature-dimension condition is convenient to verify the Poincaré inequality.

3.1 Poincaré inequality and spectral gap

For a probability measure ν on the Polish space E, define the variance of a function $f \in L^2(E,\nu)$

$$\mathrm{Var}_\nu(f) := \int_E f^2 \mathrm{d}\nu - \left(\int_E f \mathrm{d}\nu\right)^2. \tag{3.1}$$

The Poincaré inequality w.r.t. a Markov triple (E,μ,Γ) is defined as follows.

Definition 3.1 (Poincaré) *A Markov Triple (E,μ,Γ) is said to satisfy a Poincaré inequality $P(C)$ with constant $C > 0$, if for all functions $f : E \to \mathbb{R}$,*

$$\mathrm{Var}_\mu(f) \leqslant C\mathcal{E}(f) = C\int_E \Gamma(f)\mathrm{d}\mu. \tag{3.2}$$

When μ is invariant distribution, it is also convenient to define the covariance

$$\mathrm{Cov}_\mu(f,g) = \int_E fg\mathrm{d}\mu - \int_E f\mathrm{d}\mu \int_E g\mathrm{d}\mu. \tag{3.3}$$

Just as $\mathcal{E}(f,g)$, $\mathrm{Cov}(f,g)$ has good spectral structure:

- $\mathrm{Cov}_\mu(e_k, e_k) = 1$, $k \geqslant 1$;
- $\mathrm{Cov}_\mu(e_k, e_l) = 0$, $k \neq j$.

Suppose f is decomposed as

$$f(x) = \sum_{k\geqslant 0} c_k e_k(x), \tag{3.4}$$

then the Poincaré inequality $P(C)$ is equivalent to

$$\sum_{k\geqslant 1} c_k^2 \leqslant C \sum_{k\geqslant 0} \lambda_k c_k^2. \tag{3.5}$$

Therefore, the best constant C is $C = \lambda_1^{-1}$. If $P(C)$ holds, then the spectral gap $\lambda_1 \geqslant C^{-1}$.

In the overdamped Langevin dynamics, $\mathcal{E}(f)$ involves ∇f, hence $P(C)$ is about using derivatives to control function values. By adding f to a constant value, its variance and Dirichlet form won't change. However, due to the spectral structure of bilinear functionals Cov and $\mathcal{E}$, $P(C)$ describes the spectral gap of the generator $\mathcal{L}$.

Lemma 3.1 *A Markov triple (E, μ, Γ) satisfies the Poincáre inequality $P(C)$ iff for any $f : E \to \mathbb{R}$*

$$\int_E \Gamma(f)\mathrm{d}\mu \leqslant C \int_E \Gamma_2(f)\mathrm{d}\mu. \tag{3.6}$$

The inequality can be equivalently written as

$$\int_E \Gamma(f)\mathrm{d}\mu \leqslant C \int_E (\mathcal{L}f)^2\mathrm{d}\mu. \tag{3.7}$$

In the viewpoint of spectral structure, the eigenvalues of LHS and RHS are λ_k and λ_k^2 respectively. Nevertheless, this result can be proved using standard PDE techniques.

Proof Assume $\mu(f) = 0$. The method is to define the function

$$\Lambda(t) := \int_E (P_t f)^2 \mathrm{d}\mu, \tag{3.8}$$

and note that

$$\Lambda'(t) = -2\int_E \Gamma(P_t f)\mathrm{d}\mu, \qquad \Lambda''(t) = 4\int_E (\mathcal{L}P_t f)^2\mathrm{d}\mu. \tag{3.9}$$

By condition

$$\Lambda''(t) \geqslant -\frac{2}{C}\Lambda'(t), \qquad \forall t \geqslant 0, \tag{3.10}$$

hence $\Lambda'(t)$ decays to 0 as $t \to \infty$ exponentially, and $\Lambda(t)$ converges to 0 exponentially. Then

$$\mathrm{Var}_\mu(f) = -\int_0^\infty \Lambda'(t)\mathrm{d}t \leqslant \frac{C}{2}\int_0^\infty \Lambda''(t)\mathrm{d}t = -\frac{C}{2}\Lambda'(0) = C\int_E \Gamma(t)\mathrm{d}\mu. \tag{3.11}$$

Formally, this equivalent form is using second-order derivatives to control first-order derivatives. □

3.2 Tensorization, curvature-dimension and bounded perturbation

The Poincacé inequalities are preserved under tensorization.

Theorem 3.1 (tensorization) *If (E_1, μ_1, Γ_1) and (E_2, μ_2, Γ_2) satisfy Poincaré inequalities with respective constants C_1 and C_2, then the product Markov triple $(E_1 \otimes E_2, \mu_1 \otimes \mu_2, \Gamma_1 \oplus \Gamma_2)$ satisfies a Poincaré inequality with constant $C = \max(C_1, C_2)$.*

$\Gamma_1 \oplus \Gamma_2$ is defined as follows. If $\mathcal{L}_1$ has eigenvalues λ_k^1 on E_1 and $\mathcal{L}_2$ has eigenvalues λ_l^2 on E_2, then $\mathcal{L}_1 + \mathcal{L}_2$ has eigenvalues $\lambda_k^1 + \lambda_l^2$. Under the tensorization one has

$$\mathcal{E}(f) = \int_{E_2} \mathcal{E}_1(f) \mathrm{d}\mu_1 + \int_{E_1} \mathcal{E}_2(f) \mathrm{d}\mu_2. \tag{3.12}$$

The curvature-dimension condition is a sufficient condition for the Poincaré inequality.

Theorem 3.2 *Under the curvature-dimension condition $CD(\varepsilon, \infty)$, $\varepsilon > 0$, , the Markov triple (E, μ, Γ) satisfies a Poincaré inequality $P(C)$ with constant $C = \varepsilon^{-1}$.*

Corollary 3.1 *For the overdamped Langevin dynamics*

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

if $\nabla^2 V \geqslant \kappa > 0$, then the $P(\kappa^{-1})$ holds, and consequently the spectral gap $\lambda_1 \geqslant \kappa$.

The result above seems not appealing because it requires the potential function $V(x)$ to be strongly convex. However, the bounded perturbation property of the Poincáre inequality allows us to deal with more general cases.

Theorem 3.3 (perturbation) *Assume (E, μ, Γ) satisfies $P(C)$. If μ_1 is a distribution whose density h w.r.t. μ satisfies $1/b \leqslant h \leqslant b$ for some $b > 0$, then (E, μ_1, Γ) satisfies $P(b^3 C)$.*

The proof is surprisingly elementary. Just use the fact that

$$\mathrm{Var}_\nu(f) = \frac{1}{2} \int_{E \times E} [f(x) - f(y)]^2 \nu(\mathrm{d}x)\nu(\mathrm{d}y), \tag{3.13}$$

and the direct comparison between μ_1 and μ. Using this result we can deal with more general potential functions.

Corollary 3.2 *For the overdamped Langevin dynamics*

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

if $V = V_1 + V_2$, where $\nabla^2 V_1 \geqslant \kappa > 0$ and $|V_2| \leqslant M$, then $P(e^{3M}\kappa^{-1})$ holds.

3.3 Variance decay

The Poincaré inequality also yields the exponential decay of the variance in time.

