MEAN FIELD PARTICLE SYSTEMS AND THEIR LIMITS TO NONLOCAL PD'S

Li Chen

January 7, 2024

Contents

Abstract

This lecture aims to give an introduction on the mean field derivation of a family of non-local partial differential equations with and without diffusion.

Chapter 1

MODEL DESCRIPTION AND INTRODUCTION

In this chapter we will outline the setting of relevant particle models for both first and second order systems. The picture of mean-field limit problem will be illustrated in detail only for the deterministic first order system. We leave a formulation of deterministic second order system as exercises and postpone the stochastic case after the review of stochastic calculus. We consider a system of N particles and denote by $x_1(t), x_2(t), \dots, x_N(t) \in \mathcal{C}^1([0,T];\mathbb{R}^d)$, $i = 1, \dots, N$, the trajectories of the particles.

The first order system is then governed by the system of ordinary differential equations

$$\begin{cases} dx_i &= \frac{1}{N}\sum_{j=1}^N K(x_i, x_j)dt + \sigma dW_i(t), \quad 1 \le i \le N \\ x_i|_{t=0} &= x_i(0) \end{cases}.$$

where $K : \mathbb{R}^{2d} \to \mathbb{R}^d$ is a given function and $W_i(t)$ are i.i.d. Brownian motions. We will show more details in the case $\sigma > 0$ in later chapters. For the moment, we take $\sigma = 0$, which corresponds to deterministic case.

The given function K is called the pair interaction force.

Example. One example for a well-behaved K is given by the quadratic potential, i.e.

$$K(x, y) = \nabla(|x - y|^2),$$

which is a Lipschitz continuous function.

Another typical interaction force which is not continuous is gradient of the Coulomb potential (or the fundamental solution of Poisson equation), namely

$$K(x, y) = \nabla \frac{1}{|x - y|^{d-2}} = C(d)\frac{x - y}{|x - y|^d}.$$

Definition 1.0.1 (Empirical Measure)**.** For a set of particles $x_1, x_2, \dots, x_N \in \mathbb{R}^d$, the empirical measure is defined by

$$\mu^N \triangleq \frac{1}{N}\sum_{j=1}^N \delta_{x_i},$$

where δ_y is the point measure. Namely, for all measurable sets $E \subset \mathbb{R}^d$, it holds that $\delta_y(E) = 1$ if $y \in E$ and $\delta_y(E) = 0$ if $y \notin E$.

An appropriate quantity to study the limit $N \to \infty$ is the empirical measure given in 1.0.1. If the initial empirical measure converges " in some sense " to a measure $\mu(0)$ i.e.

$$\mu^N(0) \to \mu(0).$$

would $\mu^N(t)$ also converge to some measure $\mu(t)$ in the same sense? Furthermore, what is the time evolution of $\mu(t)$? How does the microscopic structure being kept?

Note. Consider the following case when the limit measure $\mu(t)$ is absolutely continuous with respect the Lebesgue measure, this means that

$$d\mu(0,x) = \rho_0(x)dx \quad \rho_0 \in L^1(\mathbb{R}^d).$$

would the limit function have the same property ?

The following Proposition serves as a motivation on which partial differential equation $\mu(t)$ should satisfy.

Proposition 1.0.1. The empirical measure $\mu^N(t)$ solves the following partial differential equation (in the sense of distribution)

$$\partial_t \mu(t,x) + \nabla \cdot \left(\mu(t,x) \int_{\mathbb{R}^d} K(x,y) d\mu(t,y) \right) = 0.$$

Namely, for any $\varphi \in C_0^\infty(\mathbb{R}^d)$, it holds

$$\langle \mu^N(t), \varphi \rangle = \langle \mu^N(0), \varphi \rangle + \int_0^t \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \int_{\mathbb{R}^d} K(x,y) d\mu^N(s,y) d\mu^N(s,x) ds$$

Proof. Take $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ and by the definition of emperical measure Definition 1.0.1 we have

$$
\begin{aligned}
\frac{d}{dt}\langle \mu^N(t), \varphi\rangle &\triangleq \frac{d}{dt}\int_{\mathbb{R}^d} \varphi(x) d\mu^N(t,x) \\
&\overset{\text{Def.}}{=} \frac{d}{dt}\int_{\mathbb{R}^d} \frac{1}{N}\sum_{i=1}^N \varphi(x) d\delta_{x_i(t)} \\
&\overset{\text{Lin.}}{=} \frac{1}{N}\sum_{i=1}^N \frac{d}{dt}\varphi(x_i(t)) \\
&= \frac{1}{N}\sum_{i=1}^N \nabla\varphi(x_i(t)) \cdot \frac{d}{dt}x_i(t) \\
&= \frac{1}{N}\sum_{i=1}^N \nabla\varphi(x_i(t)) \cdot \frac{1}{N}\sum_{j=1}^N K(x_i(t), x_j(x)) \\
&= \frac{1}{N}\sum_{i=1}^N \nabla\varphi(x_i(t)) \cdot \frac{1}{N}\sum_{j=1}^N \int_{\mathbb{R}^d} K(x_i(t), y) d\delta_{x_j(t)}(y) \\
&\overset{\text{Emp.}}{=} \frac{1}{N}\sum_{i=1}^N \nabla\varphi(x_i(t)) \cdot \int_{\mathbb{R}^d} K(x_i(t), y) d\mu^N(t,y) \\
&= \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \int_{\mathbb{R}^d} K(x,y) d\mu^N(t,y) d\delta_{x_i(t)}(x) \\
&= \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \int_{\mathbb{R}^d} K(x,y) d\mu^N(t,y) d\mu^N(t,x) \\
&\overset{\text{w. derivative}}{=} -\left\langle \nabla \cdot \left(\mu^N(t,\cdot)\int_{\mathbb{R}^d} K(\cdot, y) d\mu^N(t,y)\right), \varphi\right\rangle.
\end{aligned}
$$

The above discussion shows that μ^N is a weak solution of

$$
\partial_t \mu^N(t,x) + \nabla \cdot \left(\mu^N(t,x)\int_{\mathbb{R}^d} K(\cdot, y) d\mu^N(t,y)\right) = 0.
$$

□

Note. If the limit of μ^N as $N \to \infty$ in some sense exists, then μ should also satisfy the proposed PDE.

In the case that $\sigma > 0$, or in other words the stochastic system is considered, then we expect the limit partial differential equation to share a similar structure

$$
\partial_t \mu(t,x) + \nabla \cdot \left(\mu(t,x)\int_{\mathbb{R}^d} K(\cdot, y) d\mu(t,y)\right) = \Delta\mu(t,x).
$$

More details in this case will be described after we review the stochastic calculus in the next chapters.

The above description for mean field limit problem also works for the so called second order particle system, which is usually given in the following formulation:

Let $((x_1(t), v_1(t)), \ldots, (x_N(t), v_N(t))) \in \mathcal{C}^1([0,T]; \mathbb{R}^{2d})$ be the positions and velocities of N particles, with given initial data $(x_i(0), v_i(0))$ for $i = 1, \ldots, N$

The dynamical system for these particles are given by, according to the Newton's second law,

$$
(\text{MPS}) \begin{cases} \frac{d}{dt}x_i(t) &= v_i(t) \\ \frac{d}{dt}v_i(t) &= \frac{1}{N}\sum_{j=1}^N F(x_i(t), v_i(t); x_j(t), v_j(t)) \quad 1 \le i \le N \end{cases},
$$

where the interaction force F is given. An example for F is given through the gravitation potential, namely

$$F(x, v; y, u) = \frac{x - y}{|x - y|^d}.$$

The empirical measure from Definition 1.0.1 can be rewritten to include the velocity as well

$$\mu^N \triangleq \frac{1}{N} \sum_{j=1}^{N} \delta_{x_i(t), v_i(t)}.$$

Exercise. Try to find out the partial differential equation that μ^N should satisfy in the sense of distribution. Hint: Calculate for $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^{2d})$ the following time derivative.

$$\frac{d}{dt} \langle \mu^N(t), \varphi \rangle.$$

Arrangement of the lecture In Chapter 2, we are going to discuss the deterministic case for bounded Lipschitz continuous interaction forces. A brief review of the well-posedness theory of ordinary differential equation is given as a first step to warm up. Then we prove the mean field limit of this problem in the framework of Wasserstein-1 distance.

The stochastic case with bounded Lipschitz continuous interaction force will be studied in Chapter 3. As a preparation, we review the mandatory concepts of probability theory, the definition of the Itô integral, and the well-posedness of stochastic differential equations. Based on that, the solvability of McKean-Vlasov equation is given, and consequently, the mean field limit and propagagtion of chaos result for the stochastic system is proved in the Wasserstein-2 distance.

The alternative approach through the partial differential equation theory, to solve the McKean-Vlasov equation, is given in Chapter 4. In this chapter, we give some of the solution theory of second order partial differential equations without proofs, which is going to be presented in an independent lecture.

In Chapter 5, we briefly explain an idea to handle problems with singular interaction forces. Namely, we start with a smoothed particle system with an N dependent scalling parameter both with logarithmic scalling and algebraic scalling. Then present the convergence results for the particle processes in expectation or in the sense of probability.

The relative entropy method is presented in Chapter 6, with which one can obtain the strong L^1 convergence of the first marginal density of the N particle distribution. In the end, we show the idea of applying the convergence in probability within the relative entropy framework and obtain strong L^1 convergence in the propagation of chaoes.

Chapter 2

MEAN-FIELD LIMIT IN THE DETERMINISTIC SETTING

In this chapter we focus on the deterministic version of the mean-field limit. Namely, we start from a system of deterministic interacting particle system with mean-field structure and prove that the corresponding empirical measure converges weakly to the measure valued solution of the corresponding partial differential equation.

We are going to work only with the first order system formulation, in deterministic case the second order system can be rewritten into a first order system, for which the whole argument remains the same.

Recall a system of N particles and denote by $x_1(t), x_2(t), \dots, x_N(t) \in \mathcal{C}^1([0,T];\mathbb{R}^d)$, $i = 1,\dots,N$ the trajectories of the particles.

The so-called first order system is then governed by the system of ordinary differential equations

$$\begin{cases} dx_i(t) & = \frac{1}{N}\sum_{j=1}^{N} K(x_i, x_j)dt \quad 1 \le i \le N \\ x_i(t)|_{t=0} & = x_i(0) \in \mathbb{R}^d \end{cases}. \tag{2.1}$$

where $K : \mathbb{R}^{2d} \to \mathbb{R}^d$ is a given function.

In the case of higher dimensional vectors we sometimes use the following notation

$$X_N(t) = \left(x_1(t), x_2(t), \dots, x_N(t)\right)^T \in \mathbb{R}^{dN}.$$

2.1 Review Of ODE Theory

We review the well-posedness theory of ODE system with a general setting.

For $\forall T > 0$ and $f \;\; [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, we consider the following initial value problem

$$(\text{IVP}) \begin{cases} \frac{d}{dt}x(t) & = f(t,x) \quad t \in [0,T] \\ x|_{t=0} & = x_0 \in \mathbb{R}^d \end{cases}. \tag{2.2}$$

Assumption A. $f \in \mathcal{C}([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ and f is Lipschitz continuous in x, which means there $\exists L > 0$ such that $\forall (t,x), (t,y) \in [0,T] \times \mathbb{R}^d$

$$|f(t,x) - f(t,y)| \le L|x - y|.$$

Theorem 2.1.1 (Existence and uniqueness of IVP)**.** If Assumption A holds then the Equation 2.2 has a unique solution $x \in \mathcal{C}^1([0,T];\mathbb{R}^d)$.

Proof. We use Picard iteration to prove the existence, we can define the equivalent way of solving the Equation 2.2 by considering the integral equation

$$x(t) - x_0 = \int_0^t f(s, x(s))ds \quad \forall t \in [0, T].$$

Then our Picard iteration is given by the following

$$\begin{aligned} x_1(t) &= x_0 + \int_0^t f(s, x_0)ds \\ x_2(t) &= x_0 + \int_0^t f(s, x_1(s))ds \\ &\vdots \\ x_m(t) &= x_0 + \int_0^t f(s, x_{m-1}(s))ds. \end{aligned}$$

By Assumption A and properties of integration we have $x_m(t) \in \mathcal{C}^1([0, T]; \mathbb{R}^d)$.

Due to completeness of $\mathcal{C}([0, T]; \mathbb{R}^d)$ we only need to show that $(x_m(t))_{m\in\mathbb{N}}$ is a Cauchy sequence to get the existence. We first prove by induction that for $m \geq 2$ it holds for some constant M that

$$|x_m(t) - x_{m-1}(t)| \leq \frac{ML^{m-1}|t|^m}{m!}.$$

IA For $m = 1$ it holds

$$\begin{aligned} |x_2(t) - x_1(t)| &\overset{\text{Tri.}}{\leq} \int_0^t |f(s, x_1(s)) - f(s, x_0)|ds \\ &\leq L \int_0^t |x_1(s_0) - x_0| ds_0 \\ &\leq L \int_0^t \int_0^{s_0} |f(s_1, x_0)| ds_1 ds_0 \\ &\leq ML \int_0^t (s_0 - 0) ds_0 \\ &= \frac{MLt^2}{2}. \end{aligned}$$

where we chose $M \geq \max_{s\in[0,T]} |f(s, x_0)|$
IV Suppose for $m \in \mathbb{N}$ it holds

$$|x_m(t) - x_{m-1}(t)| \leq \frac{ML^{m-1}|t|^m}{m!}.$$

IS $m \to m + 1$

$$\begin{aligned} |x_{m+1}(t) - x_m(t)| &= \left| \int_0^t f(s, x_m(s)) - f(s, x_{m-1}(s))ds \right| \\ &\overset{\text{Tri.}}{\leq} \int_0^t |f(s, x_m(s)) - f(s, x_{m-1})(s)|ds \\ &\leq L \int_0^t |x_m(s) - x_{m-1}(s)|ds \\ &\overset{\text{IV}}{\leq} L \int_0^t \frac{ML^{m-1}|s|^m}{m!} ds \\ &= \frac{ML^m|t|^{m+1}}{(m+1)!}. \end{aligned}$$

Now take arbitrary $p, m \in \mathbb{N}$ then by triangle inequality we obtain for $\forall t \in [0, T]$ that

$$\begin{aligned}
|x_{m+p} - x_m(t)| &\leq \sum_{k=m+1}^{m+p} |x_k(t) - x_{k-1}(t)| \\
&\leq \sum_{k=m+1}^{m+p} M \frac{L^{k-1}T^k}{k!} = \frac{M}{L} \sum_{k=m+1}^{m+p} \frac{(LT)^k}{k!} \\
&\leq \frac{M}{L} \frac{(LT)^{m+1}}{(m+1)!} \sum_{k=0}^{p-1} \frac{(LT)^k}{k!} \leq \frac{M}{L} \frac{(LT)^{m+1}}{(m+1)!} \sum_{k=0}^{\infty} \frac{(LT)^k}{k!} \\
&= \frac{M}{L} \frac{(LT)^{m+1}}{(m+1)!} e^{LT} \xrightarrow{m\to\infty} 0 \text{ uniformly in } t \in [0, T].
\end{aligned}$$

Therefore $x_m(t)$ has a limit $x(t) \in \mathcal{C}([0, T]; \mathbb{R}^d)$ with

$$\max_{t\in[0,T]} |x_m(t) - x(t)| \xrightarrow{m\to\infty} 0.$$

Then by taking $m \to \infty$ in

$$x_m(t) = x_0 + \int_{t_0}^{t} f(s, x_{m-1}(s))ds.$$

we get

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds.$$

which means $x(t) \in \mathcal{C}([0, T]; \mathbb{R}^d)$ is a solution of the equivalent integral equation. Furthermore because of the continuity of f, we also obtain that $x(t) \in \mathcal{C}^1([0, T]; \mathbb{R}^d)$

To prove the uniqueness suppose we have two solutions $x(t), \tilde{x}(t) \in \mathcal{C}([0, T]; \mathbb{R}^d)$ then they satisfy

$$\begin{aligned}
x(t) &= x_0 + \int_{t_0}^{t} f(s, x(s))ds \\
\tilde{x}(t) &= x_0 + \int_{t_0}^{t} f(s, \tilde{x}(s))ds.
\end{aligned}$$

By taking the difference of these two solutions and using the Lipschitz continuity of f in x we obtain

$$\begin{aligned}
|x(t) - \tilde{x}(t)| &\leq \int_0^t |f(s, x(s)) - f(s, \tilde{x(s)})|ds + |x_0 - x_0| \\
&\leq L \int_{)}^{t} |x(s) - \tilde{x}(s)|ds \\
&\leq L \int_0^t e^{-\alpha s} |x(s) - \tilde{x}(s)| e^{\alpha s} ds.
\end{aligned}$$

For any $\alpha > 0$. By considering the quantity $P(t) = e^{-\alpha t}|x(t) - \tilde{x}(t)|$, we obtain

$$\begin{aligned}
|x(t) - \tilde{x}(t)| &\leq L \int_0^t \max_{0\leq s\leq t} \{e^{-\alpha s}|x(s) - \tilde{x}(s)|\} e^{\alpha s} ds \\
&\leq L \max_{0\leq s\leq t} \{e^{-\alpha s}|x(s) - \tilde{x}(s)|\} \int_0^t e^{\alpha s} ds.
\end{aligned}$$

This implies that

$$P(t) = e^{-\alpha t}|x(t) - \tilde{x}(t)| \leq \max_{t\in[0,T]} P(t) \leq \frac{L}{\alpha} \max_{t\in[0,T]} P(t) \quad \forall t \in [0, T].$$

By choosing $\alpha = 2L$ we have

$$\max_{t\in[0,T]} e^{-2Lt}|x(t) - \tilde{x}(t)| = 0.$$

i.e

$$x(t) = \tilde{x}(t) \quad \forall t \in [0, T].$$

This concludes the uniqueness proof □

Remark. An alternative proof for uniqueness uses Gronwall's inequality which we will give in the next lemma. Furthermore similar to the uniqueness proof, one can obtain that the solution $x(t; t_0, x_0)$ is continuously dependent on initial data

Lemma 2.1.1 (Gronwall's inequality)**.** Let $\alpha, \beta, \varphi \in \mathcal{C}([a, b]; \mathbb{R}^d)$ and $\beta(t) \geq 0$ for $\forall t \in [a, b]$ such that

$$0 \leq \varphi(t) \leq \alpha(t) + \int_a^t \beta(s)\varphi(s)ds \quad \forall t \in [a, b].$$

then

$$\varphi(t) \leq \alpha(t) + \int_a^t \beta(s)e^{\int_s^t \beta(\tau)d\tau}\alpha(s)ds \quad \forall t \in [a, b].$$

Specially if $\alpha(t) \equiv M$ then we have

$$\varphi(t) \leq Me^{\int_a^b \beta(\tau)d\tau} \quad \forall t \in [a, b].$$

Proof. Define

$$\psi(t) = \int_a^t \beta(\tau)\varphi(\tau)d\tau \quad \forall t \in [a, b].$$

because of the continuity of β and φ we get that ψ is differentiable on $[a, b]$ and

$$\psi'(t) = \beta(t)\varphi(t).$$

Since $\beta(t) \geq$ we have

$$\psi'(t) = \beta(t)\varphi(t) \leq \beta(t)(\alpha(t) + \psi(t)) \quad \forall t \in [a, b].$$

Then by multiplying both sides with $e^{-\int_a^t \beta(\tau)d\tau}$ we obtain

$$\begin{aligned}\frac{d}{dt}(e^{-\int_a^t \beta(\tau)d\tau}\psi(t)) &= e^{-\int_a^t \beta(\tau)d\tau}(\psi'(t) - \beta(t)\psi(t)) \\ &\leq \beta(t)\alpha(t)e^{-\int_a^t \beta(\tau)d\tau}.\end{aligned}$$

Integrate the above inequality from a to t to get

$$e^{-\int_a^t \beta(\tau)d\tau}\psi(t) - e^{-\int_a^t \beta(\tau)d\tau}\psi(a) \leq \int_a^t \beta(s)\alpha(s)e^{-\int_a^s \beta(\tau)d\tau}ds,$$

which implies

$$\psi(t) \leq \int_a^t \beta(s)\alpha(s)e^{\int_s^t \beta(\tau)d\tau} ds.$$

and

$$\varphi(t) \leq \alpha(t) + \psi(t) \leq \alpha(t) + \int_a^t \beta(s)\alpha(s)e^{\int_s^t \beta(\tau)d\tau} ds.$$

The case with $\alpha(t) \equiv M$ is handled by using the main theorem of Differential and Integral calculus

$$\begin{aligned}
\varphi(t) &\leq M\left(1 + \int_a^t \beta(s)e^{\int_s^t \beta(\tau)d\tau} ds\right) \\
&= M(1 - e^{\int_s^t \beta(\tau)d\tau}|_a^t) \\
&= Me^{\int_a^t \beta(\tau)d\tau}.
\end{aligned}$$

□

Exercise. In the many particle setting Equation 2.1 which assumptions for K are needed so that the interacting particle system has a unique global solution?

2.2 Mean-field particle system, well-posedness and problem setting

In the setting of first order system Equation 2.1, we give the assumption for interaction force:

Assumption B. $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is anti-symmetric i.e

$$K(x, y) = -K(y, x) \quad K(x, x) = 0.$$

Furthermore, $K \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$ and there exists some $L > 0$ such that $\forall x, y \in \mathbb{R}^d$ it holds

$$\sup_y |\nabla_x K(x, y)| + \sup_x |\nabla_y K(x, y)| \leq L.$$

Lemma 2.2.1. Suppose Assumption B holds, then $\forall T > 0$ the many particle system Equation 2.1 has a unique solution

$$X_N(t) = (x_1(t), x_2(t), \ldots, x_N(t)) \in \mathcal{C}^1([0, T]; \mathbb{R}^{dN}).$$

and for any fixed $t \in [0, T]$ the map

$$X_N(t, \cdot) : \mathbb{R}^{dN} \to \mathbb{R}^{dN} \ : \ x \mapsto X_N(t, x)$$

is a bijection.

Let $\mu^N(t) \triangleq \frac{1}{N} \sum_{j=1}^N \delta_{x_i(t)}$ be the empirical measure given in Definition 1.0.1 . In the introduction we showed that the empirical measure satisfies a partial differential equation in the following sense: $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ it holds

$$\frac{d}{dt} \langle \mu^N(t), \varphi \rangle = \langle \mu^N(t), \nabla \varphi \cdot \mathcal{K} \mu^N(t) \rangle.$$

where

$$\mathcal{K} \mu^N(\cdot) = \int_{\mathbb{R}^d} K(\cdot, y) d\mu^N(y).$$

We expect: if $\mu^N \to \mu$ in some sense, then the limiting measure μ should also satisfy

$$\begin{cases} \partial_t \mu + \nabla \cdot (\mu \mathcal{K} \mu) = 0 \\ \mu_0 = \lim_{N \to \infty} \mu^N(0) \end{cases}$$

in the same sense. This weak sense is understood in the "sense of distributions", which we define in the following section.

2.3 A short introduction for Distributions

As an conventional notation, we use $\alpha \in \mathbb{N}_0^n$ of length $|\alpha| = \sum_i \alpha_i$ as a multi-index, with which the partial derivatives in multi-dimension can be written in a short version

$$\partial^\alpha = \prod_i^n \frac{\partial_i^\alpha}{\partial x_i^{\alpha_i}}.$$

Definition 2.3.1. Let $\Omega \subset \mathbb{R}^d$ be an open subset, the space of test functions $\mathcal{D}(\Omega)$ consists of all the functions in $\mathcal{C}_0^\infty(\Omega)$ supplemented by the following convergence:

We say $\varphi_m \to \varphi \in \mathcal{C}_0^\infty(\Omega)$ iff

1. There exists a compact set $K \subset \Omega$ such that $\operatorname{supp} \varphi_m \subset K$ for all m
2. For all multi indices α it holds

$$\sup_K |\partial^\alpha \varphi_m - \partial^\alpha \varphi| \xrightarrow{m\to\infty} 0.$$

Remark. $\mathcal{D}(\Omega)$ is a linear space

Definition 2.3.2 (Distribution)**.** The space of Distributions is denoted by $\mathcal{D}'(\Omega)$ and is the dual space of $\mathcal{D}(\Omega)$ i.e. it is the linear space of all continuous linear functionals on $\mathcal{D}(\Omega)$

We say a functional $T : \mathcal{D}(\Omega) \to \mathbb{C}$ is continuous linear iff

1. $\langle T, \alpha\varphi_1 + \beta\varphi_2\rangle = \alpha\langle T, \varphi_1\rangle + \beta\langle T, \varphi_2\rangle$, $\forall \alpha, \beta \in \mathbb{R}$ and $\forall \varphi_1, \varphi_2 \in \mathcal{D}(\Omega)$
2. If $\varphi_m \to \varphi$ in $\mathcal{D}(\Omega)$ then $\langle T, \varphi_m\rangle \to \langle T, \varphi\rangle$

We can define several operations on the space of distributions but since most of them are not used in this Lecture we only define the multiplication with a smooth function

Definition 2.3.3. For a smooth function $f \in \mathcal{C}^\infty$ and a distribution $T \in \mathcal{D}'$ the product is defined as follows

$$\langle Tf, \varphi\rangle = \langle T, f\varphi\rangle \quad \forall \varphi \in \mathcal{D}.$$

Remark. Multiplication between two Distributions $T, F \in \mathcal{D}'$ is not well defined.

Next we give a list of examples for distributions.

Example. For functions $f \in L^1_{\text{loc}}(\Omega)$ we can define the associated distribution $T_f \in \mathcal{D}'(\Omega)$ by

$$\langle T_f, \varphi\rangle = \int_\Omega f(x)\varphi(x)dx \quad \forall \varphi \in \mathcal{D}(\Omega).$$

With this meaning we say $L^1_{\text{loc}}(\Omega) \subset \mathcal{D}'(\Omega)$

Similarly $L^p_{\text{loc}} \subset \mathcal{D}'(\Omega)$, using Hölder's inequality one obtains $L^p_{\text{loc}}(\Omega) \subset L^q_{\text{loc}}(\Omega)$ for $1 \le q < p \le \infty$.

Remark. The support of a distribution is also well-defined

Theorem 2.3.1. L^1_{loc} functions are uniquely determined by distributions. More precisely for two functions $f, g \in L^1_{\text{loc}}(\Omega)$ if

$$\int_\Omega f\varphi dx = \int_\Omega g\varphi dx \quad \forall \varphi \in \mathcal{D}(\Omega).$$

then $f = g$ a.e. in Ω

This proof is left as an exercise

Example. The set of probability density functions on $\mathbb{R}$ is a subset of $\mathcal{D}'(\mathbb{R})$. For any probability density function $\rho(x)$ the associated distribution $T_\rho \in \mathcal{D}'(\mathbb{R})$ is defined by

$$\langle T_\rho, \varphi\rangle = \int_\mathbb{R} \varphi(x)\rho(x)dx \quad \forall \varphi \in \mathcal{D}(\mathbb{R}).$$

Example. The set of measures $\mathcal{M}(\Omega)$ is a subset of $\mathcal{D}'(\Omega)$. For any $\mu \in \mathcal{M}(\Omega)$ the associated distribution T_μ is defined by

$$\langle T_\mu, \varphi \rangle = \int_\Omega \varphi(x) d\mu \quad \forall \varphi \in \mathcal{D}(\Omega).$$

Example. An important example of a distribution, which is not defined within the framework of locally integrable functions, is the Delta distribution $\delta_y(x)$ (concentrated on $y \in \mathbb{R}^d$)

$$\langle \delta_y, \varphi \rangle = \int_{\mathbb{R}^d} \varphi(x) d\delta_y(x) = \varphi(y) \quad \forall \varphi \in \mathcal{D}(\Omega).$$

where

$$\delta_y(E) = \begin{cases} 1, & y \in E \\ 0, & y \notin E \end{cases}.$$

The empirical measure μ^N is actually given by using the Delta distribution

$$\mu^N(t) \triangleq \frac{1}{N} \sum_{j=1}^{N} \delta_{x_i(t)} \quad \langle \mu^N, \varphi \rangle = \frac{1}{N} \sum_{j=1}^{N} \varphi(x_i(t)).$$

We define the convergence for a sequence of distributions as follows

Definition 2.3.4. For a sequence of distributions $(T_m)_{m \in \mathbb{N}} \subset \mathcal{D}'(\Omega)$ we say that it converges against a limit $T \in \mathcal{D}'(\Omega)$ if and only if

$$\langle T_m, \varphi \rangle \to \langle T, \varphi \rangle \quad \forall \varphi \in \mathcal{D}(\Omega).$$

Based on this convergence we give some examples in the approximation of $\delta_0(x)$

Example (Heat Kernel)**.** The heat kernel for $x \in \mathbb{R}$ and $t > 0$ is given by

$$f_t(x) = \frac{1}{(4\pi t)^{\frac{1}{2}}} e^{-\frac{|x|^2}{4t}}.$$

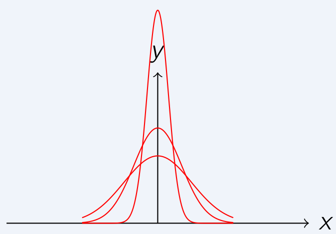


Figure 2.1: Heat Kernel for different t

Lemma 2.3.1. The sequence of distributions associated to the heat kernel converge to the Delta distribution.

Proof. We consider the limit $t \to 0^+$ and obtain $\forall \varphi \in \mathcal{C}_0^\infty(\Omega)$

$$\begin{aligned}\lim_{t\to 0^+}\int_{\mathbb{R}} f_t(x)\varphi(x) &= \lim_{t\to 0^+}\int_{\mathbb{R}} \frac{1}{(4\pi t)^{\frac{1}{2}}} e^{-\frac{|x|^2}{4t}}\varphi(x)dx \\ &= \lim_{t\to 0^+}\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-y^2}\varphi(2\sqrt{t}y)dy \\ &= \varphi(0) = \langle \delta_0, \varphi\rangle.\end{aligned}$$

where we used the substitution $x = 2\sqrt{t}y$. □

Example. For the rectangular functions

$$Q_n(x) = \begin{cases} \frac{n}{2}, & |x| \le \frac{1}{n} \\ 0, & |x| > \frac{1}{n} \end{cases}.$$

Then

$$Q_n \xrightarrow{n\to\infty} \delta_0(x).$$

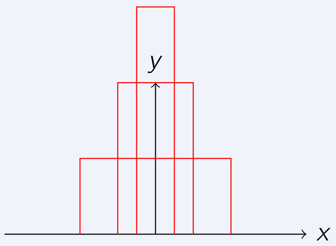


Figure 2.2: Rectangular functions for different n

We leave the proof as an exercise.

Example. The Dirichlet kernel

$$D_n(x) = \frac{\sin(n+\frac{1}{2})x}{\sin\frac{x}{2}} = 1 + 2\sum_{k=1}^{n}\cos(kx).$$

Then $D_n \xrightarrow{n\to\infty} 2\pi\delta_0(x)$ in the sense of distribution

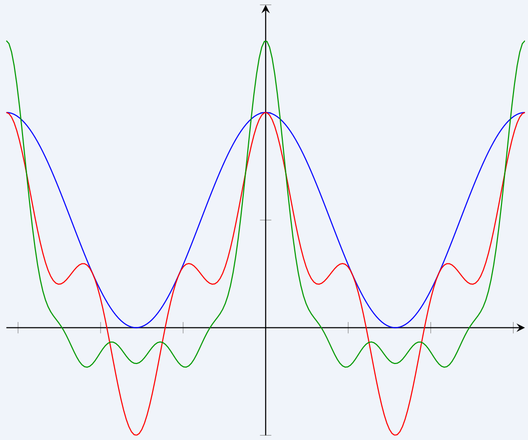


Figure 2.3: Dirichlet kernel for different n

We leave the proof also as an exercise.

2.4 Weak Derivative Of Distributions

Definition 2.4.1. For all distributions $T \in \mathcal{D}'(\Omega)$ we define the derivative $\partial_i T$ by

$$\langle \partial_i T, \varphi \rangle := -\langle T, \partial_i \varphi \rangle \quad \forall \varphi \in \mathcal{D}(\Omega)$$
$$\langle \partial_i^\alpha T, \varphi \rangle := (-1)^{|\alpha|} \langle T, \partial_i^\alpha \varphi \rangle \quad \forall \varphi \in \mathcal{D}(\Omega),$$

where we use the notation of multi index α.

Exercise. Show that $\partial_i T \in \mathcal{D}'(\Omega)$. Namely prove that the function $-\langle T, \partial_i \varphi \rangle$ is a continuous and linear functional of $\varphi \in \mathcal{D}(\Omega)$.

We give a couple examples

Example. The weak derivative of the Dirac Delta distribution is given by: $\forall \varphi \in \mathcal{D}(\Omega)$,

$$\langle \delta_0', \varphi \rangle = -\langle \delta_0, \varphi' \rangle = -\varphi(0)$$
$$\langle \delta_0^{(k)}, \varphi \rangle = (-1)^k \varphi^{(k)}(0).$$

Lemma 2.4.1. The weak derivative of the 1-D Heaviside function

$$H(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0 \end{cases}.$$

is the Dirac Delta distribution.

Proof. For $\forall \varphi \in \mathcal{D}(\Omega)$, it holds

$$\langle H', \varphi \rangle \overset{\text{Def.}}{=} -\langle H, \varphi' \rangle = -\int_{-\infty}^{\infty} H(x)\varphi'(x)dx = -\int_0^{\infty} \varphi'(x)dx$$
$$= \varphi(0) = \langle \delta_0, \varphi \rangle.$$

Therefore $H' = \delta_0$ in the sense of distribution. $\square$

2.5 Weak Formulation Of The Mean Field Partial Differential Equation

We can now go on to properly formulate the mean field partial differential equation in a weak sense. Using the notation of the empirical measure we can rewrite our earlier definition of the Equation 2.1 as follows

$$\begin{cases} \frac{d}{dt}x_i(t) & = \langle K(x_i,\cdot), \mu^N(t,\cdot)\rangle = \int_{\mathbb{R}^d} K(x_i,y)d\mu^N(t,y) \\ x_i(0) & = x_{i,0} \in \mathbb{R}^d, t \in [0,T] \end{cases}.$$

As has been discussed before, the empirical measure satisfies for $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$

$$\frac{d}{dt}\langle \mu^N, \varphi\rangle = \langle \mu^N \mathcal{K}\mu^N, \nabla\varphi\rangle = \langle -\operatorname{div}(\mu^N\mathcal{K}\mu^N), \varphi\rangle.$$

where

$$\mathcal{K}\mu^N(x) = \int_{\mathbb{R}^d} K(x,y)d\mu^N(y),$$

which means that the empirical measure μ^N satisfies the following equation in the sense of distribution

$$(\text{MPDE}) \quad \partial_t\mu^N + \operatorname{div}(\mu^N\mathcal{K}\mu^N) = 0.$$

Exercise. Show $\mu^N\mathcal{K}\mu^N$ is a distribution for smooth $K(x,y)$

Next we concentrate on the following mean field PDE

$$(\text{MFE}) \begin{cases} \partial_t\mu + \operatorname{div}(\mu\mathcal{K}\mu) & = 0 \\ \mu|_{t=0} & = \mu_0 \end{cases}, \tag{2.3}$$

where

$$\mathcal{K}\mu(x) = \int_{\mathbb{R}^d} K(x,y)d\mu(y).$$

We give the definition of the weak solution of Equation 2.3:

Definition 2.5.1 (Weak Solution of Equation 2.3)**.** For all $t \in [0,T]$, $\mu(t) \in \mathcal{M}(\mathbb{R}^d)$ is called a weak solution of Equation 2.3, where $\mathcal{M}(\mathbb{R}^d)$ denotes the space of measures on $\mathbb{R}^d$, if $\forall\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ it holds

$$\langle \mu(t), \varphi\rangle - \langle \mu_0, \varphi\rangle = \int_0^t \langle \mu(s)\mathcal{K}\mu(s), \nabla\varphi\rangle ds.$$

Remark. If $\mu_0 = \mu^N(0)$ i.e. the initial data is given by an empirical measure, then $\mu^N(t,\cdot)$ is a weak solution of the Equation 2.3.

