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

Li Chen

November 7, 2023

Contents

1 Model description and Introduction **1**

1.1 1st Order Particle Systems . . . 1
1.2 Motivation For Partial Differential Equation . . . 3
1.3 2nd Order Particle Systems . . . 4
1.4 Lecture Structure . . . 5

2 MEAN-FIELD LIMIT IN THE DETERMINISTIC SETTING **6**

2.1 Review Of ODE Theory . . . 6
2.2 Mean-field particle system, well-posedness and problem setting . . . 11
2.3 A short introduction for Distributions . . . 12
2.4 Weak Derivative Of Distributions . . . 15
2.5 Weak Formulation Of The Mean Field Partial Differential Equation . . . 16
2.5.1 Stability . . . 19
2.6 Mollification Operator . . . 22
2.6.1 Conservation of Mass . . . 23
2.7 Mean Field Limit . . . 25

3 MEAN FIELD LIMIT FOR SDE SYSTEM **27**

3.1 Basics On Probability Theory . . . 27
3.1.1 Probability Spaces and Random Variables . . . 27
3.1.2 Borel Cantelli . . . 31
3.1.3 Strong Law Of Large Numbers . . . 32
3.1.4 Conditional Expectation . . . 33
3.1.5 Stochastic Processes And Brownian Motion . . . 34
3.1.6 Brownian Motion . . . 35
3.1.7 Convergence of Measure and Random Variables . . . 37
3.2 Itô Integral . . . 38
3.2.1 Itô's Formula . . . 42
3.2.2 Multi-Dimensional Itô processes and Formula . . . 46
3.3 Relation To The Mean Field Limit . . . 48
3.4 Solving Stochastic Differential Equations . . . 49
3.5 Stochastic Mean Field Limit . . . 54
3.5.1 Stochastic Particle System . . . 55
3.5.2 I.I.D Case . . . 55
3.5.3 Toy Example . . . 58
3.5.4 Makean-Vlasov . . . 60

4 PDE Approach To Solving the Makean-Vlasov Equation **66**

4.1 Motivation . . . 66
4.2 Problem Definition . . . 67
4.3 Heat Equation and the Heat Kernel . . . 68
4.3.1 Motivation . . . 68
4.3.2 Derivation by Fourier Transform . . . 68
4.3.3 Back to the Makean-Vlasov Equation . . . 79

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

The following chapter will outline how the relevant particle models are defined, we differentiate between first and second order systems focusing here on first order systems while leaving the second order setting as exercises

1.1 1st Order Particle Systems

Definition 1.1.1 (1st Order Particle System)**.** 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.

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

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

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

For $\sigma = 0$ we say the system is deterministic

We consider the following examples for K

Example. A common example for a well-behaved K is

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

which is a locally Lipschitz continuous function.

Another typical interaction force which is not continuous is the potential field given by Coulomb potential, namely

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

Definition 1.1.2 (Empirical Measure)**.** For a set of particles $(x_1(t), x_2(t), \dots, x_N(t)) \in \mathcal{C}^1([0,T];\mathbb{R}^d),\ i = 1, \dots, N$ we define the empirical measure by

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

Our goal is the study of the limit of this system as $N \to \infty$. An appropriate quantity is to consider the empirical measure 1.1.2. 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 $\mu(t)$?

$$\mu^N(t) \xrightarrow{?} \mu(t).$$

Furthermore, can we find an equation which $\mu(t)$ satisfies and in which sense does it satisfy this equation?

> **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 ?

1.2 Motivation For Partial Differential Equation

Let the following Proposition serve as a motivation on which partial differential equation $\mu(t)$ should satisfy and consider only the deterministic case for now.