Theorem 3.4 (variance) *Given a Markov triple (E, μ, Γ) with the associated Markov semigroup $(P_t)_{t \geqslant 0}$, the following assertions are equivalent:*

1. (E, μ, Γ) *satisfies a Poincaré inequality* $P(C)$*;*

2. For every function $f : E \to \mathbb{R}$ *in* $L^2(E, \mu)$*,*

$$\mathrm{Var}_\mu(P_t f) \leqslant e^{-2t/C} \mathrm{Var}_\mu(f). \tag{3.14}$$

3. For every function $f \in L^2(E, \mu)$*, there exists a constant* $c(f) > 0$ *such that, for every* $t \geqslant 0$*,*

$$\mathrm{Var}_\mu(P_t f) \leqslant c(f) e^{-2t/C}. \tag{3.15}$$

We present a proof based on functional inequalities.

Proof Assume f satisfies $\int_E f \mathrm{d}\mu = 0$. Define

$$\Lambda(t) := \mathrm{Var}_\mu(P_t f) = \int_E (P_t f)^2 \mathrm{d}\mu, \tag{3.16}$$

then

$$\Lambda'(t) = 2 \int_E P_t f \mathcal{L} P_t f \mathrm{d}\mu = -2\mathcal{E}(P_t f). \tag{3.17}$$

The Poincaré inequality then implies

$$\Lambda(t) \leqslant -\frac{C}{2} \Lambda'(t), \tag{3.18}$$

which yields the exponential decay in $\Lambda(t)$. □

As we have seen in Section 2, the variance decay

$$\int_E |(P_t f)(x) - \mu(f)|^2 \mu(\mathrm{d}x) \leqslant e^{-2t/C} \int_E |f(x) - \mu(f)|^2 \mu(\mathrm{d}x) \tag{3.19}$$

directly follows from the fact that the spectral gap $\lambda_1 > C^{-1}$.

4 Log-Sobolev inequality

The Poincáre inequality gives exponential decay in the variance. The log-Sobolve inequality introduced is stronger and is able to give the exponential decay in entropy.

4.1 Log-Sobolev inequality and tightening

Given the positive function f on E, define the entropy of f with respect to a distribution ν by

$$\mathrm{Ent}_\nu(f) = \int_E f \log f \mathrm{d}\nu - \int_E f \mathrm{d}\nu \log \left(\int_E f \mathrm{d}\nu \right). \tag{4.1}$$

It's easy to see for any $c > 0$, $\mathrm{Ent}_\nu(cf) = c \,\mathrm{Ent}_\nu(f)$. Therefore, we can always rescale f to satisfy

$$\int_E f \mathrm{d}\nu = 1, \tag{4.2}$$

so that $f(x)$ can be viewed as the probability density in the reference measure ν. In this case, $\mathrm{Ent}_\nu(f)$ is exactly the entropy of the density function f.

Definition 4.1 (log-Sobolev) *A Markov triple (E, μ, Γ) is said to satisfy a log-Sobolev inequality $LS(C)$ with constant C, if for any function $f : E \to \mathbb{R}$,*

$$\mathrm{Ent}_\mu(f^2) \leqslant 2C\mathcal{E}(f). \tag{4.3}$$

Like the Poincaré inequality, a sufficent condition of $LS(C)$ is as follows.

Lemma 4.1 *A Markov triple (E, μ, Γ) satisfies a log-Sobolev inequality $LS(C)$ for some $C > 0$ if*

$$\int_E f\Gamma(\log f) \leqslant C \int_E f\Gamma_2(\log f)\mathrm{d}\mu \tag{4.4}$$

for every positive function f.

This result is convenient for the application of the curvature-dimension condition.

Theorem 4.1 *The log-Sobolev inequality $LS(C)$ implies the Poincaré inequality $P(C)$.*

Proof Apply $LS(C)$ with $f = 1 + \varepsilon h$ where $\int_E h\mathrm{d}\mu = 0$. As $\varepsilon \to 0$, it can be checked using the Taylor expansion

$$\mathrm{Ent}_\mu(f^2) = 2\varepsilon^2 \int_E h^2\mathrm{d}\mu + o(\varepsilon^2). \tag{4.5}$$

On the other hand, $\mathcal{E}(f) = \varepsilon^2\mathcal{E}(h)$, hence

$$2\int_E h^2\mathrm{d}\mu \leqslant 2C\mathcal{E}(h), \tag{4.6}$$

which implies the Poincaré inequality $P(C)$. □

4.2 Tensorization, curvature-dimension and bounded perturbation

Theorem 4.2 (tensorization) *If (E_1, μ_1, Γ_1) and (E_2, μ_2, Γ_2) satisfy $LS(C_1, D_1)$ and $LS(C_2, D_2)$ respectively, then the Markov triple $(E_1 \otimes E_2, \mu_1 \otimes \mu_2, \Gamma_1 \oplus \Gamma_2)$ satisfies $LS(\max(C_1, C_2), D_1 + D_2)$.*

Using the equivalent form of the log-Sobolev inequality, it is easy to derive the log-Sobolev inequality under the curvature-dimension condition.

Theorem 4.3 *Under the curvature condition $CD(\varepsilon, \infty)$, $\varepsilon > 0$, the Markov triple (E, μ, Γ) satisfies a log-Sobolev inequality $LS(C)$ with constant $C = \varepsilon^{-1}$.*

Since $CD(\varepsilon, \infty)$ implies $LS(\varepsilon^{-1})$, for every function f,

$$\mathrm{Ent}_\mu(f^2) \leqslant \frac{2}{\varepsilon}\mathcal{E}(f). \tag{4.7}$$

For the general case $CD(\varepsilon, n)$, the proof is more technical.

Corollary 4.1 *For the overdamped Langevin dynamics*

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

if $\nabla^2 V \geqslant \kappa > 0$, then $LS(\kappa^{-1})$ holds.

Since $LS(C)$ is stronger than $P(C)$, we have the following estimates [3].

Corollary 4.2 *For the overdamped Langevin dynamics*

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

assume $\nabla^2 V \geqslant \kappa > 0$. Let $\lambda(\mu)$ be the best constant satisfying the Poincaré inequality

$$\lambda(\mu)\mathrm{Var}_\mu(f) \leqslant \int_{\mathbb{R}^d} |\nabla f|^2 \mathrm{d}\mu,$$

and $\rho(\mu)$ be the best constant satisfying the log-Sobolev inequality

$$\rho(\mu)\mathrm{Ent}_\mu(f) \leqslant 2\int_{\mathbb{R}^d} |\nabla f|^2 \mathrm{d}\mu,$$

then $\lambda(\mu) \geqslant \rho(\mu) \geqslant \kappa$.

Similarly, the log-Sobolev inequality still holds under bounded perturbation.

Theorem 4.4 (perturbation) *Assume that the Markov triple (E, μ, Γ) satisfies a log-Sobolev inequality $LS(C)$. Let μ_1 be a probability measure with density h with respect to μ such that $1/b \leqslant h \leqslant b$ for some constant $b > 0$. Then μ_1 satisfies $LS(b^2 C)$.*

Corollary 4.3 *For the overdamped Langevin dynamics*

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

if $V = V_1 + V_2$, where $\nabla^2 V_1 \geqslant \kappa > 0$ and $|V_2| \leqslant M$, then $LS(e^{2M}\kappa^{-1})$ holds.

Therefore, for the overdamped Langevin dynamics, the log-Sobolev inequality can be established.

4.3 Entropy decay

The crucial property of the log-Sobolev inequality is the exponential decay in entropy.