Definition 2.5.2 (Push Forwad Measure)**.** For a measurable function X and a measure $\mu_0 \in \mathcal{M}(\mathbb{R}^d)$, the push forward measure $X_{\#}\mu_0$ is given by: for any Borel set $B \subset \mathbb{R}^d$, it holds

$$X_{\#}\mu_0 := \mu_0(X^{-1}(B)).$$

We consider the following initial value problem, the so called characteristics flow,

$$\begin{cases} \frac{d}{dt}x(t,x_0,\mu_0) = \int_{\mathbb{R}^d} K(x(t,x_0,\mu_0),y)d\mu(y,t) \\ x(0,x_0,\mu_0) = x_0, \quad \forall x_0 \in \mathbb{R}^d \\ \mu(\cdot,t) = x(t,\cdot,\mu_0)_{\#}\mu_0. \end{cases} \tag{2.4}$$

The solution flow $x(t,\cdot,\mu_0)$ gives for any time $t > 0$ a map

$$x(t,\cdot,\mu_0) \ : \ \mathbb{R}^d \to \mathbb{R}^d.$$

Remark. It can be easily checked that the push forward measure $\mu(t)$ obtained in the Weak Formulation Of The Mean Field Partial Differential Equation is a weak solution of the Equation 2.3.

The solution of the Weak Formulation Of The Mean Field Partial Differential Equation is going to be proved in the following space:

$$\mathcal{P}_1(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) \ : \ \int_{\mathbb{R}^d} |x| d\mu(x) < \infty\}.$$

where $\mathcal{P}(\mathbb{R}^d)$ is the space of all probability measures

Assumption C. We say an interaction force K is regular if $K \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$ and there exists an $L > 0$ such that

$$\sup_y |\nabla_x K(x,y)| + \sup_x |\nabla_y K(x,y)| \leq L.$$

Actually this assumption is also needed to in order to show the well-posedness of the particle system, which is a direct application of the ODE theory. We will use the same assumption to solve the Equation 2.3 and Equation 2.4.

Theorem 2.5.1 (Existence and Uniqueness of Equation 2.4)**.** Let Assumption C hold for K and $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ then the Equation 2.4 has a unique solution $x(t,x_0,\mu_0) \in \mathcal{C}^1(\mathbb{R};\mathbb{R}^d)$ and $x(t,\cdot,\mu_0)_{\#}\mu_0 \in \mathcal{P}_1$ for $\forall t > 0$.

Proof. The proof is based on Picard iteration.

Let $C_1 = \int_{\mathbb{R}^d} |x| d\mu_0(x)$ and define the following Banach space

$$X := \{v \in \mathcal{C}(\mathbb{R}^d;\mathbb{R}^d) \ \ \|v\|_X < \infty\}, \quad \|v\|_X := \sup_{x\in\mathbb{R}^d} \frac{|v(x)|}{1+|x|}..$$

As preparations we need the following estimates for the non-local term, by using Assumption C for K we have $\forall v, w \in X$

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} K(v(x),v(y))d\mu_0(y) - \int_{\mathbb{R}^d} K(w(x),w(y))d\mu_0(y) \right| \\ &\leq L \int_{\mathbb{R}^d} |v(x)-w(x)| + |v(y)-w(y)| d\mu_0(y) \\ &\leq L\|v-w\|_X(1+|x|) + L\|v-w\|_X \int_{\mathbb{R}^d} (1+|y|)d\mu_0(y) \\ &\leq L(2+C_1)\|v-w\|_X(1+|x|). \end{aligned}$$

Now define the Picard iteration for $\forall y \in \mathbb{R}^d$

$$
\begin{aligned}
x_0(t,y) &= y \\
x_1(t,y) &= y + \int_0^t \int_{\mathbb{R}^d} K(x_0(s,y), x_0(s,z)) d\mu_0(z) ds \\
&\vdots \\
x_m(t,y) &= y + \int_0^t \int_{\mathbb{R}^d} K(x_{m-1}(s,y), x_{m-1}(s,z)) d\mu_0(z) ds.
\end{aligned}
$$

Then we can bound the difference between x_1 and x_0 by

$$
\begin{aligned}
|x_1(t,y) - x_0(t,y)| &= \left| \int_0^t \int_{\mathbb{R}^d} K(x_0(s,y), x_0(s,z)) d\mu_0(z) ds \right| \\
&= \left| \int_0^t \int_{\mathbb{R}^d} K(y,z) d\mu_0(z) ds \right| \\
&\leq \int_0^{|t|} \int_{\mathbb{R}^d} L(|y| + |z|) d\mu_0(z) ds \\
&= \int_0^{|t|} L(|y| + C_1) ds \\
&\leq L(1 + C_1)(1 + |y|)|t|.
\end{aligned}
$$

Furthermore for $\forall m \geq 1$ we have

$$
\begin{aligned}
&|x_m(t,y) - x_{m-1}(t,y)| \\
&= \left| \int_0^t \int_{\mathbb{R}^d} \left(K(x_{m-1}(s,y), x_{m-1}(s,z)) - K(x_{m-2}(s,y), x_{m-2}(s,z)) \right) d\mu_0(z) ds \right| \\
&\leq L(2 + C_1) \int_0^{|t|} \|x_{m-1}(s,\cdot) - x_{m-2}(s,\cdot)\|_X (1 + |y|) ds.
\end{aligned}
$$

hence by dividing both sides by $1 + |y|$ we have

$$
\begin{aligned}
\|x_m(t,\cdot) - x_{m-1}(t,\cdot)\|_X &\leq L(2 + C_1) \int_0^{|t|} \|x_{m-1}(s,\cdot) - x_{m-2}(s,\cdot)\|_X ds \\
&\leq \frac{((2 + C_1)L|t|)^d}{(m-1)!}.
\end{aligned}
$$

which implies for $\forall m > n \to \infty$

$$
\|x_m(t,\cdot) - x_n(t,\cdot)\|_X \leq \sum_{i=n}^{m-1} \|x_{i+1}(t,\cdot) - x_i(t,\cdot)\|_X \to 0.
$$

Therefore for $T > 0$

$$
x_m(t,\cdot) \to x(t,\cdot) \text{ in } X \text{ uniformly in } [-T,T].
$$

and $x \in \mathcal{C}(\mathbb{R};\mathbb{R}^d)$ satisfies that, after taking the limit in Picard iteration $\forall y \in \mathbb{R}^d$

$$
x(t,y) = y + \int_0^t \int_{\mathbb{R}^d} K(x(s,y), x(s,z)) d\mu_0(z) ds.
$$

By the fundamental theorem of calculus and Assumption C we know that for $y \in \mathbb{R}^d$ and $x(t,y) \in \mathcal{C}^1(\mathbb{R};\mathbb{R}^d)$

$$\frac{d}{dt}x(t,y) = \int_{\mathbb{R}^d} K(x(t,y),x(t,z))d\mu_0(z) = \int_{\mathbb{R}^d} K(x(t,y),z')d\mu(z',t).$$

where $\mu(\cdot,t)$ is the push forward measure of μ_0 along $x(t,\cdot)$

For uniqueness, we consider two solutions $x, \tilde{x}$. By taking the difference we have

$$x(t,y) - \tilde{x}(t,y) = \int_0^t \int_{\mathbb{R}^d} \left(K(x(s,y),x(s,z)) - K(\tilde{x}(s,y),\tilde{x}(s,z))\right) d\mu_0(z)ds.$$

Using estimates similarly to before we obtain

$$\|x(t,\cdot) - \tilde{x}(t,\cdot)\|_X \leq L(2+C_1)\int_0^{|t|} \|x(s,\cdot)-\tilde{x}(s,\cdot)\|_X ds.$$

By applying Gronwall's inequality we get

$$\|x(t,\cdot)-\tilde{x}(t,\cdot)\|_X = 0,$$

considering the fact that $\|x(0,\cdot)-\tilde{x}(0,\cdot)\|_X = 0$. □

2.5.1 Stability

Let's remind us of the N-particle system Equation 2.1, the Mean field equation Equation 2.3 and its weak solution as defined in Definition 2.5.1. We have thus far done the following things

1. If $\mu_0 = \mu_N(0)$ then $\mu_N(t)$ is a weak solution of Equation 2.3
2. If $\mu_0 = \mathcal{P}_1(\mathbb{R}^d)$ and the assumption on hold for K,then $x(t,\cdot,\mu_0)_{\#}\mu_0 \in \mathcal{P}_1$ is the solution of Equation 2.3.

We will prove the stability of the mean field PDE, which means directly that

$$\mu_N(0) \to \mu(0) \Rightarrow \mu_N(t) \to \mu(t).$$

by using the so called Monge-Kantorovich distance (or Wasserstein distance)

Definition 2.5.3 (Monge-Kantorovich Distance). Let

$$\mathcal{P}_p(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) \ : \ \int_{\mathbb{R}^d} |x|^p d\mu(x) < \infty\}, \quad p \geq 1.$$

For two measures $\mu,\nu \in \mathcal{P}_p(\mathbb{R}^d)$, the Monge-Kantorovich distance $\text{dist}_{\text{MK},p}(\mu,\nu)$ or $W^p(\mu,\nu)$ is defined by

$$\operatorname*{dist}_{\text{MK},p}(\mu,\nu) = W^p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\pi(x,y)\right)^{\frac{1}{p}}.$$

where

$$\Pi(\mu,\nu) = \left\{\pi \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d) \ : \ \int_{\mathbb{R}^d} \pi(\cdot,dy) = \mu(\cdot) \text{ and } \int_{\mathbb{R}^d} \pi(dx,\cdot) = \nu(\cdot)\right\}..$$

Remark. For $\forall \varphi,\psi \in \mathcal{C}(\mathbb{R}^d)$ such that $\varphi(x) \sim O(|x|^p)$ for $|x| \gg 1$ and $\psi(y) \sim O(|y|^p)$ for

$|y| \gg 1$, for $\pi \in \Pi(\mu, \nu)$ it holds

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(x) + \psi(y)) d\pi(x, y) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y).$$

Remark (Kantorovich-Rubinstein duality). It can be shown that the W^1 distance can be computed by

$$\operatorname*{dist}_{\mathrm{MK},1}(\mu, \nu) = W^1(\mu, \nu) = \sup_{\varphi \in \mathrm{Lip}(\mathbb{R}^d), \mathrm{Lip}(\varphi) \leq 1} \left| \int_{\mathbb{R}^d} \varphi(x) d\mu(x) - \int_{\mathbb{R}^d} \varphi(x) d\nu(x) \right|.$$

Theorem 2.5.2 (Dobrushin's stability). Let $\mu_0, \overline{\mu}_0 \in \mathcal{P}_1(\mathbb{R}^d)$ and $(x(t, \cdot, \mu_0), \mu(\cdot, t))$, $(x(t, \cdot, \overline{\mu}_0), \overline{\mu}_0(\cdot, t))$ be solutions of Theorem 2.5.1. Then $\forall t > 0$ it holds

$$\operatorname*{dist}_{\mathrm{MK},1}(\mu(\cdot, t), \overline{\mu}(\cdot, t)) \leq e^{2|t|L} \operatorname*{dist}_{\mathrm{MK},1}(\mu_0, \overline{\mu}_0).$$

Proof. Let (x_0, μ_0) and $(\overline{x}_0, \overline{\mu}_0)$ be two initial data pairs of problem Equation 2.4 and $\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)$. By taking the difference of these two problems, we have

$$\begin{aligned}
&x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0) \\
&= x_0 - \overline{x}_0 \ + \int_0^t \int_{\mathbb{R}^d} K(x(s, x_0, \mu_0), y) d\mu(s, y) ds \\
&\qquad\qquad - \int_0^t \int_{\mathbb{R}^d} K(x(s, \overline{x}_0, \overline{\mu}_0), y) d\overline{\mu}(s, y) ds.
\end{aligned}$$

where $\mu(\cdot, t) = x(t, \cdot, \mu_0)_{\#}\mu_0$ and $\overline{\mu}(\cdot, t) = x(t, \cdot, \overline{\mu}_0)_{\#}\overline{\mu}_0$. Now we compute further and get

$$\begin{aligned}
&x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0) \\
&= x_0 - \overline{x}_0 \ + \int_0^t \int_{\mathbb{R}^d} K(x(s, x_0, \mu_0), x(s, z, \mu_0)) d\mu_0(z) ds \\
&\qquad\qquad - \int_0^t \int_{\mathbb{R}^d} K(x(s, \overline{x}_0, \overline{\mu}_0), x(s, \overline{z}, \overline{\mu}_0)) d\overline{\mu}_0(\overline{z}) ds \\
&= x_0 - \overline{x_0} \ + \int_0^t \iint_{\mathbb{R}^d \times \mathbb{R}^d} (K(x(s, x_0, \mu_0), x(s, z, \mu_0)) \\
&\qquad\qquad - K(x(s, \overline{x}_0, \overline{\mu}_0), x(s, \overline{z}, \overline{\mu}_0)))\, d\pi_0(z, \overline{z}) ds.
\end{aligned}$$

There for by assumption on Assumption C for K, we have

$$\begin{aligned}
&|x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0)| \\
&\leq |x_0 - \overline{x}_0| + L \int_0^t \iint_{\mathbb{R}^d \times \mathbb{R}^d} \Big(|x(s, x_0, \mu_0) - x(s, \overline{x}_0, \overline{\mu}_0)| \\
&\qquad\qquad + |x(s, z, \mu_0) - x(s, \overline{z}, \overline{\mu}_0)| \Big) d\pi_0(z, \overline{z}) ds \\
&\leq |x_0 - \overline{x}_0| + L \int_0^t |x(s, x_0, \mu_0) - x(s, \overline{x}_0, \overline{\mu}_0)| ds \\
&\qquad\qquad + L \int_0^t \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x(s, z, \mu_0) - x(s, \overline{z}, \overline{\mu}_0)| d\pi_0(z, \overline{z}) ds.
\end{aligned}$$

Next we integrate both sides in $x_0, \overline{x}_0$ with respect to the measure π_0

$$
\begin{aligned}
&\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x(t,x_0,\mu_0) - x(t,\overline{x}_0,\overline{\mu}_0)| d\pi_0(x_0,\overline{x}_0) \\
&\le \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x_0 - \overline{x}_0| d\pi_0(x_0,\overline{x}_0) \\
&+ L\int_0^t \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x(s,x_0,\mu_0) - x(s,\overline{x}_0,\overline{\mu}_0)| d\pi_0(x_0,\overline{x}_0) ds \\
&+ L\int_0^t \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x(s,z,\mu_0) - x(s,\overline{z},\overline{\mu}_0)| d\pi_0(z,\overline{z}) ds
\end{aligned}
$$

.

By denoting

$$D[\pi_0](t) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x(s,z,\mu_0) - x(s,\overline{z},\overline{\mu}_0)| d\pi_0(z,\overline{z}).$$

we have obtained the estimate

$$D[\pi_0](t) \le D[\pi_0](0) + 2L\int_0^t D[\pi_0](s)ds.$$

which implies by Gronwall's inequality that

$$D[\pi_0](t) \le D[\pi_0](0)e^{2Lt}.$$

Now let $\varphi_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ be the map such that

$$\varphi_t(x_0,\overline{x}_0) = (x(t,x_0,\mu_0), x(t,\overline{x}_0,\overline{\mu}_0)),$$

and for arbitrary $\pi_0 \in \Pi(\mu_0,\nu_0)$, $\pi_t := (\varphi_t)_\#\pi_0$ be the push forward measure of π_0 by φ_t. It is obvious that

$$\pi_t = (\varphi_t)_\#\pi_0 \in \Pi(\mu(\cdot,t),\overline{\mu}(\cdot,t))$$

Therefore

$$
\begin{aligned}
\operatorname*{dist}_{\mathrm{MK},1}(\mu(\cdot,t),\overline{\mu}(\cdot,t)) &= \inf_{\pi\in\Pi(\mu(\cdot,t),\overline{\mu}(\cdot,t))} \iint_{\mathbb{R}^d\times\mathbb{R}^d} |z-\overline{z}| d\pi(z,\overline{z}) \\
&\le \inf_{\pi_0\in\Pi(\mu_0,\overline{\mu}_0)} \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x(t,z,\mu_0) - x(t,\overline{z},\overline{\mu}_0)| d\pi(z,\overline{z}) \\
&= \inf_{\pi_0\in\Pi(\mu_0,\overline{\mu}_0)} D[\pi_0](t) \\
&\le \inf_{\pi_0\in\Pi(\mu_0,\overline{\mu}_0)} D[\pi_0](0)e^{2Lt} \\
&= e^{2Lt}\operatorname*{dist}_{\mathrm{MK},1}(\mu_0,\overline{\mu}_0).
\end{aligned}
$$

□

2.6 Mollification Operator

In this section, we give a short introduction of the mollification operator. It will be used to prove the conservation of mass on the PDE level, instead of using the concept of push-forward measure.

Definition 2.6.1 (Mollification-Kernel)**.** A function $j(x) \in \mathcal{C}_0^\infty$ is called a mollification kernel if it satisfies the following properties

1. $j(x) \geq 0$
2. $\operatorname{supp} j \subset \overline{B_1(0)}$
3. $\int_{\mathbb{R}^d} j(x)dx = 1$

A typical example of a smooth kernel is given by

Example.

$$j(x) = \begin{cases} k\exp(-\frac{1}{1-|x|^2}) & \text{if } |x| < 1 \\ 0 & \text{if otherwise} \end{cases}.$$

where k is given s.t the integral is 1

Remark. Based on the given function j it is easy to prove that its rescaled sequence converges to the Dirac Delta distribution in the weak sense

$$j_\varepsilon(x) = \frac{1}{\varepsilon^d} j(\frac{x}{\varepsilon}) \xrightarrow{\varepsilon\to 0} \delta_0.$$

Exercise. Prove that for $\varphi(x) \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ it holds that $\forall x \in \mathbb{R}^d$

$$\lim_{\varepsilon\to 0} j_\varepsilon \star \varphi(x) = \varphi(x).$$

Definition 2.6.2 (Mollification Operator)**.** For $\forall u \in L^1_{\text{loc}}(\mathbb{R}^d)$ we define the following function as its mollification

$$J_\varepsilon(u)(x) \triangleq j_\varepsilon(x) \star u(x) = \int_{\mathbb{R}^d} j_\varepsilon(x-y)u(y)dy.$$

where J_ε is called the mollification operator

Remark. Notice that $\operatorname{supp} j_\varepsilon(x) \subset \overline{B_\varepsilon(0)}$ we obtain

$$J_\varepsilon(u)(x) = \int_{B_\varepsilon(0)} j_\varepsilon(x-y)u(y)dy < \infty.$$

Lemma 2.6.1. The following statements are true for the mollification operator:

1. If $u(x) \in L^1(\mathbb{R}^d)$ and $\operatorname{supp} u(x)$ is compact in $\mathbb{R}^d$ then $\forall \varepsilon > 0$ it holds

$$J_\varepsilon(u) = j_\varepsilon \star u \in \mathcal{C}_0^\infty.$$

2. if $u \in C_0(\mathbb{R}^d)$ then

$$J_\varepsilon(u) \xrightarrow{\varepsilon\to 0} u \text{ uniformly on } \operatorname{supp} u.$$

Proof. 1. Let $K = \operatorname{supp} u \subset \mathbb{R}^d$, which is compact by assumption, then we have

$$\operatorname{supp} j_\varepsilon \star u = \{x \in \mathbb{R}^d \mid \operatorname{dist}(x, K) \leq \varepsilon\}.$$

is also compact. For the differentiability it is enough to show the first order partial differentiability at any given point, as the argument for higher order differentiability is analog

Now for $\forall x \in \operatorname{supp} j_\varepsilon \star u$ we have that $\forall i = 1, 2, \ldots, d$

$$\frac{\partial}{\partial x_i} \int_{\mathbb{R}^d} j_\varepsilon(x-y)u(y)dy = \int_K \frac{\partial}{\partial x_i} j_\varepsilon(x-y)u(y)dy.$$

where we have used the fact that

$$|\frac{\partial}{\partial x_i} j_\varepsilon(x-y)u(y)| \leq |\frac{\partial}{\partial x_i} j_\varepsilon(x-y)|\Big|_K \|u\|_{L^1} \leq \frac{Cj'}{\varepsilon^d}.$$

to show the uniform integrability of $\frac{\partial}{\partial x_i} j_\varepsilon(x-y)u(y)$

2. we need to prove that for $u \in \mathcal{C}_0(\mathbb{R}^d)$ it holds

$$\|J_\varepsilon(u) - u\|_{L^\infty(\operatorname{supp} u)} \xrightarrow{\varepsilon \to 0} 0.$$

Actually $\forall x \in \operatorname{supp} u$ we have the following estimate

$$\begin{aligned}
|j_\varepsilon \star u(x) - u(x)| &= \left| \int_{\mathbb{R}^d} j_\varepsilon(x-y)(u(y) - u(x))dy \right| \\
&= \left| \int_{\operatorname{supp} u} j_\varepsilon(x-y)(u(y) - u(x))dy \right| \\
&\leq \max_{\substack{x,y \in \operatorname{supp} u \\ |x-y|<\varepsilon}} |u(y) - u(x)| \int_{\mathbb{R}^d} j_\varepsilon(x-y)dy \xrightarrow{\varepsilon \to 0} 0.
\end{aligned}$$

where we have used the fact that $u \in \mathcal{C}(\operatorname{supp} u)$ which means u is uniformly continuous to obtain the limit in the last step above. □

2.7 Conservation of Mass

Let $\mu(t)$ be the push forward measure obtained from Equation 2.4, one can check that is is a weak solution of Equation 2.3 by using test functions.

As a consequence of Theorem 2.5.1, we obtain if the initial measure has a probability density, then the solution is also integrable for any fixed time t.

Corollary. Let f_0 be a probability density of μ_0 on $\mathbb{R}^d$ with $\int_{\mathbb{R}^d} |x| f_0(x)dx < \infty$, then the Cauchy problem

$$\begin{cases} \partial_t f + \nabla \cdot (f \mathcal{K} f) = 0 \\ f|_{t=0} = f_0 \end{cases}$$

has a unique weak solution $f(t, \cdot) \in L^1(\mathbb{R}^d)$ and $\|f(t, \cdot)\|_{L^1(\mathbb{R}^d)} = 1$.

Proof. We need to prove $\forall t \in \mathbb{R}$ and $\mu_t \in \mathcal{P}_1(\mathbb{R}^d)$ absolutely continuous with respect to the Lebesgue measure i.e. $\forall B \in \mathcal{B}$ and $\int_B d\lambda = 0$, where λ denotes the Lebesgue measure on $\mathbb{R}^d$, it holds $\mu_t(B) = 0$. This comes from that the solution flow of Equation 2.4 is bijective. The mass conservation property, $\|f(t, \cdot)\|_{L^1(\mathbb{R}^d)} = 1$ comes from the definition of probability measures. □

In the next we give an alternative proof of the conservation of mass without using the characteristics presentation and instead only use the definition of a weak solution

Proof. The weak solution in the sense of distribution means that $\forall \varphi \in \mathcal{C}_0^\infty$ it holds for all

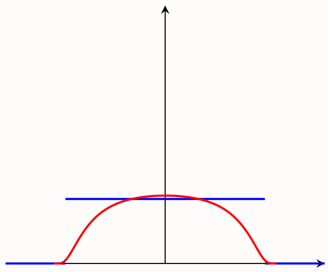
$0 \le \tilde{t} < t < \infty$

$$\int_{\mathbb{R}^d} \varphi(x) f(t,x) dx = \int_{\mathbb{R}^d} \varphi(x) f(\tilde{t},x) dx + \int_{\tilde{t}}^{t} \iint_{\mathbb{R}^{2d}} f(s,x) K(x,y) f(s,y) \nabla\varphi(x) dx dy ds.$$

Now we take a sequence of test functions defined as follows. For $R > 0$

$$\varphi_R(x) = \begin{cases} 1, & |x| \le R \\ \text{smooth}, & R < |x| < 2R, . \\ 0, & |x| \ge R \end{cases}$$

An example of this is the mollification of a step function i.e. $\varphi_R = j_{\frac{R}{2}} \cdot \mathbb{1}_{B_{\frac{3R}{2}}}$



One obtains directly for the gradient estimate $|\nabla\varphi_R(x)| \le \frac{C}{R}$. Therefore with this test function, we obtain from the weak solution formula that

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} f(t,x)\varphi_R(x) dx - \int_{\mathbb{R}^d} f(\tilde{t},x)\varphi_R(x) dx \right| \\ &= \left| \int_{\tilde{t}}^{t} \iint_{\mathbb{R}^{2d}} f(s,x) K(x,y) f(s,y) \nabla\varphi_R(x) dx dy ds \right| \\ &\le \frac{CL}{R} \int_{\tilde{t}}^{t} \iint_{\mathbb{R}^{2d}} (1 + |x| + |y|) f(s,x) f(s,y) |\nabla\varphi_R(x)| dx dy ds \\ &\le \frac{C}{R} |t - \tilde{t}|. \end{aligned}$$

Where C depends on $\|(1 + |\cdot|) f(t,\cdot)\|_{L^1(\mathbb{R}^d)}$. Since

$$|f(t,x)\varphi_R(x)| \le |f(t,x)| \quad \forall x \in \mathbb{R}^d.$$

we can use the dominant convergence theorem to obtain

$$\int_{\mathbb{R}^d} f(t,x)\varphi_R(x) dx \xrightarrow{R\to\infty} \int_{\mathbb{R}^d} f(t,x) dx > 0.$$

Therefore passing to the limit $R \to \infty$ we have

$$\int_{\mathbb{R}^d} f(t,x) dx = \int_{\mathbb{R}^d} f(\tilde{t},x) dx = \int_{\mathbb{R}^d} f_0(x) dx.$$

□

2.8 Mean Field Limit

Theorem 2.8.1 (Mean Field Limit). For $f_0 \in L^1(\mathbb{R}^d)$, let $\mu_0^N = \frac{1}{N} \sum_{i=1}^N \delta_{x_{i,0}}$ such that

$$\operatorname*{dist}_{\text{MK},1} (\mu_0^N, f_0) \xrightarrow{N\to\infty} 0.$$

Let $X_N(t)$ be the solution of the N particle system Equation 2.1 with its empirical measure

$$\mu^N(t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i(t,X_{N,0})}.$$

Then

$$\operatorname*{dist}_{\mathrm{MK},1}(\mu^N(t), f(t,\cdot)) \leq e^{2Lt} \operatorname*{dist}_{\mathrm{MK},1}(\mu_0^N, f_0) \xrightarrow{N\to\infty} 0.$$

And $\mu^N(t) \rightharpoonup f(t,\cdot)$ weakly in measures , i.e for $\forall\varphi \in \mathcal{C}_b(\mathbb{R}^d)$ it holds

$$\int_{\mathbb{R}^d} \varphi(x) d\mu^N(t,x) \xrightarrow{N\to\infty} \int_{\mathbb{R}^d} \varphi(x) f(t,x) dx.$$

Proof. The stability result from Theorem 2.5.2 gives us already the convergence rate estimate. We are left to prove the weak convergence in measure. Note $\forall\varphi \in \mathrm{Lip}(\mathbb{R}^d)$ we have

$$\begin{aligned}\left|\int_{\mathbb{R}^d} \varphi(x) d\mu^N(t,x) - \int_{\mathbb{R}^d} \varphi(x) f(t,x) dx\right| &= |\iint_{\mathbb{R}^d\times\mathbb{R}^d} (\varphi(x)-\varphi(y)) d\pi_t(x,y)| \\ &\leq \mathrm{Lip}(\varphi) \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y| d\pi_t(x,y) \\ &\to 0.\end{aligned}$$

where $\pi_t \in \Pi(\mu^N(t), f(t,\cdot))$

Since $\mathrm{Lip}(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$ and because the total mass is 1, the above also holds for test functions in $\mathcal{C}_b(\mathbb{R}^d)$. Hence the weak convergence in measure is true. The fact that $\mathrm{Lip}(\mathbb{R}^d)$ is dense in $\mathcal{C}_0(\mathbb{R}^d)$ can be obtained by using the mollification operator introduced in Definition 2.6.2. More precisely we have to show that $\forall\varphi \in C_b^\infty$ it holds

$$\int_{\mathbb{R}^d} \varphi(x) d\mu^N(t,x) \xrightarrow{N\to\infty} \int_{\mathbb{R}^d} \varphi(X) f(t,x) dx.$$

Notice we have shown that the above convergence holds for all $\varphi \in \mathrm{Lip}(\mathbb{R}^d)$. Then for any fixed $\varphi \in \mathcal{C}_b^\infty$ and $\forall\varepsilon > 0$ we choose $R > 1$ s.t.

$$\frac{2\|\varphi\|_{L^\infty(\mathbb{R}^d)} M_1}{R} \leq \frac{\varepsilon}{2}.$$

where $M_1 = \int_{\mathbb{R}^d} |x| d\mu^N(t,x)$. Let $\varphi_m \in \mathcal{C}_0^\infty(B_{2R})$ be the approximation of φ on $B_{\frac{3R}{2}}$. This means that $\exists M > 0$ such that for $\forall m > M$ it holds

$$\|\varphi_m - \varphi\|_{L^\infty(B_R)} < \frac{\varepsilon}{4}.$$

Now we take $\varphi_{M+1} \in \mathcal{C}_0^\infty(B_{2R})$ which is obviously Lipschitz continuous. Therefore the convergence holds. Then $\exists N_1 > 0$ such that $\forall N > N_1$ we have

$$\left|\int_{\mathbb{R}^d} \varphi_{M+1}(x)(d\mu^N(t,x) - f(t,x)) dx\right| < \frac{\varepsilon}{4}.$$

To summarize we obtain that

$$
\begin{aligned}
&\left|\int_{\mathbb{R}^d} \varphi(x)d\mu^N(t,x) - \int_{\mathbb{R}^d} \varphi(x)f(t,x)dx\right| \\
\leq &\left|\int_{B_R} \varphi(x)(d\mu^N(t,x) - f(t,x)dx)\right| + \left|\int_{B_R^c} \varphi(x)(d\mu^N(t,x) - f(t,x)dx)\right| \\
\leq &\left|\int_{B_R} \varphi_{M+1}(x)(d\mu^N(t,x) - f(t,x)dx)\right| + \left|\int_{B_R} (\varphi_{M+1}(x) - \varphi(x))(d\mu^N(t,x) - f(t,x)dx)\right| \\
&+ \left|\int_{B_R^c} |\varphi(x)|\frac{|x|}{R}(d\mu^N(t,x) + f(t,x)dx)\right| \\
&< \frac{\varepsilon}{4} + \|\varphi_{M+1} - \varphi\|_{L^\infty(B_R)} + \frac{2}{R}\|\varphi\|_{L^\infty(\mathbb{R}^d)}M_1 < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{2} \leq \varepsilon.
\end{aligned}
$$

□

This concludes the chapter on the Mean-field Limit in the deterministic setting, we have thus far reviewed the basics of relevant ODE Theory, introduced the Mean-Field particle system Equation 2.1 and the associated Mean-Field equation Equation 2.3 and finished by proving a convergence result for the Mean-Field Limit.

Chapter 3

MEAN FIELD LIMIT FOR SDE SYSTEM

3.1 Basics On Probability Theory

This section is dedicated to a short review of basic concepts in probability theory.

3.1.1 Probability Spaces and Random Variables

Definition 3.1.1 (σ-Algebra)**.** Let Ω be a given set, then a $\sigma-$algebra $\mathcal{F}$ on Ω is a family of subsets of Ω s.t.

1. $\emptyset \in \mathcal{F}$
2. $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$
3. If $A_1, A_2, \ldots \in \mathcal{F}$, then $A = \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$.

Definition 3.1.2 (Measure Space)**.** A tuple $(\Omega, \mathcal{F})$ is called a measurable space. The elements of $\mathcal{F}$ are called measurable sets.

Definition 3.1.3 (Probability Measure)**.** A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that the following properties hold

1. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$
2. If $A_1, A_2, \ldots \in \mathcal{F}$ s.t. $A_i \cap A_j = \emptyset \ \forall i \neq j$, then $\mathbb{P}(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$.

Definition 3.1.4 (Probability Space)**.** The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. $F \in \mathcal{F}$ is called an event. We say the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete, if $\mathcal{F}$ contains all zero-measure sets i.e. if

$$\inf\{\mathbb{P}(F) \ : \ F \in \mathcal{F}, G \subset F\} = 0.$$

then $G \in \mathcal{F}$ and $\mathbb{P}(G) = 0$.

Without loss of generality we use in this lecture $(\Omega, \mathcal{F}, \mathbb{P})$ as complete probability space.

Definition 3.1.5 (Almost Surely)**.** If for some $F \in \mathcal{F}$ it holds $\mathbb{P}(F) = 1$, then we say that F happens with probability 1 or happens almost surely (a.s.).

Remark. Let $\mathcal{U}$ be a family of subsets of Ω, then there exists a smallest $\sigma-$algebra of Ω which contains $\mathcal{H}$, it is denoted by $\mathcal{U}_{\mathcal{H}} = \bigcap\{\mathcal{H} : \mathcal{H}$ is a $\sigma-$ algebra of Ω, and $\mathcal{U} \subset \mathcal{H}\}$.

Example. The $\sigma-$algebra generated by a topology τ of Ω , $\mathcal{U}_\tau \triangleq \mathcal{B}$ is called the Borel $\sigma-$algebra, the elements $B \in \mathcal{B}$ are called Borel sets.

Definition 3.1.6 (Measurable Functions)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, a function $Y : \Omega \to \mathbb{R}^d$ is called measurable if and only if $Y^{-1}(B) \in \mathcal{F}$ holds for all $B \in \mathcal{B}$ or equivalent for all $B \in \tau$.

Example. Let $X : \Omega \to \mathbb{R}^d$ be a given function, then the $\sigma-$algebra $\mathcal{U}(X)$ generated by X is

$$\mathcal{U}(X) = \{X^{-1}(B) \ : \ B \in \mathcal{B}\}.$$

Lemma 3.1.1 (Doob-Dynkin)**.** If $X, Y : \Omega \to \mathbb{R}^d$ are given functions, then Y is $\mathcal{U}(X)$ measurable if and only if there exists a Boreal measurable function $g : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$Y = g(x)..$$

Exercise. Proof the above lemma

Definition 3.1.7 (Random Variable)**.** A random variable $X : \Omega \to \mathbb{R}^d$ is a $\mathcal{F}-$measurable function. Every random variable induces a probability measure or $\mathbb{R}^d$

$$\mu_X(B) = \mathbb{P}(X^{-1}(B)) \quad \forall B \in \mathcal{B}.$$

This measure is called the distribution of X.

Remark (L^p spaces)**.** Let $X : \Omega \to \mathbb{R}^d$ be a random variable and $p \in [1, \infty)$. With norm

$$\|X\|_p = \|X\|_{L^p(\mathbb{P})} = \left(\int_\Omega |X(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}}, \text{ for } 1 \le p|\infty,$$
$$\|X\|_\infty = \inf\{N \in \mathbb{R} : |X(\omega)| \le N \text{ a.s.}\}, \text{ for } p = \infty,$$

it can be shown that the space $L^p(\Omega) = \{X : \Omega \to \mathbb{R}^d | \|X\|_p \le \infty\}$ is a Banach space.