Proposition 1.2.1. We say $\mu(t)$ solves the following partial differential equation (in the sense of distribution)

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

Proof. Take $\varphi \in C_0^\infty(\mathbb{R}^d)$ and calculate

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

i.e μ^N is a solution to

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

If we can now take the limit $N \to \infty$ we obtain that μ should satisfy the proposed PDE □

Corollary. If $\sigma > 0$ i.e our system is stochastic then we expect the limit partial differential equation to share a similar structure

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

We define the stochastic case in detail later

1.3 2nd Order Particle Systems

We define a second order particle system as follows

Definition 1.3.1. Given the N particles

$$((x_1(t), v_1(t)), \dots, (x_N(t), v_N(t))) \in \mathcal{C}^1([0, T]; \mathbb{R}^{2d}).$$

with initial values $x_i(0)$ for $i = 1, \dots, N$

Then our second order system is then governed by

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

In this setting $(x_i(t), v_i(t))$ mean the position and velocity of the i-th particle respectively. An example for F would be

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

The empirical measure from Definition 1.1.2 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. Calculate for $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^{2d})$ the following in the second order case

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

1.4 Lecture Structure

In Chapter 1, we are going to discuss the deterministic case for "Good" interaction forces (2-3 weeks) while giving a brief review of the well-posedness theory of ordinary differential equation. And prove the mean field limit in the framework of 1-Wasserstein distance.

The stochastic case will be studied in Chapter 2. Where we first review the mandatory concepts of probability theory, the definition of the Itô integral, and the well-posedness of stochastic differential equations. Then the propagation of chaos result of the interacting SDE system is studied, where the well-posedness of Mckean-Vlasov equation plays an important role. If time allows, we will study non-smooth interaction forces in chapter 3.

The first result is the convergence in probability, which implies the weak convergence of propagation of chaos. The second topic is to introduce the relative entropy method to get the convengence in L^1 space.

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, recall

Definition (1st Order Particle System)**.** 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.

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

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

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

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

2.1 Review Of ODE Theory

Definition 2.1.1 (Intial Value Problem)**.** For $\forall T > 0$ and
$f \; [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ we consider the initial value problem given by

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

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 $\forall (t,x), (t,y) \in [0,T] \times \mathbb{R}^d$

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

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

$$
\begin{aligned}
|x_{m+p} - x_m(t)| &\leq \sum_{k=m+1}^{m+p} |x_k(t) - x_{k-1}(t)| \\
&\leq \sum_{k=m+1}^{m+p} M \frac{L^{k-1}T^k}{k!} \\
&= \frac{M}{L} \sum_{k=m+1}^{m+p} \frac{(LT)^k}{k!}
\end{aligned}
$$

.

Continuing on the next page

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

Therefore $x_m(t)$ has a limit $x(t) \in \mathcal{C}^1([0, T]; \mathbb{R}^d)$ 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 $x(t) \in \mathcal{C}^1([0, T]; \mathbb{R}^d)$ is a solution of the equivalent integral equation.

To prove the uniqueness suppose we have two solutions $x(t), \tilde{x}(t) \in \mathcal{C}^1([0, T]; \mathbb{R}^d)$ 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

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

We obtain

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

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 give in the following. 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

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

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

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

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

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

□

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

Let us again give the model and problem setting

Definition 2.2.1 (1st Order Particle System)**.** 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.

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

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

where $K : \mathbb{R}^{2d} \to \mathbb{R}^d$ is anti-symmetric i.e

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

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

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

Lemma 2.2.1. When Assumption B holds for K then for $\forall T > 0$ the (MPS) has a unique solution

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

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

In the introduction we saw that the empirical measure satisfies a partial differential equation

Definition 2.2.2 (PDE Problem)**.** Let $\mu^N(t)$ be the empirical measure

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

Then from the introduction we know that for $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ the empirical measure satisfies

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

where

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

Idea. If $\mu^N \to \mu$ in some sense, then the limiting measure μ should also satisfy

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

in the sense weak sense i.e. the "sense of distributions" which we define in the following section

2.3 A short introduction for Distributions

Definition 2.3.1. Let $\Omega \subset \mathbb{R}^d$ be an open subset then 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 $\exists K \subset \Omega$ such that $\operatorname{supp} \varphi_m \subset K$ for $\forall m$
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 (Multi-Index)**.** A multi-index $\alpha \in \mathbb{N}_0^n$ of length $|\alpha| = \sum_i \alpha_i$ for example $\alpha = (0, 2, 1) \in \mathbb{N}_0^3$ can be used to denote partial derivatives of higher order as such :

$$\partial^\alpha = \prod_i (\frac{\partial}{\partial x_i})^{\alpha_i}.$$

Definition 2.3.3 (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 functions on $\mathcal{D}(\Omega)$