Theorem 4.5 (entropy) *The log-Sobolev inequality $LS(C)$ for the probability measure μ is equivalent to saying that for every positive function f in $L^1(E, \mu)$,*

$$\mathrm{Ent}_\mu(P_t f) \leqslant e^{-2t/C}\mathrm{Ent}_\mu(f), \qquad \forall t \geqslant 0. \tag{4.8}$$

Proof Define the Fisher information of f w.r.t. ν by

$$I_\nu(f) = \int_E \frac{\Gamma(f)}{f}\mathrm{d}\mu, \tag{4.9}$$

then by direct calculation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}_\mu(P_t f) = -I_\mu(P_t f). \tag{4.10}$$

Hence we only need to verify

$$\mathrm{Ent}_\mu(f) \leqslant \frac{C}{2} I_\mu(f), \tag{4.11}$$

which is equivalent to

$$\mathrm{Ent}_\mu(f) \leqslant \frac{C}{2}\int_E \frac{\Gamma(f)}{f}\mathrm{d}\mu = 2C\int_E \Gamma(\sqrt{f})\mathrm{d}\mu. \tag{4.12}$$

Let $f = g^2$, then it becomes

$$\mathrm{Ent}_\mu(g^2) \leqslant 2C\int_E \Gamma(g)\mathrm{d}\mu, \tag{4.13}$$

which is exacly the log-Sobolev inequality. □

The convergence in entropy is a strong result. Let $\nu \ll \mu$ with density f, then $H(\nu|\mu) = \mathrm{Ent}_\mu(f)$.

$$\|\mu - \nu\|_{\mathrm{TV}}^2 \leqslant \frac{1}{2}H(\nu|\mu), \tag{4.14}$$

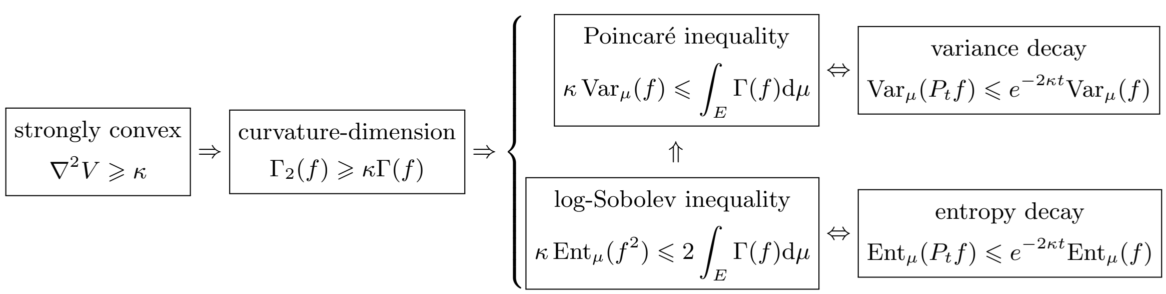
the exponential convergence in total variation can be obtained.

5 Summary

In this part we introduce two important inequalities: Poincaré inequality and log-Sobolev inequality to establish the ergodic properties of the Markov process. These functional inequalities yield exponential convergence of the distribution law due to the following simple observation: the time derivative of the function integrals automatically involve the derative integrals.

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_E (P_t f)^2\mathrm{d}\mu = 2\int_E (P_t f)(\mathcal{L}P_t f)\mathrm{d}\mu = -2\int_E \Gamma(P_t f)\mathrm{d}\mu. \tag{5.1}$$

The curvature-dimension provides a sufficient condition to build the functional inequalities, which estimates the convergence in the strongly convex case. The bounded perturbation of these inequalities allow us to deal with more general potentials. The theory functional inequalities theory is briefly summarized in the flow chart below.



Part II

Ergodicity of Particle Systems

We introduce several approaches to study the ergodicity of interacting particle systems using functional inequalities. Our goal is to prove that the exponential convergence the rate of the interacting particle system does not depend on N, when the nonlinear interaction part is not strong enough.

6 Mean-field interacting particle system

Consider the interacting particle system (IPS) of N-particles in $\mathbb{R}^d$:

$$\mathrm{d}X_t^i = -\nabla V(X_t^i)\mathrm{d}t - \frac{1}{N-1}\sum_{j\neq i}\nabla W(X_t^i - X_t^j)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^i, \quad i = 1,\cdots,N, \tag{6.1}$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is the external potential and $W : \mathbb{R}^d \to \mathbb{R}$ is the pairwise interaction potential, with $W(x) = W(-x)$, i.e., $W(x)$ is an even function on $\mathbb{R}^d$. The invariant distribution of the N-particle system can be explicitly written as

$$\mu(\mathrm{d}x) \propto \exp\bigg(-\sum_{i=1}^N V(x^i) - \frac{1}{N-1}\sum_{1\leqslant i<j\leqslant N}\nabla W(x^i - x^j)\bigg)\mathrm{d}x. \tag{6.2}$$

Note that the condition $W(x) = W(-x)$ ensures that μ is symmetric in all N particles. Our question is, does the distribution law of the IPS (6.1) converges to its invariant distribution μ^N with a convergence rate independent of N? This question is closed related to the uniform-in-time propagation of chaos of dynamical particle systems, and one may refer to [4] for further discussion.

Formally, as the number of particles $N \to \infty$, the IPS (6.1) converges to the mean-field limit, which is a McKean-Vlasov process (MVP):

$$\begin{cases} \mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t - (\nabla W * \mu_t)(X_t)\mathrm{d}t + \sqrt{2}B_t, \\ \quad \mu_t = \mathrm{Law}(X_t). \end{cases} \tag{6.3}$$

If the IPS has a convergence c uniform in N, then it reasonable to deduce that the MVP also has a convergence rate c. Conversely, if the convergence rate of IPS is not uniform in N, then the MVP cannot converge to the invariant distribution exponentially. It is worth pointing that the MVP can have multiple invariant distributions (e.g., V is double-well, see [5]), hence it is necessary to require the nonlinear interaction part to be not strong enough.

Now we present a historical review of the study of the ergodicity of the MVP (6.3).

1. Carrillo, McCann & Villani [6] (2003): Seminal work in the study of ergodic granular media equations. When $V(x)$ is uniformly convex and $W(x)$ is convex, the free energy

   $$\mathcal{F}(\rho) = \int_{\mathbb{R}^d} V(x)\rho(x)\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y + \int_{\mathbb{R}^d}\rho\log\rho\mathrm{d}x$$

   minus its minimum has exponential decay.

2. Bolley, Gentil & Guillin [7] (2013): Extension of the work above. When $V(x)$ is uniformly convex andd $W(x)$ is convex outside a finite region, the exponential convergence in Wasserstein-2 distance is proved. Note that by Talagrand's inequality this is weaker than entropy decay.

3. Eberle, Guillin, etc. [8,9] (2011-2020): The reflection coupling is used to derive the geometric ergodicity for the mean-field IPS (6.1) and the MVP (6.3) in the Wasserstein-1 distance. The unifrom-in-time propagation of chaos can also be derived [11].

4. Guillin, Liu, Wu & Zhang [3] (2019): The functional inequalities are used to derive the geometric ergodicity for the mean-field IPS (6.1) and the MVP (6.3) in the free energy and Wasserstein-2 distance. Stronger than the results above.