Remark. If $p = 2$ then $L^2(\mathbb{P})$ is a Hilbert space with inner product

$$\langle X, Y\rangle = \mathbb{E}[X(\omega) \cdot Y(\Omega)] = \int_\Omega X(\omega) \cdot Y(\omega) d\mathbb{P}(\omega).$$

Definition 3.1.8 (Expectation and Variance)**.**
Let X be a random variable, if $\int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty$, then

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} x d\mu_X(x)$$

is called the expectation of X (w.r.t. $\mathbb{P}$);

$$\mathbb{V}[X] = \int_\Omega |X - \mathbb{E}[X]|^2 d\mathbb{P}(\omega)$$

is called variance and the simple relation holds

$$\mathbb{V}[X] = \mathbb{E}[|X - \mathbb{E}[X]|^2] = \mathbb{E}[|X|^2] - \mathbb{E}[X]^2.$$

Remark. If $f : \mathbb{R}^d \to \mathbb{R}$ is measurable and $\int_\Omega |f(X(\omega))| d\mathbb{P}(\Omega) < \infty$, then it holds

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} f(x) d\mu_X(x).$$

Definition 3.1.9 (Distribution Functions)**.** Note for $x, y \in \mathbb{R}^d$ we write $x \leq y$ if $x_i \leq y_i$ for $\forall i$

1. $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^d$ is a random variable, its distribution function $F_x : \mathbb{R}^d \to [0, 1]$ is defined by

$$F_X(x) = \mathbb{P}(X \leq x) \quad x \in \mathbb{R}^d.$$

2. If $X_1, \ldots, X_m : \Omega \to \mathbb{R}^d$ are random variables, then their joint distribution function is

$$\begin{aligned} &F_{X_1,\ldots,X_m} : (\mathbb{R}^d)^m \to [0, 1] \\ &F_{X_1,\ldots,X_m} = \mathbb{P}(X_1 \leq x_1, \ldots, X_m \leq x_m) \quad \forall x_i \in \mathbb{R}^d. \end{aligned}$$

Definition 3.1.10 (Density Function of random variables)**.** If there exists a non-negative function $f(x) \in L^1(\mathbb{R}^d; \mathbb{R})$ such that

$$F(x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_d} f(y) dy \quad y = (y_1, \ldots, y_d).$$

then f is called the density function of X and

$$\mathbb{P}(X^{-1}(B)) = \int_B f(x) dx \quad \forall B \in \mathcal{B}.$$

Example. Let X be a random variable with density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|x-m|^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

then we say that X has a Gaussian (or Normal) distribution with mean m and variance σ^2. In this case, we use the notation $X \sim \mathcal{N}(m, \sigma^2)$. Obviously

$$\int_{\mathbb{R}} x f(x) dx = m \quad \text{and} \quad \int_{\mathbb{R}} |x - m|^2 f(x) dx = \sigma^2.$$

Definition 3.1.11 (Independent Events)**.** Events $A_1, \ldots, A_n \in \mathcal{F}$ are called independent if $\forall 1 \leq k_1 < \ldots < k_m \leq n$ it holds

$$\mathbb{P}(A_{k_1} \cap A_{k_2} \cap \ldots \cap A_{k_m}) = \mathbb{P}(A_{k_1})\mathbb{P}(A_{k_2}) \ldots \mathbb{P}(A_{k_m}).$$

Definition 3.1.12 (Independent $\sigma-$Algebra)**.** Let $\mathcal{F}_j \subset \mathcal{F}$ be $\sigma-$algebras for $j = 1, 2, \ldots$.

Then we say $\mathcal{F}_j$ are independent if for $\forall 1 \leq k_1 < k_2 < \ldots < k_m$ and $\forall A_{k_j} \in \mathcal{F}_{k_j}$ it holds

$$\mathbb{P}(A_{k_1} \cap A_{k_2} \cap \ldots \cap A_{k_m}) = \mathbb{P}(A_{k_1})\mathbb{P}(A_{k_2}) \ldots \mathbb{P}(A_{k_m}).$$

Definition 3.1.13 (Independent Random Variables)**.** We say random variables $X_1, \ldots, X_m : \Omega \to \mathbb{R}^d$ are independent if for $\forall B_1, \ldots, B_m \subset \mathcal{B}$ in $\mathbb{R}^d$ it holds

$$\mathbb{P}(X_{j_1} \in B_{j_1}, \ldots, X_{j_k} \in B_{j_k}) = \mathbb{P}(X_{j_1} \in B_{j_1}) \ldots \mathbb{P}(X_{j_k} \in B_{j_k}).$$

which is equivalent to proving that $\mathcal{U}(X_1), \ldots, \mathcal{U}(X_k)$ are independent

Theorem 3.1.1. $X_1, \ldots, X_m : \Omega \to \mathbb{R}^d$ are independent if and only if

$$F_{X_1, \ldots, X_m}(x_1, \ldots, x_m) = F_{X_1}(x_1) \ldots F_{x_m}(x_m) \quad \forall x_i \in \mathbb{R}^d.$$

Theorem 3.1.2. If $X_1, \ldots, X_m : \Omega \to \mathbb{R}$ are independent and $\mathbb{E}[|X_i|] < \infty$ then

$$\mathbb{E}[|X_1, \ldots, X_m|] < \infty.$$

and

$$\mathbb{E}[X_1 \ldots X_m] = \mathbb{E}[X_1] \ldots \mathbb{E}[X_m].$$

Theorem 3.1.3. $X_1, \ldots, X_m : \Omega \to \mathbb{R}$ are independent and $\mathbb{V}[X_i] < \infty$ then

$$\mathbb{V}[X_1 + \ldots + X_m] = \mathbb{V}[X_1] + \ldots + \mathbb{V}[X_m].$$

Exercise. Proof the above theorems

3.1.2 Borel Cantelli

Definition 3.1.14. Let $A_1, \ldots, A_m, \ldots \in \mathcal{F}$, the set

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega : \omega \text{ belongs to infinite many } A_m\text{'s}\}.$$

is called A_m infinitely often or A_m i.o.

Lemma 3.1.2 (Borel Cantelli)**.** If $\sum_{m=1}^{\infty} \mathbb{P}(A_m) < \infty$, then $\mathbb{P}(A_m \text{ i.o. }) = 0$

Proof. By definition we have

$$\mathbb{P}(A_m \text{ i.o. }) \leq \mathbb{P}(\bigcup_{m=n}^{\infty}) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \xrightarrow{m \to \infty} 0.$$

□

Definition 3.1.15 (Convergence in Probability)**.** We say a sequence of random variables $(X_k)_{k=1}^{\infty}$ converges in probability to X if $\forall \varepsilon > 0$ it holds

$$\lim_{k \to \infty} \mathbb{P}(|X_k - X| > \varepsilon) = 0.$$

Theorem 3.1.4 (Application of Borel Cantelli)**.** If $X_k \to X$ in probability, then there exists a subsequence $(X_{k_j})_{j=1}^{\infty}$ such that

$$X_{k_j}(\omega) \to X(\omega) \text{ for almost surely } \omega \in \Omega.$$

This means that $\mathbb{P}(|X_{k_j} - X| \to 0) = 1$.

Proof. Due to the convergence in probability, for each j, there exists k_j with $k_j > k_{j-1}$ such that

$$\mathbb{P}(|X_{k_j} - X| > \frac{1}{j}) \leq \frac{1}{j^2},$$

which implies that

$$\sum_{j=1}^{\infty} \mathbb{P}(|X_{k_j} - X| > \frac{1}{j}) = \sum_{j=1}^{\infty} \frac{1}{j^2} < \infty.$$

Let $A_j = \{\omega \ : \ |X_{k_j} - X| > \frac{1}{j}\}$, then by Borel Cantelli we have $\mathbb{P}(A_j \text{ i.o.}) = 0$. This means for almost surely $\omega \in \Omega$, there exists a $J \in \mathbb{N}$ such that $\forall j > J$, it holds

$$|X_{k_j}(\omega) - X(\omega)| \leq \frac{1}{j}.$$

□

3.1.3 Strong Law of Large Numbers

Definition 3.1.16. A sequence of random variables $X_1, \ldots, X_n$ is called identically distributed if

$$F_{X_1}(x) = F_{X_2}(x) = \ldots = F_{X_n}(x) \quad \forall x \in \mathbb{R}^d.$$

If additionally $X_1, \ldots, X_n$ are independent then we say they are identically-independent-distributed i.i.d.

Theorem 3.1.5 (Strong Law Of Large Numbers)**.** Let $X_1, \ldots, X_N$ be a sequence of i.i.d integrable random variables on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then

$$\mathbb{P}(\lim_{N\to\infty} \frac{X_1 + \ldots + X_N}{N} = \mathbb{E}[X_i]) = 1.$$

where $\mathbb{E}[X_i] = \mathbb{E}[X_j]$.

Proof. Suppose for simplicity $\mathbb{E}[X_i^4] < \infty$ for $\forall i = 1, 2, \ldots$. Then without loss of generality we may assume $\mathbb{E}[X_i] = 0$ otherwise we use $X_i - \mathbb{E}[X_i]$ as our new sequence. Consider

$$\mathbb{E}[(\sum_{i=1}^{N} X_i)^4] = \sum_{i,j,k,l} \mathbb{E}[X_i X_j X_k X_l].$$

If $i \neq j, k, l$ then because of independence it follows that

$$\mathbb{E}[X_i X_j X_k X_l] = \mathbb{E}[X_i]\mathbb{E}[X_j X_k X_l] = 0.$$

Then

$$\mathbb{E}[(\sum_{i=1}^{N} X_i)^4] = \sum_{i=1}^{N} \mathbb{E}[X_i^4] + 3\sum_{i\neq j} \mathbb{E}[X_i^2 X_j^2]$$
$$= N\mathbb{E}[X_1^4] + 3(N^2 - N)\mathbb{E}[X_1^2]^2$$
$$\leq N^2 C.$$

Therefore for fixed $\varepsilon > 0$

$$\mathbb{P}(|\frac{1}{N}\sum_{i=1}^{N} X_i| \geq \varepsilon) = \mathbb{P}(|\sum_{i=1}^{N} X_i|^4 \geq (\varepsilon N)^4)$$
$$\overset{\text{Mrkv.}}{\leq} \frac{1}{(\varepsilon N)^4}\mathbb{E}[|\sum_{i=1}^{N} X_i|^4]$$
$$\leq \frac{C}{\varepsilon^4}\frac{1}{N^2}.$$

Then by Borel Cantelli we get

$$\mathbb{P}(|\frac{1}{N}\sum_{i=1}^{N} X_i| \geq \varepsilon \text{ i.o.}) = 0.$$

because

$$\sum_{N=1}^{\infty} \mathbb{P}(A_N) = \sum_{N=1}^{\infty} \frac{C}{\varepsilon^4}\frac{1}{N^2} < \infty.$$

where

$$A_N = \{\omega \in \Omega \ : \ |\frac{1}{N}\sum_{i=1}^{N} X_i| \geq \varepsilon\}.$$

Now we take $\varepsilon = \frac{1}{k}$ then the above gives

$$\lim_{N\to\infty} \sup \frac{1}{N}\sum_{i=1}^{N} X_i(\omega) \leq \frac{1}{k}.$$

holds except for $\omega \in B_k$ with $\mathbb{P}(B_k) = 0$. Let $B = \bigcup_{k=1}^{\infty} B_k$ then $\mathbb{P}(B) = 0$ and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i(\omega) = 0 \text{ a.s.}.$$

□

3.1.4 Conditional Expectation

We use in this part $(\Omega, \mathcal{F}, \mathbb{P})$ as the probability space.

Definition 3.1.17. Let Y be random variable, then $\mathbb{E}[X|Y]$ is defined as a $\mathcal{U}(Y)-$measurable random variable s.t for $\forall A \in \mathcal{U}(Y)$ it holds

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}[X|Y] d\mathbb{P}.$$

Definition 3.1.18. Let $\mathcal{U} \subset \mathcal{F}$ be a $\sigma-$algebra, if $X : \Omega \to \mathbb{R}^d$ is an integrable random variable then $\mathbb{E}[X|\mathcal{U}]$ is defined as a random variable on Ω s.t. $\mathbb{E}[X|\mathcal{U}]$ is $\mathcal{U}-$measurable and for $A \in \mathcal{U}$

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{U}] d\mathbb{P}.$$

Exercise. Proof the following equalities

1. $\mathbb{E}[X|Y] = \mathbb{E}[X|\mathcal{U}(Y)]$
2. $\mathbb{E}[\mathbb{E}[X|\mathcal{U}]] = \mathbb{E}[X]$
3. $\mathbb{E}[X] = \mathbb{E}[X|\mathcal{W}]$, where $\mathcal{W} = \{\emptyset, \Omega\}$

Remark. One can define the conditional probability similarly. Let $\mathcal{V} \subset \mathcal{U}$ be a $\sigma-$algebra then for $A \in \mathcal{U}$ the conditional probability is defined as follows

$$\mathbb{P}(A|\mathcal{V}) = \mathbb{E}[\mathbb{1}_A|\mathcal{V}],$$

where $\mathbb{1}_A$ is the indication function of set A.

Theorem 3.1.6. Let X be an integrable random variable, then for all $\sigma-$algebras $\mathcal{U} \subset \mathcal{F}$ the conditional expectation $\mathbb{E}[X|\mathcal{U}]$ exists and is unique up to $\mathcal{U}-$measurable sets of probability zero.

Proof. Omit □

Theorem 3.1.7 (Properties Of Conditional Expectation)**.**

1. If X is $\mathcal{U}-$measurable then $\mathbb{E}[X|\mathcal{U}] = X$ a.s.
2. $\mathbb{E}[aX + bY|\mathcal{U}] = a\mathbb{E}[X|\mathcal{U}] + b\mathbb{E}[Y|\mathcal{U}]$
3. If X is $\mathcal{U}-$measurable and XY is integrable then
$$\mathbb{E}[XY|\mathcal{U}] = X\mathbb{E}[Y|\mathcal{U}].$$
4. If X is independent of $\mathcal{U}$ then $\mathbb{E}[X|\mathcal{U}] = \mathbb{E}[X]$ a.s.
5. If $\mathcal{W} \subset \mathcal{U}$ are two $\sigma-$algebras then
$$\mathbb{E}[X|\mathcal{W}] = \mathbb{E}[\mathbb{E}[X|\mathcal{U}]|\mathcal{W}] = \mathbb{E}[\mathbb{E}[X|\mathcal{W}]|\mathcal{U}] \text{ a.s.}$$
6. If $X \le Y$ a.s. then $\mathbb{E}[X|\mathcal{U}] \le \mathbb{E}[Y|\mathcal{U}]$ a.s.

Exercise. Proof the above properties

Lemma 3.1.3 (Conditional Jensen's Inequality)**.** Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}[\varphi(x)] < \infty$, then

$$\varphi(\mathbb{E}[X|\mathcal{U}]) \le \mathbb{E}[\varphi(X)|\mathcal{U}].$$

Exercise. Proof the above Lemma

3.1.5 Stochastic Processes And Brownian Motion

Definition 3.1.19 (Stochastic Process)**.** A stochastic process is a parameterized collection of random variables

$$X \; : \; [0,T] \times \Omega \to \mathbb{R}^d \; : \; (t,\omega) \mapsto X(t,\omega).$$

For $\omega \in \Omega$, the map

$$X(\cdot,\omega) \; : \; [0,T] \to \mathbb{R}^d \; : \; t \mapsto X(t,\omega).$$

is called sample path.

Definition 3.1.20 (Modification and Indistinguishable)**.** Let $X(\cdot)$ and $Y(\cdot)$ be two stochastic processes, then we say they are modifications of each other if

$$\mathbb{P}(X(t) = Y(t)) = 1 \qquad \forall t \in [0,T].$$

We say that they are indistinguishable if

$$\mathbb{P}(X(t) = Y(t) \; \forall t \in [0,T]) = 1.$$

Remark. Note that if two stochastic processes are indistinguishable then they are also always a modification of each other, the reverse is not always true.

Definition 3.1.21 (History)**.** Let $X(t)$ be a real valued process. The $\sigma-$algebra

$$\mathcal{U}(t) := \mathcal{U}(X(s) \mid 0 \leq s \leq t).$$

is called the history of X until time $t \geq 0$.

Definition 3.1.22 (Martingale)**.** Let $X(t)$ be a real valued process and $\mathbb{E}[|X(t)|] < \infty$ for $\forall t \geq 0$

1. If $X(s) = \mathbb{E}[X(t)|\mathcal{U}(s)]$ a.s. $\forall t \geq s \geq 0$, then $X(\cdot)$ is called a martingale
2. If $X(s) \lesseqqgtr \mathbb{E}[X(t)|\mathcal{U}(s)]$ a.s. $\forall t \geq s \geq 0$, then $X(\cdot)$ is called a (super) sub-martingale

Lemma 3.1.4. Suppose $X(\cdot)$ is a real-valued martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function. If $\mathbb{E}[|\varphi(X(t))|] < \infty$ for $\forall t \geq 0$ then $\varphi(X(\cdot))$ is a sub-martingale

We leave the proof of this lemma as an exercise. Hint: apply Jensen's inequality.

Theorem 3.1.8 (Martingale-Inequalities)**.** Assume $X(\cdot)$ is a process with continuous sample paths a.s.

1. If $X(\cdot)$ is a sub-martingale then $\forall \lambda > 0$, $t \geq 0$ it holds

$$\mathbb{P}(\max_{0 \leq s \leq t} X(s) \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}[X(t)^+].$$

2. If $X(\cdot)$ is a martingale and $1 < p < \infty$ then

$$\mathbb{E}[\max_{0 \leq s \leq t} |X(s)|^p] \leq (\frac{p}{p-1})^p \mathbb{E}[|X(t)|^p].$$

Proof. Omit □

3.1.6 Brownian Motion

Definition 3.1.23 (Brownian Motion)**.** A real valued stochastic process $W(\cdot)$ is called a Brownian motion or Wiener process if

1. $W(0) = 0$ a.s.
2. $W(t)$ is continuous a.s.
3. $W(t) - W(s) \sim \mathcal{N}(0, t-s)$ for $\forall t \geq s \geq 0$
4. $\forall\, 0 < t_1 < t_2 < \ldots < t_n$, $W(t_1), W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$ are independent

Remark. One can derive directly that

$$\mathbb{E}[W(t)] = 0, \quad \mathbb{E}[W^2(t)] = t, \quad \mathbb{E}[W(t)W(s)] = t \wedge s \qquad \forall t, s \geq 0.$$

The first two facts are trivial from the normal distribution, the third one can obtained inthe following, for $t \geq s \geq 0$

$$\begin{aligned}\mathbb{E}[W(t)W(s)] &= \mathbb{E}[(W(t) - W(s))(W(s))] + \mathbb{E}[(W(s)w(s))] \\ &= \mathbb{E}[W(t) - W(s)]\mathbb{E}[W(s)] + \mathbb{E}[W(s)W(s)] = s.\end{aligned}$$

Definition 3.1.24. An $\mathbb{R}^d$ valued process $W(\cdot) = (W^1(\cdot), \ldots, W^d(\cdot))$ is a $d-$dimensional Wiener process (or Brownian motion) if

1. $W^k(\cdot)$ is a 1-D Wiener process for $\forall k = 1, \ldots, d$
2. $\mathcal{U}(W^k(t)\ ,\ t \geq 0)$, the $\sigma-$algebras, are independent $k = 1, \ldots, d$.

Remark. If $W(\cdot)$ is a $d-$Dimensional Brownian motion, then $W(t) \sim \mathcal{N}(0, t)$ and for any Borel set $A \subset \mathbb{R}^d$

$$\mathbb{P}(W(t) \in A) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \int_A e^{-\frac{|x|^2}{2t}}\, dx.$$

Theorem 3.1.9. If $X(\cdot)$ is a given stochastic process with a.s. continuous sample paths and

$$\mathbb{E}[|X(t) - X(s)|^\beta] \leq C|t-s|^{1+\alpha}.$$

Then for $0 < \gamma < \frac{\alpha}{\beta}$, $T > 0$, and a.s. ω, there exists $K = K(\omega, \gamma, T)$ s.t.

$$|X(t,\omega) - X(s,\omega)| \leq K|t-s|^\gamma \quad \forall 0 \leq s, t \leq T.$$

Proof. Omit □

An application of this result on Brownian motion is interesting since

$$\mathbb{E}[|W(t) - W(s)|^{2m}] \leq C|t-s|^m \text{ holds for all } m \in \mathbb{N}.$$

we get immediately

$$W(\cdot, \omega) \in \mathcal{C}^\gamma([0,T]) \quad 0 < \gamma < \frac{m-1}{2m} < \frac{1}{2}, \forall m \gg 1.$$

This means that Brownian motions is a.s. path Hölder continuous up to exponent $\frac{1}{2}$.

Remark. One can also further prove that the path wise smoothness of Brownian motion can not be better than Hölder continuous. Namely

1. $\forall \gamma \in (\frac{1}{2}, 1]$ and a.s. ω, $t \mapsto W(t, \omega)$ is nowhere Hölder continuous with exponent γ
2. $\forall$ a.s. $\omega \in \Omega$ the map $t \mapsto W(t, \omega)$ is nowhere differentiable and is of infinite variation on each subinterval.

Definition 3.1.25 (Markov Property)**.** An $\mathbb{R}^d-$valued process $X(\cdot)$ is said to have the Markov property, if $\forall 0 \le s \le t$ and $\forall B \subset \mathbb{R}^d$ Borel set, it holds

$$\mathbb{P}(X(t) \in B | \mathcal{U}(s)) = \mathbb{P}(X(t) \in B | X(s)) \text{ a.s..}$$

Remark. The $d-$Dimensional Wiener Process $W(\cdot)$ has Markov property and it holds almost surely that

$$\mathbb{P}(W(t) \in B | W(s)) = \frac{1}{(2\pi(t-s))^{\frac{n}{2}}} \int_B e^{-\frac{|x-W(s)|^2}{2(t-s)}} dx.$$

3.1.7 Convergence of Measure and Random Variables

In the following we include a couple of definitions for the convergence of measures and random variables

Theorem 3.1.10 (Weak convergence of measures)**.** The following statements are equivalent

1. $\mu_n \rightharpoonup \mu$
2. For $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$ it holds $\int f d\mu_n \to \int f d\mu$.
3. For $\forall B \in \mathcal{B}$, it holds $\mu_n(B) \to \mu(B)$.
4. For $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$ uniform continuous it holds $\int f d\mu_n \to \int f d\mu$.

Theorem 3.1.11 (Weak convergence of Random variable)**.** The following statements are equivalent

1. X_n converges weakly in Law to X, i.e. $X_n \rightharpoonup X$.
2. For $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$ it holds $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(x)]$.

Exercise. Prove that

$$X_n \to X \text{ a.s.} \Rightarrow \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \to \infty} 0 \Rightarrow X_n \rightharpoonup X.$$

Definition 3.1.26 (Tightness)**.** A set of probability measures $S \subset \mathcal{P}(\mathbb{R}^d)$ is called tight, if for $\forall\ \varepsilon > 0$ there exists $\exists\ K \subset \mathbb{R}^d$ compact such that

$$\sup_{\mu \in S} \mu(K^c) \le \varepsilon.$$

Theorem 3.1.12 (Prokhorov's theorem)**.** A sequence of measures $(\mu_n)_{n \in \mathbb{N}}$ is tight in $\mathcal{P}(\mathbb{R}^d)$ iff any subsequence has a weakly convergences subsequence.

Proof. Refer to literature ☐

3.2 Itô Integral

From now on we denote by $W(\cdot)$ the $1-D$ Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$

Definition 3.2.1.

1. $\mathcal{W}(t) = \mathcal{U}(W(s)|0 \leq s \leq t)$ is called the history up to t
2. The $\sigma-$algebra

$$\mathcal{W}^{+}(t) := \mathcal{U}(W(s) - W(t)|s \geq t).$$

is called the future of the Brownian motion beyond time t.

Definition 3.2.2 (Non-Anticipating Filtration)**.** A family $\mathcal{F}(\cdot)$ of $\sigma-$algebras is called non-anticipating (w.r.t $W(\cdot)$) if

1. $\mathcal{F}(t) \supseteq \mathcal{F}(s)$ for $\forall t \geq s \geq 0$
2. $\mathcal{F}(t) \supseteq \mathcal{W}(t)$ for $\forall t \geq 0$
3. $\mathcal{F}(t)$ is independent of $\mathcal{W}^{+}(t)$ for $\forall t \geq 0$

A primary example of this is

$$\mathcal{F}(t) := \mathcal{U}(W(s), 0 \leq s \leq t, X_0).$$

where X_0 is a random variable independent of $\mathcal{W}^{+}(0)$

Definition 3.2.3 (Non-Anticipating Process)**.** A real-valued stochastic process $G(\cdot)$ is called non-anticipating (w.r.t. $\mathcal{F}(\cdot)$) if for $\forall t \geq 0$, $G(t)$ is $\mathcal{F}(t)-$measurable.

From now on we use $(\Omega, \mathcal{F}, \mathcal{F}(t), \mathbb{P})$ as a filtered probability space with right continuous filtration $\mathcal{F}(t) = \bigcap_{s \geq t} \mathcal{F}(s)$. Note we also use the convention that $\mathcal{F}(t)$ is complete .

Definition 3.2.4.

1. A stochastic process is adapted to $(\mathcal{F}(t))_{t \geq 0}$ if X_t is $\mathcal{F}(t)$ measurable for $\forall t \geq 0$
2. A stochastic process is progressively measurable w.r.t. $\mathcal{F}(t)$ if

$$X(s, \omega) \ : \ [0, t] \times \Omega \to \mathbb{R}.$$

is $\mathcal{B}([0, t]) \times \mathcal{F}(t)$ measurable for $\forall t > 0$.

Definition 3.2.5. We denote $\mathbb{L}^2([0, T])$ the space of all real-valued progressively measurable stochastic processes $G(\cdot)$ s.t.

$$\mathbb{E}[\int_0^T G^2 dt] < \infty.$$

We denote $\mathbb{L}^1([0, T])$ the space of all real-valued progressively measurable stochastic processes

$F(\cdot)$ s.t.

$$\mathbb{E}[\int_0^T |F|dt] < \infty.$$

Definition 3.2.6 (Step-Process)**.** $G \in \mathbb{L}^2([0,T])$ is called a step process if there exists a partition of the interval $[0,T]$ i.e. $P = \{(t_0, t_1, \ldots, t_m) : 0 = t_0 < t_1 < \ldots < t_m = T\}$ s.t.

$$G(t) = G_k \quad \forall t_k \le t < t_{k+1} \quad k = 0, \ldots, m-1,$$

where G_k is an $\mathcal{F}(t_k)$ measurable random variable.

Remark. Note that the above definition directly yields the following representation for any step process $G \in \mathbb{L}^2([0,T])$

$$G(t,\omega) = \sum_{k=0}^{m-1} G_k(\omega) \cdot \mathbb{1}_{[t_k,t_{k+1})}(t).$$

Definition 3.2.7 ((Simple) Itô Integral)**.** Let $G \in \mathbb{L}^2([0,T])$ be a step process. Then we define

$$\int_0^T G(t,\omega)dW_t := \sum_{k=0}^{m-1} G_k(\omega) \cdot (W(t_{k+1},\omega) - W(t_k,\omega)).$$

Proposition 3.2.1. Let $G, H \in \mathbb{L}^2([0,T])$ be two step processes, then for $\forall a, b \in \mathbb{R}$ it holds

1. $\int_0^T (aG + bH)dW_t = a\int_0^T GdW_t + b\int_0^T HdW_t$
2. $\mathbb{E}[\int_0^T GdW_t] = 0$.

Proof. (1). This case is easy. Set

$$\begin{aligned} G(t) &= G_k \quad t_k \le t < t_{k+1} \quad k = 0, \ldots, m_1 - 1 \\ H(t) &= H_l \quad t_l \le t < t_{l+1} \quad l = 0, \ldots, m_2 - 1. \end{aligned}$$

Let $0 \le t_0 < t_1 < \ldots \le t_n = T$ be the collection of t_k's and t_k's which together form a new partition of $[0,T]$ then obviously $G, H \in \mathbb{L}^2([0,T])$ are again step processes on this new partition. We have directly the linearity by definition on the Itô integral for step processes

$$\int_0^T (G+H)dW_t = \sum_{j=0}^{n-1}(G_j + H_j) \cdot (W(t_{j+1}) - W(t_j)).$$

(2). By definition we have

$$\mathbb{E}[\int_0^T GdW_t] = \mathbb{E}[\sum_{k=0}^{m-1} G_k(W(t_{k+1}) - W(t_k))] = \sum_{k=0}^{m-1} \mathbb{E}[G_k(W(t_{k+1}) - W(t_k))].$$

Notice that G_k by definition is $\mathcal{F}_{t_k}$ measurable and $W(t_{k+1}) - W(t_k)$ is measurable in $\mathcal{W}^+(t_k)$. Since $\mathcal{F}_{t_k}$ is independent of $\mathcal{W}^+(t_k)$, we can deduce that G_k is independent of $W(t_{k+1}) - W(t_k)$

which implies

$$\sum_{k=0}^{m-1} \mathbb{E}[G_k(W(t_{k+1}) - W(t_k))] = \sum_{k=0}^{m-1} \mathbb{E}[G_k] \cdot \mathbb{E}[W(t_{k+1}) - W(t_k)] = 0.$$

□

Lemma 3.2.1 ((Simple) Itô isometry)**.** For step processes $G \in \mathbb{L}^2([0,T])$ we have

$$\mathbb{E}[(\int_0^T G dW_t)^2] = \mathbb{E}[\int_0^T G^2 dt].$$

Proof. By definition we can write

$$\mathbb{E}[\left(\int_0^T G dW_t\right)^2] = \sum_{k,j=0}^{m-1} \mathbb{E}[G_k G_j (W(t_{k+1}) - W(t_k))(W(t_{j+1}) - W(t_j))].$$

If $j < k$, then $W(t_{k+1}) - W(t_k)$ is independent of $G_k G_j (W(t_{j+1}) - W(t_j))$. Therefore

$$\sum_{j<k} \mathbb{E}[\dots] = 0 \quad \text{and} \quad \sum_{j>k} \mathbb{E}[\dots] = 0.$$

Then we have

$$\begin{aligned}\mathbb{E}[\left(\int_0^T G dW_t\right)^2] &= \sum_{k=0}^{m-1} \mathbb{E}[G_k^2 (W(t_{k+1}) - W(t_k))^2] \\ &= \sum_{k=0}^{m-1} \mathbb{E}[G_k^2]\mathbb{E}[(W(t_{k+1}) - W(t_k))^2] = \sum_{k=0}^{m-1} \mathbb{E}[G_k^2](t_{k+1} - t_k) = \mathbb{E}[\int_0^T G^2 dt].\end{aligned}$$

□

For general $\mathbb{L}^2([0,T])$ processes we use approximation by step processes to define the Itô integral

Lemma 3.2.2. If $G \in \mathbb{L}^2([0,T])$ then there exists a sequence of bounded step processes $G^n \in \mathbb{L}^2([0,T])$ s.t.

$$\mathbb{E}[\int_0^T |G - G^n|^2 dt] \xrightarrow{n \to \infty} 0.$$

We roughly sketch the Idea here and refer the rigorous proof to stochastic calculus lecture.

If $G(\cdot, \omega)$ is a.s. continuous then we can take

$$G^n(t) := G(\frac{k}{n}) \quad \frac{k}{n} \le t < \frac{k+1}{n} \quad k = 0, \dots, \lfloor nT \rfloor.$$

For general $G \in \mathbb{L}^2([0,T])$ let

$$G^m(t) := \int_0^t m e^{m(s-t)} G(s) ds.$$

Then $G^m \in \mathbb{L}^2([0,T])$, $t \mapsto G^m(t,\omega)$ is continuous for a.s. ω and

$$\int_0^T |G - G^m|^2 dt \to 0 \text{ a.s..}$$

Definition 3.2.8 (Itô Integral)**.** If $G \in \mathbb{L}^2([0,T])$. Let step processes G^n be an approximation of G. Then we define the Itô integral by using the limit

$$I(G) = \int_0^T G dW_t := \lim_{n\to\infty} \int_0^T G^n dW_t,$$

where the limit exists in $L^2(\Omega)$.

In order to derive the validity of this definition, one has to check

1. Existence of the limit. This can be obtained by showing that it is a Cauchy sequence, namely by Itô isometry we have

$$\mathbb{E}[\Big(\int_0^T (G^m - G^n) dW_t\Big)^2] = \mathbb{E}[\int_0^T |G^m - G^n|^2 dt] \xrightarrow{n,m\to\infty} 0.$$

   This implies $\int_0^T G^n dW_t$ has a limit in $L^2(\Omega)$ as $n \to \infty$

2. The limit is independent of the choice of approximation sequences. Let $\tilde{G}^n$ be another step process which converges to G. Then we have

$$\mathbb{E}[\int_0^T |\tilde{G}^n - G^n|^2 dt] \le \mathbb{E}[\int_0^T |G^n - G|^2 dt] + \mathbb{E}[\int_0^T |\tilde{G}^n - G|^2 dt].$$

   it follows that

$$\mathbb{E}[\Big(\int_0^T \tilde{G}^n dW_t - \int_0^T G^n dW_t\Big)^2] = \mathbb{E}[\int_0^T |\tilde{G}^n - G^n|^2 dt] \to 0.$$

By using this approximation, all the properties for step processes can be obtained for general $\mathbb{L}^2([0,T])$ processes

Theorem 3.2.1 (Properties of the Itô Integral)**.** For $\forall a, b \in \mathbb{R}$ and $\forall G, H \in \mathbb{L}^2([0,T])$ it holds

1. $\int_0^T (aG + bH) dW_t = a\int_0^T G dW_t + b\int_0^T H dW_t$
2. $\mathbb{E}[\int_0^T G dW_t] = 0$
3. $\mathbb{E}[\int_0^T G dW_t \cdot \int_0^T H dW_t] = \mathbb{E}[\int_0^T GH dt]$

Lemma 3.2.3 (Itô Isometry)**.** For general $G \in \mathbb{L}^2([0,T])$ we have

$$\mathbb{E}[\Big(\int_0^T G dW_t\Big)^2] = \mathbb{E}[\int_0^T G^2 dt].$$

Proof. Choose step processes $G_n \in \mathbb{L}^2([0,T])$ such that $G_n \to G$ (in the sense previously defined) then by Definition 3.2.8 we get

$$\|I(G) - I(G_n)\|_{L^2} \xrightarrow{n\to\infty} 0.$$

Then using the simple version of Itô isometry one obtains

$$\mathbb{E}[\Big(\int_0^T G dW_t\Big)^2] = \lim_{n\to\infty} \mathbb{E}[\Big(\int_0^T G_n dW_t\Big)^2] = \lim_{n\to\infty} \mathbb{E}[\int_0^T (G_n)^2 dt] = \mathbb{E}[\int_0^T (G)^2 dt].$$

$\square$

Remark. The Itô integral is a map from $\mathbb{L}^2([0,T])$ to $L^2(\Omega)$

Remark. For $G \in \mathbb{L}^2([0,T])$ the Itô integral $\int_0^\tau G dW_t$ with $0 \leq \tau \leq T$ is a martingale. We also refer the proof of this statement to stochastic calculus lecture.

3.2.1 Itô's Formula

Definition 3.2.9 (Itô Process)**.** Let $X(\cdot)$ be a real-valued process given by

$$X(r) = X(s) + \int_s^r F dt + \int_s^r G dW_t.$$

for some $F \in \mathbb{L}^1([0,T])$ and $G \in \mathbb{L}^2([0,T])$ for $0 \leq s \leq r \leq T$, then $X(\cdot)$ is called Itô process. Furthermore we say $X(\cdot)$ has a stochastic differential.

$$dX = F dt + g dW_t \quad \forall 0 \leq t \leq T.$$

Theorem 3.2.2 (Itô's Formula)**.** Let $X(\cdot)$ be an Itô process given by $dX = F dt + G dW_t$ for some $F \in \mathbb{L}^1([0,T])$ and $G \in \mathbb{L}^2([0,T])$. Assume $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ is continuous and $\frac{\partial u}{\partial t}, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}$ exists and are continuous. Then $Y(t) := u(X(t),t)$ satisfies

$$\begin{aligned} dY &= \frac{\partial u}{\partial t} dt + \frac{\partial u}{\partial x} dX + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} G^2 dt \\ &= (\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} F + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} G^2) dt + \frac{\partial u}{\partial x} G dW_t. \end{aligned}$$

Note that the differential form of the Itô formula is understood as an abbreviation of the following integral form, for all $0 \leq s < r \leq T$

$$\begin{aligned} & u(X(r),r) - u(X(s),s) \\ &= \int_s^r (\frac{\partial u}{\partial t}(X(t),t) + \frac{\partial u}{\partial x}(X(t),t)F(t) + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(X(t),t)G^2(t))dt + \int_s^r \frac{\partial u}{\partial x}(X(t),t)G(t)dW_t. \end{aligned}$$

Proof. The proof is split into five steps

Step 1. First we prove two simple cases. If $X(t) = W_t$ then

(1) $d(W_t)^2 = 2W_t dW_t + dt$

(1) $d(tW_t) = W_t dt + t dW_t$

The integral version of (1) is $W_t^2 - W_0^2 = \int_0^t 2W_s dW_s + t$ a.s.