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

1. $\langle T, \alpha\varphi_1 + \beta\varphi_2\rangle = \alpha\langle T, \varphi_1\rangle + \beta\langle T, \varphi_2\rangle$
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 2.3.4. 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, instead the convolution of two Distributions is defined

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

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

and 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 $1 < q < p < \infty$

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

Theorem 2.3.1. L^1_{loc} functions are uniquely deterimed 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

Proof. 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 $P(x)$ the associated distribution $T_P \in \mathcal{D}'(\mathbb{R})$ is defined by

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

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 in the above way 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 2.3.5. For a sequence of distributions $(T_m)_{m\in\mathbb{N}} \subset \mathcal{D}'(\Omega)$ we was it converges against a limit $T \in \mathcal{D}'(\Omega)$ iff

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

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

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

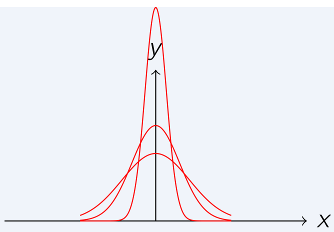


Figure 2.1: Heat Kernel for different t

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 $\forall \varphi \in \mathcal{C}_0^\infty(\Omega)$

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

where we used $x = 2\sqrt{t}y$ $\square$

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

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

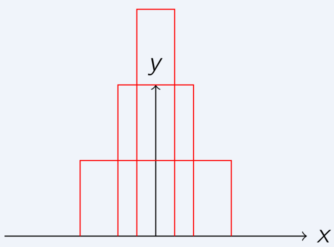


Figure 2.2: Rectangular functions for different n

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

$$D_n \xrightarrow{n\to\infty} 2\pi\delta_0(x).$$

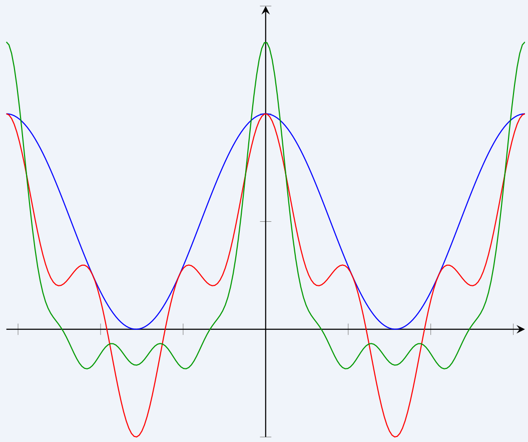


Figure 2.3: Dirichlet kernel for different n

2.4 Weak Derivative Of Distributions

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

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

For multi index α

Exercise. Prove the function $-\langle T, \partial_i\varphi\rangle$ is a continuous and linear function

Hint: Consider the case where $T := T_f$ for $f \in L^1_{\text{loc}}$

We give a couple examples

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

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

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

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

Therefore

$$H' = \delta_0.$$

□

We can now go on to properly formulate the mean field partial differential equation in a weak sense

2.5 Weak Formulation Of The Mean Field Partial Differential Equation

Using the notation of the empirical measure we can rewrite our earlier definition of the (MPS) 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 PDE

Definition 2.5.1 (Mean Field Equation (MFE))**.** Define the mean field equation as

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

where

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

We give the definition of the weak solution of (MFE)

Definition 2.5.2 (Weak Solution of MFE)**.** For all $t \in [0, T]$, $\mu(t) \in \mathcal{M}(\mathbb{R}^d)$ is called a weak solution of (MFE), where $\mathcal{M}(\mathbb{R}^d)$ denotes the space of measures on $\mathbb{R}^d$

For $\forall \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ it holds

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

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 (MFE)

We define the following initial value problem, the so called characteristics equation