7 Convergence rate by Poincaré inequality

For convenience, define the total energy of the particle system by

$$H(x^1, \cdots, x^N) = \sum_{i=1}^{N} V(x^i) + \frac{1}{N-1} \sum_{1 \leqslant i < j \leqslant N} W(x^i - x^j), \tag{7.1}$$

then if $\nabla^2 V \geqslant \kappa > 0$ and $\nabla^2 W \geqslant 0$, it can be verified that

$$\nabla^2 H \geqslant \kappa I, \quad \text{in } \mathbb{R}^{Nd \times Nd}, \tag{7.2}$$

hence the curvature-dimension condition $CD(\kappa, \infty)$ holds for μ. Without these convexity conditions, more meticulous estimates are required.

With the total energy $H(x)$ defined in (7.1), the generator of the IPS is

$$\mathcal{L} f = -\nabla H \cdot \nabla f + \Delta f, \tag{7.3}$$

and our goal is to estimate the spectral gap of $\mathcal{L}$ in $L^2(\mathbb{R}^{Nd}, \mu)$ without the convexity assumption. To achieve this goal, we derive the Poincaré inequality for the one-particle system, then generalize the result to the N-particle system. The results in this part are mainly from [3].

7.1 Spectral gap of the single particle system

Consider to the overdamped Langevin dynamics in $\mathbb{R}^d$ given by

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \tag{7.4}$$

whose invariant distribution is $\mathrm{d}\mu \propto e^{-V}\mathrm{d}x$. If the potential function $V(x)$ is strongly convex, the spectral gap can be estimated under the curvature-dimension condition. Under the general dissipation condition, the result is stated as follows. Define the function $b_0 : (0, +\infty) \to \mathbb{R}$

$$b_0(r) = \sup_{|x-y|=r} -\frac{1}{r} \langle x - y, \nabla V(x) - \nabla V(y) \rangle. \tag{7.5}$$

For example, if $V(x) = |x|^2/2$, then $b_0(r) = -r$. If the Lipschitzian constant

$$c_{\mathrm{Lip}} = \frac{1}{4} \int_0^{\infty} \exp\left(\frac{1}{4} \int_0^s b_0(u)\mathrm{d}u \right) s\mathrm{d}s < +\infty, \tag{7.6}$$

then we have the following estimate of the spectral gap (Theorem 1.1, [12]):

Theorem 7.1 *The spectral gap of the generator $\mathcal{L} = -\nabla V \cdot \nabla + \Delta$ on $L^2(\mathbb{R}^d, \mu)$ satisfies*

$$\lambda_1(\mu) \geqslant \frac{1}{c_{\mathrm{Lip}}}. \tag{7.7}$$

It is worthing noting that the assumption on $b_0(r)$ is very similar to the function $\kappa(r)$ in [8]. Both results require the drift force to be dissipative to obtain contractivity.

In the IPS, given the positions of the other $N-1$ particles, the conditional distribution of $\mu \in \mathcal{P}(\mathbb{R}^{Nd})$ in the i-th particle is

$$\mathrm{d}\mu^i(x^i|\hat{x}^i) \propto \exp\Big(- V(x^i) - \frac{1}{N-1}\sum_{j\neq i} W(x^i - x^j)\Big)\mathrm{d}x^i, \tag{7.8}$$

and the associated energy function is

$$H^i(x^i|\hat{x}^i) = V(x^i) + \frac{1}{N-1}\sum_{j\neq i} W(x^i - x^j). \tag{7.9}$$

Note that μ^i is the conditional distribution rather than the marginal distribution μ, thus its definition relies on $\hat{x}^i$, positions of other $N-1$ particles. Define the conditional generator

$$\mathcal{L}^i = -\nabla_i H^i \cdot \nabla + \Delta_i, \tag{7.10}$$

then $\mathcal{L}^i$ is self-adjoint in $L^2(\mathbb{R}^d, \mu^i)$, and define $\lambda_1(\mu^i)$ by its spectral gap. Under appropriate conditions, we can derive uniform lower bounds for μ^i.

For the IPS, define the function $b_0(r)$ and the constant $c_{\mathrm{Lip},m}$ as follows:

$$b_0(r) = \sup_{|x-y|=r,z} -\frac{1}{r}\langle x-y, \nabla V(x) - \nabla V(y) + \nabla W(x-z) - \nabla W(y-z)\rangle, \tag{7.11}$$

and

$$c_{\mathrm{Lip},m} = \frac{1}{4}\int_0^{\infty} \exp\left(\frac{1}{4}\int_0^s b_0(u)\mathrm{d}u\right) s\mathrm{d}s < +\infty. \tag{7.12}$$

Using Theorem 7.1, we immediately obtain

Corollary 7.1 *The spectral gap of the generator $\mathcal{L}^i = -\nabla_i H^i + \Delta_i$ on $L^2(\mathbb{R}^d, \mu^i)$ satisfies*

$$\lambda_1(\mu^i) \geqslant \frac{1}{c_{\mathrm{Lip},m}}. \tag{7.13}$$

Therefore, we define the uniform upper bound of the spectral gap:

$$\lambda_{1,m} = \inf_{1\leqslant i\leqslant N, \hat{x}^i} \lambda_1(\mu^i) \geqslant \frac{1}{c_{\mathrm{Lip},m}}. \tag{7.14}$$

7.2 Spectral gap of the interacting particle system

To begin with, consider the interacting particle system in a general form,

$$H(x^1, \cdots, x^N) = \sum_{i=1}^N V(x^i) + U(x^1, \cdots, x^N), \tag{7.15}$$

where all interactions are embedded in $U(x)$. The corresponding invariant distribution is

$$\mu(\mathrm{d}x) \propto \exp\big(- H(x)\big)\mathrm{d}x^1 \cdots \mathrm{d}x^N, \tag{7.16}$$

and its conditional invariant distributions are

$$\mu^i(x^i|\hat{x}^i) \propto \exp\big(-V(x^i) - U(x)\big)\mathrm{d}x^i, \qquad i = 1, \cdots, N. \tag{7.17}$$

If there is a uniform lower bound on the spectral gaps of the conditional distributions μ^i, then the spectral gap of the particle system can be obtained.

Theorem 7.2 *Assume* $Z = \int_{\mathbb{R}^{Nd}} e^{-H}\mathrm{d}x$ *and* $Z^i(\hat{x}^i) = \int_{\mathbb{R}^d} e^{-H^i}\mathrm{d}x^i$ *are finite for any* $\hat{x}^i$*. Suppose*

1. *The conditional distributions* μ^i *satisfy the uniform Poincaré inequality, i.e.,*

$$\lambda_{1,m} := \inf_{1\leqslant i\leqslant N, \hat{x}^i} \lambda_1(\mu^i) > 0. \tag{7.18}$$

2. *For some constant* $h \in \mathbb{R}$,

$$\big(1_{i\neq j}\nabla^2_{ij}U\big)_{Nd\times Nd} \geqslant h. \tag{7.19}$$

Then the spectral gap $\lambda_1(\mu) \geqslant \lambda_{1,m} + h$.