By definition of Itô integral, for a.s. $\omega \in \Omega$ we have

$$\begin{aligned}
\int_0^t 2W_s dW_s &= 2\lim_{n\to\infty}\sum_{k=0}^{n-1} W(t_k^n)\left(W(t_{k+1}^n) - W(t_k^n)\right) \\
&= \lim_{n\to\infty}\Bigg[\sum_{k=0}^{n-1} W(t_k^n)\left(W(t_{k+1}^n) - W(t_k^n)\right) - \sum_{k=0}^{n-1}\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 \\
&\qquad + \sum_{k=0}^{n-1} W(t_{k+1}^n)\left(W(t_{k+1}^n) - W(t_k^n)\right)\Bigg] \\
&= -\lim_{n\to\infty}\left[\sum_{k=0}^{n-1}\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 - \sum_{k=0}^{n-1}\left(W(t_k^n)\right)^2 + \sum_{k=0}^{n-1}\left(W(t_{k+1}^n)\right)^2\right] \\
&= -\lim_{n\to\infty}\sum_{k=0}^{n-1}\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 + (W(t))^2 - (W(0))^2.
\end{aligned}$$

where for any fixed n, the partition of $[0,T]$ is given by $0 \le t_0^n < t_1^n < \ldots < t_n^n = T$ and $t_k^n - t_{k+1}^n = \frac{1}{n}$. It remains to prove that the limit

$$\lim_{n\to\infty}\sum_{k=0}^{n-1}\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 - t = 0.$$

Actually, by writing out the square we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{k=0}^{n-1}\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 - \left(t_{k+1}^n - t_k^n\right)\right)^2\right] \\
=&\mathbb{E}\left[\sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\left(\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 - \left(t_{k+1}^n - t_k^n\right)\right)\cdot\left(\left(W(t_{l+1}^n) - W(t_l^n)\right)^2 - \left(t_{l+1}^n - t_l^n\right)\right)\right].
\end{aligned}$$

The terms with $k \neq l$ vanish because of the independence. Therefore

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=0}^{n-1}\left(\left(W(t_{k+1}^n) - W(t_k^n)\right)^2 - \left(t_{k+1}^n - t_k^n\right)\right)^2\right] \\
&= \sum_{k=0}^{n-1}(t_{k+1}^n - t_k^n)^2\mathbb{E}\left[\left(\frac{\left(W(t_{k+1}^n) - W(t_k^n)\right)^2}{t_{k+1}^n - t_k^n} - 1\right)^2\right] \\
&= \sum_{k=0}^{n-1}(t_{k+1}^n - t_k^n)^2\mathbb{E}\left[\left(\left(\frac{W(t_{k+1}^n) - W(t_k^n)}{\sqrt{t_{k+1}^n - t_k^n}}\right)^2 - 1\right)^2\right] \\
&\le C\cdot\frac{t^2}{n} \to 0,
\end{aligned}$$

where we have used the fact that $Y = \frac{W(t_{k+1}^n) - W(t_k^n)}{\sqrt{t_{k+1}^n - t_k^n}} \sim \mathcal{N}(0,1)$. Hence $\mathbb{E}[(Y^2-1)^2]$ is bounded by a constant C

We will prove the integral form of (2) : $tW_t - 0W_0 = \int_0^t W_s ds + \int_0^t s dW_s$. Actually we have that

$$\begin{aligned}
\int_0^t s dW_s + \int_0^t W_s ds &= \lim_{n\to\infty}\sum_{k=0}^{n-1} t_k^n\left(W(t_{k+1}^n) - W(t_k^n)\right) + \lim_{n\to\infty}\sum_{k=0}^{n-1} W(t_{k+1}^n)(t_{k+1}^n - t_k^n) \\
&= W(t)\cdot t - 0\cdot W(0).
\end{aligned}$$

Step 2. Now let us prove the Itô product rule. Namely, if

$$dX_1 = F_1 dt + G_1 dW_t \quad \text{and} \quad dX_2 = F_2 dt + G_2 dW_t.$$

for some $G_i \in \mathbb{L}^2([0,T])$ and $F_i \in \mathbb{L}^1([0,T])$ $i = 1, 2$, then

$$d(X_1X_2) = X_2 dX_1 + X_1 dX_2 + G_1G_2 dt \quad = (X_2F_1 + X_1F_2 + G_1G_2)dt + (X_2G_1 + X_1G_2)dW_t.$$

where the above should be understood as the integral equation.

(1) We prove the case F_i, G_i are time independent. Assume for simplicity $X_1(0) = X_2(0)$ then it follows that

$$X_i(t) = F_i t + G_i W(t), \quad i = 1, 2..$$

Then it holds almost surely that

$$\begin{aligned}
&\int_0^t (X_2 dX_1 + X_1 dX_2 + G_1G_2 ds) \\
=&\int_0^t (X_2F_1 + X_1F_2)ds + \int_0^t (X_2G_1 + X_1G_2)dW_s + \int_0^t G_1G_2 ds \\
=&\int_0^t \left(F_1(F_2 s + G_2W(s)) + F_2(F_1 s + G_1W(s))\right) ds + G_1G_2 t \\
&+ \int_0^t \left(G_1(F_2 s + G_2W(s)) + G_2(F_1 s + G_1W(s))\right) dW_s \\
=&G_1G_2 t + F_1F_2 t^2 + (F_1G_2 + F_2G_1)\left(\int_0^t W(s)ds + \int_0^t s dW_s\right) \\
&+ 2G_1G_2 \int_0^t W(s) dW_s \\
=&G_1G_2(W(t))^2 + F_1F_2 t^2 + (F_1G_2 + F_2G_1) tW(t) = X_1(t) + X_2(t),
\end{aligned}$$

where we have used the results from Step 1. Therefore Itô formula is true when F_i, G_i are time independent random variables.

(2) If F_i, G_i are step processes, then we apply the above formula in each sub-interval.

(3) For $F_i \in \mathbb{L}^1([0,T])$ and $G_i \in \mathbb{L}^2([0,T])$, we take the step process approximation of them, namely

$$\mathbb{E}\Big[\int_0^T |F_i^n - F_i| dt\Big] \to 0 \quad \mathbb{E}\Big[\int_0^T |G_i^n - G_i|^2 dt\Big] \to 0 \qquad (n \to \infty), i = 1, 2.$$

Notice that for each Itô process given by step processes

$$X_i^n(t) = X_i(0) + \int_0^t F_i^n ds + \int_0^t G_i^n dW_s.$$

the product rule holds, i.e.

$$X_1^n(t)X_2^n(t) - X_1(0)X_2(0) = \int_0^t \left(X_1^n(s)dX_2^n(s) + X_2^n(s)dX_1^n(s) + G_1G_2 ds\right).$$

Therefore after taking the limit, we obtain that the product rule holds for Itô processes.

Step 3. If $u(X) = X^m$ for $m \in \mathbb{N}$ then we claim

$$d(X^m) = mX^{m-1}dX + \frac{1}{2}m(m-1)X^{m-2}G^2dt.$$

We prove this statement by induction.
IA Note that $m = 2$ is given by the product rule.
IV Suppose the formula holds for $m - 1 \in \mathbb{N}$
IS $m - 1 \to m$.
By using the product rule, we have that

$$\begin{aligned} d(X^m) = d(XX^{m-1}) &= Xd(X^{m-1}) + X^{m-1}dX + (m-1)X^{m-2}G^2dt \\ &\overset{\text{IV}}{=} X\left((m-1)X^{m-2}dX + \frac{1}{2}(m-1)(m-2)X^{m-3}G^2dt\right) \\ &\quad + X^{m-1}dX + (m-1)X^{m-2}G^2dt \\ &= mX^{m-1}dX + (m-1)(\frac{m}{2} - 1 + 1)X^{m-2}G^2dt. \end{aligned}$$

Thus the statement holds for all $m \in \mathbb{N}$.

Step 4. If $u(X,t) = f(X)g(t)$ where f and g are polynomials $f(X) = X^m$, $g(t) = t^n$. Then by the product rule we have

$$d(u(X,t)) = d(f(X)g(t)) = f(X)dg + gdf(X) + (G_1 \cdot 0)dt.$$

by step 3 this is equal to

$$f(X)g'(t)dt + g(t)f'(X)dX + \frac{1}{2}g(t)f''(X)G^2dt = \frac{\partial u}{\partial t}dt + \frac{\partial u}{\partial X}dX + \frac{1}{2}\frac{\partial^2 u}{\partial X^2}G^2dt.$$

Furthermore, by superposition, we know that the Itô formula is also true if $u(X,t) = \sum_{i=1}^m g_m(t)f_m(X)$ where f_m and g_m are polynomials

Step 5. For u continuous such that $\frac{\partial u}{\partial t}, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}$ exists and are also continuous, then there exists polynomial sequences u^n s.t.

$$u^n \to u \quad \frac{\partial u^n}{\partial t} \to \frac{\partial u}{\partial t}, \quad \frac{\partial u^n}{\partial x} \to \frac{\partial u}{\partial x}, \quad \frac{\partial^2 u}{\partial x^2} \to \frac{\partial^2 u}{\partial x^2}.$$

uniformly on compact $K \subset \mathbb{R} \times [0,T]$.

By using the fact that

$$u^n(X(t),t) - u^n(X(0),0) = \int_0^t \left(\frac{\partial u^n}{\partial t} + \frac{\partial u^n}{\partial x}F + \frac{1}{2}\frac{\partial^2 u^n}{\partial x^2}G^2\right)dr + \int_0^t \frac{\partial u^n}{\partial x}G dW_r \quad \text{a.s.}$$

Itô's formula is proven by taking the limit $n \to \infty$. □

Remark. One can get the existence of the polynomial sequence in Step 5, by using Hermetian polynomials

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}.$$

Exercise. If $u \in \mathcal{C}^\infty$, $\frac{\partial u}{\partial x} \in \mathcal{C}_b$ then prove Step 4 $\Rightarrow$ Step 5

Use Taylor expansion and use the uniform convergence of the Taylor series on compact support

3.2.2 Multi-Dimensional Itô processes and Formula

We shortly extend the definition of Itô processes and the Itô Formula to the multi-dimensional case, we include the dimensionality as a subscript for clearness.

Definition 3.2.10 (Multi-Dimensional Itô's Integral)**.** We the define the $n-$dimensional Itô integral for $G \in \mathbb{L}^2_{n\cdot m}([0,T])$, $G_{ij} \in \mathbb{L}^2([0,T])$ $1 \le i \le n$, $1 \le j \le m$

$$\int_0^T G dW_t = \begin{pmatrix} \vdots \\ \int_0^T G_{ij} dW_t^j \\ \vdots \end{pmatrix}_{n\times 1} .$$

Remark. It is a direct consequence from $1-$ Brownian motion that

$$\mathbb{E}\Big[\int_0^T G dW_t\Big] = 0$$
$$\mathbb{E}\Big[\Big(\int_0^T G dW_t\Big)^2\Big] = \mathbb{E}\Big[\int_0^T |G|^2 dt\Big],$$

where $|G|^2 = \sum_{i,j}^{n,m} |G_{ij}|^2$

Definition 3.2.11 (Multi-Dimensional Itô process)**.** We define the $n-$dimensional Itô process as

$$X(t) = X(s) + \int_s^t F_{n\times 1}(r) dr + \int_0^t G_{n\times m}(r) dW_{m\times 1}(r)$$

or differential version $\quad dX^i = F^i dt + \sum_{j=1}^m G^{ij} dW_t^i \qquad 1 \le i \le n.$

Theorem 3.2.3 (Multi Dimensional Itô's formula)**.** We define the $n-$dimensional Itô's formula for $u \in \mathcal{C}^{2,1}(\mathbb{R}^n \times [0,T], \mathbb{R})$ by

$$\begin{aligned} du(X(t),t) &= \frac{\partial u}{\partial t}(X(t),t)dt + \nabla u(X(t),t) \cdot dX(t) \\ &+ \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 u}{\partial x_i \partial x_j}(X(t),t) \sum_{l=1}^m G^{il} G^{il} dt. \end{aligned}$$

Proposition 3.2.2. For real valued processes X_1, X_2

$$\begin{cases} dX_1 & = F_1 dt + G_1 dW_1 \\ dX_2 & = F_2 dt + G_2 dW_2 \end{cases} \Rightarrow d(X_1X_2) = X dX_2 + X_2 dX_1 + \sum_{k=1}^m G_1^k G_2^k dt.$$

Remark (Multiplication Rules)**.** The following formal multiplication rules are frequently used in computation:

$$(dt)^2 = 0 \ , \ dt dW^k = 0 \ , \ dW^k dW^l = \delta_{kl} dt$$

.

Using the above we can write the Itô's formula into a short version as follows

$$
\begin{aligned}
du(X,t) &= \frac{\partial u}{\partial t}dt + \nabla u \cdot dX + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 u}{\partial x_i \partial x_j} dX^i dX^j \\
&= \frac{\partial u}{\partial t}dt + \sum_{i=1}^{n} \frac{\partial u}{\partial x^i} F^i dt + \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \sum_{i=1}^{m} G^{ik} dW_k \\
&+ \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 u}{\partial x_i \partial x_j} \left(F^i dt + \sum_{k=1}^{m} G^{ik} dW_k \right) \left(F^j dt + \sum_{l=1}^{m} G^{i;} dW_l \right) \\
&= (\frac{\partial u}{\partial t} + F \cdot \nabla u + \frac{1}{2} H \cdot D^2 u) dt + \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \sum_{k=1}^{m} G^{ik} dW_k,
\end{aligned}
$$

where

$$
\begin{aligned}
dX^i &= F^i dt + \sum_{k=1}^{m} G^{ik} dW_k \\
H_{ij} &= \sum_{k=1}^{m} G^{ik} G^{jk} \ , \ A \cdot B = \sum_{i,j=1}^{m} A_{ij} B_{ij}.
\end{aligned}
$$

Example. A typical example for G is

$$
G^T G = \sigma I_{n \times n}.
$$

Remark. If F and G are deterministic

$$
dX = F(t)dt + G(t)dW_t.
$$

Then for arbitrary test function $u \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ we have by Itô's formula

$$
\begin{aligned}
u(X(t)) - u(X(0)) &= \int_0^t \nabla u(X(s)) \cdot F(s) ds + \int_0^t \frac{1}{2}(G^T G) : D^2 u(X(s)) ds \\
&+ \int_0^t \nabla u(X(s)) \cdot G(s) dW_s.
\end{aligned}
$$

Let $\mu(s,\cdot)$ be the law of $X(s)$ then by taking the expectation of the above integral

$$
\begin{aligned}
\int_{\mathbb{R}^n} u(x) d\mu(s,x) - \int_{\mathbb{R}^n} u(x) d\mu_0(x) &= \int_0^t \int_{\mathbb{R}^n} \nabla u(x) \cdot F(s) d\mu(s,x) \\
&+ \int_0^t \int_{\mathbb{R}^n} \frac{1}{2}(G^T(s)G(s)) : D^2 u(x) \cdot d\mu(s,x) \\
&+ 0.
\end{aligned}
$$

Remember the weak derivative of measures, we obtain that the law $\mu(s,\cdot)$ satisfies the following second order parabolic equation in the sense of distribution,

$$
\partial_t \mu - \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} \Big(\sum_{k=1}^{m} G^{ik} G^{kj} \mu \Big) + \nabla \cdot (F\mu) = 0.
$$

Example. If $F = 0$ $m = n$ and $G = \sqrt{2} I_{n \times n}$, then the law μ of $X(t) = \sqrt{2} dW_t$ fulfills the heat equation

$$
\partial_t \mu - \Delta \mu = 0..
$$

3.3 PDE Version of Many particle system by Itô's formula

Recall the particle system in the following:

$$\begin{cases} dX_i & = \frac{1}{N}\sum_{j=1}^N K(x_i,x_j)dt + \sqrt{2}dW_t^i \qquad 1 \le i \le N, \\ X_i(0) & = x_{0,i} \end{cases}.$$

Notice that in the deterministic case, we consider the empirical measure $\mu_N(t) = \frac{1}{N}\sum_{i=1}^N \delta_{X_i(t)}$ and expect that the empirical measure converges in the weak sense to the corresponding solution of nonlocal PDE. Actually by using the Itô's formula, one can consider the time evolution of the joint measure and study the limit of its one-particle marginal. More details in this topic will be explained in the chapter with relative entropy later.

We derive within this part only the higher order linear PDE. By denoting

$$\mathbb{X}_N(t) = (X_1(t), \dots, X_N(t)),$$
$$F(\mathbb{X}_N) = \begin{pmatrix} \vdots \\ \frac{1}{n}\sum_{j=1}^N K(X_i,X_j) \\ \vdots \end{pmatrix},$$

the particle system is rewritten into

$$d\mathbb{X}_N(t) = F(\mathbb{X}_N(t))dt + \sqrt{2}d\mathbb{W}_t,$$

where $\mathbb{W}_t$ is the dN dimensional Brownian motion.

At time $t = 0$ the X_i are independent random variables, at any time $t > 0$ they are dependent and the particles have joint law

$$\mathbb{X}_N(t) \sim u^N(\cdot, t).$$

Here $u^N(\cdot,t) \in \mathcal{M}(\mathbb{R}^{dN})$, and by Itô's formula we get for arbitrary test function $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^{dN})$

$$\begin{aligned} \varphi(\mathbb{X}_N(t)) = \varphi(\mathbb{X}_N(0)) &+ \int_0^t \nabla\varphi(\mathbb{X}_N(s)) \cdot F(\mathbb{X}_N(s))ds \\ &+ \int_0^t \Delta\varphi(\mathbb{X}_N(s))ds + \int_0^t \sqrt{2}\nabla\varphi(\mathbb{X}_N(s))d\mathbb{W}_s^i. \end{aligned}$$

Taking the expectation on both sides, then the last term disappears by definition of Itô processes, we obtain that u^N satisfies the following partial differential equation in the weak sense,

$$\partial_t u^N - \sum_{i=1}^N \Delta_{x_i} u^N + \sum_{i=1}^N \nabla_{x_i}\left(\frac{1}{N}\sum_{j=1}^N K(x_i,x_j)u^N\right) = 0.$$

3.4 Solving Stochastic Differential Equations

In this section, we review the existence and uniqueness results in solving stochastic differential equation in the general setting. The main techniques is again the Picard iteration.

We use $(\Omega, \mathcal{F}, \mathbb{P})$ as the probability space, with $W(\cdot)$ as a given m-D dimensional Brownian motion. Let X_0 be an n-D dimensional random variable independent of $W(0)$, in the rest of this section, we use the Filtration given by

$$\mathcal{F}_t = \mathcal{U}(X_0) \cup \mathcal{U}(W(s), 0 \leq s \leq t).$$

Given the above basic setup we review the solution theory of stochastic differential equations of the type for $X : (t, \omega) \to \mathbb{R}^n$:

$$\begin{cases} d \underbrace{X(t)}_{n\times 1} & = \underbrace{b(X(t), t)}_{n\times 1} dt + \underbrace{B(X(t), t)}_{n\times m} d \underbrace{W_t}_{m\times 1} \quad 0 \leq t \leq T \\ X_t|_{t=0} & = X_0 \end{cases}. \tag{3.1}$$

Where

$$b : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$$
$$B : \mathbb{R}^n \times [0, T] \to M^{n\times m}.$$

Definition 3.4.1 (Solution)**.** We say an $\mathbb{R}^n$-valued stochastic process $X(\cdot)$ is a solution of the Equation 3.1 if

1. $X(t)$ is progressively measurable w.r.t $\mathcal{F}_t$
2. (Drift) $F := b(X(t), t) \in \mathbb{L}^1_n([0, T]) \;\Leftrightarrow\; \int_0^t \mathbb{E}[F(s)]ds < \infty$
3. (Diffusion) $G := B(X(t), t) \in \mathbb{L}^2_{n\times m}([0, T]) \Leftrightarrow \int_0^t \mathbb{E}[|G(s)|^2]ds < \infty$
4. (Equation) $X(t) - X(0) = \int_0^t b(X(s), s)ds + \int_0^t B(X(s), s)dW_s$.

Remark. "$X(t)$ is progressively measurable w.r.t $\mathcal{F}_t$" means that for any given $t \in [0, T]$, $X(t)$ is random variable measurable with respect to $\mathcal{F}_t$.

The goal from now on is to prove the existence and uniqueness of such solution, for that we first define what it means for a solution to be unique

Definition 3.4.2. For two solution $X, \tilde{X}$ we say they are unique if

$$\mathbb{P}(X(t) = \tilde{X}(t),\ \forall t \in [0, T]) = 1 \Leftrightarrow \max_{0\leq t\leq T} |X(t) - \tilde{X}(t)| = 0 \text{ a.s.}.$$

i.e they are indistinguishable.

The following assumptions for b, B are needed:

Assumption D.
Let $b : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ and $B : \mathbb{R}^n \times [0, T] \to M^{n\times m}$, be continuous and Lipschitz continuous with respect to x with Lipschitz constant $L > 0$, namely $\forall x, \tilde{x} \in \mathbb{R}^n$, it holds

$$|b(x, t) - b(\tilde{x}, t)| + |B(x, t) - B(\tilde{x}, t)| \leq L|x - \tilde{x}|.$$

Remark. It is obvious that b, B fulfill the linear growth condition

$$|b(x,t)| + |B(x,t)| \leq |b(0,t)| + |B(0,t)| + L(|x|) \leq L(1+|x|),$$

where for simplicity, we use the same notation L for the bound of $|b(0,t)| + |B(0,t)|$.

Theorem 3.4.1 (Existence and Uniqueness of Solution)**.** Let Assumption D hold for Equation 3.1 and assume the initial data X_0 is square integrable and independent of $W(0)$. Then there exists a unique solution $X \in \mathbb{L}_n^2([0,T])$ of the Equation 3.1.

Proof. We begin with the uniqueness proof.

Suppose there are two solutions X and $\tilde{X}$ of the SDE, then by using the definition of solution Definition 3.4.1 we have

$$X(t) - \tilde{X}(t) = \int_0^t (b(X(s),s) - b(\tilde{X}(s),s))ds + \int_0^t B(X(s),s) - B(\tilde{X}(s),s)dW_s.$$

If the diffusion term B were 0 (i.e. the deterministic case) we could use a Grönwall type inequality and get the uniqueness directly.

In stochastic setting, we consider the square integrable solutions and apply Itôs isometry. Note that $|a+b|^2 \leq 2(a^2+b^2)$

$$|X(t) - \tilde{X}(t)|^2 \leq 2\Big|\int_0^t (b(X(s),s) - b(\tilde{X}(s),s))ds\Big|^2 + \Big|\int_0^t B(X(s),s) - B(\tilde{X}(s),s)dW_s\Big|^2.$$

Now consider the following

$$\begin{aligned}
\mathbb{E}\Big[|X(t) - \tilde{X}(t)|^2\Big] &\leq 2\mathbb{E}\Big[\Big|\int_0^t (b(X(s),s) - b(\tilde{X}(s),s))ds\Big|^2\Big] \\
&\quad + 2\mathbb{E}\Big[\Big|\int_0^t B(X(s),s) - B(\tilde{X}(s),s)dW_s\Big|^2\Big] \\
&\overset{\text{Hold.}}{\leq} 2t\mathbb{E}\Big[\int_0^t \Big|b(X(s),s) - b(\tilde{X})((s),s)\Big|^2 ds\Big] + 2\mathbb{E}\Big[\int_0^t \Big|B(X(s),s) - B(\tilde{X}(s),s)\Big|^2 ds\Big] \\
&\overset{\text{Lip.}}{\leq} 2(t+1)L^2\mathbb{E}\Big[\int_0^t |X(s) - \tilde{X}(s)|^2 ds\Big] \\
&= 2(t+1)L^2 \int_0^t \mathbb{E}\Big[|X(s) - \tilde{X}(s)|^2\Big] ds
\end{aligned}$$

,

where the following Hölder's inequality was used

$$\Big(\int_0^t 1|f|ds\Big)^2 \leq \Big(\int_0^t 1^2 ds\Big)^{\frac{1}{2}\cdot 2} \cdot \Big(\int_0^t |f|^2 ds\Big)^{\frac{1}{2}\cdot 2} \leq t\int_0^t |f|^2 ds.$$

Notice that $\mathbb{E}\Big[|X(0) - \tilde{X}(0)|^2\Big] = 0$, by Gronwall's inequality we have

$$\mathbb{E}\Big[|X(t) - \tilde{X}(t)|^2\Big] = 0.$$

i.e $X(t)$ and $\tilde{X}(t)$ are modifications of each other and it remains to show that they are actually indistinguishable.

The above results shows that for any $t \in [0,T]$,

$$\mathbb{P}(A_t) = 0, \text{ where } A_t = \{\omega \in \Omega \mid |X(t) - \tilde{X}(t)| > 0\}.$$

This implies further that

$$\mathbb{P}(\max_{t\in\mathbb{Q}\cap[0,T]}|X(t)-\tilde{X}(t)|>0)=\mathbb{P}(\bigcup_{k=1}^{\infty}A_{t_k})=0.$$

Since for any fixed ω, $X(t,\omega)$ is continuous in t, we can extend the maximum over the entire interval $[0,T]$ for any fixed ω,

$$\max_{t\in\mathbb{Q}\cap[0,T]}|X(t)-\tilde{X}(t)|=\max_{t\in[0,T]}|X(t)-\tilde{X}(t)|.$$

Then the probability over the entire interval must also be 0

$$\mathbb{P}(\max_{t\in[0,T]}|X(t)-\tilde{X}(t)|>0)=0 \quad \text{i.e. } X(t)=\tilde{X}(t)\ \forall t \text{ a.s..}$$

This concludes the uniqueness proof.

For the existence, we use first the Picard iteration imilar to the deterministic case given in the following

$$\begin{aligned} X_0(t)&=X(0)\\ &\vdots\\ X_{n+1}(t)&=X(0)+\int_0^t b(X_n(s),s)ds+\int_0^t B(X_n(s),s)dW_s. \end{aligned}$$

Let $d_n(t)=\mathbb{E}[|X_{n+1}(t)-X_n(t)|^2]$, then we claim by induction that

$$d_n(t)\le\frac{(Mt)^{n+1}}{(n+1)!} \text{ for some } M>0.$$

IA: For $n=0$ we have

$$\begin{aligned} d_0(t)=\mathbb{E}[|X_1(t)-X_0(t)|^2]&\le\mathbb{E}\Big[2\Big(\int_0^t b(X_0(s),s)ds\Big)^2+2\Big(\int_0^t B(X_0(s),s)dW_s\Big)^2\Big]\\ &\le 2t\mathbb{E}\Big[\int_0^t L^2(1+X(0)^2)ds\Big]+2\mathbb{E}\Big[\int_0^t L^2(1+X(0))ds\Big]\\ &\le tM \qquad \text{where } M\ge 2L^2(1+\mathbb{E}[X(0)^2])+2L^2(1+T). \end{aligned}$$

IV: suppose the assumption holds for $n-1\in\mathbb{N}$
IS: We will prove it holds also for $n\in\mathbb{N}$:

$$\begin{aligned} d_n(t)&=\mathbb{E}[|X_{n+1}(t)-X_n(t)|^2]\\ &\le 2L^2T\mathbb{E}\Big[\int_0^t|X_n(s)-X_{n-1}(s)|^2ds\Big]+2L^2\mathbb{E}\Big[\int_0^t|X_n(s)-X_{n-1}(s)|^2ds\Big]\\ &\overset{\text{IV}}{\le} 2L^2(1+T)\int_0^t\frac{(Ms)^n}{n!}ds\le\frac{M^{n+1}t^{n+1}}{(n+1)!}. \end{aligned}$$

Different from the deterministic case, because of the additional dependence on variable ω which has only the measurability, the completeness argument works only pathwisely. Instead, we will use a similar argument as in the uniqueness proof.

$$
\begin{aligned}
&\mathbb{E}\Big[\max_{0\le t\le T}|X_{n+1}(t)-X_n(t)|^2\Big]\\
&\le \mathbb{E}\Big[\max_{0\le t\le T}\Big(2\Big|\int_0^t b(X_n(s),s)-b(X_{n-1}(s),s)ds\Big|^2+2\Big|\int_0^t B(X_n(s),s)-B(X_{n-1}(s),s)dW_s\Big|^2\Big)\Big]\\
&\le 2TL^2\mathbb{E}\Big[\int_0^T |X_n(s)-X_{n-1}(s)|^2ds\Big]+2\mathbb{E}\Big[\max_{0\le t\le T}\Big|\int_0^t B(X_n(s),s)-B(X_{n-1}(s),s)dW_s\Big|^2\Big]\\
&\le 2TL^2\mathbb{E}\Big[\int_0^T |X_n(s)-X_{n-1}(s)|^2ds\Big]+8\mathbb{E}\Big[\int_0^T |B(X_n(s),s)-B(X_{n-1}(s),s)|^2ds\Big]\\
&\le C\cdot\mathbb{E}\Big[\int_0^T |X_n(s)-X_{n-1}(s)|^2ds\Big],
\end{aligned}
$$

where we used the following Doobs martingales L^p inequality add a reference.

$$\mathbb{E}\Big[\max_{0\le s\le t}|X(s)|^p\Big]\le\Big(\frac{p}{p-1}\Big)^p\mathbb{E}[|X(t)|^p].$$

By Picard iteration we know the distance $d_n(t)=\mathbb{E}[|X_{n+1}(s)-X_n(s)|^2]$ is bounded by

$$d_n(t)\le C\cdot\mathbb{E}\Big[\int_0^T |X_n(s)-X_{n-1}(s)|^2ds\Big]\le C\frac{M^nT^{n+1}}{(n+1)!}.$$

Therefore we obtain that

$$\mathbb{E}\Big[\max_{0\le t\le T}|X_{n+1}(t)-X_n(t)|^2\Big]\le C\cdot\mathbb{E}\Big[\int_0^T |X_n(s)-X_{n-1}(s)|^2ds\Big]\le C\frac{M^nT^{n+1}}{(n+1)!}.$$

By defining

$$A_n=\Big\{\omega\in\Omega:\max_{0\le t\le T}|X_{n+1}(t)-X_n(t)|^2>\frac{1}{2^n}\Big\},$$

the Markov's inequality implies that

$$\mathbb{P}(A_n)\le 2^{2n}\mathbb{E}\Big[\max_{0\le t\le T}|X_{n+1}(t)-X_n(t)|^2\Big]\le 2^{2n}\frac{CM^nT^{n+1}}{(n+1)!}.$$

Then by Borel Cantelli, we have

$$\sum_{n=0}^{\infty}\mathbb{P}(A_n)\le C\sum_{n=0}^{\infty}2^{2n}\frac{(MT)^n}{(n+1)!}<\infty\Rightarrow\mathbb{P}(\bigcap_{n=0}^{\infty}\bigcup_{m=n}^{\infty}A_m)=0.$$

i.e $\exists B\subset\Omega$ with $\mathbb{P}(B)=1$ s.t $\forall\ \omega\in B$, $\exists\ N(\omega)>0$ s.t

$$\max_{0\le t\le T}|X_{n+1}(t,\omega)-X_n(t,\omega)|\le 2^{-n}.$$

In fact we can give B directly by

$$\Big(\bigcap_{n=0}^{\infty}\bigcup_{m=n}^{\infty}A_m\Big)^C=\bigcup_{n=0}^{\infty}\bigcap_{m=n}^{\infty}A_m^C=B.$$

then for each $\omega\in B$ we can make a Cauchy sequence argument by

$$\max_{0\le t\le T}|X_{n+k}(t)-X_n(t)|\le\sum_{j=1}^{k}\max|X_{n+j}(t)-X_{n+(j-1)}(t)|\le\sum_{j=1}^{k}\frac{1}{2^{n+j-1}}<\frac{1}{2^{n-1}}..$$

Therefore, we get

$$X_n(t,\omega) \to X(t,\omega) \qquad \text{uniform in } t \in [0,T].$$

Finaly, for a.s. ω , we can take the limit in the iteration and obtain

$$X(t) = X_0 + \int_0^t b(X(s),s)ds + \int_0^t B(X(s),s)dW_s.$$

It remains to show that $X_t \in \mathbb{L}^2([0,T])$ note that $X_0 \in \mathbb{L}^2([0,T])$ already and

$$\begin{aligned}\mathbb{E}[|X_{n+1}(t)|^2] &\leq C(1+\mathbb{E}[|X_0|^2]) + C\int_0^t \mathbb{E}[|X_n(s)|^2]ds \\ &\leq C\sum_{j=0}^{n} C^{j+1}\frac{t^{j+1}}{(j+1)!}(1+\mathbb{E}[|X_0|^2]) \leq C\cdot e^{Ct}.\end{aligned}$$

Where we used $\mathbb{E}[X_0]=0$,the linear growth condition for the first integral and Itô isometry for the second integral .

Using the above we conclude by Fatous's lemma

$$\mathbb{E}[|X(t)|^2] = \mathbb{E}[\lim_{n\to\infty}|X_{n+1}(t)|] \leq \liminf_{n\to\infty}\mathbb{E}[|X_{n+1}(t)|^2] \leq C\cdot e^{Ct}.$$

Therefore

$$\int_0^T \mathbb{E}[|X(t)|^2] \leq CT\cdot e^{CT}.$$

□

Theorem 3.4.2 (Higher Moments Estimate)**.** Let the Assumptions for b , B be given in Assumption D and X_0 satisfies

$$\mathbb{E}[|X_0|^{2p}] < \infty.$$

for some $p \geq 1$ then $\forall t \in [0,T]$

$$\mathbb{E}[|X(t)|^{2p}] \leq C(1+\mathbb{E}[|X_0|^{2p}])e^{Ct}.$$

and $\mathbb{E}[|X(t)-X_0|^{2p}] \leq C(1+\mathbb{E}[|X_0|^{2p}])e^{Ct}t^p$

Proof. Left as an exercise.

□

3.5 Stochastic Mean Field Limit

This part is basically taken from Lacker's lecture notes in 2018.

3.5.1 Convergence of the empirical measure for i.i.d. Random Variables

In the discussion of the mean field limit of a many particle interacting system, we will use again the so-called Wasserstein distance (or Kantorovich–Rubinstein metric) between two measures:

Definition 3.5.1 (Wasserstein Distance (or Kantorovich–Rubinstein metric))**.** For all $\mu,\nu \in \mathcal{P}_p(\mathbb{R}^d)$, $(p \geq 1)$ the Wasserstein Distance (or Kantorovich–Rubinstein metric) of μ and ν is given by

$$W^p(\mu,\nu) = \underset{MK,p}{\text{dist}}(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \left(\int\int_{\mathbb{R}^{2d}} |x-y|^p \pi(dxdy)\right)^{\frac{1}{p}},$$

where

$$\Pi(\mu,\nu) = \Bigg\{\pi \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d) : \int_{\mathbb{R}^d\times E} \pi(dx,dy) = \nu(E),$$
$$\int_{E\times\mathbb{R}^d} \pi(dx,dy) = \mu(E), \forall \text{ Borel set } E\subset\mathbb{R}^d\Bigg\}.$$

Remark. It is easy to show that

$$W_1(\mu,\tilde{\mu}) \le W_2(\mu,\tilde{\mu}).$$

follows automatically by using the Hölders inequality. Similarly, this holds also for all $p > q \ge 1$ that

$$W_q(\mu,\tilde{\mu}) \le W_p(\mu,\tilde{\mu}).$$

Lemma 3.5.1. Cite Villani's book Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_p(\mathbb{R}^d)$ be a sequence of measures,then following are equivalent

1. $W_p(\mu_n,\mu)\to 0$
2. For $\forall f\in\mathcal{C}(\mathbb{R}^d)$ such that $|f(x)|\le C(1+|x|^p)$, it holds $\int f d\mu_n \to \int f d\mu$.
3. $\mu_n \rightharpoonup \mu$, namely $\int f d\mu_n \to \int f d\mu$ is true for all $f\in\mathcal{C}_b(\mathbb{R}^d)$ and $\int |x|^p d\mu_n \to \int |x|^p d\mu$
4. $\mu_n \rightharpoonup \mu$ and $\lim\limits_{r\to\infty} \sup_n \int_{|x|\ge r} |x|^p d\mu_n = 0$.

Definition 3.5.2 (Empirical Measure (Stochastic version))**.** For random variables $(X_i)_{i\le N}$ we define the (stochastic) empirical measure by

$$\mu_N(\omega) = \frac{1}{N}\sum_{i=1}^N \delta_{X_i(\omega)}.$$

In the following we first discuss the convergence of empirical measure with i.i.d random variables. Actually, a direct corollary of law of large numbers gives that

Proposition 3.5.1. If $(X_i)_{i\in\{1,\dots,N\}}$ are i.i.d random variables with law μ_X then $\forall\ f\in\mathcal{C}_b(\mathbb{R}^d)$ it holds that

$$\mathbb{P}(\lim_{N\to\infty}\int f d\mu_N = \int f d\mu) = 1.$$

One can actually prove the stronger statement that the choice of $f\in\mathcal{C}_b$ does not influence the convergence, which means one can pull the function selection into the probability.

Proposition 3.5.2. If $(X_i)_{i\in\{1,\dots,N\}}$ are i.i.d random variables with law μ_X then it holds that

$$\mathbb{P}(\mu_N \rightharpoonup \mu) = 1.$$

i.e

$$\mathbb{P}(\forall f \in \mathcal{C}_b(\mathbb{R}^d) \ : \int f \, d\mu_N \to \int f \, d\mu) = 1.$$

We omit the technical proof here.

Lemma 3.5.2 (General Dominated Convergence)**.** Let $(X_n)_{n\in\mathbb{N}} \subset L^p$ be a sequence of random variables then the following are equivalent

1. $(X_n)_{n\in\mathbb{N}}$ are uniformly integrable and $X_n \to X$ $\mathbb{P}$-a.s.
2. $\|X_n - X\| \to 0$ for some $X \in L^p$

We leave the proof as exercise.

Definition 3.5.3. A sequence of random variables $(X_i)_{i\in\mathbb{N}}$ is called uniform integrable if

$$\lim_{r\to\infty} \sup_{i\in\mathbb{N}} \mathbb{E}\left[|X_i| \cdot \mathbb{1}_{|X_i|\geq r}\right] = 0.$$

Lemma 3.5.3 (De la Vallèe Poussin Criterion)**.** A sequence of random variables (X_i) is uniformly integrable if and only if there $\exists\varphi$ convex with $\lim\limits_{x\to\infty} \frac{\varphi(x)}{x} = \infty$, such that

$$\sup_i \mathbb{E}\left[\varphi(|X_i|)\right] < \infty.$$

We omit the technical proof here.

Proposition 3.5.3. If $(X_i)_{i\in\{1,\dots,N\}}$ are i.i.d random variables with law $\mu \in \mathcal{P}^p(\mathbb{R}^d)$ for $p \geq 1$, μ_N denotes its empirical measure, then it holds

$$W_p(\mu_N, \mu) \to 0 \qquad \text{a.s.} \quad \text{ and } \mathbb{E}\left[W_p^p(\mu_N, \mu)\right] \to 0.$$

Proof. Remember that the following convergences are equivalent

1. $W_p(\mu_N, \mu) \to 0$
2. $\mu_N \rightharpoonup \mu$ and $\int |x|^p d\mu_N \to \int |x|^p d\mu$
3. $\mu_n \rightharpoonup \mu$ and $\lim\limits_{r\to\infty} \sup_N \int_{|x|\geq r} |x|^p d\mu_N = 0$

Note that if we fix a.s. ω then we can treat this as the deterministic case.

We already know that

$$\mu_N \rightharpoonup \mu \text{ a.s.}.$$

since (X_i) are i.i.d then $|X_i|^p$ is also i.i.d and we use the Law of large numbers

$$\int |x|^p d\mu_N = \frac{1}{N}\sum_{i=1}^N |X_i|^p \xrightarrow{L.L.N.} \mathbb{E}[|X_i|^p] < \infty.$$

And we get a.s. that $W_p(\mu_N, \mu) \to 0$

For the stronger statement

$$\mathbb{E}\left[W_p^p(\mu_N, \mu)\right] \to 0.$$

we first note that

$$\begin{aligned} W_p^p(\mu_N, \mu) &\leq 2^{p-1}(W_p^p(\mu_N, \delta_0) + W_p^p(\delta_0, \mu)) \\ &= 2^{p-1}(\frac{1}{N}\sum_{i=1}^N |X_i|^p + W_p^p(\delta_0, \mu)). \end{aligned}$$

Then it is sufficient to show the uniform integrability of the first part

$$\frac{1}{N}\sum_{i=1}^{N}|X_i|^p.$$

Since $|X_i|^p$ is integrable then there exists a convex function φ with $\lim\limits_{x\to\infty}\frac{\varphi(x)}{x}=\infty$ and

$$\mathbb{E}[\varphi(|X_i|^p)]<\infty.$$

Since φ is convex we apply Jensen's inequality to get

$$\sup_N \mathbb{E}\Big[\varphi\Big(\frac{1}{N}\sum_{i=1}^{N}|X_i|^p\Big)\Big] \overset{\text{Jen.}}{\leq} \sup_N \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\varphi(|X_i|^p)] = \mathbb{E}[\varphi(|X_i|^p)] < \infty.$$

Finally Lemma 3.5.3 implies the uniform integrability and we conclude by Lemma 3.5.2

$$\mathbb{E}\big[W_p^p(\mu_N,\mu)\big]\to 0.$$

□

Remark. All the above statement only apply to arbitrary i.i.d sequences of random variables, but in our Mean-Field-Limit we only get the i.i.d property at $t=0$ such that we seek to prove that even as $N\to\infty$ we nonetheless get a convergence.

3.5.2 Setting of the Stochastic Particle System

We will study a general version of the mean field type stochastic particle system in this part.

The family of N interacting particles $X_1^N(\omega,t), X_2^N(\omega,t), \dots, X_N^N(\omega,t) \in \mathbb{R}^d$ with i.i.d initial data $(X_i^N(0))_{i\in\{1,\dots,N\}} \subset L^2(\Omega)$, with given law μ_0, satisfies the following stochastic system

$$
\text{(SDEN)} \begin{cases} dX_i^N(t) &= b(X_i^N(t), \mu_N(t))dt + \sigma(X_i^N(t), \mu_N(t))dW_t^i \\ X_i^N(0) &= X_{i,0}^N \end{cases}, \tag{3.2}
$$

where $\mu_N(t)$ is the stochastic empirical measure of $(X_1^N(t), X_2^N(t), \dots, X_N^N(t))$. The results from previous section imply that the initial data converges:

$$
\mathbb{E}\left[W_2^2(\mu_N(0), \mu_0)\right] \to 0.
$$

However since for the solution of interacting particle system are for $t > 0$ are no longer independent, the corresponding result for solution in $N \to \infty$ needs to be investigated.

We give the following assumptions within this problem.

Assumption E. Assume drift $b : \mathbb{R}^d \times \mathcal{P}^2(\mathbb{R}^d) \to \mathbb{R}^d$ and diffusion $\sigma : \mathbb{R}^d \times \mathcal{P}^2(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are Lipschitz continuous i.e. $\exists L > 0$ s.t.

$$
|b(X,\mu) - b(\tilde{X},\tilde{\mu})| + |\sigma(X,\mu) - \sigma(\tilde{X},\tilde{\mu})| \leq L\left(|X - \tilde{X}| + W_2(\mu,\tilde{\mu})\right).
$$

Example (Stochastic Toy Model)**.** Let our particle system be given as in Equation 3.2 with drift and diffusion for $\nabla V \in \text{Lip}$

$$
\begin{aligned}
b(X,\mu) &= \nabla V \star \mu(X) \\
\sigma(X,\mu) &= \sigma_0 > .
\end{aligned}
$$

Within this toy model one can easily check that Assumption E holds:

$$
\begin{aligned}
|b(X,\mu) - b(\tilde{X},\tilde{\mu})| &= \left|\int \nabla V(X-y)d\mu(y) - \int \nabla V(\tilde{X}-y)d\tilde{\mu}(y)\right| \\
&\geq \int |\nabla V(X-y) - \nabla V(\tilde{X}-y)|d\mu(y) + \left|\int \nabla V(\tilde{X}-y)(d\mu(y) - d\tilde{\mu}(y))\right| \\
&\overset{\text{Lip.}}{\leq} L\cdot|X-\tilde{X}| + LW_1(\mu,\tilde{\mu}) \\
&\leq L\cdot(|X-\tilde{X}| + W_2(\mu,\tilde{\mu})).
\end{aligned}
$$

Exercise. Think about what happens if the initial data is i.i.d but the diffusion coefficient $\sigma = 0$, can one obtain a convergence ?

Theorem 3.5.1 (Sovalbility of N particle problem)**.** Let the assumption Assumption E holds, then the stochastic many particle system (SDEN) has a unique strong solution in $\mathbb{L}^2_{dN}([0,T])$.

Proof. By using the notation $\mathbb{X} = (X_1^N, \dots, X_N^N) \in \mathbb{R}^{dN}$, $\mathbb{W} = (W^1, \dots, W^N) \in \mathbb{R}^{mN}$, and

$$
B(\mathbb{X}) = \begin{pmatrix} \vdots \\ b(X_i^N, \frac{1}{N}\sum_{k=1}^N \delta_{X_k^N}) \\ \vdots \end{pmatrix}_{dN},
$$

$$
\Sigma(\mathbb{X})_{dN\times mN} \;:\; \text{diag}(\Sigma(\mathbb{X})) = \left(\sigma(X_1^N, \tfrac{1}{N}\sum_{k=1}^N \delta_{X_k^N}), \dots \sigma(X_N^N, \tfrac{1}{N}\sum_{k=1}^N \delta^N_{X_k})\right)
$$

the stochastic many particle system (SDEN) is rewritten into

$$d\mathbb{X}(t) = B(\mathbb{X}(t))dt + \Sigma(\mathbb{X}(t))d\mathbb{W}_t.$$

Now if B and Σ satisfy Assumption D we get the existence and uniqueness of solution in $\mathbb{L}^2_{dN}([0,T])$ by applying Theorem 3.4.1. Actually,

$$\begin{aligned}\left|B\left(\mathbb{X}\right) - B\left(\mathbb{Y}\right)\right|^2_{\mathbb{R}^{dN}} &= \sum_{j=1}^N \left|b\left(X_j, \frac{1}{N}\sum_{k=1}^N \delta_{X_k}\right) - b\left(Y_j, \frac{1}{N}\sum_{k=1}^N \delta_{Y_k}\right)\right| \\ &\leq \sum_{j=1}^N 2L^2 \left(|X_j - Y_j|^2 + W_2^2\left(\frac{1}{N}\sum_{k=1}^N \delta_{X_k}, \frac{1}{N}\sum_{k=1}^N \delta_{Y_k}\right)\right) \\ &\leq 4L^2 |\mathbb{X} - \mathbb{Y}|^2,\end{aligned}$$

where we have used the definition of W_2 distance in the last estimate:

$$\begin{aligned}W_2^2\left(\frac{1}{N}\sum_{k=1}^N \delta_{X_k}, \frac{1}{N}\sum_{k=1}^N \delta_{Y_k}\right) &\leq \iint_{\mathbb{R}^d} |X - Y|^2 \left(\frac{1}{N}\sum_{k=1}^N \delta_{(X_k, Y_k)}\right) \\ &= \frac{1}{N}\sum_j^N |X_j - Y_j|^2 = \frac{1}{N}|\mathbb{X} - \mathbb{Y}|^2.\end{aligned}$$

For the diffusion coefficient Σ, the argument is analog. Then by Theorem 3.4.1 we get a unique solution for fixed N. □

As $N \to \infty$ we expect that the limiting equation is

$$\begin{cases} dY^i(t) &= b(Y^i(t), \mu(t))dt + \sigma(Y^i(t), \mu(t))dW^i_t \\ Y^i(0) &= X^N_{i,0} \in L^2(\Omega) \text{ i.i.d}, \mu(t) \sim law(Y^i) \end{cases}.$$

In fact since the above system beyond the initial data is independent of N, we may consider the equation without particle index. This equation is called Mckean-Vlasov equation which is a non-linear non-local SDE.

3.5.3 Well-posedness of Mckean-Vlasov equation

We consider in this part the solvalbility of Makean-Vlasov Equation

$$\text{(MVE)} \begin{cases} dY(t) &= b(Y(t), \mu(t))dt + \sigma(Y(t), \mu(t))dW_t \\ Y(0) &= \xi \in L^2(\Omega; \mathbb{R}^d) \text{ i.i.d}, \mu \sim law(Y) \end{cases}, \tag{3.3}$$

under the condition that b and σ satisfy Assumption E. Since the equation Equation 3.3 includes also the information of law which depends on the solution, for convenience we consider the law on space of the sample paths.

Definition 3.5.4 (Space of Continuous Sample Paths)**.** The Space $\mathcal{C}^d = \mathcal{C}([0,T]; \mathbb{R}^d)$ is called the continuous sample path space with norm $\|X\|_T = \sup_{0\leq t\leq T}|X(t)|$. this norm $\|\cdot\|_T$ induces a σ-algebra on $\mathcal{C}^d$. Then a random Variable with $\mathcal{C}^d$ as its range is a map $X : \Omega \to \mathcal{C}^d$.

Definition 3.5.5 (Measure on the Space of Sample Paths)**.** Since the norm $\|\cdot\|_T$ induces a σ-algebra on $\mathcal{C}^d$, we work on the probability measures on $\mathcal{C}^d$ with finite second moment, namely the space $\mathcal{P}^2(\mathcal{C}^d)$. With the map, for any $t \in [0,T]$,

$$I_t : \mathcal{C}^d \to \mathbb{R}^d, \quad X \mapsto X(t),$$

we call the corresponding push forward measure $\mu_t := I_{t\#}\mu \in \mathcal{P}^2(\mathbb{R}^d)$ the t-marginal of

$\mu \in \mathcal{P}^2(\mathcal{C}^d)$, which is given by: for all Borel sets $A \subset \mathbb{R}^d$, it holds

$$\mu_t(A) = \mu(I_t^{-1}(A))..$$

Proposition 3.5.4 (Wasserstein Distance for the measures $\mathcal{C}^d$)**.** For arbitrary measures $\mu, \tilde{\mu} \in \mathcal{P}^2(\mathcal{C}^d)$ the following inequality holds

$$\sup_{t\in[0,T]} W_{\mathbb{R}^d,2}(\mu(t), \tilde{\mu}(t)) \le W_{\mathcal{C}^d,2}(\mu, \tilde{\mu}),$$

where

$$W_{\mathcal{C}^d,2}(\mu, \tilde{\mu}) = \inf_{\pi\in\Pi(\mu,\tilde{\mu})} \int_{\mathcal{C}^d\times\mathcal{C}^d} \|x-y\|_T^2 d\pi(x,y).$$

Proof. For $\pi \in \Pi(\mu, \tilde{\mu})$, we choose concretely $\pi_t = I_t^{\otimes 2}{}_{\#}\pi$. Let μ_t and $\tilde{\mu}_t$ be the t-marginals of μ and $\tilde{\mu}$ separately, then we have

$$\begin{aligned}
\sup_{t\in[0,T]} W_2(\mu_t, \tilde{\mu}_t) &\le \sup_{t\in[0,T]} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\pi_t(x,y) \\
&= \sup_{t\in[0,T]} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d(I_t^{\otimes 2}{}_{\#}\pi)(x,y) \\
&= \sup_{t\in[0,T]} \int_{\mathcal{C}^d\times\mathcal{C}^d} |x(t)-y(t)|^2 d\pi(x,y) \\
&\le \int_{\mathcal{C}^d\times\mathcal{C}^d} \sup_{t\in[0,T]} |x(t)-y(t)|^2 d\pi(x,y) \\
&= \int_{\mathcal{C}^d\times\mathcal{C}^d} \|x-y\|_T^2 d\pi(x,y).
\end{aligned}$$

It remains to check that $\pi_t \in \Pi(\mu_t, \tilde{\mu}_t)$. To show this, we take arbitrary $A \in \mathcal{B}(\mathbb{R}^d)$ and obtain

$$\begin{aligned}
\pi_t(A\times\mathbb{R}^d) &= \pi((I_t^{\otimes 2})^{-1}(A\times\mathbb{R}^d)) \\
&= \pi(\{(x,y)\in\mathcal{C}^d\times\mathcal{C}^d \;:\; I_t(x)\in A, I_t(y)\in\mathbb{R}^d\}) \\
&= \pi(\{(x,y)\in\mathcal{C}^d\times\mathcal{C}^d \;:\; I_t(x)\in A, y\in\mathcal{C}^d\}) \\
&= \pi(I_t^{-1}(A)\times\mathcal{C}^d) = \mu(I_t^{-1}(A)) = I_{t\#}\mu(A) = \mu_t(A).
\end{aligned}$$

this shows that μ_t is the first marginal of π_t. Similarly, one can show that $\tilde{\mu}_t$ is the other marginal. ☐

Remark. With the above notation we have that for any measurable function f on $\mathcal{C}^d$, it holds

$$\int_{\mathcal{C}^d} f(x)d\mu(x) = \int_{\mathbb{R}^d} f(x(t))d\mu_t.$$

Theorem 3.5.2 (Uniqueness and Existence of Solution for Mckean-Vlasov)**.** If b and σ satisfy Assumption E, then Equation 3.3 has a unique strong solution $Y \in \mathbb{L}^2([0,T])$ and $\mu = law(Y) \in \mathcal{P}^2(\mathcal{C}^d)$.

Proof. We use the notation

$$d_t^2(\mu, \tilde{\mu}) = \inf_{\pi\in\Pi(\mu,\tilde{\mu})} \int_{\mathcal{C}^d\times\mathcal{C}^d} \|x-y\|_t^2 d\pi(x,y).$$

For any given $\mu \in \mathcal{P}^2(\mathcal{C}^d)$ we consider the following SDE

$$\begin{cases} dY^\mu(t) & = b(Y^\mu(t), \mu(t))dt + \sigma(Y^\mu(t), \mu(t))dW_t \\ Y(0) & = \xi \in L^2(\Omega) \end{cases}.$$

Let $\varphi(\mu) = \mathcal{L}(Y^\mu)$ be the law of Y^μ.

For the existence and the uniqueness of Y^μ we need to check

$$|b(x, \mu(t)) - b(\tilde{x}, \mu(t))| + |\sigma(x, \mu(t)) - \sigma(\tilde{x}, \mu(t))| \leq L|x - \tilde{x}|.$$

Since it is the same measure the Wasserstein distance is 0 and the above is true by Assumption E.

In the next, we will prove that φ has a fixpoint $\overline{\mu}$, then according to the definition of φ, we know that $\overline{\mu}$ is the solution of Equation 3.3.

We prove this fact by Picard iteration method.

First we do the estimate for the difference between two measures. Let $\mu, \tilde{\mu}$ be arbitrary given measure in $\mathcal{P}^2(\mathcal{C}^d)$, the integral version of the equation is

$$Y^\mu(t) - \xi = \int_0^t b(Y^\mu(s), \mu(s))ds + \int_0^t \sigma(Y^\mu(s), \mu(s))dW_s \qquad \mu = \mu, \tilde{\mu}.$$

Now we take the difference for these two equations and proceed the sup in t, and use Hölder's inequality,

$$\begin{aligned} &\sup_{0\leq t\leq \tau} |Y^\mu(t) - Y^{\tilde{\mu}}(t)|^2 \\ &= \sup_{0\leq t\leq \tau} \left| \int_0^t b(Y^\mu(s), \mu(s)) - b(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))ds + \int_0^t \sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))dW_s \right|^2 \\ &\leq \sup_{0\leq t\leq \tau} 2t \int_0^t |b(Y^\mu(s), \mu(s)) - b(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))|^2 ds \\ &+ \sup_{0\leq t\leq \tau} 2 \left| \int_0^t \sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))dW_s \right|^2. \end{aligned}$$

Next by taking the expectation and using the assumption Assumption E, we obtain that

$$\begin{aligned} &\mathbb{E}[\sup_{0\leq t\leq \tau} |Y^\mu(t) - Y^{\tilde{\mu}}(t)|^2] \\ &\leq 4\tau L^2 \mathbb{E}\left[\int_0^\tau |Y^\mu(s) - Y^{\tilde{\mu}}(s)|^2 + W_2^2(\mu(s), \tilde{\mu}(s))ds\right] \\ &+ 16L^2 \mathbb{E}\Big[\int_0^\tau |Y^\mu(s) - Y^{\tilde{\mu}}(s)|^2 + W_2^2(\mu(s), \tilde{\mu}(s))ds\Big], \end{aligned}$$

where we used Doobs-L^p inequality for the second term.

$$\begin{aligned} &\mathbb{E}\Big[\sup_{0\leq t\leq \tau} \Big| \int_0^t \sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))dW_s \Big|^2\Big] \\ &\leq 8\mathbb{E}\Big[\int_0^\tau |\sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))|^2 ds\Big] \\ &\leq 8L^2 \mathbb{E}\Big[\int_0^\tau |Y^\mu(s) - Y^{\tilde{\mu}}(s)|^2 + W_2^2(\mu(s), \tilde{\mu}(s))ds\Big]. \end{aligned}$$

Therefore we obtained

$$\mathbb{E}\big[\|Y^\mu - Y^{\tilde\mu}\|_\tau^2\big] \le C\int_0^\tau \mathbb{E}\big[\|Y^\mu - Y^{\tilde\mu}\|_s^2\big]ds + C\int_0^\tau \mathbb{E}\big[W_2^2(\mu(s),\tilde\mu(s))\big]ds$$

,

which implies, by Gronwall inequality, that

$$\begin{aligned}\mathbb{E}\big[\|Y^\mu - Y^{\tilde\mu}\|_\tau^2\big] &\le C(\tau)\cdot\int_0^\tau \mathbb{E}\big[W_2^2(\mu(s),\tilde\mu(s))\big]ds\\ &\le C(\tau)\cdot\int_0^\tau \mathbb{E}\big[\sup_{0\le t\le s} W_2^2(\mu(t),\tilde\mu(t))\big]ds\\ &\le C(\tau)\int_0^\tau d_s(\mu,\tilde\mu)ds.\end{aligned}$$

using the inequality Proposition 3.5.4.

Remember that $\varphi(\mu) = \mathcal{L}(Y^\mu)$ and $\varphi(\tilde\mu) = \mathcal{L}(Y^{\tilde\mu})$, we have

$$\begin{aligned}d_\tau^2(\varphi(\mu),\varphi(\tilde\mu)) &= \inf_{\pi\in\Pi(\varphi(\mu),\varphi(\tilde\mu))}\int_{\mathcal{C}^d\times\mathcal{C}^d}\|x-y\|_\tau^2 d\pi(x,y)\\ &\le \int_{\mathcal{C}^d\times\mathcal{C}^d}\|x-y\|_\tau^2 d\pi_1(x,y)\\ &=\mathbb{E}\big[\|Y^\mu - Y^{\tilde\mu}\|_\tau^2\big] \le C(\tau)\int_0^\tau d_s(\mu,\tilde\mu)ds.\end{aligned}$$

where we denote π_1 to be the joint distribution of Y^μ and $Y^{\tilde\mu}$. To summarize, we have that for $\forall \mu,\tilde\mu\mathcal{P}^2(\mathcal{C}^d)$, it holds

$$d_t(\varphi(\mu),\varphi(\tilde\mu)) \le C(t)\int_0^t d_s(\mu,\tilde\mu)ds. \tag{3.4}$$

This estimate implies immediately the uniqueness of the solution. Namely, if we have two solutions $\mu,\tilde\mu$ such that $\varphi(\mu)=\mu$ and $\varphi(\tilde\mu)=\tilde\mu$, then the above estimate Equation 3.4 says

$$d_t(\mu,\tilde\mu) \le C(t)\int_0^t d_s(\mu,\tilde\mu)ds \Rightarrow d_t(\mu,\tilde\mu)=0.$$

To prove the existence. Take arbitrary $\mu_0\in\mathcal{P}^2(\mathcal{C}^d)$, (for example $\mu_0 = \mathcal{L}(\xi)$), we define the following iteration

$$\begin{aligned}\varphi(\mu_0) &= \mu_1\\ \varphi(\mu_1) &= \mu_2\\ &\vdots\\ \varphi(\mu_k) &= \mu_{k+1}\\ &\vdots,\end{aligned}$$

the estimate Equation 3.4 means that (μ_k) is Cauchy in $\mathcal{P}^2(\mathcal{C}^d)$.

Exercise. Prove that (μ_k) is Cauchy in $\mathcal{P}^2(\mathcal{C}^d)$.

Therefore, there exists a $\mu\in\mathcal{P}^2(\mathcal{C}^d)$ such that

$$W_{\mathcal{C}^d,2}^2(\mu_k,\mu)\to 0.$$

□

In the many particle setting, the empirical measure μ_N is not exactly the law of X^N, it is a sequence of stochastic measure. Therefore unlike the deterministic, the proof for existence and uniqueness does not hold exactly for our Equation 3.2.

For the initial data, because they are i.i.d. random variables, we have proved that

$$\mathbb{E}[W_2^2(\mu_N(0), \mu_0)] \xrightarrow{N\to\infty} 0.$$

We expect that the mean field limit holds also in the same meaning, namely

$$\mathbb{E}[W_{\mathcal{C}^d,2}^2(\mu_N, \mu)] \to 0.$$

Theorem 3.5.3 (Mean-Field-Limit)**.** Let b and σ fulfill Assumption E and use μ_N to be the empirical measure of the solution of Equation 3.2, then there exists a measure $\mu \in \mathcal{P}^2(\mathcal{C}^d)$ such that

$$\lim_{N\to\infty} \mathbb{E}\big[W_{\mathcal{C}^d,2}^2(\mu_N, \mu)\big] = 0, .$$

and for any fixed $k \in \mathbb{N}$ it holds

$$(X_1^N, \dots, X_k^N) \xRightarrow{(D)} (Y_1, \dots, Y_k).$$

where $Y_1, \dots, Y_k$ are independent copies of the solution of the Mckean-Vlasov equation Equation 3.3.

Proof. The proof is similar to what we have done in the Theorem 3.5.2, the critical part is to work with our stochastic empirical measure, we do so by introducing an intermediate empirical measure. We compute

$$\begin{aligned} |X_i^N(t) - Y_i(t)|^2 \leq 2t \int_0^t |b(X_i^N(s), \mu_N(s)) - b(Y_i(s), \mu(s))|^2 \\ + 2\Big|\int_0^t \sigma(X_i^N(s), \mu_N(s)) - \sigma(Y_i(s), \mu(s)) dW_s^i\Big|^2. \end{aligned}$$

Similar to the previous cases in handling the stochastic term, we use the Doob's inequality and obtain

$$\mathbb{E}\Big[\sup_{0\leq r\leq t} |X_i^N(r) - Y_i(r)|^2\Big] \leq 2(8+2t)L^2\mathbb{E}\Big[\int_0^t \sup_{0\leq r\leq s} |X_i^N(r) - Y_i(r)|^2 + W_2^2(\mu_N(s), \mu(s))ds\Big].$$

The by Gronwall's inequality, we have

$$\mathbb{E}\Big[\sup_{0\leq r\leq t} |X_i^N(r) - Y_i(r)|^2\Big] \leq C \cdot \mathbb{E}\Big[\int_0^t d_r^2(\mu_N, \mu) dr\Big]. \tag{3.5}$$

Let $\overline{\mu}_N$ be the empirical measure of Y_i, i.e. $\overline{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{Y_i}$ And denote $\mu \sim \mathcal{L}(Y_i)$ for $\forall t > 0$ then by Proposition 3.5.3, we have

$$\mathbb{E}\big[W_2^2(\overline{\mu}_N, \mu)\big] \to 0.$$

Now we consider for $\forall$ a.s. $\omega \in \Omega$

$$d_t^2(\overline{\mu}_N, \mu_N) = \inf_{\pi\in\Pi(\mu_N, \overline{\mu}_N)} \int_{\mathcal{C}^d\times\mathcal{C}^d} \|x-y\|_t^2 d\pi(x,y) \leq \frac{1}{N}\sum_{i=1}^N \|X_i^N - Y_i\|_t^2,$$

where we have taken $\pi = \mu_N \otimes \overline{\mu}_N$. We continue by taking the expectation and applying Equation 3.5

$$\mathbb{E}\big[d_t^2(\mu_N, \overline{\mu}_N)\big] \leq \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big[\sup_{0\leq r\leq t}|X_i^N(r) - Y_i(r)|^2\Big] \leq 2C\int_0^t \mathbb{E}[d_r^2(\mu_N, \mu)]dr.$$

Therefore, we obtain by triangle inequality that

$$\begin{aligned}\mathbb{E}\big[d_t^2(\mu_N, \mu)\big] &\leq 2\mathbb{E}\big[d_t^2(\mu_N, \overline{\mu}_N)\big] + 2\mathbb{E}\big[d_t^2(\overline{\mu}_N, \mu)\big] \\ &\leq C\int_0^t \mathbb{E}\big[d_r^2(\mu_N, \mu)\big]dr + C\mathbb{E}\big[d_t^2(\overline{\mu}_N, \mu)\big].\end{aligned}$$

Then by Grönwall

$$\mathbb{E}\big[d_t^2(\mu_N, \mu)\big] \leq e^{CT}\mathbb{E}\big[W_2^2(\mu_{N,0}, \mu_0)\big] + e^{CT}\mathbb{E}\big[d_t^2(\overline{\mu}_N, \mu)\big] \xrightarrow{N\to\infty} 0,$$

where the last two limits are obtained by Proposition 3.5.3. Therefore, for $\forall 1 \leq k < \infty$, it holds that

$$\begin{aligned}\mathbb{E}\Big[\max_{1\leq i\leq k}\sup_{0\leq r\leq t}\|X_i^N(r) - Y_i(r)\|^2\Big] &\leq \max_{1\leq i\leq k}\frac{1}{N}\sum_{i=1}^{k}\mathbb{E}[\|X_i^N - Y_i\|_t^2] \\ &\leq C\cdot k\mathbb{E}[d_t^2(\mu_N, \mu)] \xrightarrow{N\to\infty} 0.\end{aligned}$$

This concludes that the set of random variables converges weakly to the solution of Mckean-Vlasov equation. $\square$

Chapter 4

PDE Approach for the Mckean-Vlasov Equation

In this chapter, we give an alternative approach in solving the Mckean-Vlasov equation Equation 3.3 based on the PDE theory. Due to the time limitation, we will use in this chapter the simplest framework, i.e. L^2 theory, in the study of the weak solution to second order parabolic PDEs. More complete version of the solution theory will be given in the lecture of PDEs in the spring semester.

To describe the method more explicitly, we use a simplified model other than the general setting in Equation 3.3. More precisely, we take b as a given operator in the convolution form with kernel F in the following

$$b(Y(t), u) = \int F(Y(t) - y)u(y)dy = \int F(y)u(Y(t) - y)dy.$$

In addition, the diffusion coefficient is simply taken to be a constant, i.e. $\sigma = \sqrt{2}$. Notice that with this convolution structure, the condition that b need to be Lipschitz continuous for solving the stochastic differential equation requires that the interaction kernel F to be Lipschitz continuous if one only search for measure valued u. On the other hand, u satisfies a second order parabolic PDE, where more regularity of the solution can be obtained. Therefore, it allows the opportunity to weaken the assumption of F with the help of the convolution structure. It will further give the chance to get mean field limit for singular interaction forces.

We explain the connection with PDE again in the following: with the simpler setting, the (MVE) can be rewritten as

$$\text{(MVE*)} \begin{cases} dY(t) & = F \star \mu(t)(Y(t))dt + \sqrt{2}dW_t \\ Y(0) & = \xi \in L^2(\Omega), \quad \mu(t) \sim law(Y(t)) \; . \\ \mu_0 & \sim law(\xi) \end{cases} \tag{4.1}$$

By applying Itô's formula we have, for $\forall \varphi \in \mathcal{C}_0^\infty([0, T) \times \mathbb{R}^d)$ it holds

$$\begin{aligned} \varphi(Y(t), t) - \varphi(Y(0), 0) = & \int_0^t \frac{\partial \varphi}{\partial t}(Y(s), s) + \nabla\varphi(Y(s), s) \cdot F \star \mu(s)(Y(s)) \\ & + \frac{1}{2} \underbrace{(\sqrt{2} \cdot \sqrt{2})}_{tr(\sigma \cdot \sigma^T)} \Delta\varphi(Y(s), s)ds + \int_0^t \nabla\varphi(Y(s), s)\sqrt{2}dW_s. \end{aligned}$$

Then by taking the expectation on both sides, notice that the last term disappears, we have

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi(x, t)d\mu(t) - \int_{\mathbb{R}^d} \varphi(x, 0)d\mu_0 \\ & = \int_0^t \int_{\mathbb{R}^d} \Big(\frac{\partial \varphi}{\partial t}(x, s) + \nabla\varphi(x, s) \cdot F \star \mu(s)(x) + \Delta\varphi(x, s) \Big) d\mu(s)ds. \end{aligned}$$

This shows that $\mu(t)$ satisfies the following diffusion equation in the weak sense.

$$\begin{cases} \partial_t \mu - \Delta\mu + \nabla \cdot (F \star \mu\mu) = 0 \\ \mu(0) = \mu_0 \end{cases} . \tag{4.2}$$

Compare this weak PDE to the one we got in the discrete case Equation 2.3, there is only one additional diffusion term. From PDE point of view, this additional diffusion provides more regularity of the solution. We refer also more theories on this kind of equation to the PDE lecture in spring.

Strategy of proving the Well-posedness of Mckean-Vlasov equation by PDE approach. Suppose that we have obtained a solution μ of Equation 4.2 with density u, we first plug it into the Equation 4.1 to get

$$\begin{cases} dY(t) = F \star u(t)(Y(t))dt + \sqrt{2}dW_t \\ Y(t) = \xi \in L^2(\Omega) \quad \mathcal{L}(\xi) = \mu_0, \quad u_0 \text{ is the density of } \mu_0 \end{cases}.$$

Now if $F \star u(t)$ is bounded and Lipschitz continuous, then we get a solution $Y(t)$ from the SDE theory. We denote $\overline{u}$ to be density of the Law of $Y(t)$. Then by Itô's formula we have that for $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$, it holds

$$\begin{aligned} &\int_{\mathbb{R}^d} \varphi(x,t)\overline{u}(x,t) - \int_{\mathbb{R}^d} \varphi(x,0)u_0(x)dx \\ &= \int_0^t \int_{\mathbb{R}^d} \left(\frac{\partial \varphi}{\partial t}(x,s) + \nabla\varphi(x,s) \cdot F \star u(x,s) + \Delta\varphi(x,s) \right) \overline{u}(x,s)dxds. \end{aligned}$$

This means that $\overline{\mu}$ satisfies the following linear PDE in the weak sense:

$$\begin{cases} \partial_t \overline{u} - \Delta \overline{u} + \nabla \cdot (F \star u \overline{u}) = 0 \\ \overline{u}|_{t=0} = u_0 \end{cases}. \tag{4.3}$$

If we can prove $\overline{u} = u$, then we get a solution to the Equation 4.1.