Proof The generator of the particle system is $\mathcal{L} = -\nabla H \cdot \nabla + \Delta$, which is a self-adjoint operator on $L^2(\mathbb{R}^{Nd}, \mu)$. Our goal is to establish the (dual) Poincaré inequality

$$(\lambda_{1,m} + h)\int_{\mathbb{R}^{Nd}} |\nabla f|^2 \mathrm{d}\mu \leqslant \int_{\mathbb{R}^{Nd}} (\mathcal{L}f)^2 \mathrm{d}\mu \tag{7.20}$$

for all functions $f : E \to \mathbb{R}$. Using Bakry-Emery's formula,

$$\begin{aligned}
\int_{\mathbb{R}^{Nd}} (\mathcal{L}f)^2\mathrm{d}\mu &= \int_{\mathbb{R}^{Nd}} \Gamma_2(f)\mathrm{d}\mu \\
&= \int_{\mathbb{R}^{Nd}} \Big(|\nabla^2 f|^2 + \nabla^2 H(\nabla f, \nabla f)_{\mathbb{R}^{Nd}}\Big)\mathrm{d}\mu \\
&= \int_{\mathbb{R}^{Nd}} \Big(|\nabla^2 f|^2 + \sum_{i=1}^N \nabla^2 V(x^i)(\nabla_i f, \nabla_i f)_{\mathbb{R}^d} + \nabla^2 U(\nabla f, \nabla f)_{\mathbb{R}^{Nd}}\Big)\mathrm{d}\mu
\end{aligned}$$

Note that

$$|\nabla^2 f|^2 = \sum_{i,j=1}^N |\nabla^2_{ij} f|^2 \geqslant \sum_{i=1}^N |\nabla^2_i f|^2. \tag{7.21}$$

Define $\hat{\mu}^i \in \mathcal{P}(\mathbb{R}^{(N-1)d})$ to be the marginal distribution $\hat{x}^i = (x^1, \cdots, x^{i-1}, x^{i+1}, \cdots, x^N)$ (excluding x^i), then $\mu(x) = \mu^i(x^i|\hat{x}^i)\hat{\mu}^i(\hat{x}^i)$. Now we have the estimate

$$\begin{aligned}
\int_{\mathbb{R}^{Nd}} (\mathcal{L}f)^2\mathrm{d}\mu \geqslant \sum_{i=1}^N \int_{\mathbb{R}^{(N-1)d}} \int_{\mathbb{R}^d} \Big(|\nabla^2_i f|^2 + \nabla^2 V(x^i)(\nabla_i f, \nabla_i f)_{\mathbb{R}^d} + \nabla^2_i j U(\nabla f, \nabla f)_{\mathbb{R}^{Nd}}\Big)\mathrm{d}\mu^i \mathrm{d}\mu(\hat{x}^i) \\
+ \int_{\mathbb{R}^{Nd}} \sum_{i\neq j} \nabla^2_{ij} U(\nabla f, \nabla f)\mathrm{d}\mu.
\end{aligned}$$

Using the uniform Poincaré inequality, one has

$$\int_{\mathbb{R}^d} \Big(|\nabla^2_i f|^2 + \nabla^2 V(x^i)(\nabla_i f, \nabla_i f)_{\mathbb{R}^d} + \nabla^2_i j U(\nabla f, \nabla f)_{\mathbb{R}^{Nd}}\Big)\mathrm{d}\mu^i \geqslant \lambda_{1,m}\int_{\mathbb{R}^d} |\nabla_i f|^2 \mathrm{d}\mu_i. \tag{7.22}$$

Moreover, using the assumption (7.19) one has

$$\int_{\mathbb{R}^{Nd}} \sum_{i\neq j} \nabla^2_{ij} U(\nabla f, \nabla f) \mathrm{d}\mu \geqslant h \int_{\mathbb{R}^{Nd}} |\nabla f|^2 \mathrm{d}\mu. \tag{7.23}$$

Combining (7.22)(7.23) one obtains the desired result. □

Now write U in the form of pairwise interactions W. The main theorem is as follows:

Theorem 7.3 *For the IPS* (6.1), *assume the following conditions:*

(H1) *The external potential $V : \mathbb{R}^d \to \mathbb{R}$ is C^2-smooth, its Hessian $\nabla^2 V$ in $\mathbb{R}^{d\times d}$ is bounded from below, and there exists $c_1, c_2 > 0$ such that*

$$x \cdot \nabla V(x) \geqslant c_1 |x|^2 - c_2, \qquad \forall x \in \mathbb{R}^d. \tag{7.24}$$

(H2) *The pairwise poetntial $W : \mathbb{R}^d \to \mathbb{R}$ is even and C^2-smooth, its Hessian is bounded and*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \exp(-[V(x) + V(y) + \lambda W(x,y)]) \mathrm{d}x\mathrm{d}y < +\infty, \qquad \forall \lambda > 0. \tag{7.25}$$

(H3) *For the function $b_0(r)$ defined by*

$$b_0(r) := \sup_{|x-y|=r,z} -\frac{1}{r}\langle x-y, \nabla V(x) - \nabla V(y) + \nabla W(x-z) - \nabla W(y-z)\rangle, \tag{7.26}$$

the Lipschitzian constant $c_{\mathrm{Lip},m}$ is finite,

$$c_{\mathrm{Lip},m} = \frac{1}{4}\int_0^\infty \exp\left(\frac{1}{4}\int_0^s b_0(u)\mathrm{d}u\right) s\mathrm{d}s < +\infty. \tag{7.27}$$

Then the spectral gaps of the conditional distributions μ^i has a uniform lower bound,

$$\lambda_{1,m} := \inf_{1\leqslant i\leqslant N, \hat{x}^i} \lambda_1(\mu^i) \geqslant \frac{1}{c_{\mathrm{Lip},m}}. \tag{7.28}$$

If there exists a constant $h > -\lambda_{1,m}$ such that

$$\frac{1}{N-1}\big(1_{i\neq j}\nabla^2_{ij} W(x^i - x^j)\big)_{Nd\times Nd} \geqslant h, \qquad \forall x \in \mathbb{R}^{Nd}, \tag{7.29}$$

then the spectral gap of the particle system satisfies

$$\lambda_1(\mu) \geqslant \lambda_{1,m} + h \geqslant \frac{1}{c_{\mathrm{Lip},m}} + h. \tag{7.30}$$

7.3 Explicit estimate of convergence rate

Surprisingly, the spectral gap estimate (Theorem 7.3) also tells the chaos property of the invariant distribution μ.

Corollary 7.2 *Under the conditions of Theorem 7.3, for any two bounded Lipschitz functions f, g on $\mathbb{R}^d$ and $i \neq j$,*

$$\mathrm{Cov}_\mu(f(x^i), g(x^j)) \leqslant \frac{c_{\mathrm{Lip},m}}{(1 + c_{\mathrm{Lip},m}h)(N-1)}(\|f\|_{\mathrm{Lip}}^2 + \|g\|_{\mathrm{Lip}}^2). \tag{7.31}$$

Roughly speaking, two particles x^i and x^j become asymptotically independent at the rate $1/N$.

We derive the explicit estimates of the spectral gap $c_{\mathrm{Lip},m}$.

Corollary 7.3 *Assume there exists constants $c_V, c_W, c_1, c_2 \in \mathbb{R}$ and $R \geqslant 0$ such that $c_V + c_W > 0$*

$$\begin{aligned}\langle \nabla V(x) - \nabla V(y), x - y\rangle &\geqslant c_V|x-y|^2 - c_1|x-y|1_{|x-y|\leqslant R},\\ \langle \nabla W(x-z) - \nabla W(y-z), x - y\rangle &\geqslant c_W|x-y|^2 - c_2|x-y|1_{|x-y|\leqslant R},\end{aligned}$$

then $b_0(r) \leqslant -(c_V + c_W)r + (c_1 + c_2)1_{r\leqslant R}$, hence

$$c_{\mathrm{Lip},m} \leqslant \frac{1}{c_V + c_W}\exp\left(\frac{1}{4}(c_1 + c_2)R\right). \tag{7.32}$$

8 Convergence rate by log-Sobolev inequality

Compared to the Poincaré inequality, the log-Sobolev inequality yields stronger convergence in relative entropy. We establish the log-Sobolev inequality with a constant unirom in the number of particles N in this section. Similar with the Poincaré inequality, we start from the log-Sobolev inquality of the conditional distribution μ^i.