This chapter is arranged in the following: in the first section we review the heat kernel to give a first impression in solving a diffusion type of PDE. It will also be used in writing the weak solution formulation of Equation 4.2. In the second section, we prove that Equation 4.2 has a unique solution by using Leray-Schauder fixed point theorem, the exact definition of the solution is given in detail later. Then in the last section, we show that $u = \overline{u}$ by using the theory of backward PDE, which complete the proof of solving Equation 4.1. The whole chapter is aim to introduce the framework, therefore assumptions on the interaction kernel are not optimal. Actually, the method is more powerful to solve the Equation 4.1 with singular interaction kernels.

4.1 Heat Equation and the Heat Kernel

The Cauchy problem of heat equation with source term f is formulated in the following:

$$(\text{HE}) \begin{cases} \partial_t u(x,t) - \Delta u(x,t) & = f(x,t) \\ u|_{t=0} & = u_0 \end{cases}. \tag{4.4}$$

We show that this problem can be solve explicitly by using Fourier transform: $\mathcal{F} : L^1(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d)$ given by

$$\hat{u}(k) := \mathcal{F}(u)(k) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} u(x)e^{ix\cdot k}dx.$$

We will omit the details in proving that the Fourier transform can be extended to a map from $L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and its list of properties.

Exercise. Prove formally that $\widehat{\nabla u} = \frac{k}{i}\hat{u}(k)$ and $-\widehat{\Delta u} = |k|^2\hat{u}(k)$.

Using the Fourier transformation we can transform the hear equation Equation 4.4 into the following ODE, namely for $k \in \mathbb{R}^d$ we have that is

$$\begin{cases} \partial_t \hat{u}(k) + |k|^2\hat{u}(k) = \hat{f}(k) \\ \hat{u}_0(k) = \hat{u}_0 \end{cases}.$$

Then by the solution representation of linear ODE, we have $\forall k \in \mathbb{R}^d$ that

$$\hat{u}(k,t) = e^{-|k|^2 t}\hat{u}_0(k) + \int_0^t e^{-|k|^2(t-\tau)}\hat{f}(k,\tau)d\tau.$$

We use the inverse Fourier transform to get the solution formula for Equation 4.4:

$$u(x,t) = \frac{1}{(4\pi t)^{\frac{d}{2}}}\int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{4t}}u_0(y)dy + \int_0^t\int_{\mathbb{R}^d}\frac{1}{(4\pi(t-\tau))^{\frac{d}{2}}}e^{\frac{-|x-y|^2}{4(t-\tau)}}f(y,\tau)dyd\tau. \tag{4.5}$$

In the above formula, function $\frac{1}{(4\pi t)^{\frac{d}{2}}}e^{-\frac{|x|^2}{4t}}$ plays an important role. It is called the heat kernel, we will use the notation

$$K(x,t) = \frac{1}{(4\pi t)^{\frac{d}{2}}}e^{-\frac{|x|^2}{4t}}.$$

It is easy to prove that $K \xrightarrow{t\to 0^+} \delta$ in the sense of distribution. Furthermore, direct computation shows that $\forall t > 0$ it holds

$$\partial_t K - \Delta K = 0.$$

Since the above discussions are rather formal, in the next we prove rigorously that Equation 4.5 gives exactly the solution of Equation 4.4.

Theorem 4.1.1 (Solution to the Heat Equation)**.** Let the initial data $u_0 \in \mathcal{C}_b(\mathbb{R}^d)$ and $f \in \mathcal{C}^{2,1}(\mathbb{R}^d \times [0,T])$ with compact support, then

$$u(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy + \int_0^t\int_{\mathbb{R}^d} K(x-y,t-s)f(y,s)dyds$$

is a solution to the heat equation

,

Proof. By denoting

$$u_1(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy$$
$$u_2(x,t) = \int_0^t\int_{\mathbb{R}^d} K(x-y,t-s)f(y,s)dyds$$

and the superposition principle, we only need to prove that u_1 and u_2 are solutions to

$$(\text{P1})\begin{cases}\partial_t u_1 - \Delta u_1 = 0\\ u_1(0) = u_0\end{cases} \qquad (\text{P2})\begin{cases}\partial_t u_2 - \Delta u_1 = f\\ u_2(0) = 0\end{cases}.$$

respectively.

We begin with showing that u_1 is a solution to (P1). Obviously, $\partial_t u_1 - \Delta u_1 = 0$, it remains to show that it satisfies the initial data. Actually, for $x \in \mathbb{R}^d$ it holds

$$\begin{aligned}\lim_{t\to 0^+} u_1(x,t) \lim_{t\to 0^+} &= \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy\\ &= \lim_{t\to 0^+}\int_{\mathbb{R}^d}\frac{1}{(4\pi t)^{\frac{d}{2}}}e^{-\frac{|x-y|^2}{4t}}u_0(y)dy\\ &= \lim_{t\to 0^+}\int_{\mathbb{R}^d}\frac{1}{\pi^{\frac{d}{2}}}e^{-|z|^2}u_0(x+2\sqrt{t}z)dz\\ &= \int_{\mathbb{R}^d}\lim_{t\to 0^+}\frac{1}{\pi^{\frac{d}{2}}}e^{-|z|^2}u_0(x+2\sqrt{t}z)dz = u_0(x),\end{aligned}$$

where we used the change of variables $\frac{x-y}{2\sqrt{t}} = -z$.

For $u_2(x,t)$, first note that the initial condition is satisfied $\lim_{t\to 0^+} u_2(x,t) = 0$. Next, we show that it satisfies the equation.

$$
\begin{aligned}
(\partial_t - \Delta)u_2 &= \int_0^t \int_{\mathbb{R}^d} K(y,s)(\partial_t - \Delta_x) f(x-y,t-s)dyds \\
&+ \int_{\mathbb{R}^d} K(y,t)f(x-y,0)dy \\
&= \left(\int_0^{\varepsilon}\int_{\mathbb{R}^d} + + \int_{\varepsilon}^{t}\int_{\mathbb{R}^d}\right) K(y,s)\left(-\partial_s - \Delta_y\right) f(x-y,t-s)dyds \\
&+ \int_{\mathbb{R}^d} K(y,t)f(x-y,0)dy \\
&=: I_\varepsilon + J_\varepsilon + L.
\end{aligned}
$$

We are allowed to exchange the order of differential and integral operators, since the Heat-Kernel decays exponentially in the space variable, which gives uniform integrability of the integrands.

Since $f \in C_b^{2,1}$ and the space integral of the heat kernel is 1, we get that

$$
|I_\varepsilon| \le C \cdot \varepsilon
$$

For J_ε, since it is away from the singular point of the heat kernel and f has compact support, we do integral by parts and obtain

$$
\begin{aligned}
J_\varepsilon &= \int_\varepsilon^t \int_{\mathbb{R}^d} K(y,s)(-\partial_t - \Delta_y) f(x-y,t-s)dyds \\
&= \int_\varepsilon^t \int_{\mathbb{R}^d} \underbrace{(-\partial_t - \Delta_y)K(y,s)}_{=0} f(x-y,t-s)dyds \\
&+ \int_{\mathbb{R}^d} K(y,\varepsilon) - f(x-y,t-\varepsilon)dy \\
&\underbrace{- \int_{\mathbb{R}^d} K(y,t)f(x-y,0)dy}_{=L}.
\end{aligned}
$$

Thus we have proved that

$$
\partial_t u_2 - \Delta u_2 = \lim_{\varepsilon\to 0}\left(\int_{\mathbb{R}^d} \underbrace{K(y,\varepsilon)}_{\to\delta} f(x-y,t-\varepsilon)dy + \underbrace{C\varepsilon}_{\to 0}\right) = f(x,t).
$$

□

4.2 Well-posedness of nonlocal PDE Equation 4.2

We will only use the L^2 weak solution framework to show the existence and uniqueness of weak solution.

As a necessary tool, we introduce the H^1 Space and the definition of L^2 weak solution of Equation 4.2.

Definition 4.2.1 (H^1 Sobolev Spaces). We define

$$
H^1(\mathbb{R}^d) = \{u \in L^2(\mathbb{R}^d) : \nabla u \in L^2(\mathbb{R}^d)\}
$$
with norm $\|u\|_{H^1} = \|u\|_{L^2} + \|\nabla u\|_{L^2}$,

where the gradient is defined in the sense of distribution, namely, for $\forall \varphi \in C_0^\infty$,

$$\langle \nabla u, \varphi \rangle = -\langle u, \nabla \varphi \rangle.$$

We denote its dual space as

$$H^{-1}(\mathbb{R}^d) = (H^1(\mathbb{R}))' = \{l : l \text{ is bounded linear functional of } H^1(\mathbb{R}^d)\}.$$

We will also use the space

$$L^2(0,T;H^1(\mathbb{R}^d)) = \left\{u : \int_0^T \|u(t)\|^2_{H^1(\mathbb{R}^d)} dt < \infty\right\}$$
$$L^\infty(0,T;L^2(\mathbb{R}^d)) = \left\{u : \operatorname*{ess\,sup}_{t\in(0,T)} \|u(t)\|_{L^2(\mathbb{R}^d)} < \infty\right\}.$$

Remark. The Sobolev space H^1 is a separable Hilbert space

We will also use the following compactness result:

Lemma 4.2.1 (Compact Embedding)**.** Let

$$V(\mathbb{R}^d) := \{u \in H^1(\mathbb{R}^d) : u \in L^1(\mathbb{R}^d) \text{ and } \int_{\mathbb{R}^d} |x|^2|u|(x)dx < \infty\},$$

then $V(\mathbb{R}^d)$ embedded in $L^p(\mathbb{R}^d)$, $\forall p \in [1, \frac{dp}{d-p})$ compactly. Namely for any sequence $(u_n)_{n\in\mathbb{N}}$, if $\exists C > 0$ such that

$$\|u_n\|_{H^1(\mathbb{R}^d)} + \int_{\mathbb{R}^d} (1+|x|^2)|u(x)|dx \le C,$$

then there exists a subsequence $(u_{n_j})_{j\in\mathbb{N}}$ and a function $u \in V(\mathbb{R}^d)$ such that

$$\|u_{n_j} - u\|_{L^p(\mathbb{R}^d)} \to 0, \text{ for } p \in [1, \frac{dp}{d-p}).$$

The proof can be done by the compact embedding in bounded domains and the bounded control of moment.

Lemma 4.2.2 (Aubin-Lions Lemma, Simon Page 87, Cor. 6)**.** Let Banach spaces satisfy $X \subset B \subset Y$, and the embedding $X \to B$ be compact. Let $1 < q \le \infty$. If a set of functions F is bounded in $L^q(0,T;B) \cap L^1_{loc}(0,T;X)$ and $\partial_t F$ is bounded in $L^1_{loc}(0,T;Y)$. The F is relatively compact in $L^p(0,T;B)$, $\forall p < q$.

Then as a corollary of the above two lemmata, take spaces $X = V(\mathbb{R}^d)$, $B = L^2(\mathbb{R}^d)$, $Y = H^{-1}(\mathbb{R}^d)$, we obtain the following corollary that we are going to use in the lecture.

Lemma 4.2.3 (Aubin-Lions Lemma, special version)**.** If $(u_n)_{n\in\mathbb{N}}$ satisfies

1. $\sup_t \int (1+|x|^2)|u_n(t,x)|dx \le C$
2. $\|u_n\|_{L^2(0,T;H^1(\mathbb{R}^d))} + \|u_n\|_{L^\infty(0,T;L^2(\mathbb{R}^d))} \le C$
3. $\|\partial_t u_n\|_{L^1(0,T;H^{-1}(\mathbb{R}^d))} \le C$

Then (u_n) is relatively compact in $L^q(0,T;L^2(\mathbb{R}^d))$ i.e. there $\exists u \in L^q(0,T;L^2(\mathbb{R}^d))$, $\forall 1 \le q < \infty$ s.t. there is a subsequence converge to it:

$$\|u_{n_j} - u\|_{L^q(0,T;L^2(\mathbb{R}^d))} \to 0.$$

Definition 4.2.2 (Weak Solution)**.** We say that a function

$$u \in L^2(0,T;H^1(\mathbb{R}^d) \cap L^\infty(0,T;L^2(\mathbb{R}^d))),$$

with $\partial_t u \in L^2(0,T;H^{-1}(\mathbb{R}^d))$ is a weak solution of the Equation 4.2 if for $\forall \varphi \in L^2(0,T;H^1(\mathbb{R}^d))$ it holds

$$\int_0^T \langle \partial_t u, \varphi \rangle_{(H^{-1},H^1)} dt = \int_0^T \int_{\mathbb{R}^d} \nabla\varphi \cdot F \star u u dxdt - \int_0^T \int_{\mathbb{R}^d} \nabla u \cdot \nabla \varphi dxdt.$$

The main theorem of this section is:

Theorem 4.2.1. Assume that $F \in L^2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, $0 \le u_0 \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} |x|^2 u_0(x)dx$ is finite, then Equation 4.2 has a unique weak solution as defined in Definition 4.2.2. In addition, $u \in L^\infty(0,T;L^1(\mathbb{R}^2))$ and $\int_{\mathbb{R}^d} |x|^2 u(x,t)dx$ is finite for any given $t > 0$.

We will use the Leray-Schauder fixed point theorem to prove that.

Theorem 4.2.2 (Leray-Schauder Fixed Point Theorem)**.** Let U be a Banach Space and

$$T : (u,\sigma) \in U \times [0,1] \to U.$$

If

1. T is compact
2. $T(u,0) = 0$ for $\forall u \in U$
3. $\exists C > 0$ s.t for $\forall u \in U$ with $u = T(u,\sigma)$ for some $\sigma \in [0,1]$ it holds

$$\|u\|_U \le C.$$

Then the map $T(\cdot,1)$ has a fixed point

Proof strategy for Theorem 4.2.1. Consider the Banach space $U = L^q(0,T;L^2(\mathbb{R}^d))$ for a fixed $2 < q < \infty$. We define a map $T : (u,\sigma) \in U \times [0,1] \to U$ by solving the linearized PDE, i.e. $\forall v \in U$, let $u = T(v,\sigma)$ be the weak solution of

$$\begin{cases} u_t - \Delta u + \sigma \nabla \cdot (F \star v \ u) = 0 \\ u|_{t=0} = \sigma u_0 \end{cases}. \tag{4.6}$$

The following points should be checked:

1. The existence and uniqueness of u of problem Equation 4.6. This is given in Theorem 4.2.4. And the compactness of operator T is also given in Theorem 4.2.4 by the uniform estimates and the Aubin-Lions Lemma 4.2.3.
2. For $\sigma = 0$, the problem reduces to heat equation which can be solved by using the fundamental solution representation $u = T(v,0) = 0$.
3. The uniform estimate for arbitrary fix point can be done directly by using the same estimates in Theorem 4.2.4.

This means we can apply Leray-Schauder Fixed Point Theorem and get a fix point $T(\cdot,1)$ which is a solution to Equation 4.2.

As a preparation, we first study a linear equation with given drift term, with which we can build an iteration and prove that the map has a fixed point to solve Equation 4.2. The linear equation reads:

$$\text{(LDE)}\begin{cases} \partial_t - \Delta u + \nabla \cdot (\overline{b}(x,t)u) = 0 \\ u|_{t=0} = u_0 \end{cases} \tag{4.7}$$

With the help of heat kernel representation we give a version of weak solution of Equation 4.7:

Definition. $u \in L^\infty(0,T;L^1(\mathbb{R}^d))$ is called a mild solution of Equation 4.7 if

$$u(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy + \int_0^t \int_{\mathbb{R}^d} \nabla K(x-y,t-\tau)\cdot(\overline{b}(y,\tau)u(y,\tau))dyd\tau.$$

Theorem 4.2.3 (Well-posedness of Equation 4.7)**.** For any given T, if $\overline{b} \in L^q(0,T;L^\infty(\mathbb{R}^d))$, $2 < q \leq \infty$ and $u_0 \in L^1(\mathbb{R}^d)$, then the (LDE) has a unique mild solution $u \in L^\infty(0,T;L^1(\mathbb{R}^d))$.

Proof. We build up an iteration and prove it has a fixed point. Namely, consider a map

$$\begin{aligned} &\mathcal{T} : L^\infty(0,T;L^1(\mathbb{R}^d)) \to L^\infty(0,T;L^1(\mathbb{R}^d)) \\ &u \mapsto \mathcal{T}(u) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy + \int_0^t \int_{\mathbb{R}^d} \nabla K(x-y,t-\tau)\cdot(\overline{b}(y,\tau)u(y,\tau))dyd\tau. \end{aligned}$$

We first show that $\mathcal{T}(u) \in L^\infty(0,T;L^1(\mathbb{R}^d))$. Actually, for $\forall t > 0$, it holds

$$\begin{aligned} \int_{\mathbb{R}^d} |\mathcal{T}(u)(x,t)|dx \leq& \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(x-y,t)|u_0(y)|dydx \\ &+ \int_0^t d\tau \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} dy |\nabla K(x-y,t-\tau)\overline{b}(y,\tau)u(y,\tau)| \\ &= I + II. \end{aligned}$$

Since we have fixed $t > 0$ we use Fubini and obtain $I \leq \|u_0\|_{L^1(\mathbb{R}^d)}$.

Considering the fact that

$$\int_{\mathbb{R}^d} |\nabla K(x,s)|dx = \frac{1}{(4\pi s)^{\frac{d}{2}}}\int_{\mathbb{R}^d} \frac{1}{\sqrt{s}}\left|\frac{x}{2\sqrt{s}}\right| e^{-\frac{|x|^2}{4s}}dx \leq \frac{1}{\sqrt{s}}C.$$

Then for the second term we get for $1 \leq q' < 2$

$$\begin{aligned} II \leq& \int_0^t d\tau \|\overline{b}(\cdot,\tau)\|_{L^\infty(\mathbb{R}^d)} \int_{\mathbb{R}^d} |\nabla K(x,t-\tau)|dx \int_{\mathbb{R}^d} u(y,\tau)dy \\ \leq& \|u\|_{L^\infty(L^1)} C \cdot \int_0^t \frac{1}{\sqrt{t-\tau}} \|\overline{b}(\cdot,\tau)\|_{L^\infty(\mathbb{R}^d)}d\tau \leq C\|\overline{b}\|_{L^q(0,t;L^\infty(\mathbb{R}^d))}\frac{1}{1-\frac{q'}{2}}t^{1-\frac{q'}{2}} \leq C^\star t^{1-\frac{q'}{2}}. \end{aligned}$$

Note that we do not need to consider y in K since we can use a translation, $L^\infty(L^1) = L^\infty(0,T;L^1(\mathbb{R}^d))$.

This shows that the map $\mathcal{T}$ is indeed well defined.

Next we prove that $\mathcal{T}(u)$ is a contraction.

In fact, $\forall u_1, u_2 \in L^\infty(L^1)$, for $t^\star$ s.t. $C^\star t^{\star 1-\frac{q'}{2}} < \frac{1}{2}$, by doing similar argument as before, we have

$$
\begin{aligned}
&\|\mathcal{T}(u_1)-\mathcal{T}(u_2)\|_{L^\infty(L^1)}\\
&=\operatorname*{ess\,sup}_{0\le t\le t^\star}\int_{\mathbb{R}^d}|\mathcal{T}(u_1)-\mathcal{T}(u_2)|(x,t)dx\\
&\le \operatorname*{ess\,sup}_{0\le t\le t^\star}\int_0^t d\tau\int_{\mathbb{R}^d}dy|\nabla K(x-y,t-\tau)(\overline{b}(y,\tau)(u_1-u_2))(y,\tau)|\\
&\le \operatorname*{ess\,sup}_{0\le t\le t^\star}\|u_1-u_2\|_{L^\infty(L^1)}\int_0^t\frac{1}{\sqrt{t-\tau}}\|\overline{b}(\cdot,\tau)\|_{L^\infty(\mathbb{R}^d)}d\tau\\
&\le C^\star {t^\star}^{1-\frac{q'}{2}}\|u_1-u_2\|_{L^\infty(L^1)}\le\frac{1}{2}\|u_1-u_2\|_{L^\infty(L^1)}.
\end{aligned}
$$

Then $\mathcal{T}$ is a contraction on space $L^\infty(0,t^\star;L^1(\mathbb{R}^d))$.

Since $t^\star$ only depends on $\|\overline{b}\|_{L^q(0,T;L^\infty(\mathbb{R}^d))}$ and dimension d, then for the given $T>0$, we can repeat the above argument finite times and obtain that $u\in L^\infty([0,T];L^1(\mathbb{R}^d))$ satisfies

$$
u(x,t)=\int_{\mathbb{R}^d}K(x-y,t)u_0(y)dy+\int_0^t\int_{\mathbb{R}^d}\nabla K(x-y,t-\tau)\cdot(\overline{b}(y,\tau)u(y,\tau))dyd\tau.
$$

□

Theorem 4.2.4. Assume that $F\in L^2(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)$, $v\in L^q(0,T;L^2(\mathbb{R}^d))$ $(q>2)$, $u_0\in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d}|x|^2u_0(x)dx$ is finite, then Equation 4.6 has a unique weak solution in $L\infty(0,T;L^2(\mathbb{R}^d))\cap L^2(0,T;H^1(\mathbb{R}^d))$, namely for $\forall\varphi\in L^2(0,T;H^1(\mathbb{R}^d))$ it holds

$$
\int_0^T\langle\partial_t u,\varphi\rangle_{\langle H^{-1},H^1\rangle}dt=-\int_0^T\int_{\mathbb{R}^d}\nabla u\nabla\varphi-F\star v\cdot u\nabla\varphi dxdt.
$$

In addition, $u\in L^\infty(0,T;L^1(\mathbb{R}^2))$ and $\int_{\mathbb{R}^d}|x|^2u(x,t)dx$ is bounded for any given $t>0$. If in addition $u_0\ge0$, then $u\ge0$ a.e..

Proof. We introduce first a mollified problem and prove that its mild solution exists, then proceed the uniform estimate and do compactness argument to show that the weak solution exists.

Notice that $v\in L^q(0,T;L^2(\mathbb{R}^d))$, we consider the following problem

$$
(\text{PDE})_\varepsilon\begin{cases} u_t^\varepsilon-\Delta u^\varepsilon+\nabla\cdot(\tilde{j}_\varepsilon\star(F\star v(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}u^\varepsilon))=0\\ u^\varepsilon|_{t=0}=j_\varepsilon\star(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}u_0)\end{cases}. \tag{4.8}
$$

where j_ε is the mollification kernel in x and $\tilde{j}_\varepsilon$ is mollification in x and t. Here without loss of generality, for the mollification of a function in t variable, we do the mollification of its zero extension of this function. This fact will not be explicitly mentioned.

Notice that these mollification operations are aimed to show that u^ε is not only a mild solution, but also a classical solution for fixed ε. The we can choose u^ε legally as test function to proceed further the L^2 estimates.

By modifying slightly of the proof of Theorem 4.2.3, we obtain that Equation 5.5 has a unique mild solution

$$
\begin{aligned}
u^\varepsilon(x,t)=&\int_{\mathbb{R}^d}K(x-y,t)j_\varepsilon\star(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}u_0)(y)dy \qquad (4.9)\\
&+\int_0^t\int_{\mathbb{R}^d}K(x-y,t-s)\nabla\cdot(\tilde{j}_\varepsilon\star(F\star v(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}u^\varepsilon))(y,s)dyds.
\end{aligned}
$$

By Theorem 4.1.1, then u^ε satisfies Equation 5.5 in the classical sense.

In the next, we do the uniform in ε estimates for u^ε.

1. $L^\infty(0,T;L^1(\mathbb{R}^d))$ **estimate**

   Similar to the estimates given in Theorem 4.2.3, we can obtain

$$\begin{aligned}\|u^\varepsilon(\cdot,t)\|_{L^1(\mathbb{R}^d)} \le \|K(t,\cdot)\|_{L^1(\mathbb{R}^d)} \cdot \|j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}} u_0)\|_{L^1(\mathbb{R}^d)} \\ + \int_0^t \frac{1}{\sqrt{t-s}} \|\tilde{j}_\varepsilon \star (F \star v(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}\ u^\varepsilon))(\cdot,s)\|_{L^\infty(\mathbb{R}^d)} ds \\ \le \|u_0\|_{L^1(\mathbb{R}^d)} + \int_0^t \frac{1}{\sqrt{t-s}} \|\tilde{j}_\varepsilon \star (F\star v)(\cdot,s)\|_{L^\infty(\mathbb{R}^d)} \|u^\varepsilon(\cdot,s)\|_{L^1(\mathbb{R}^d)} ds \\ \le \|u_0\|_{L^1(\mathbb{R}^d)} + \int_0^t \frac{1}{\sqrt{t-s}} \|F\|_{L^2(\mathbb{R}^d)} \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \|u^\varepsilon(\cdot,s)\|_{L^1(\mathbb{R}^d)} ds \\ \le \|u_0\|_{L^1(\mathbb{R}^d)} + C\int_0^t \frac{1}{\sqrt{t-s}} \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \|u^\varepsilon(\cdot,s)\|_{L^1(\mathbb{R}^d)} ds.\end{aligned}$$

   Since $\int_0^t \frac{1}{\sqrt{t-s}} \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} ds \le C\|v\|_{L^q(0,T;L^2(\mathbb{R}^d))} t^{1-\frac{q'}{2}}$ for $2 < q \le \infty$ (i.e. $1 \le q' < 2$), then by Gronwall's inequality, we have

$$\|u^\varepsilon(\cdot,t)\|_{L^1(\mathbb{R}^d)} \le C(t, \|v\|_{L^q(0,T;L^2(\mathbb{R}^d)}, \|u_0\|_{L^1(\mathbb{R}^d)}).$$

2. L^2 **estimate**

   we multiply the Equation 5.5 by u^ε and integrate on $\mathbb{R}^d$, by using integration by parts we have

$$\int_{\mathbb{R}^d} \partial_t u^\varepsilon \cdot u^\varepsilon - \int_{\mathbb{R}^d} \Delta u^\varepsilon \cdot u^\varepsilon = -\int \nabla \cdot (\tilde{j}_\varepsilon \star (F \star v(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}\ u^\varepsilon)))u^\varepsilon$$

   therefore by using Hölder's inequality, we obtain the standard $L^2(\mathbb{R}^d)$ estimates

$$\frac{d}{dt}\int_{\mathbb{R}^d} |u^\varepsilon|^2 dx + \int_{\mathbb{R}^d} |\nabla u^\varepsilon|^2 dx \le \int_{\mathbb{R}^d} \left|\tilde{j}_\varepsilon \star (F \star v(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}\ u^\varepsilon)\right|^2 dx.$$

   The using similar tricks as in the last step we have

$$\begin{aligned}\int_{\mathbb{R}^d} |u^\varepsilon(x,t)|^2 dx + \int_0^t \int_{\mathbb{R}^d} |\nabla u^\varepsilon(x,s)|^2 dx dx \\ \le \int_{\mathbb{R}^d} |u_0|^2 dx + C\int_0^t \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \int_{\mathbb{R}^d} |u^\varepsilon(x,s)|^2 dx.\end{aligned}$$

   After applying Grönwall's inequality we have

$$\sup_{0\le t\le T} \|u^\varepsilon\|_{L^2}^2 + \int_0^t \int_{\mathbb{R}^d} |\nabla u^\varepsilon|^2 \le C(\|u_0\|_{L^2(\mathbb{R}^d)}, \|v\|_{L^q(0,T;L^2(\mathbb{R}^d))}).$$

We need to further derive the estimate for $\partial_t u^\varepsilon$. This can be obtained immediately by using the equation itself. Namely, $\forall \varphi \in C_0^\infty([0,T);\mathbb{R}^d)$

$$\begin{aligned}&\langle \partial_t u^\varepsilon, \varphi\rangle = \langle \Delta u^\varepsilon - \nabla\cdot(\tilde{j}_\varepsilon \star (F\star v(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}}\ u^\varepsilon))), \varphi\rangle \\ \le& \|\nabla u^\varepsilon\|_{L^2(0,T;L^2(\mathbb{R}^d))} \cdot \|\nabla\varphi\|_{L^2(0,T;L^2\mathbb{R}^d)} \\ &+ \|F\|_{L^2(\mathbb{R}^d)} \|v\|_{L^q(0,T;L^2(\mathbb{R}^d))} \|u^\varepsilon\|_{L^2(0,T;L^2(\mathbb{R}^d))} \cdot \|\nabla\varphi\|_{L^{q'}(0,T;L^2(\mathbb{R}^d))}, \quad q \in (2,\infty).\end{aligned}$$

It means

$$\|\partial_t u^\varepsilon\|_{L^{q'}(0,T;H^{-1}(\mathbb{R}^d))} \le C.$$

Now we take a convergent subsequence (not relabeled), there exists u s.t.

$$u_\varepsilon \overset{\star}{\rightharpoonup} u \quad \text{in } L^\infty(0,T;L^2(\mathbb{R}^d)) \cap L^2(0,T;H^1(\mathbb{R}^d)) \cap L^{q'}(0,T;L^2(\mathbb{R}^d)).$$

and u satisfies the weak version of the PDE, which means that for any text function φ it holds

$$\int_0^T \langle \partial_t u, \varphi \rangle_{(H^1,H^1)} dt = -\int_0^T \int_{\mathbb{R}^d} (\nabla u - F \star v \cdot u) \cdot \nabla \varphi dx dt.$$

Here we omit the tedious weak convergence argument in taking the limit in Equation 5.5, since the equation is linear.

Next we prove that $u_0 \geq 0$ implies $u \geq 0$ a.e.

Choose $\varphi = u^- = \min\{0, -u\}$, $(u \in L^2(0,T;H^1(\mathbb{R}^d)) \Rightarrow u^- \in L^2(0,T;H^1(\mathbb{R}^d)))$, we have

That u_- lies in the space is in fact non trivial and is part of the PDE lecture

$$\int_0^t \partial_t u \cdot u^- ds = \int_0^t \int_{\mathbb{R}^d} -\nabla u \cdot \nabla u^- dx ds - \int_0^t \int_{\mathbb{R}^d} F \star v \cdot u \nabla u^- dx ds.$$

We proceed further and get

$$\begin{aligned} &\frac{1}{2} \int_0^t \int_{\mathbb{R}^d} \partial_t |u^-|^2 dx ds + \int_0^t \int_{\mathbb{R}^d} |\nabla u^-|^2 dx ds \\ \leq &\frac{1}{2} \int_0^t \int_{\mathbb{R}^d} |\nabla u^-|^2 dx ds + \frac{C}{2} \int_0^t \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \int_{\mathbb{R}^d} |u^-|^2 dx ds. \end{aligned}$$

Then Gronwall's inequality implies

$$\int_{\mathbb{R}^d} |u^-|^2 dx \leq e^{C(\|v\|_{L^q(0,T;L^2(\mathbb{R}^d))})} \int_{\mathbb{R}^d} |u_{0-}|^2 dx = 0,$$

where the nonnegative inital data menas that $u_{0-} = 0$

The uniqueness of the solution follows also from the L^2 estimate similarly.

Based on the nonnegativity of the solution, we obtain further the second moment estimate. One can choose a mollified version of $|x|^2$ as a test function and obtain the boundedness through Gronwall's inequality. In the next, we do a second moment estimate from the mild solution formulation:

Second Moment estimate We do direct estimate by using the mild solution representation of u

$$\begin{aligned} \int_{\mathbb{R}^d} |x|^2 u(x,t) dx &= \int_{\mathbb{R}^d} |x|^2 \int_{\mathbb{R}^d} K(x-y,t) u_0(y) dy dx \\ &+ \int_0^t \int_{\mathbb{R}^d} dx |x|^2 \int_{\mathbb{R}^d} \nabla K(x-y,t-s) \cdot (F \star vu)(s,y) dy ds \\ &= I + II. \end{aligned}$$

We bound them again individually.

$$\begin{aligned} I &\leq \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} (|x-y|^2 + |y|^2) K(x-y,t) u_0(y) dy dx \\ &\leq \int_{\mathbb{R}^d} |x|^2 K(x,t) dx \cdot \int_{\mathbb{R}^d} u_0(y) dy + \int_{\mathbb{R}^d} |y|^2 u_0(y) dy \\ &\leq C \cdot t \|u_0\|_{L^1(\mathbb{R}^d)} + \int_{\mathbb{R}^d} |y|^2 u_0(y) dy, \end{aligned}$$

where the second moment of heat kernel is computed in the following:

$$\int_{\mathbb{R}^d} |x|^2 K(x,t)dx = 4t \int_{\mathbb{R}^d} \frac{|x|^2}{4t} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} dx \le Ct.$$

For II, we use the following estimate

$$\begin{aligned} II &\le \int_0^t \int_{\mathbb{R}^d} dx \underbrace{2|x|}_{|x|\le|x-y|+|y|} \int_{\mathbb{R}^d} K(x-y,t-s)\cdot(F\star vu)(s,y)dyds \\ &\le \int_0^t \int_{\mathbb{R}^d} dx|x|K(x,t-s) \int_{\mathbb{R}^d} (F\star vu)(s,y)dyds \\ &\quad + C\cdot \int_0^t \int_{\mathbb{R}^d} K(x,t-s)dx \cdot \int |y|(F\star vu)(y,s)dy \\ &\le C\cdot \int_0^t \sqrt{t-s}\|F\|_{L^2(\mathbb{R}^d)}\|v(\cdot,s)\|_{L^2(\mathbb{R}^d)}\|u(\cdot,s)\|_{L^1(\mathbb{R}^d)}ds \\ &\quad + C\cdot \int_0^t \|F\|_{L^2(\mathbb{R}^d)}\|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \int_{\mathbb{R}^d} |y|u(y)dy \\ &\le C(t) + C\int_0^t \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \int_{\mathbb{R}^d} |y|^2 u(y,s)dyds, \end{aligned}$$

where we have used the first moment of heat kernel

$$\int_{\mathbb{R}^d} |x|K(x,t)dx = 4t\int_{\mathbb{R}^d} \frac{|x|}{2\sqrt{t}} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} dx \le C\sqrt{t}.$$

Therefore we have obtained

$$\int_{\mathbb{R}^d} |x|^2 u(x,t)dx \le C(t) + C\int_0^t \|v(\cdot,s)\|_{L^2(\mathbb{R}^d)} \int_{\mathbb{R}^d} |y|^2 u(y,s)dyds,$$

which implies by Gronwall's inequality that

$$\int_{\mathbb{R}^d} |x|^2 u(x,t)dx \le C(t, \|v\|_{L^q(0,T;L^2(\mathbb{R}^d)}, \|u_0\|_{L^1(\mathbb{R}^d)}, \||\cdot|^2 u_0\|_{L^1(\mathbb{R}^d)}).$$

This finished the proof of Theorem 4.2.4. □

Now we are ready to finish the proof of page 69.

Proof. of page 69. To prove the existence, as described in item 4.2, we only need to show that the map $v \in U \to u$ is compact. By Theorem 4.2.4, we know that the estimates for u satisfies the assumptions of Aubin-Lions lemma (4.2.3), therefore, the operator T is compact.

Furthermore, one can proceed the same estimates as has been done in Theorem 4.2.4 to obtain the estimate for and fixed point of map T. Therefore Theorem 4.2.2 gives us that the solution of (4.2) exists.