8.1 Zegarlinski's condition

For a given distribution μ, recall that the log-Sobolev inequality is given by

$$\mathrm{Ent}_\mu(f^2) \leqslant 2C\mu(|\nabla f|^2), \tag{8.1}$$

or equivalently,

$$\mu(f^2 \log|f|) \leqslant C\mu|\nabla f|^2 + \mu f^2 \log(\mu f^2)^{\frac{1}{2}}. \tag{8.2}$$

The best constant $\rho_{\mathrm{LS}}(\mu) = 1/C$ satisfying the log-Sobolev inequality is called the log-Sobolev constant. For the IPS, suppose the conditional distrubutions $\mu^i \in \mathcal{P}(\mathbb{R}^d)$ satisfy the uniform log-Sobolev inequality, can we derive the log-Sobolev inequality for the particle system? This is answered by the Zegarlinski's condition.

Definition 8.1 (Zegarlinski) *Given the distribution $\mu \in \mathcal{P}(\mathbb{R}^{Nd})$, let $\mu^i = \mu(x^i|\hat{x}^i) \in \mathcal{P}(\mathbb{R}^d)$ be the conditional distributions of μ. Define c_{ij}^Z to be the best nonnegative constant to satisfy*

$$|\nabla_i(\mu^j(f^2))^{\frac{1}{2}}| \leqslant (\mu^j(|\nabla_i f|^2))^{\frac{1}{2}} + c_{ij}^Z(\mu^j(|\nabla_j f|^2))^{\frac{1}{2}} \tag{8.3}$$

for all smooth functions $f(x^1, \cdots, x^N)$.

It is easy to see $c^Z_{ii} = 0$. The matrix $c^Z := (c^Z_{ij})_{N\times N}$ is called Zegarlinski's matrix. c^Z is uniquely determined by the joint distribution μ. A sufficient condition of the Zegarlinski's condition is

Lemma 8.1 *If for any function $g = g(x^j) \in C^1_b(\mathbb{R}^d)$ on the single particle x^j,*

$$|\nabla_i \mu^j(g)| \leqslant c_{ij}\mu^j(|\nabla g|), \tag{8.4}$$

then $c^Z_{ij} \leqslant c_{ij}$.

The proof is the direct application of the Cauchy inequality. It is an interesting question to ask what the inequality (8.4) tells about the conditional distribution μ^i. As an example, let (x, y) be a joint random variable and consider the following inequality,

$$\frac{\partial}{\partial x}\mathbb{E}[g(y)|x] \leqslant c\mathbb{E}[|\nabla g(y)||x], \quad \forall g \in C^1_b(\mathbb{R}^d), \tag{8.5}$$

what does it tells about the conditional distribution $p(y|x)$? Assume $p(y|x)$ has a density function $\rho(y - \lambda x)$, where $\lambda \in \mathbb{R}$ is a parameter, then the inequality is equivalent to $|\lambda| \leqslant c$. That is to say, the constant c characterizes how the conditional distribution $p(y|x)$ is sensitive to the value of x. The less c is, the less $p(y|x)$ is sensitive to x. If the whole Zegarlinski matrix c^Z is small, then the N particles in the distribution μ become statistically irrelevant. In this sense, the Zegarlinski controls the chaos property of the distribution, which allows us to derive the log-Sobolev inequality of the N-particle system from the one-particle conditional distribution.

The main result of Zegarlinski is as follows (Theorem 0.1, [13]).

Theorem 8.1 (Zegarlinski) *Let μ^i be the conditional distribution of μ in N particles. If*

(1) *μ^i satisfies a uniform log-Sobolev inequality, i.e.,*

$$\rho_{\mathrm{LS},m} = \inf_{1\leqslant i\leqslant N, \hat{x}^i} \rho_{\mathrm{LS}}(\mu^i) > 0. \tag{8.6}$$

(2) *The following Zegarlinski's condition is verified*

$$\gamma := \sup_{1\leqslant i\leqslant N} \max\left\{\sum_{j=1}^N c^Z_{ji}, \sum_{j=1}^N c^Z_{ij}\right\} < 1. \tag{8.7}$$

Then the Gibbs measure μ satisfies the log-Sobolev inequality

$$\rho_{\mathrm{LS},m}(1-\gamma)^2\mathrm{Ent}_\mu(f^2) \leqslant 2\mu(|\nabla f|^2) \tag{8.8}$$

for all smooth bounded functions f on $\mathbb{R}^{Nd}$, i.e.,

$$\rho_{\mathrm{LS}}(\mu) \geqslant \rho_{\mathrm{LS},m}(1-\gamma)^2. \tag{8.9}$$

The proof is elementary algebra but a bit tediuous due to repeated usage of Cauchy inequalities. We only breifly describe the proof here.

Proof Using the uniform log-Sobolev inequality repeatedly, we only need to prove

$$\sum_{k=0}^{N-1} \mu|\nabla_{k+1}(\mathbb{E}^{1:k} f^2)^{\frac{1}{2}}|^2 \leqslant (1-\gamma)^{-2}\mu|\nabla f|^2. \tag{8.10}$$

Using the Zegarlinski's condition repeatedly, one has

$$|\nabla_{k+1}(\mathbb{E}^{1:k} f^2)^{\frac{1}{2}}| \leqslant \sum_{i=1}^{N} \lambda_{k+1,i}^{(k+1)} (\mathbb{E}^{1:k}|\nabla_i f|^2)^{\frac{1}{2}}, \tag{8.11}$$

where $\lambda_{j,i}^{(k+1)}$ are defined by

$$0 \leqslant \lambda_{j,k}^{(k+1)} \leqslant \delta_{j,i} + \sum_{n=1}^{k} (c^Z)_{j,i}^n. \tag{8.12}$$

The Cauchy's inequality gives

$$\mu|\nabla_{k+1}(\mathbb{E}^{1:k} f^2)^{\frac{1}{2}}|^2 \leqslant \sum_{i=1}^{N} \lambda_{k+1,i}^{(k+1)} \sum_{i=1}^{N} \lambda_{k+1,i}^{(k+1)} \mu|\nabla_i f|^2 \leqslant (1-\gamma)^{-1} \sum_{i=1}^{N} \lambda_{k+1,i}^{(k+1)} \mu|\nabla_i f|^2. \tag{8.13}$$

Therefore,

$$\begin{aligned}\sum_{k=0}^{n-1} \mu|\nabla_{k+1}(\mathbb{E}^{1:k} f^2)^{\frac{1}{2}}|^2 &\leqslant (1-\gamma)^{-1} \sum_{i=1}^{N} \Big(\sum_{k=0}^{N-1} \lambda_{k+1,i}^{(k+1)} \Big) \mu|\nabla_i f|^2 \\ &\leqslant (1-\gamma)^{-2} \sum_{i=1}^{N} \mu|\nabla_i f|^2,\end{aligned}$$

yielding the desired result. □

Lemma 8.2 *For the Gibbs distribution μ, the Zegarlinski coefficient c_{ji}^Z satisfies*

$$c_{ji}^Z \leqslant \frac{1}{N-1} c_{\mathrm{Lip},m} \|\nabla^2 W\|_\infty. \tag{8.14}$$

Proof We briefly descibe the proof. By direct calculation,

$$\begin{aligned}\nabla_j \mu^i(g) &= \mathrm{Cov}_{\mu^i}(g, -\nabla_j H) \\ &= \mathrm{Cov}_{\mu^i}(g, -\frac{1}{N-1}(\nabla W)(x^i - x^j)) \\ &= -\frac{1}{N-1}\langle(-\mathcal{L}^i g), (-\mathcal{L}^i g)^{-1}\Big((\nabla W)(\cdot - x^j) - \mu^i\big((\nabla W)(\cdot - x^j)\big)\Big)\rangle_{\mu^i} \\ &= -\frac{1}{N-1}\int \nabla_i g \cdot \nabla_i(-\mathcal{L}^i g)^{-1}\Big((\nabla W)(\cdot - x^j) - \mu^i\big((\nabla W)(\cdot - x^j)\big)\Big)\mathrm{d}\mu^i.\end{aligned}$$

Since the operator $(\mathcal{L}^i)^{-1}$ has an upper bound $c_{\mathrm{Lip},m}$, we have

$$\|\nabla_i(-\mathcal{L}^i g)^{-1}\Big((\nabla W)(\cdot - x^j) - \mu^i\big((\nabla W)(\cdot - x^j)\big)\Big)\|_\infty \leqslant c_{\mathrm{Lip},m}\|\nabla^2 W\|_\infty. \tag{8.15}$$

□

8.2 Log-Sobolev inequality of the interacting particle system

Now we state the main theorem, which establishes the log-Sobolev inequality for the interacting particle system.

Theorem 8.2 *Assume that*

(1) *For some best constant $\rho_{\mathrm{LS},m} > 0$, the conditional marginal distributions $\mu^i = \mu(x^i|\hat{x}^i)$ on $\mathbb{R}^d$ satisfy the log-Sobolev inequality*

$$\rho_{\mathrm{LS},m}\,\mathrm{Ent}_{\mu^i}(f^2) \leqslant 2\int_{\mathbb{R}^d} |\nabla f|^2 \mathrm{d}\mu^i, \qquad \forall f \in C_b^1(\mathbb{R}^d) \tag{8.16}$$

for all i and $\hat{x}^i$.

(2) *Zegarlinski's condition*

$$\gamma := \sup_{1\leqslant i\leqslant N} \max\left\{\sum_{j=1}^N c_{ji}^Z, \sum_{j=1}^N c_{ij}^Z\right\} < 1. \tag{8.17}$$

Then μ satisfies

$$\rho_{\mathrm{LS},m}(1-\gamma)^2 \mathrm{Ent}_\mu(f^2) \leqslant 2\int_{\mathbb{R}^{Nd}} |\nabla f|^2 \mathrm{d}\mu, \qquad \forall f \in C_b^1(\mathbb{R}^d), \tag{8.18}$$

i.e., the log-Sobolev constant of μ satisfies

$$\rho_{\mathrm{LS}}(\mu) \geqslant \rho_{\mathrm{LS},m}(1-\gamma)^2. \tag{8.19}$$

A sufficient condition of uniform log-Sobolev inequality is as follows. Assume $\nabla^2 W(x) \geqslant -K_0$ and

$$\nabla^2 V(x) \geqslant K, \qquad |x| \geqslant R. \tag{8.20}$$

Write $V = V_c + V_b$, where V_c is strongly convex and V_b is bounded. By Bakry-Emery's Γ_2 criterion,

$$\exp\bigg(-V_c(x^i) - \frac{1}{N-1}\sum_{j\neq i} W(x^i - x^j)\bigg) \tag{8.21}$$

satisfies a log-Sobolev inequality with constant $K_1 > 0$. The log-Sobolev inequality holds for μ^i due to bounded perturbation.

8.3 Ergodicity of McKean-Vlasov process

We have proved that the IPS (6.1) has a uniform log-Sobolev constant

$$\rho_{\mathrm{LS}} \geqslant \rho_{\mathrm{LS},m}(1 - c_{\mathrm{Lip},m}\|\nabla^2 W\|_\infty)^2. \tag{8.22}$$

Consequently, one has the log-Sobolev inequality

$$\rho_{\mathrm{LS}}\mathrm{Ent}_\mu(f) \leqslant 2\int_{\mathbb{R}^d} |\nabla\sqrt{f}|^2 \mathrm{d}\mu, \tag{8.23}$$

for all positive functions f. Assume $\mu(f) = 1$ and define ν by $\nu(\mathrm{d}x) = f(x)\mu(\mathrm{d}x)$, then the relative entropy and the Fisher information of ν are

$$H(\nu|\mu) = \mathrm{Ent}_\mu(f) = \int f \log f \mathrm{d}\mu, \quad I(\nu|\mu) = \mathcal{E}(\sqrt{f}) = \int \frac{|\nabla f|^2}{4f}\mathrm{d}\mu = \frac{1}{4}\int |\nabla \log f|^2 \mathrm{d}\nu. \tag{8.24}$$

Then the log-Sobolev inequality can be equivalently written as

$$\rho_{\mathrm{LS}} H(\nu|\mu) \leqslant 2I(\nu|\mu), \qquad \forall \nu \ll \mu. \tag{8.25}$$

As $N \to \infty$, we expect the inequality has a mean-field limit, which tells the ergodicity of the McKean-Vlasov process. To be clear, let $\nu \in \mathcal{P}(\mathbb{R}^d)$ be a distribution in $\mathbb{R}^d$, we want to know the mean-field limit of $H(\nu^{\otimes N}|\mu)$ and $I(\nu^{\otimes N}|\mu)$.

Mean-field limit of $H(\nu^{\otimes N}|\mu)$ Let $\alpha(\mathrm{d}x) \propto e^{-V(x)}\mathrm{d}x$ be the reference distribution on $\mathbb{R}^d$. Define the free energy of $\nu \in \mathcal{P}(\mathbb{R}^d)$ by

$$E_f(\nu) = H(\nu|\alpha) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\nu(\mathrm{d}x)\nu(\mathrm{d}y) \tag{8.26}$$

$$= H(\nu) + \int_{\mathbb{R}^d} V(x)\nu(\mathrm{d}x) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\nu(\mathrm{d}x)\nu(\mathrm{d}y) + \mathrm{const.} \tag{8.27}$$

Moreover, define the mean-field entropy by

$$H_W(\nu) = E_f(\nu) - \inf_{\bar{\nu}} E_f(\bar{\nu}). \tag{8.28}$$

then we have the following:

Lemma 8.3 *If $\nu \in \mathcal{P}(\mathbb{R}^d)$ satisfies $H(\nu|\alpha) < +\infty$, then*

$$\frac{1}{N}H(\nu^{\otimes N}|\mu) \to H_W(\nu). \tag{8.29}$$

Proof Define

$$\tilde{Z}_N := \int \exp\Big(-\frac{1}{2(N-1)}\sum_{i\neq j} W(x^i - x^j)\Big)\mathrm{d}\alpha^{\otimes N}, \tag{8.30}$$

then

$$\mathrm{d}\mu = \frac{1}{\tilde{Z}_N}\exp\Big(-\frac{1}{2(N-1)}\sum_{i\neq j} W(x^i - x^j)\Big)\mathrm{d}\alpha^{\otimes N}. \tag{8.31}$$