Finally, we prove the uniqueness. Suppose there are two (weak) solutions u_1, u_2, then we consider the difference

$$w = u_1 - u_2.$$

which satisfies (in a weak sense)

$$\begin{cases} \partial_t w - \Delta w + \nabla\cdot(F\star u_1\cdot w) + \nabla\cdot(F\star w\cdot u_2) = 0 \\ w|_{t=0} = 0 \end{cases}.$$

take w as a test function

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^d} w^2 dx + \int_{\mathbb{R}^d} |\nabla w|^2 dx \\
\leq& \int_{\mathbb{R}^d} |\nabla w \cdot (F \star u_1) w| dx + \int_{\mathbb{R}^d} |\nabla w \cdot (F \star w) u_2| dx \\
\leq& \frac{1}{2}\int_{\mathbb{R}^d} |\nabla w|^2 dx + C\int_{\mathbb{R}^d} (|F \star u_1|^2 |w|^2 + |F \star w|^2 |u_2|^2) dx \\
\leq& \frac{1}{2}\int_{\mathbb{R}^d} |\nabla w|^2 dx + \|F \star u_1\|^2_{L^\infty(\mathbb{R}^d)} \int_{\mathbb{R}^d} |w|^2 dx + \|F \star w\|^2_{L^\infty(\mathbb{R}^d)} \|u_2\|^2_{L^2(\mathbb{R}^d)} \\
\leq& \frac{1}{2}\int_{\mathbb{R}^d} |\nabla w|^2 dx + C\int_{\mathbb{R}^d} |w|^2 dx.
\end{aligned}
$$

where we used

$$
\operatorname*{ess\,sup}_x \left| \int F(x-y) w(y) dy \right|^2 \leq \|F\|_2^2 \|w\|_2^2.
$$

then by Gronwall's inequality

$$
\frac{d}{dt}\int |w|^2 \leq C \cdot \int |w|^2 \Rightarrow \int_{\mathbb{R}^d} |w(x,t)^2| dx \leq e^{Ct} \int_{\mathbb{R}^d} |w(x,0)|^2 dx = 0.
$$

it follows $u_1 = u_2$ a.e. □

4.3 Solvability of the Makean-Vlasov Equation

As has been discussed in the beginning of this chapter. we want to prove in this section that the problem Equation 4.3 has a unique solution $\overline{u} = u$ where u is the unique solution of Equation 4.2.

Since the equation is linear, it is enough to show that the problem

$$
\begin{cases} \partial_t \overline{u} - \Delta \overline{u} + \nabla \cdot (F \star u \overline{u}) = 0 \\ \overline{u}|_{t=0} = 0 \end{cases} \tag{4.10}
$$

has only solution $\overline{u} = 0$.

Theorem 4.3.1. The solution of Equation 4.10 is $\overline{u} = 0$.

Proof. The weak formulation of Equation 4.10 is that for any test function $\varphi \in \mathcal{C}_0^\infty([0,T) \times \mathbb{R}^d)$ it holds

$$
\int_{\mathbb{R}^d} \varphi(x,t)\overline{u}(x,t) dx = \int_0^t \int_{\mathbb{R}^d} (\partial_t \varphi + \Delta \varphi - \nabla \varphi F \star u) \overline{u}(x,s) dx ds. \tag{4.11}
$$

In order to show that $\overline{u} = 0$, we have to show that $\forall a.e. \tilde{t} \in (0,T]$, $\forall g \in \mathcal{C}_0^\infty(\mathbb{R}^d)$

$$
\int_{\mathbb{R}^d} g(x)\overline{u}(x,t) dx = 0,
$$

which means that, $\overline{u}(\cdot, \tilde{t}) = 0$.

This means that it is enough to show that for any given $g \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ the following problem

$$
\begin{cases} \partial_t \varphi + \Delta \varphi - \nabla \varphi \cdot (F \star u) = 0 \\ \varphi(x,t) = g(x) \end{cases}.
$$

has a solution $\varphi \in \mathcal{C}_0^\infty([0,T) \times \mathbb{R}^d)$, which can be viewed as a test function to be plugged into Equation 4.11.

Without loss of generality, suppose $\operatorname{supp} g \subset B_{\frac{R}{2}}$, then we consider the corresponding problem

$$\begin{cases} \partial_t w_{R,\varepsilon} + \Delta w_{R,\varepsilon} - \nabla w_{R,\varepsilon} \cdot j_\varepsilon \star (F \star u) = 0 \\ w_{R,\varepsilon}|_{\partial B_R} = 0 \\ w_{R,\varepsilon}|_{t=\tilde{t}} = g \end{cases} . \tag{4.12}$$

This is an initial boundary value problem of linear diffusion equation, the solution theory will be given in the PDE lecture. Here we direclty use the result i.e. there exists a unique solution

$$w_{R,\varepsilon} \in C^\infty(\overline{B_R} \times [0, \tilde{t}]) \text{ with } w_{R,\varepsilon}|_{\partial B_R} = 0.$$

The function $w_{R,\varepsilon}$ should be a candidate for φ, however it is not in $C_0^\infty(\mathbb{R}^d \times [0, \tilde{t}])$. To overcome this difficulty we introduce a $C_0^\infty(\mathbb{R}^d)$ cutoff function

$$\psi_R = \begin{cases} 1 & |x| \le \frac{R}{2} \\ \text{smooth in between} & \\ 0 & |x| \ge R \end{cases},$$

with the property that $|\nabla \psi_R| \le \frac{C}{R}$, $|\Delta \psi_R| \le \frac{C}{R}$. We use $\psi_R \cdot w_{R,\varepsilon}$ as a test function in the weak formulation of Equation 4.10, then it holds

$$\psi_R \cdot w_{R,\varepsilon}|_{t=\tilde{t}} = g.$$

and

$$\int_{\mathbb{R}^d} g(x)\overline{u}(x, \tilde{t})dx - 0 = \int_0^{\tilde{t}} \int_{\mathbb{R}^d} (\partial_t + \Delta - F \star j_\varepsilon \star u \cdot \nabla)(w_{R,\varepsilon}\psi_R)\overline{u}dxdt$$

.

using that $w_{R,\varepsilon}$ is solution to Equation 4.12, we have

$$\begin{aligned} &\int_0^{\tilde{t}} \int_{\mathbb{R}^d} \underbrace{(\partial_t + \Delta - F \star j_\varepsilon \star u \cdot \nabla) w_{R,\varepsilon}}_{=0} \cdot \psi_R d\mu^Y \\ &+ \int_0^{\tilde{t}} \int_{\mathbb{R}^d} (2\nabla \psi_R \cdot \nabla w_{R,\varepsilon} + w_{R,\varepsilon}\Delta \psi_R - F \star u \cdot w_{R,\varepsilon} \cdot \nabla \psi_R) d\mu^Y \\ &+ \int_0^{\tilde{t}} \int_{\mathbb{R}^d} (F \star j_\varepsilon \star u - F \star u) \cdot \nabla w_{R,\varepsilon} \psi_R d\mu^Y \\ &= I + II + III. \end{aligned}$$

if we have that $\|\nabla w_{R,\varepsilon}\|_{L^\infty} + \|w_{R,\varepsilon}\|_{L^\infty} \le C$ uniformly in R and ε (we refer this estimate again to the PDE lecture), then II can be bounded as follows

$$|II| \le \|\overline{u}\|_{L^\infty(L^1)}(\|\nabla w_{R,\varepsilon}\|_\infty \cdot \|\nabla \psi_R\|_\infty + \|w_{R,\varepsilon}\|_\infty \cdot \|\Delta \psi_R\|_\infty + \|F \star u\|_\infty \cdot \|w_{R,\varepsilon}\| \|\nabla \psi_R\|_\infty).$$

For III

$$\begin{aligned} |III| &\le C \cdot \|F \star j_\varepsilon \star u - F \star u\|_\infty \\ &= \left| \int j_\varepsilon(x-y)(F \star u(y) - F \star u(x))dy \right| \le \|D^2 V \star u\|_\infty C\varepsilon^2 \to 0. \end{aligned}$$

This complete the proof. ☐

Chapter 5

Mean Field Limits for Non Lipschitz Interaction

In this chapter we introduce the mollification method to study the interaction particle system with non-Lipschitz force. Let's remember the particle system we proposed

$$(\mathrm{SDE})\begin{cases} dX_i^N = \nabla V \star \mu_N(X_i^N)dt + \sqrt{2}dW_t^i \\ X_i^N(0) = \xi_i \text{ i.i.d } u_0 = \mathcal{L}(\xi_i) \end{cases}.$$

where μ_N is the empirical measure given by the solution of the above SDE, i.e. $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i^N}$.

Notice that the strong solution theory in SDE needs to assume that ∇V is Lipschitz continuous. However, according the modern PDE theory, the expected mean field limit problem

$$\begin{cases} \partial_t u - \Delta u + \nabla\cdot(\nabla V \star u \cdot u) = 0 \\ u(0) = u_0 \end{cases}.$$

has solution even for some singular interaction, which means the Lipschitz continuity of ∇V is not necessary. Actually, in many of the models in applied sciences, ∇V does have singularity. A natural question arises, would it be possible to find its microscopic version in the mean field scaling?

Here is a list of possible methods (not complete):

- One might be able to start with the weak solution theory of the SDE system, which is out of the scope of this lecture.
- One could also start with a higher dimensional linear PDE for the joint law of these N particles, i.e. the so called Liouville equation. Then derive a BBGKY hierarchy and study the limiting hierarchy, this is a classical method in statistical physics, which works also for bounded Lipschitz force.
- Alternatively, one can derive the relative entropy estimate for the Liouville equation and use the superadditivity property of the relative entropy to study the mean-field limit. This framework will be demonstrated in the last chapter of this lecture.
- Another possibility is to start from a smoothed version of the SDE system, where the smoothing effect should vanish in the $N \to \infty$ limit.

The main scope of this chapter is to proceed rigorous analysis for the smoothed interacting system.

More precisely, if ∇V is not Lipschitz continuous, we consider its mollification

$$V_\varepsilon = j_\varepsilon \star V, \quad j_\varepsilon \text{ is a standard mollification kernel,}$$

and consider particle system with this smoothed interaction

$$(\mathrm{SDE}_\varepsilon)\begin{cases} dX_i^{N,\varepsilon}(t) = \nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon}(t))dt + \sqrt{2}dW_t^i \\ X_i^{N,\varepsilon}(0) = \xi_i \text{ i.i.d } u_0 = \mathcal{L}(\xi_i) \end{cases}. \tag{5.1}$$

The regulartization parameter ε should be chosen such that $\varepsilon(N) \to 0$ when $N \to \infty$. Namely, the expected limiting problem is still like before, we list here also the corresponding Mckean-Vlasov SDE.

$$\begin{cases} \partial_t u - \Delta u + \nabla \cdot (\underbrace{\nabla V \star u}_{\text{Lip}} u) = 0 \\ u|_{t=0} = u_0 \end{cases} \tag{5.2}$$

$$\widehat{(\text{SDE})} = \begin{cases} d\hat{X}_i(t) = \nabla V \star u(\hat{X}_i(t), t)dt + \sqrt{2}dW_t^i \\ \hat{X}_i = \xi_i \\ u = \mathcal{L}(\hat{X}_i) \end{cases}. \tag{5.3}$$

We assume that the problem Equation 5.2 has a solution u, so that $\nabla V \star u$ is bounded Lipschitz continuous. Then the SDE Equation 5.3 has an $\mathbb{L}^2$ strong solution. One can proceed the same argument as in the previous chapter, the PDE approach to solve Mckean-Vlasov equation, to prove that u is exactly the law of $\hat{X}$. This gives the solvability of the Mckean-Vlasov equation with non Lipschitz force.

In the whole analysis, the following so called intermediate problem plays an important role to bridge Equation 5.1 and Equation 5.3 (or Equation 5.2),

This intermediate problem is exactly the McKean-Vlasov equation as a mean field limit of Equation 5.1 for fixed ε,

$$\overline{(\text{SDE})} = \begin{cases} d\overline{X}_i^\varepsilon(t) = \nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon(t), t)dt + \sqrt{2}dW_t^i \\ \overline{X}_i^\varepsilon = \xi_i \\ u^\varepsilon = \mathcal{L}(\overline{X}_i^\varepsilon) \end{cases}. \tag{5.4}$$

where the law of $\overline{X}_i^\varepsilon$ satisfies the following PDE:

$$\begin{cases} \partial_t u^\varepsilon - \Delta u^\varepsilon + \nabla \cdot (\nabla V_\varepsilon \star u^\varepsilon u^\varepsilon) = 0 \\ u^\varepsilon|_{t=0} = u_0 \end{cases}. \tag{5.5}$$

Strategy is to proceed the following limit with $\varepsilon(N) \xrightarrow{N\to\infty}$ that

$$\text{SDE}_\varepsilon - \overline{\text{SDE}} \to 0$$
$$\overline{\text{SDE}} - \widehat{\text{SDE}} \to 0.$$

We will use two different scaling parameters in the connection of ε and N in the combined limit $\varepsilon(N) \xrightarrow{N\to\infty} 0$, which are going to be explained in separate sections in this chapter.

- $\varepsilon \sim (\frac{1}{\ln N})^a$. With this scaling the strategy is almost the same as in the bounded Lipschitz continuous case in chapter 3, we will estimate the difference only on the stochastic level. This scaling can be understood as a very strong cut-off such that the whole procedure don't feel the singularity of the potential.
- $\varepsilon \sim N^{-\beta}$, $\beta > 0$. This is called an algebraic scaling. For appropriate choice of β, the estimate we are going to show in this chapter shows that one has to work with the singularity of the potential.

Assumption F. Assume that Equation 5.2 (together with Equation 5.3) and Equation 5.5 (together with Equation 5.4) have solutions u and u_ε separately so that $\nabla V \star u(x,t)$ and $\nabla V_\varepsilon \star u^\varepsilon(x,t)$ are continuous and in x Lipschitz continuous. Furthermore, the following uniform in ε estimates holds: $\exists C \geq 0$

- $\|D^2 V_\varepsilon \star u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \leq C$
- $\|\nabla V_\varepsilon \star (u^\varepsilon - u)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \leq C\varepsilon$

- $\|(V_\varepsilon - V) \star \nabla u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\varepsilon$

Remark. The above assumptions depend on the uniform estimates to the solution u of Equation 5.2 and the structure of V. We take $V = \delta_0$ as an example to show that if $\|D^2 u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}$ is bounded and $\|u - u^\varepsilon\|_{L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d))} \le C\varepsilon$ then all the assumptions above are satisfied.

- $\|D^2 V_\varepsilon \star u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le \|V_\varepsilon\|_{L^1(\mathbb{R}^d)} \|D^2 u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C$
- $\|\nabla V_\varepsilon \star (u^\varepsilon - u)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le \|V_\varepsilon\|_{L^1(\mathbb{R}^d)} \|\nabla(u^\varepsilon - u)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\varepsilon$
- It holds for $x \in R^d$ and $t \in [0,T]$ that

$$\begin{aligned}|(V_\varepsilon - V) \star \nabla u(x,t)| &\le \frac{1}{\varepsilon^d}\int_{\mathbb{R}^d} \Big| j(\frac{x-y}{\varepsilon})\,((V\star\nabla u)(y) - (V\star\nabla u)(x))\Big| dy \\ &\le \|V \star D^2 u\|_\infty \varepsilon \cdot \frac{1}{\varepsilon^d}\int_{\mathbb{R}^d} j(\frac{x-y}{\varepsilon}) \cdot \frac{|x-y|}{\varepsilon} dy \le C\varepsilon.\end{aligned}$$

This assumption mainly provides that the corresponding Mckean-Vlasov equations Equation 5.4 and Equation 5.3 have strong $\mathbb{L}^2$ solutions and the control of the difference between them. Actually, under appropriate conditions on V (even for singular potentials) and initial data, one can show that this assumption fulfills by using PDE techniques, which are more advance than the approach discussed in the previous sections. We will not present these theories in this course.

For simplicity, we will always keep the following two examples of V in mind:

Example.

1. $V = \pm\delta_0$. In this case, the corresponding PDE is of local type: $\partial_t u - \Delta u \pm \nabla \cdot (\nabla u u) = 0$.
2. We take $V(x) = \frac{1}{|x|^\lambda}$ with $\lambda \in (0, d-2)$. Actually, if the potential can be written into $V = V_1 + V_2$ where $|D^2 V_1| \in L^1(\mathbb{R}^d)$ and $|D^2 V_2| \in L^\infty(\mathbb{R}^d)$, then the argument in the algebraic case will also work.

The following assumption is mainly needed for the result of convergence in probability.

Assumption G. Assume that $\exists C \ge 0$ such that $\||D^2 V_\varepsilon| \cdot u^\varepsilon\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C$, where C doesn't depend on ε.

For the potential $V = V_1 + V_2$, where $|D^2 V_1| \in L^1(\mathbb{R}^d)$ and $|D^2 V_2| \in L^\infty(\mathbb{R}^d)$, if we can obtain that $\|u^\varepsilon\|_{L^\infty(0,T;L^1(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d))}$ is bounded, then we can easily obtain that the assumption Assumption G holds.

5.1 Logarithmic Scaling, Convergence in Expectation

Theorem 5.1.1 (Logarithmic Scaling)**.** Let the Assumption F holds, for a given potential V in , there exists $\alpha > 0$ such that for $\varepsilon \sim (\ln N)^{-\alpha}$, it holds

$$\max_{1\le i\le N} \mathbb{E}\big[\sup_{0\le t\le T} |X_i^{N,\varepsilon}(t) - \hat{X}_i(t)|\big] \to 0, \qquad \text{for } N \to \infty.$$

Remark. The choice of α in the above theorem depends on the integrability of V, we will give two examples in the proof to show how to determine α.

Proof. First of all, the mollification kernel V_ε with any given ε implies that the interaction force is bounded Lipschitz continuous. Therefore, the strong $\mathbb{L}^2$ theory gives a unique solution of the system Equation 6.1, $X_i^{N,\varepsilon}$. Parallel, under the Assumption F, we know that $\overline{X}_i^\varepsilon$ and $\hat{X}_i$ also exist. In the next, we do directly the estimates among those three stochastic processes.

Step 1. Estimate for $X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)$.

Since they all share the same initial data, we have that by Hölder's and triangle inequalities,

$$\begin{aligned}
|X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)|^2 &= \left|\int_0^t \nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon}(s)) - \nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon(s)) ds\right|^2 \\
&\le t\int_0^t |\nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon}(s)) - \nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon(s))|^2 ds \\
&\le t\int_0^t |\nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon}(s)) - \nabla V_\varepsilon \star \mu_N(\overline{X}_i^\varepsilon(s))|^2 ds \\
&+ t\int_0^t |\nabla V_\varepsilon \star \mu_N(\overline{X}_i^\varepsilon(s)) - \nabla V_\varepsilon \star \overline{\mu}_N(\overline{X}_i^\varepsilon(s))|^2 ds \\
&+ t\int_0^t |\nabla V_\varepsilon \star \overline{\mu}_N(\overline{X}_i^\varepsilon(s)) - \nabla V_\varepsilon \star u_\varepsilon(\overline{X}_i^\varepsilon(s))|^2 ds \\
&\le t\cdot \|\nabla^2 V_\varepsilon \star \mu_N\|_\infty^2 \int_0^t |X_i^{N,\varepsilon}(s) - \overline{X}_i^\varepsilon(s)|^2 ds \\
&+ t\int_0^t \|\nabla^2 V_\varepsilon\|_\infty^2 \Big(\frac{1}{N}\sum_{j=1}^N |X_j^{N,\varepsilon}(s) - \overline{X}_j^\varepsilon(s)|\Big)^2 ds + III,
\end{aligned}$$

where we used the notation $\overline{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\overline{X}_i^\varepsilon}$ to be the empirical measure of $\overline{X}_i^\varepsilon$'s. The following estimate has also been used in the above estimates

$$\begin{aligned}
&|\nabla V_\varepsilon \star \mu_N(\overline{X}_i^\varepsilon(s)) - \nabla V_\varepsilon \star \overline{\mu}_N(\overline{X}_i^\varepsilon(s))| \\
=&\left|\frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - X_j^{N,\varepsilon}(s)) - \frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s))\right| \\
\le&\|\nabla^2 V_\varepsilon\|_\infty \frac{1}{N}\sum_{j=1}^N |X_j^{N,\varepsilon}(s) - \overline{X}_j^\varepsilon(s)|.
\end{aligned}$$

The third term above needed to be done by using a mollified version of the law of large number theorem, we leave it as an exercise.

> **Exercise.** Show that the III term above can be bounded as follows
>
> $$III \le \frac{2t^2}{N}\|\nabla V_\varepsilon\|_\infty^2.$$
>
> *Hint:* Law of Large numbers

Now we take the superimum in $\tilde{t} \in [0, T]$, the expectation, the maximum in $i = 1, \dots, N$, and obtain

$$\begin{aligned}
&\max_{1\le i\le N} \mathbb{E}\Big[\sup_{0\le t\le \tilde{t}} |X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)|^2\Big] \\
\le&\|D^2 V_\varepsilon\|_\infty^2 \tilde{t}\int_0^{\tilde{t}} \max_{1\le j\le N} \mathbb{E}\Big[\sup_{0\le t\le s} |X_j^{N,\varepsilon}(t) - \overline{X}_j^\varepsilon(t)|^2\Big] ds + \frac{2\tilde{t}^2}{N}\|\nabla V_\varepsilon\|_\infty^2.
\end{aligned}$$

By Grönwall we can obtain

$$\max_{1\le i\le N} \mathbb{E}\Big[\sup_{0\le t\le T} |X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)|^2\Big] \le e^{\|D^2 V_\varepsilon\|_\infty^2 T}\cdot \frac{2T^2}{N}\|\nabla V_\varepsilon\|_\infty^2. \tag{5.6}$$

In the next, for two examples, we will choose $\varepsilon(N)$ such that the above quantity converges to zero. Notice that

$$V_\varepsilon = \int_{\mathbb{R}^d} \frac{1}{\varepsilon^d} j(\frac{x-y}{\varepsilon}) V(y) dy.$$

- For $V = \delta_0$, we have that

  $$\|D^2 V_\varepsilon\|_\infty \leq \frac{C}{\varepsilon^{d+2}}, \qquad \|\nabla V_\varepsilon\|_\infty \leq \frac{C}{\varepsilon^{(d+1)}}.$$

  Then for $\forall \eta \in (0,1)$ we can choose $\varepsilon > 0$ such that

  $$e^{\frac{C}{\varepsilon^{2(d+2)}}} = N^\eta \quad \text{which means that } \varepsilon = (\frac{C}{\eta \ln N})^{\frac{1}{2(d+2)}}.$$

  then we plug in this choice in Equation 5.6 and obtain that

  $$\max_{1\leq i\leq N} \mathbb{E}\Big[\sup_{0\leq t\leq T} |X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)|^2\Big] \leq e^{\frac{C}{\varepsilon^{(d+1)2}}} \frac{\frac{C}{\varepsilon^{(d+1)2}}}{N} \leq \frac{C}{N^{1-\eta}}.$$

- $V = \frac{1}{|x|^\lambda}$ for $0 < \lambda < d-2$. We do the following estimates for its mollification. For any $x \in \mathbb{R}^d$, we have

  $$\begin{aligned}
  |D^2 V_\varepsilon(x)| &\leq \int_{\mathbb{R}^d} \frac{1}{\varepsilon^d} |j(\frac{x-y}{\varepsilon}) D^2 V(y)| dy \\
  &= \int_{|x-y|\leq\varepsilon} \frac{1}{\varepsilon^d} |j(\frac{x-y}{\varepsilon}) D^2 V(y)| dy \\
  &\leq \int_{|y|\leq|x|+\varepsilon} \frac{1}{\varepsilon^d} |j(\frac{x-y}{\varepsilon}) D^2 V(y)| dy \\
  &\leq \int_{|y|\leq R} \frac{1}{\varepsilon^d} |j(\frac{x-y}{\varepsilon}) \frac{1}{|y|^{\lambda+2}}| dy + \int_{R\leq|y|\leq|x|+\varepsilon} \frac{1}{\varepsilon^d} |j(\frac{x-y}{\varepsilon}) \frac{1}{|y|^{\lambda+2}}| dy \\
  &\leq \frac{C}{\varepsilon^d} R^{d-\lambda-2} + \frac{C}{R^{\lambda+2}} \leq \frac{C}{\varepsilon^{\lambda+2}}
  \end{aligned}$$

  where we have optimized the value of R in the last step. Similarly, one obtains that

  $$\|DV_\varepsilon\|_\infty \leq C\frac{C}{\varepsilon^{\lambda+1}}.$$

  Then we can do the same discussion as in the first example, and obtain that for reasonably chosen $\varepsilon \sim (\ln N)^{-\alpha}$ it holds

  $$\max_{1\leq i\leq N} \mathbb{E}\Big[\sup_{0\leq t\leq T} |X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)|^2\Big] \leq \frac{C}{N^{1-\eta}} \to 0.$$

This concludes the first step

$$\varepsilon(N) \xrightarrow{N\to\infty} 0 \quad \text{SDE}_\varepsilon - \overline{\text{SDE}} \to 0.$$

Step 2. Estimate for $\overline{X}_i^\varepsilon(t) - \hat{X}_i(t)$. Here we use the Assumption F an obtain directly

$\forall t \in [0, T]$,

$$
\begin{aligned}
|\overline{X}_i^\varepsilon(t) - \hat{X}_i(t)|^2 &= \left| \int_0^t \nabla V_\varepsilon \star u^\varepsilon(\overline{X}^\varepsilon(s)) - \nabla V \star u(\hat{X}(s)) ds \right|^2 \\
&\leq t \int_0^t |\nabla V_\varepsilon \star u^\varepsilon(\overline{X}^\varepsilon(s)) - \nabla V_\varepsilon \star u(\overline{X}^\varepsilon(s))|^2 \\
&+ |\nabla V_\varepsilon \star u(\overline{X}^\varepsilon(s)) - \nabla V_\varepsilon \star u(\hat{X}(s))|^2 + |\nabla V_\varepsilon \star u(\hat{X}(s)) - \nabla V \star u(\hat{X}(s))|^2 ds \\
&\leq C \|D^2 V_\varepsilon \star u\|^2 \int_0^t |\overline{X}_i^\varepsilon(s) - \hat{X}_i(s)|^2 ds + \|\nabla V_\varepsilon \star (u^\varepsilon - u)\|_\infty^2 + \|(V_\varepsilon - V) \star \nabla u\|_\infty^2 \\
&\leq C \int_0^t |\overline{X}_i^\varepsilon(s) - \hat{X}_i(s)|^2 ds + C\varepsilon^2,
\end{aligned}
$$

where in the last step we have used Assumption F. Then By Gronwall's inequality, we have

$$
\max_{1 \leq i \leq N} \mathbb{E}\Big[\sup_{0 \leq t \leq T} |\overline{X}_i^\varepsilon(t) - \hat{X}_i(t)|^2 \Big] \leq C(T)\varepsilon^2. \tag{5.7}
$$

Finally, the results hold with combing the two steps together. □

Exercise. Do the log scaling by repeating the proof in the framework of $W_{2,\mathcal{C}^d}$

5.2 The Algebraic Scaling, Convergence in Probability

As has been shown, if the scaling parameter in the mollification kernel is of logarithmic, expect requiring more assumptions from the PDE solution, the whole estimate in the expectation are basically the same as in the bounded Lipschitz case. One doesn't observe the structure of the singular kernel. In this section, we consider the same problem setting but instead we use the so called algebraic scaling $\varepsilon = N^{-\beta}$ for $\beta \in (0, \frac{1}{2})$ to be determined.

As a preparation, we introduce first a version of law of large numbers.

Lemma 5.2.1 (A version of l.l.n). Suppose $(\overline{Y}_i)_{i \leq N}$ is a collection of i.i.d random variables with law $v \in L^1(\mathbb{R}^d)$, and $U \in L^\infty$ be a given function, define the following subset of Ω

$$
A_\theta^i(U, v) = \left\{ \omega \in \Omega : \left| \frac{1}{N} \sum_{1 \leq j \leq N} U(\overline{Y}_i - \overline{Y}_j) - U \star v(\overline{Y}_i) \right| \geq \frac{1}{N^\theta} \right\} \text{ and } A_\theta^N = \bigcup_{i=1}^N A_\theta^i(U, v).
$$

then $\forall \tilde{k} \in \mathbb{N}$, $\theta \in (0, \frac{1}{2})$ it holds

$$
\mathbb{P}(A_\theta^N(U, v)) \leq N \max_{1 \leq i \leq N} \mathbb{P}(A_\theta^i(U, v)) \leq N \cdot N^{2\tilde{k}(\theta - \frac{1}{2})} C(\tilde{k}) \|U\|_{L^\infty}^{2\tilde{k}}.
$$

Proof. Markovs inequality implies that

$$
\begin{aligned}
\mathbb{P}(A_\theta^i(U, v)) &\leq N^{2\tilde{k}\theta} \mathbb{E}\Big[\Big| \frac{1}{N} \sum_{j=1}^N U(\overline{Y}_i - \overline{Y}_j) - U \star v(\overline{Y}_i) \Big|^{2\tilde{k}} \Big] \\
&= N^{2\tilde{k}\theta} \mathbb{E}\Big[\Big(\frac{1}{N^2} \sum_{j,k=1}^N h(\overline{Y}_i, \overline{Y}_j) h(\overline{Y}_i, \overline{Y}_k) \Big)^{\tilde{k}} \Big],
\end{aligned}
$$

where $h(\overline{Y}_i, \overline{Y}_j) = U(\overline{Y}_i - \overline{Y}_j) - U \star v(\overline{Y}_i)$. In the terms where j appeared only once, it holds that

$$\mathbb{E}\Big[h(\overline{Y}_i,\overline{Y}_j)\prod_{\substack{m=1\\m\neq j}}^{2\tilde{k}-1}h(\overline{Y}_i,\overline{Y}_{l_m})\Big]=\int dxv(x)\int dyv(y)h(x,y)\mathbb{E}\Big[\prod_{\substack{m=1\\m\neq j}}^{2\tilde{k}-1}h(x,\overline{Y}_{l_m})\Big]=0,$$

where we have used the fact that for fixed $x\in\mathbb{R}^d$

$$\int dyv(y)h(x,y)=\int dyv(y)(U(x,y)-U\star v(x))=0.$$

Denote the set $\mathcal{P}$ to be the term that might not vanishes.

$$\mathcal{N}:=\Big\{\prod_{j=1}^{2\tilde{k}}h(\overline{Y}_i,\overline{Y}_{i_j}):i_j\in\{1,\dots,N\}\text{ s.t. } i_j \text{ appeared at least twice}\Big\}.$$

It is easy to obtain that the number of the elements in set $\mathcal{N}$ can be bounded by $C(\tilde{k})N^{\tilde{k}}$. On the other hand, the each of the nonvanishing can be bounded in the following

$$\mathbb{E}\Big[\prod_{j=1}^{2\tilde{k}}h(\overline{Y}_i,\overline{Y}_{i_j})\Big]\le C(\tilde{k})\cdot\|U\|_\infty^{2\tilde{k}}.$$

Therefore, we obtain that

$$\mathbb{P}(A^i_\theta(U,v))\le N^{2\tilde{k}\theta}\frac{C(\tilde{k})}{N^{2\tilde{k}}}\cdot N^{\tilde{k}}\|U\|_\infty^{2\tilde{k}}=C(\tilde{k})N^{2\tilde{k}(\theta-\frac{1}{2})}\|U\|_\infty^{2\tilde{k}}.$$

This complete the proof of this lemma. □

From microscopic to the intermediate level, due to the singularity, it is not realistic to study the convergence in expectation. However, one can expect that the probability that the particle trajectory is away from the mean field trajectory is small. More precisely, we will prove the following result

Theorem 5.2.1. Let the Assumption F and Assumption G hold, for a given V in , $\varepsilon=N^{-\beta}$, then $\forall\alpha\in(0,\frac{1}{2})$, there exists β_1 which depdends on α, if $\beta\in(0,\beta_1)$, it holds $\forall\gamma>0$ that

$$\sup_{0\le t\le T}\mathbb{P}\Big(\max_{1\le i\le N}|X_i^{N,\varepsilon}(t)-\overline{X}_i^\varepsilon(t)|\ge N^{-\alpha}\Big)\le\frac{C(\gamma)}{N^\gamma}.$$

where $C(\gamma)$ is a constant depends on γ.

Proof. First remember that $X_i^{N,\varepsilon}$ and $\overline{X}_i^N$ are solutions to Equation 6.1 and Equation 5.4 separately, so they are path continuous. We define a stopping time

$$\tau(\omega)=\inf\{t\in(0,T]:\max_{1\le i\le N}|X_i^{N,\varepsilon}(t)-\overline{X}_i^\varepsilon(t)|\ge N^{-\alpha}\}.$$

Then using the stopping time we define , the cut-off process

$$S(\omega,t)=N^{2\alpha k}\max_{1\le i\le N}|(X_i^{N,\varepsilon}-\overline{X}_i^\varepsilon)(t\wedge\tau(\omega))|^{2k}\le 1,$$

where $k\in\mathbb{N}$ to be determined. We have by Markovs inequality

$$\sup_{0<t\le T}\mathbb{P}\Big(\max_{1\le i\le T}|X_i^{N,\varepsilon}(t)-\overline{X}_i^\varepsilon(t)|\ge N^{-\alpha}\Big)\le\sup_{0<t\le T}\mathbb{P}(S_t\equiv 1)\overset{\text{Mrkv.}}{\le}\sup_{0<t\le T}\mathbb{E}[S_t].$$

Notice that X_i^N and $\overline{X}_i^N$ satisfies $\forall t \wedge \tau$ that

$$X_i^{N,\varepsilon}(t\wedge\tau) - X_i^{N,\varepsilon}(0) = \int_0^{t\wedge\tau} \nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon}(s))ds + \int_0^{t\wedge\tau} \sqrt{2}dW_s^i$$
$$\overline{X}_i^\varepsilon(t\wedge\tau) - \overline{X}_i^\varepsilon(0) = \int_0^{t\wedge\tau} \nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon(s))ds + \int_0^{t\wedge\tau} \sqrt{2}dW_s^i.$$

Then by taking the difference, using the Itô's formula and taking the expectation, we have

$$\begin{aligned}
&\mathbb{E}[S(t)] \\
=&N^{2\alpha k}\mathbb{E}\Big[\max_{1\le i\le N}|(X_i^{N,\varepsilon} - \overline{X}_i^\varepsilon)(t\wedge\tau(\omega))|^{2k}\Big] \\
\le&N^{2\alpha}C(t,k)\mathbb{E}\left[\max_{1\le i\le N}\int_0^{t\wedge\tau}\left|\nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon}(s)) - \nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon(s))ds\right|S(s)^{\frac{2k-1}{2k}}ds\right] \\
=&N^{2\alpha}C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big|\frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(X_i^{N,\varepsilon}(s) - X_j^{N,\varepsilon}(s)) - \frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s))\Big|S(s)^{\frac{2k-1}{2k}}ds\Big] \\
&+ N^{2\alpha}C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big|\frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s)) - \nabla V_\varepsilon \star u_\varepsilon(\overline{X}_i^\varepsilon(s),s)\Big|S(s)^{\frac{2k-1}{2k}}ds\Big] \\
=&I + II.
\end{aligned}$$

We bound the above term individually,

$$\begin{aligned}
I \le\ & N^{2\alpha}C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big|\frac{1}{N}\sum_{j=1}^N D^2V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s))\cdot(X_i^{N,\varepsilon} - \overline{X}_i^\varepsilon - (X_j^{N,\varepsilon} - \overline{X}_j^\varepsilon))\Big|S(s)^{\frac{2k-1}{2k}}ds\Big] \\
&+ N^{2\alpha}C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\|D^3V_\varepsilon\|_\infty\frac{1}{N}\sum_j|X_i^{N,\varepsilon} - \overline{X}_i^\varepsilon - (X_j^{N,\varepsilon} - \overline{X}_j^\varepsilon)|^2S(s)^{\frac{2k-1}{2k}}ds\Big] \\
=\ & I_1 + I_2.
\end{aligned}$$

where

$$I_2 \le \frac{C(t,k)}{N^{2\alpha}}\|D^3V_\varepsilon\|_\infty\int_0^t \mathbb{E}[S(s)]ds.$$

For I_1 we get

$$\begin{aligned}
I_1 \le\ & N^{2\alpha}C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\frac{1}{N}\sum_{j=1}^N\Big|D^2V_\varepsilon(\overline{X}_i^\varepsilon - \overline{X}_j^\varepsilon)\Big|\cdot 2\max_{1\le i\le N}|X_i^{N,\varepsilon}(s) - \overline{X}_i^\varepsilon(s)|S(s)^{\frac{2k-1}{2k}}ds\Big] \\
\le\ & C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}(D^2V_\varepsilon \star u^\varepsilon)(\overline{X}_i^\varepsilon(s),s)S(s)ds\Big] \\
&+ C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big(\frac{1}{N}\sum_{j=1}^N|D^2V_\varepsilon|(\overline{X}_i^\varepsilon - \overline{X}_j^\varepsilon) - |D^2V_\varepsilon| \star u^\varepsilon(\overline{X}_i^\varepsilon(s),s)\Big)S(s)ds\Big] \\
\le\ & C(t,k)\||D^2V_\varepsilon| \star u^\varepsilon\|_\infty\int_0^t \mathbb{E}[(S(s))]ds + I_{12}.
\end{aligned}$$

By adding in the one term we can use the regularity of the SDE solution u^ε to bound

Using Lemma 5.2.1 with $U = |D^2V_\varepsilon|$ and $v = u^\varepsilon(\cdot,s)$, define

$$A_\theta(|D^2V_\varepsilon|,u^\varepsilon) = \bigcup_{i=1}^N A_\theta^i \Rightarrow A_\theta^c(|D^2V_\varepsilon|,u^\varepsilon) = \bigcap_{i=1}^N (A_\theta^i)^c.$$

$$I_{12} = C(t,k)\mathbb{E}\Big[(\mathbb{1}_{A_\theta} + \mathbb{1}_{A_\theta^c}) \max_{1\le i\le N} \int_0^{t\wedge\tau} \Big(\frac{1}{N}\sum_{j=1}^N |D^2V_\varepsilon|(\overline{X}_i^\varepsilon - \overline{X}_j^\varepsilon) - |D^2V_\varepsilon| \star u^\varepsilon(\overline{X}_i^\varepsilon(s),s)\Big)S(s)ds\Big]$$

$$\le C(t,k)\frac{1}{N^\theta}\int_0^t \mathbb{E}[S(s)]ds + C(t,k)\|D^2V_\varepsilon\|_\infty \mathbb{P}(A_\theta^c)$$

$$\le C(t,k)\int_0^t \mathbb{E}[S(s)]ds + C(t,k)\|D^2V_\varepsilon\|_\infty \cdot \|D^2V_\varepsilon\|_\infty^{2\tilde{k}} C(\tilde{k})N^{2\tilde{k}(\theta-\frac{1}{2})+1}.$$

Actually, we can take $\theta = 0$ above.