By direct calculation,

$$\frac{1}{N}H(\nu^{\otimes N}|\mu) = E_f(\nu) + \frac{1}{N}\log \tilde{Z}_N. \tag{8.32}$$

As $N \to \infty$, (3.30) in [14] gives

$$\lim_{N\to\infty}\frac{1}{N}\log \tilde{Z}_N = -\inf_{\bar{\nu}} E_f(\bar{\nu}), \tag{8.33}$$

yielding the desired result. □

The definition of the mean-field entropy $H_W(\nu)$ first arises in [6]. The rigorous proof of (8.33) requires the large deviation theory. In the viewpoint of propagation of chaos, (8.33) can be understood as follows: if ν is the invariant distribution of the MVP, then uniform-in-time propagation of chaos implies

$$\lim_{N\to\infty}\frac{1}{N}H(\nu^{\otimes N}|\mu)=0. \tag{8.34}$$

Therefore, the infimum of the RHS of (8.33) is 0.

Mean-field limit of $I(\nu^{\otimes N}|\mu)$ Define the mean-field Fisher information by

$$I_W(\nu):=\int_{\mathbb{R}^d}\left|\frac{\nabla f(x)}{f(x)}+\nabla V(x)+(\nabla W*\nu)(x)\right|^2\mathrm{d}\nu(x), \tag{8.35}$$

where $f=\mathrm{d}\nu/\mathrm{d}\mu$. We have the following result:

Lemma 8.4 *If $\nu\in\mathcal{P}(\mathbb{R}^d)$ satisfies $I(\nu|\alpha)<+\infty$, then*

$$\frac{1}{N}I(\nu^{\otimes N}|\mu)\to I_W(\nu). \tag{8.36}$$

Proof By direct calculation,

$$\frac{1}{N}I(\nu^{\otimes N}|\mu)=\int\frac{1}{4}\left|\nabla\log\frac{\mathrm{d}\nu}{\mathrm{d}\alpha}(x^1)+\frac{1}{N-1}\sum_{j=2}^N\nabla W(x^1-x^j)\right|^2\mathrm{d}\nu^{\otimes N} \tag{8.37}$$

The law of large numbers implies

$$\frac{1}{N}I(\nu^{\otimes N}|\mu)\to\int\frac{1}{4}\left|\nabla\log\frac{\mathrm{d}\nu}{\mathrm{d}\alpha}(x^1)+\int\nabla W(x^1-y)\mathrm{d}\nu(y)\right|^2\mathrm{d}\nu(x^1)=I_W(\nu). \tag{8.38}$$

□

Now we state the main theorem.

Theorem 8.3 *Assume the uniform marginal log-Sobolev inequality, i.e., $\rho_{\mathrm{LS},m}>0$. Then*

(1) *There exists a unique minimizer ν_∞ of $H_W(\nu)$ over $\mathcal{M}_1(\mathbb{R}^d)$;*

(2) *The following (nonlinear) log-Sobolev inequality*

$$\rho_{\mathrm{LS}}H_W(\nu)\leqslant 2I_W(\nu),\qquad \nu\in\mathcal{M}_1(\mathbb{R}^d) \tag{8.39}$$

holds, where

$$\rho_{\mathrm{LS}}=\varlimsup_{N\to\infty}\rho(\mu)\geqslant\rho_{\mathrm{LS},m}(1-\gamma)^2. \tag{8.40}$$

(3) *The following Talagrand's transportation inequality holds*

$$\rho_{\mathrm{LS}}W_2^2(\nu,\nu_\infty)\leqslant 2H_W(\nu),\qquad \nu\in\mathcal{M}_1(\mathbb{R}^d). \tag{8.41}$$

(4) *For the solution of the McKean-Vlasov equation with the given initial distribution ν_0 of finite second moment,*

$$H_W(\nu_t) \leqslant e^{-t\rho_{\mathrm{LS}}/2} H_W(\nu_0). \tag{8.42}$$

In particular,

$$W_2(\nu_t, \nu_\infty) \leqslant \frac{2}{\rho_{\mathrm{LS}}} e^{-t\rho_{\mathrm{LS}}/2} H_W(\nu_0). \tag{8.43}$$

Proof The log-Sobolev inequality gives

$$\rho_{\mathrm{LS}} H(\nu^{\otimes N}|\mu) \leqslant 2I(\nu^{\otimes N}|\mu). \tag{8.44}$$

As $N \to \infty$, one obtains

$$\rho_{\mathrm{LS}} H_W(\nu) \leqslant 2I_W(\nu). \tag{8.45}$$

Using

$$\frac{\mathrm{d}}{\mathrm{d}t} H_W(\nu_t) = -4I_W(\nu), \tag{8.46}$$

one obtains the exponential convergence of ν_t. □

A Subjective commentary

This paper provides an alternative approach to establish the ergodicity of the IPS and the MVP, connecting the geometric ergodicity and the propagation of chaos in the non-convex case. The idea of proving the ergodicity of the IPS from uniform functional inequalities for conditional distributions is very inspiring. Open question: is it possible to extend the conditional distribution approach to more general interacting systems? Or systems with singular potentials?

References

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

[2] Joe Neeman. Notes on markov diffusions. *Probability and Stochastic Analysis*, 2016.

[3] Arnaud Guillin, Wei Liu, Liming Wu, and Chaoen Zhang. Uniform poincar {\'e} and logarithmic sobolev inequalities for mean field particles systems. *arXiv preprint arXiv:1909.07051*, 2019.

[4] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. *arXiv preprint arXiv:2106.14812*, 2021.

[5] Florent Malrieu. Uniform propagation of chaos for mckean-vlasov equations. `https://www.ceremade.dauphine.fr/~mischler/expo/FlorentMalrieu`. Accessed: 2022-04-17.

[6] José A Carrillo, Robert J McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018, 2003.

[7] François Bolley, Ivan Gentil, and Arnaud Guillin. Uniform convergence to equilibrium for granular media. *Archive for Rational Mechanics and Analysis*, 208(2):429–445, 2013.

[8] Andreas Eberle. Reflection coupling and wasserstein contractivity without convexity. *Comptes Rendus Mathematique*, 349(19-20):1101–1104, 2011.

[9] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative harris-type theorems for diffusions and mckean–vlasov processes. *Transactions of the American Mathematical Society*, 371(10):7135–7173, 2019.

[10] Shi Jin, Lei Li, Xuda Ye, and Zhennan Zhou. Ergodicity and long-time behavior of the random batch method for interacting particle systems. *arXiv preprint arXiv:2202.04952*, 2022.

[11] Alain Durmus, Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. An elementary approach to uniform in time propagation of chaos. *Proceedings of the American Mathematical Society*, 148(12):5387–5398, 2020.

[12] Liming Wu. Gradient estimates of poisson equations on riemannian manifolds and applications. *Journal of Functional Analysis*, 257(12):4015–4033, 2009.

[13] Boguslaw Zegarlinski. Dobrushin uniqueness theorem and logarithmic sobolev inequalities. *Journal of functional analysis*, 105(1):77–111, 1992.

[14] Wei Liu and Liming Wu. Large deviations for empirical measures of mean-field gibbs measures. *Stochastic Processes and their Applications*, 130(2):503–520, 2020.