For II we again apply Lemma 5.2.1

$$A_{\theta_1}(\nabla V_\varepsilon, u_\varepsilon) = \bigcup_{i=1}^N A_{\theta_1}^i.$$

By inserting the indicator functions of this set and its compliment, and notice that $S(s) \le 1$, we have

$$II = N^{2\alpha}C(t,k)\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big|\frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s)) - \nabla V_\varepsilon \star u_\varepsilon(\overline{X}_i^\varepsilon(s),s)\Big| S(s)^{\frac{2k-1}{2k}}ds\Big]$$

$$\le C(t,k)\int_0^t \mathbb{E}[S(s)]ds + C(t,k)N^{2\alpha k}\mathbb{E}\Big[\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big|\frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s)) - \nabla V_\varepsilon \star u_\varepsilon(\overline{X}_i^\varepsilon(s),s)\Big|^{2k}ds\Big]$$

$$\le C(t,k)\int_0^t \mathbb{E}[S(s)]ds$$

$$+ C(t,k)N^{2\alpha k}\mathbb{E}\Big[(\mathbb{1}_{A_{\theta_1}^c} + \mathbb{1}_{A_{\theta_1}})\max_{1\le i\le N}\int_0^{t\wedge\tau}\Big|\frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon(s) - \overline{X}_j^\varepsilon(s)) - \nabla V_\varepsilon \star u_\varepsilon(\overline{X}_i^\varepsilon(s),s)\Big|^{2k}ds\Big]$$

$$\le C(t,k)\int_0^t \mathbb{E}[S(s)]ds + C(t,k)N^{2\alpha k}\Big[\frac{1}{N^{2\theta_1 k}} + \|\nabla V_\varepsilon\|_\infty^{2k}\mathbb{P}(A_{\theta_1})\Big]$$

$$\le C(t,k)\int_0^t \mathbb{E}[S(s)]ds + C(t,k)(N^{(\alpha-\theta_1)2k} + N^{2\alpha k}\|\nabla V_\varepsilon\|_\infty^{2k}\|\nabla V_\varepsilon\|^{2\tilde{k}}C(\tilde{k})N^{2\tilde{k}(\theta_1-\frac{1}{2})+1}).$$

We can put everything together now

$$\begin{aligned}\mathbb{E}[S(t)] &\le C(1 + N^{-2\alpha}\|D^3V_\varepsilon\|_\infty + \||D^2V_\varepsilon| \star u^\varepsilon\|_\infty)\int_0^t \mathbb{E}[S(s)]ds \\ &+ C(t,k)N^{(\alpha-\theta_1)2k} + C(t,k,\tilde{k})N^{2\alpha k + 2\tilde{k}(\theta_1-\frac{1}{2})+1}\|\nabla V_\varepsilon\|_\infty^{2(k+\tilde{k})} \\ &+ C(t,k,\tilde{k})N^{2\tilde{k}(\theta-\frac{1}{2})+1}\|D^2V_\varepsilon\|_\infty^{2\tilde{k}+1}.\end{aligned}$$

Remember that $\varepsilon = N^{-\beta}$ and $\||D^2V_\varepsilon| \star u^\varepsilon\|_\infty^{2k} \le C$ by Assumption G, we have that

$$\mathbb{E}[S(t)] \le C(1 + N^{-2\alpha}\|D^3V_\varepsilon\|_\infty + \||D^2V_\varepsilon| \star u^\varepsilon\|_\infty)\int_0^t \mathbb{E}[S(s)]ds \tag{5.8}$$

$$+ C(t,k)N^{(\alpha-\theta_1)2k} + C(t,k,\tilde{k})N^{2\alpha k + 2\tilde{k}(\theta_1-\frac{1}{2})+1}\|\nabla V_\varepsilon\|_\infty^{2(k+\tilde{k})} \tag{5.9}$$

$$+ C(t,k,\tilde{k})N^{2\tilde{k}(\theta-\frac{1}{2})+1}\|D^2V_\varepsilon\|_\infty^{2\tilde{k}+1}. \tag{5.10}$$

- For $V = \delta_0$, we have that

$$\|D^3V_\varepsilon\|_\infty \le \frac{C}{\varepsilon^{d+3}} \quad \|D^2V_\varepsilon\|_\infty \le \frac{C}{\varepsilon^{d+2}}, \qquad \|\nabla V_\varepsilon\|_\infty \le \frac{C}{\varepsilon^{(d+1)}}.$$

- $V = \frac{1}{|x|^\lambda}$ for $0 < \lambda < d-2$.

$$\|D^3V_\varepsilon\|_\infty \le \frac{C}{\varepsilon^{\lambda+3}} \quad \|D^2V_\varepsilon\|_\infty \le \frac{C}{\varepsilon^{\lambda+2}}, \qquad \|\nabla V_\varepsilon\|_\infty \le \frac{C}{\varepsilon^{(\lambda+1)}}.$$

Then we can put these estimates in Equation 5.8, since $\theta_1 > \alpha$, for any given $\gamma > 0$, we first adjust the value of $\tilde{k}$ and then the value of k to complete the proof by Gronwall's inequality. □

Now we can conbine the estimate between $\overline{X}_i^\varepsilon$ and $\hat{X}_i^N$, which is exactly the same as in Equation 5.7 and obtain that

Corollary. For $\beta \in (0, \beta_1)$ and $\forall \eta \in (0,1)$, it holds

$$\sup_{0\le t\le T} \mathbb{E}[\max_{1\le i\le N} |X_i^{N,\varepsilon} - \hat{X}_i^N| > N^{-\eta\beta}] \le C \cdot N^{-(1-\eta)\beta}.$$

Proof. Notice that the proof of Equation 5.7 also implies that

$$\sup_{0\le t\le T} \mathbb{E}[\max_{1\le i\le N} |\overline{X}_i^\varepsilon - \hat{X}_i^N|] \le C \cdot \varepsilon.$$

Then we obtain $\forall \eta \in (0,1)$

$$\sup_{0\le t\le T} \mathbb{P}(\max_{1\le i\le N} |\overline{X}_i^\varepsilon - \hat{X}_i^N| > \varepsilon^\eta) \le \varepsilon^{-\eta} \sup_{0\le t\le T} \mathbb{E}[\max_{1\le i\le N} |\overline{X}_i^\varepsilon - \hat{X}_i^N|] \le C\varepsilon^{1-\eta}.$$

Therefore,

$$\begin{aligned}
&\sup_{0\le t\le T} \mathbb{P}(\max_{1\le i\le N} |X_i^{N,\varepsilon} - \hat{X}_i^N| > \varepsilon^\eta) \\
&\le \sup_{0\le t\le T} \mathbb{P}(\max_{1\le i\le N} |X_i^{N,\varepsilon} - \overline{X}_i^\varepsilon| > \varepsilon^\eta) + \sup_{0\le t\le T} \mathbb{P}(\max_{1\le i\le N} |\overline{X}_i^N - \hat{X}_i^N| > \varepsilon^\eta) \\
&\le \frac{C}{N^\gamma} + C \cdot \varepsilon^{1-\eta}.
\end{aligned}$$

Then the results is obtained by taking $\varepsilon = N^{-\beta}$. □

Chapter 6

Relative Entropy Method

In this chapter, we introduce the newly developed method from PDE point of view in the study of mean field limit. Relative entropy has been long investigated in the PDE theory to study the long time behaviour or singular limits of solutions, it has been introduced in the multi particle setting to handle non-Lipschitz interaction forces only recently. To describe the main idea, we review the N-particle system

$$\text{(SDE)} \begin{cases} dX_i^N(t) & = \frac{1}{N}\sum_{j=1}^N \nabla V(X_i^N(t) - X_j^N(t))dt + \sqrt{2}dW_t^i \\ X_i^N(0) & \text{are given i.i.d random variables with law } u_0(x) \in L^1((1+|x|^2)dx) \end{cases} \tag{6.1}$$

where the whole stochastic setting are the same as in Chapter 3. We know that if ∇V is bounded Lipschitz continuous, then the $\mathbb{L}^2$ solution exists. Instead of direct studying the stochastic equation Equation 6.1, we study the PDE which the joint law of $(X_1^N(t), \ldots, X_N^N(t))$ satisfies.

Now let $u_N(\cdot, t)$ be the corresponding joint law, then by Itô s formula and taking the expectation we know that $u_N(t, x_1, \ldots, x_N)$ satisfies the following higher dimensional PDE in the weak sense,

$$\partial_t u_N = \sum_{i=1}^N \Delta_{x_i} u_N + \sum_{i=1}^N \nabla_{x_i} \cdot (u_N \cdot \frac{1}{N}\sum_{j=1}^N \nabla V(x_i - x_j)) \tag{6.2}$$

$$u_N|_{t=0} = u_0^{\otimes N} = u_0(x_1)u_0(x_2)\ldots u_0(x_N). \tag{6.3}$$

Notice that if $\|\nabla V\|_\infty \le C$, we have proved this linear transport equation with bounded drift has a unique solution $u_N \in L^\infty([0,T]; L^1(\mathbb{R}^{dN}))$ in chapter 4.

Remark. By classical theory of linear PDE, which will be demonstrated in the PDE lecture, ∇V has enough regularity, then u_N is a classical solution.

Goal. We want to study the limit $N \to \infty$. The expect limiting PDE, as in the particle setting, should be the corresponding PDE of the Mckean-Vlasov equation, namely

$$\partial_t u - \Delta u = \nabla \cdot (u \nabla V \star u).$$

The question is how to express the limit of u_N to be u. The idea in the relative entropy method is to "upgrade" u into the higher dimension version, i.e. $u^{\otimes N}$ and study the limit

$$u_N - u^{\otimes N} \xrightarrow{N\to\infty} 0.$$

One can obtain directly that $u^{\otimes N} = u(t, x_1)u(t, x_2)\ldots u(t, x_N)$ satisfies

$$\partial_t u^{\otimes N} = \sum_{i=1}^N \Delta_{x_i} u^{\otimes N} + \sum_{i=1}^N \nabla_{x_i} \cdot (u^{\otimes N} \nabla V \star u(t, x_i)). \tag{6.4}$$

with the same initial data as in Equation 6.3. The remaining task is to compare these two higher dimensional equations Equation 6.2 and Equation 6.4.

In the analysis of diffusion type of equations, the key technique is to get *a priori* estimates of the solution. There are in general many possible "free energy"s for a given equation. As one can see from the heat equation, $\partial_t u - \Delta u = 0$, the following standard $L^p, p > 1$ estimates holds:

$$L^2: \qquad \frac{d}{dt}\int_{\mathbb{R}^d} u^2 dx + 2\int_{\mathbb{R}^d} |\nabla u|^2 dx = 0.$$

$$L^p: \qquad \frac{d}{dt}\int_{\mathbb{R}^d} |u|^p dx + \frac{4(p-1)}{p}\int_{\mathbb{R}^d} |\nabla u^{\frac{p}{2}}|^2 dx = 0..$$

These can be achieved by using using $2u$ and pu^{p-1} with $p > 1$ as test function. Actually one can obtain the so called entropy estimate by choosing the test function $\varphi = \log u$, namely

$$\int_{\mathbb{R}^d} \partial_t u \log u dx - \int_{\mathbb{R}^d} \Delta u \log u dx = 0.$$

By using the product rule we have $\partial_t u \log u = \partial_t(u \log u) - \partial_t u$. Since $\frac{d}{dt}\int_{\mathbb{R}^d} u dx = 0$ due to the conservation of mass, we know that

$$\frac{d}{dt}\int_{\mathbb{R}^d} u \log u dx + \int_{\mathbb{R}^d} \frac{|\Delta u|^2}{u} dx = 0.$$

The term $\int_{\mathbb{R}^d} \frac{|\Delta u|^2}{u} dx$ is called the entropy production, which can also be written as $\int_{\mathbb{R}^d} |\nabla\sqrt{u}|^2 dx = 0$ or $\int_{\mathbb{R}^d} |\nabla \log u|^2 u dx$.

In the next we concentrate to study so called relative entropy for two density functions: it is generally defined

Definition 6.0.1 (Relative Entropy)**.** The relative entropy between two integrable functions u_1 and u_2 are given by

$$\int_{\mathbb{R}^d} \frac{u_1}{u_2}\log\frac{u_1}{u_2} \cdot u_2 dx = \int_{\mathbb{R}^d} u_1 \log\frac{u_1}{u_2} dx$$

Remark. It is not hard to compute that if u_1 and u_2 are both solutions of the heat equation, then

$$\begin{aligned}
\partial_t \int_{\mathbb{R}^d} u_1 \log\frac{u_1}{u_2} dx &= \int_{\mathbb{R}^d} \partial_t u_1 \log\frac{u_1}{u_2} dx + \int_{\mathbb{R}^d} u_1 \frac{u_2}{u_1}(\partial_t \frac{u_1}{u_2}) dx \\
&= \int_{\mathbb{R}^d} (\partial_t u_1 \log\frac{u_1}{u_2} + \partial_t u_1 - \frac{u_1}{u_2}\partial_t u_2) dx \\
&= -\int_{\mathbb{R}^d} \nabla u_1 \nabla \log\frac{u_1}{u_2} dx + \int_{\mathbb{R}^d} \nabla\frac{u_1}{u_2}\nabla u_2 dx \\
&= -\int_{\mathbb{R}^d} \nabla\log\frac{u_1}{u_2}(\nabla u_1 - \frac{u_1}{u_2}\nabla u_2) dx \\
&= -\int_{\mathbb{R}^d} |\nabla\log\frac{u_1}{u_2}|^2 u_1 dx.
\end{aligned}$$

6.1 Relative entropy inequality for high dimensional PDE

Now we come back to the two higher order PDEs Equation 6.2 and Equation 6.4, and obtain the following relative entropy identity.

Lemma 6.1.1. Let

$$\mathcal{H}(u_N|u^{\otimes N}) = \frac{1}{N}\mathcal{H}_N(u_N|u^{\otimes N}) = \frac{1}{N}\int_{\mathbb{R}^{dN}} u_N \log\frac{u_N}{u^{\otimes N}} dx_1 \dots dx_N \qquad (6.5)$$

If u_N and $u^{\otimes N}$ are solutions of Equation 6.2 and Equation 6.4 separately, then the following identity and inequality hold

$$\frac{d}{dt}\mathcal{H}(u_N|u^{\otimes N}) = -\frac{1}{N}\int_{\mathbb{R}^{dN}}\sum_{i=1}^{N}\Big|\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}\Big|^2 u_N dx_1\dots dx_N \tag{6.6}$$

$$-\frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^{dN}}\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}\left[u_N\left(\nabla V\star u(x_i)-\frac{1}{N}\sum_{j=1}^{N}\nabla V(x_i-x_j)\right)\right]dx_1\dots dx_N$$

$$\leq -\frac{1}{2N}\int_{\mathbb{R}^{dN}}\sum_{i=1}^{N}|\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}|^2 u_N dx_1\dots dx_N \tag{6.7}$$

$$+\frac{1}{2}\int_{\mathbb{R}^{dN}}u_N\frac{1}{N}\sum_{i=1}^{N}\Big|\nabla V\star u(x_i)-\frac{1}{N}\nabla V(x_i-x_j)\Big|^2 dx_1\dots dx_N$$

Proof. We directly compute the time derivative, similar to the case of Heat equation, and obtain

$$\frac{d}{dt}\mathcal{H}(u_N|u^{\otimes N}) = \frac{1}{N}\int_{\mathbb{R}^{dN}}\left(\partial_t u_N\log\frac{u_N}{u^{\otimes N}}+\partial_t u_N-u_N\frac{u^{\otimes N}}{(u^{\otimes N})^2}\partial_t u^{\otimes N}\right)dx_1\dots dx_N.$$

Now we input the equation for $\partial_t u_N$ and $\partial_t u^{\otimes N}$ and do integral by parts,

$$\begin{aligned}
&\frac{d}{dt}\mathcal{H}(u_N|u^{\otimes N})\\
=&\frac{1}{N}\int_{\mathbb{R}^{dN}}-\sum_{i=1}^{N}(\nabla_{x_i}u_N+u_N\frac{1}{N}\sum_{j=1}^{N}\nabla V(x_i-x_j))\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}dx_1\dots dx_N\\
&-\frac{1}{N}\int_{\mathbb{R}^{dN}}-\sum_{i=1}^{N}(\nabla_{x_i}u^{\otimes N}+u^{\otimes N}\nabla V\star u(x_i))\underbrace{\nabla_{x_i}(\frac{u_N}{u^{\otimes N}})}_{=\frac{u_N}{u^{\otimes N}}\cdot\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}}dx_1\dots dx_N\\
=&\frac{1}{N}\int_{\mathbb{R}^{dN}}\sum_{i=1}^{N}(-\nabla_{x_i}u_N+\nabla_{x_i}u^{\otimes N}\cdot\frac{u_N}{u^{\otimes N}})\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}dx_1\dots dx_N\\
&+\frac{1}{N}\int_{\mathbb{R}^{dN}}\left(\sum_{i=1}^{N}\nabla V\star u(x_i)-\frac{1}{N}\sum_{j=1}^{N}\nabla V(x_i-x_j)\right)u_N\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}dx_1\dots dx_N.
\end{aligned}$$

Then the equation Equation 6.6 follows directly by rewriting the entropy production term, and inequality Equation 6.7 follows from the Hölder's inequality. □

Notice that the second term on the right hand side of Equation 6.7 needs directly the mean field limit estimates. We will apply the estimates obtained in the previous chapters for bounded Lipschitz forces and forces with singular potential and obtain separately the convergence rate estimate for the relative entropy. More precisely, we can rewrite the second term by using the stochastic setting:

$$\begin{aligned}
&\frac{d}{dt}\mathcal{H}(u_N|u^{\otimes N}) \qquad\qquad (6.8)\\
&\leq -\frac{1}{2N}\int_{\mathbb{R}^{dN}}\sum_{i=1}^{N}\Big|\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}\Big|^2 u_N dx_1\dots dx_N+\frac{1}{2}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|\nabla V\star u(X_i^N)-\nabla V\star\mu_N(X_i^N)|^2\right]\\
&\leq -\frac{1}{2N}\int_{\mathbb{R}^{dN}}\sum_{i=1}^{N}\Big|\nabla_{x_i}\log\frac{u_N}{u^{\otimes N}}\Big|^2 u_N dx_1\dots dx_N+\frac{1}{2}\mathbb{E}[\langle\mu_N,|\nabla V\star(\mu_N-u)|^2\rangle].
\end{aligned}$$

Remember the notation of $X_i^N(t)$ are solutions of (6.1) and μ_N is the corresponding random empirical measure.

Exercise. Calculate the relative entropy for the second order system

$$\begin{cases} dX_t^i = V_t^i dt \\ dV_t^i = \frac{1}{N} = \sum_{j=1}^N \nabla V(X_i - X_j)dt + \sqrt{2}dW_t^i \end{cases}.$$

6.2 From Relative Entropy to Strong L1 norm

Before we proceed further for relative entropy estimates, let's mention two lemmata.

Lemma 6.2.1 (Csiszàr-Kullback-Pinsker)**.** For $0 \leq f, g \in L^1(\mathbb{R}^d)$, let the relative entropy between f and g be given by

$$\mathcal{H}(f|g) = \int_{\mathbb{R}^d} f \log \frac{f}{g}.$$

Then the negative part of $\mathcal{H}(f|g)$ is bounded.

If furthermore $\|f\|_{L^1(\mathbb{R}^d)} = \|g\|_{L^1(\mathbb{R}^d)}$ then it holds

$$\|f-g\|_{L^1(\mathbb{R}^d)}^2 \leq 2\mathcal{H}(f|g).$$

Proof. First of all, the negative part of $\mathcal{H}(f|g)$ is bounded in the following:

$$\int_{f\leq g} f \cdot \left|\log \frac{f}{g}\right| = \int_{f\leq g} g\frac{f}{g}\left|\log\frac{f}{g}\right| \leq \frac{1}{e}\|g\|_{L^1(\mathbb{R}^d)}, \text{ where } r\log r \geq -\frac{1}{e} \text{ has been applied.}$$

If $\|f\|_{L^1} = \|g\|_{L^1}$, then

$$\int_{\mathbb{R}^d} \frac{f}{g}\log\frac{f}{g} gdx \geq \int_{\mathbb{R}^d}\frac{f}{g}\cdot gdx - \int_{\mathbb{R}^d} gdx = 0 \text{ where } r\log r \geq r-1 \text{ has been applied.}$$

this means for two probability densities ($\|\mu\|_{L^1} = \|\nu\|_{L^1} = 1$) the entropy is always nonnegative.

Furthermore, by direct computation we obtain, considering $\|f\|_{L^1} = \|g\|_{L^1} = 1$, that

$$\begin{aligned}
\left(\int_{\mathbb{R}^d}|f-g|dx\right)^2 &= \left(\int_{\mathbb{R}^d}\left|\frac{f}{g}-1\right|gdx\right)^2 \\
&= \left(\int_{\mathbb{R}^d}\frac{|\frac{f}{g}-1|}{\sqrt{\frac{f}{g}+2}}g^{\frac{1}{2}}(\sqrt{\frac{f}{g}+2})g^{\frac{1}{2}}dx\right)^2 \\
&\overset{\text{Höld.}}{\leq} \int_{\mathbb{R}^d}\frac{|\frac{f}{g}-1|^2}{(\frac{f}{g}+2)}gdx\cdot\int_{\mathbb{R}^d}(\frac{f}{g}+2)gdx \\
&\leq \frac{2}{3}\left(\int_{\mathbb{R}^d}\frac{f}{g}\log\frac{f}{g}gdx - \int_{\mathbb{R}^d}(\frac{f}{g}-1)gdx\right)\cdot 3 \\
&= 2\int_{\mathbb{R}^d} f\log\frac{f}{g}dx \\
&= 2\mathcal{H}(f|g),
\end{aligned}$$

where we used

$$r\log r - r + 1 \geq \frac{3}{2}\frac{(r-1)^2}{r+2} \quad \forall r \geq 0.$$

□

Exercise. Prove that $r\log r - r + 1 \geq \frac{3}{2}\frac{(r-1)^2}{r+2}$, $\forall r \geq 0$ holds.

Lemma 6.2.2 (super-Additivity of Relative Entropy)**.** Let F_N be symmetric probability density on $\mathbb{R}^{dN}$ and denote by $F_N^{(k)}$ its k-th marginal density, let u be any probability density on $\mathbb{R}^d$, and the relative entropy for the k-th marginal density is defined by

$$\mathcal{H}_k(F_N^{(k)}, u^{\otimes k}) = \int_{\mathbb{R}^{dk}} F_N^{(k)} \log \frac{F_N^{(k)}}{u^{\otimes k}} dx_1 \dots dx_k.$$

Then it holds that

$$\mathcal{H}_k(F_N^{(k)}, u^{\otimes k}) + \mathcal{H}_{N-k}(F_N^{(N-k)}|u^{\otimes(N-k)}) \le \mathcal{H}_N(F_N|u^{\otimes N}).$$

and

$$\frac{1}{N}\mathcal{H}_N(F_N|u^{\otimes N}) \ge \frac{1}{N}\lfloor \frac{N}{k} \rfloor \mathcal{H}_k(F_N^{(k)}|u^{\otimes k}) \ge \frac{1}{2k}\mathcal{H}_k(F_N^{(k)}|u^{\otimes k})$$

where $\lfloor \cdot \rfloor$ means the Gauss bracket.

Proof. By direct computation, we obtain

$$\begin{aligned}
&\mathcal{H}_N - \mathcal{H}_{N-k} - \mathcal{H}_k \\
&= \int_{\mathbb{R}^{dN}} F_N \log \frac{F_N}{u^{\otimes N}} dx_1 \dots dx_N - \int_{\mathbb{R}^{dk}} \Big(\int_{\mathbb{R}^{d(N-k)}} F_N \log \frac{F_N^{(N-k)}}{u^{\otimes(N-k)}} dx_{k+1} \dots dx_N \Big) dx_1 \dots dx_k \\
&\quad - \int_{\mathbb{R}^{d(N-k)}} \Big(\int_{\mathbb{R}^{dk}} F_N \log \frac{F_N^{(k)}}{u^{\otimes k}} dx_1 \dots dx_k \Big) dx_{k+1} \dots dx_N \\
&= \int_{\mathbb{R}^{dN}} F_N (\log \frac{F_N}{u^{\otimes N}} - \log \frac{F_N^{(N-k)}}{u^{\otimes N-k}} - \log \frac{F_N^{(k)}}{u^{\otimes k}}) dx_1, \dots dx_N \\
&= \int_{\mathbb{R}^{dN}} F_N(x_1, \dots, x_N) \log \frac{F_N(x_1, \dots, x_N)}{F_N^{(N-k)}(x_{k+1}, \dots, x_N) F_N^{(k)}(x_1, \dots, x_k)} dx_1, \dots dx_N \ge 0,
\end{aligned}$$

where in the last step we use the fact that

$$\int_{\mathbb{R}^{dN}} F_N dx_1, \dots dx_N = \int_{\mathbb{R}^{dN}} F_N^{(k)}(x_1, \dots, x_k) F_N^{(N-k)}(x_{(k+1)}, \dots, x_N) dx_1, \dots dx_N.$$

$\square$

As a summary from the superadditivity of the relative entropy and the Csiszàr-Kullback-Pinsker, we obtain that

Corollary.

$$\|u_N^{(1)} - u\|^2_{L^1(\mathbb{R}^d)} \le 2\mathcal{H}_1(u_N^{(1)}|u) \le \frac{4}{N}\mathcal{H}(u_N|u^{\otimes N}).$$

6.3 Completion the Relative Entropy Estimate by Mean Field Limit

If we can estimate the rest term in the relative entropy inequality Equation 6.8,

$$\mathbb{E}[\langle \mu_N, |\nabla V \star (\mu_N - u)|^2 \rangle] \xrightarrow{N\to\infty} 0$$

we will obtain the mean field limit in strong L^1 sense.

We first show a complete result for bounded Lipschitz force, and then describe the idea in how to handling the case with singular force.

Theorem 6.3.1. If $\|D^2V\|_{L^\infty(\mathbb{R}^d)} \le C$ then

$$\|u_N^{(1)} - u\|_{L^1(\mathbb{R}^d)} \to 0, \text{ as } N \to \infty..$$

Proof. According to the previous discussions, it is enough to estimate the mean field term appeared in Equation 6.8. We review the notation Y_i^N to be the solution of Mckean-Vlasov equation

$$(\text{MVE*}) \begin{cases} Y_i^N(t) & = \nabla V \star u(Y_i^N(t), t)dt + \sqrt{2}dW_t \\ Y_i^N(0) & = X_i^N(0) \in L^2(\Omega) \text{ given i.i.d. random variables} \\ u_0 & = \mathcal{L}(X_i^N(0)). \end{cases}$$

Then the mean field term in Equation 6.8 can be estimated by using the results from Chapter 3 or 4,

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N |\nabla V \star u(X_i^N) - \frac{1}{N}\sum_{j=1}^N \nabla V(X_i^N - X_j^N)|^2\Big] \\
\le& \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N |\nabla V \star u(X_i^N) - \nabla V \star u(Y_i^N)|^2\Big] + \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N |\nabla V \star u(Y_i^N) - \frac{1}{N}\sum_{j=1}^N \nabla V(Y_i^N - Y_j^N)|^2\Big] \\
&+ \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N |\frac{1}{N}\sum_{j=1}^N \nabla V(Y_i^N - X_j^N) - \nabla V(X_i^N - X_j^N)|^2\Big] \\
\le& \|D^2V \star u\|^2_{L^\infty(\mathbb{R}^d)} \max_{1\le i\le N} \mathbb{E}\big[|X_i^N - Y_i^N|^2\big] + \frac{\|\nabla V\|^2_{L^\infty(\mathbb{R}^d)}}{N} + \|D^2V\|^2_{L^\infty(\mathbb{R}^d)} \max_{1\le i\le N} \mathbb{E}\big[|Y_i^N - X_i^N|^2\big] \\
\le& C \cdot \frac{1}{N}.
\end{aligned}$$

In the above estimates we use the law of large number result for the second term, and the mean field limit result for the first and last term. □

Remark. Actually the above estimate also shows the convergence rate of the relative entropy

$$\|u_N^{(1)} - u\|_{L^1(\mathbb{R}^d)} \le \sqrt{\frac{4}{N}\mathcal{H}(u_N|u^{\otimes N})} \le \frac{C(t)}{\sqrt{N}}.$$

If the interaction potential V has singularity, we will explain the idea by using example $V = \frac{1}{|x|^\lambda}$ where $\lambda \in (0, d-2)$. Notice that in the previous chapter, we have proved that where we started with a mollified version of the many particle system, if we use the notation u_N^ε as the joint law of $(X_i^{N,\varepsilon}(t))$, $i = 1, \dots, N$, from Equation 5.1. Then the relative entropy inequality work exactly

$$\begin{aligned}
&\frac{d}{dt}\mathcal{H}(u_N^\varepsilon|u^{\otimes N}) \qquad (6.9)\\
&\le -\frac{1}{2N}\int_{\mathbb{R}^{dN}} \sum_{i=1}^N \Big|\nabla_{x_i} \log \frac{u_N^\varepsilon}{u^{\otimes N}}\Big|^2 u_N^\varepsilon dx_1 \dots dx_N + \frac{1}{2}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \big|\nabla V \star u(X_i^{N,\varepsilon}) - \nabla V_\varepsilon \star \mu_N(X_i^{N,\varepsilon})\big|^2\right] \\
&\le -\frac{1}{2N}\int_{\mathbb{R}^{dN}} \sum_{i=1}^N \Big|\nabla_{x_i} \log \frac{u_N^\varepsilon}{u^{\otimes N}}\Big|^2 u_N^\varepsilon dx_1 \dots dx_N + \frac{1}{2}\mathbb{E}[\langle \mu_N, |\nabla V_\varepsilon \star (\mu_N - u^\varepsilon)|^2\rangle] \\
&\qquad + \frac{1}{2}\|V_\varepsilon \star \nabla(u^\varepsilon - u)\|^2_\infty + \frac{1}{2}\|(V - V_\varepsilon) \star \nabla u\|^2_\infty.
\end{aligned}$$

We will show the idea how to estimate the mean field rest term in Equation 6.8 with the help of convergence in probability result in Chapter 5. The last two terms above can be easily controlled by ε^2 based on the Assumption F.

For the potential $V = \frac{1}{|x|^\lambda}$ where $\lambda \in (0, d-2)$ we have already shown that, for $\varepsilon = \frac{1}{N^\beta}$, it holds

$$\sup_{0\leq t\leq T} \mathbb{P}(\max_{1\leq i\leq N} |X_i^{N,\varepsilon} - \overline{X}_i^\varepsilon| > N^{-\alpha}) \leq \frac{C}{N^\gamma} \quad \forall \gamma > 0.$$

Then we have that for the first term on the right hand side of Equation 6.9,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left|\nabla V_\varepsilon \star u(X_i^{N,\varepsilon}) - \frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(X_i^{N,\varepsilon} - X_j^{N,\varepsilon})\right|^2\right] \\
&\leq \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left|\nabla V_\varepsilon \star u(X_i^{N,\varepsilon}) - \nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon)\right|^2\right] \\
&\quad + \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left|\nabla V_\varepsilon \star u^\varepsilon(\overline{X}_i^\varepsilon) - \frac{1}{N}\sum_{j=1}^N \nabla V_\varepsilon(\overline{X}_i^\varepsilon - \overline{X}_j^\varepsilon)\right|^2\right] \\
&\quad + \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left|\frac{1}{N}\sum_{j=1}^N \Big(\nabla V_\varepsilon(\overline{X}_i^\varepsilon - \overline{X}_j^\varepsilon) - \nabla V_\varepsilon(X_i^{N,\varepsilon} - X_j^{N,\varepsilon})\Big)\right|^2\right] \\
&= I + II + III.
\end{aligned}$$

We again consider the subset of Ω

$$A(t) = \left\{\omega \in \Omega | \max_{1\leq i\leq N} |X_i^{N,\varepsilon}(t) - \overline{X}_i^\varepsilon(t)| > N^{-\alpha}\right\} \Rightarrow \mathbb{P}(A(t)) \leq \frac{C}{N^\gamma}, \forall \gamma > 0.$$

Then the first and third term $I + III$ can be bounded by splitting the domain Ω. ($1 = \mathbb{1}_{A^c} + \mathbb{1}_A$),

$$\begin{aligned}
I + III &\leq 2\Big(\|\nabla V_\varepsilon \star u\|_\infty^2 + \|\nabla V_\varepsilon\|_\infty^2\Big)\mathbb{E}[\mathbb{1}_A] + 2\Big(\|D^2 V_\varepsilon \star u\|_\infty^2 + \|D^2 V_\varepsilon\|_\infty^2\Big)\frac{1}{N^{2\alpha}} \\
&\leq \frac{4\|\nabla V_\varepsilon\|_\infty^2}{N^\gamma} + \frac{4\|D^2 V_\varepsilon\|_\infty^2}{N^{2\alpha}} \\
&\leq C(\gamma) N^{2\beta(\lambda+1)-\gamma} + C N^{2\beta(\lambda+2)-2\alpha} \to 0,
\end{aligned}$$

where we have used $\|D^2 V_\varepsilon\|_\infty \leq \frac{C}{\varepsilon^{\lambda+2}}, \|\nabla V_\varepsilon\|_\infty \leq \frac{C}{\varepsilon^{(\lambda+1)}}$. And the second term can be bounded simply by using the law of large numbers. We arrive at in the end by choosing $0 < \beta < \min \beta_1, \alpha/(\lambda+2)$ that

$$\frac{d}{dt}\mathcal{H}(u_N^\varepsilon | u^{\otimes N}) + \frac{1}{2N}\int_{\mathbb{R}^{dN}} \sum_{i=1}^N \left|\nabla_{x_i} \log \frac{u_N^\varepsilon}{u^{\otimes N}}\right|^2 u_N^\varepsilon dx_1 \dots dx_N \leq C N^{-\beta}.$$