Definition 2.5.3 (Push Forwad Measure)**.** For a measurable function X and a measure $\mu_0 \in \mathcal{M}(\mathbb{R}^d)$ denote the push forward measure for any Borel set $B \subset \mathbb{R}^d$ by

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

Definition 2.5.4 (Characteristics equation)**.**

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

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 Characteristics equation is a weak solution of the (MFE)

Remark. The solution space of the Characteristics equation is given by

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

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

Assumption C (Regularity)**.** 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 has already been used in order to show the well-posedness of the particle system

Theorem 2.5.1 (Existence and Uniqueness of Characteristics Equation)**.** Let Assumption C hold for K and $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ then the Characteristics equation 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

$$X := \{v \in \mathcal{C}(\mathbb{R}^d) \ \ \|v\|_X < \infty\}.$$

Where

$$\|v\|_X := \sup_{x \in \mathbb{R}^d} \frac{|v(x)|}{1 + |x|}.$$

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

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

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 consider two solutions $x, \tilde{x}$ then 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

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

where clearly $\|x(0,\cdot) - \tilde{x}(0,\cdot)\|_X = 0$ □

2.5.1 Stability

Let's remind us of the N-particle system (MPS), the Mean field equation (MFE) and its weak solution as defined in Definition 2.5.2. We have thus far done the following things

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

$x(t,\cdot,\mu_0)\#\mu_0 \in \mathcal{P}_1$ is the solution of (MFE)

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 2.5.5 (Monge-Kantorovich Distance)**.** For two measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ $p \geq 1$ with

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

the Monge-Kantorovich distance $\mathrm{dist}_{\mathrm{MK},p}(\mu,\nu)$ or $W^p(\mu,\nu)$ is defined by

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

where

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

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 hold

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

Proof. Let (x_0,μ_0) and $(\overline{x}_0,\overline{\mu}_0)$ be two initial data pairs of problem Theorem 2.5.1 and $\pi_0 \in \Pi(\mu_0,\overline{\mu}_0)$ taking the difference of these two problems, we have

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

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 Regularity for K, we have

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

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

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

□

2.6 Mollification Operator

Definition 2.6.1 (Mollification-Kernel)**.** A function $j(x) \in C_0^\infty$ 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 $\varphi(x) \in C_0^\infty(\mathbb{R}^d)$ 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 $\forall u \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ 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.

1. If $u(x) \in L^1(\mathbb{R}^d)$ and $\operatorname{supp} u(x)$ is compact in $\mathbb{R}^d$ then

$$J_\varepsilon(u) = j_\varepsilon \star u \in \mathcal{C}_0^\infty \quad \forall \varepsilon > 0.$$

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 $K = \operatorname{supp} u \subset \mathbb{R}^d$ be compact , 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

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

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

For (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 ☐

2.6.1 Conservation of Mass

Let $u(t)$ be the push forward measure obtained from Definition 2.5.4, one can check that is is a weak solution of (MFE) by using test functions. Furthermore we obtain that 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 $\|f(t, \cdot)\|_{L^1(\mathbb{R}^d)} = 1$. The weak solution in the sense of distribution means that $\forall \varphi \in \mathcal{C}_0^\infty$ it holds for all $0 \leq \tilde{t} < t < \infty$

$$\int_{\mathbb{R}^d} \varphi(x) f(t, x) dx - \int_{\mathbb{R}^d} \varphi(x) f(\tilde{t}, x) dx = \int_{\tilde{t}}^t \int_{\mathbb{R}^d} f(s, x) \mathcal{K} f(s, x) dx ds.$$

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 $\int_B d\lambda = 0$ it holds $\mu_t(B) = 0$. The mass conservation property, $\|f(t, \cdot)\|_{L^1(\mathbb{R}^d)} = 1$ comes from the definition of probability measures □

Exercise. Let $\mu_t \in \mathcal{P}_1(\mathbb{R}^d)$ be an absolutely continuous measure with respect to the Lebesgue measure, then proof that for $\forall B \in \mathcal{B}$ such that

$$\int_B d\lambda = 0.$$

it holds that $\mu_t(B) = 0$

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. In the weak solution formulation it holds

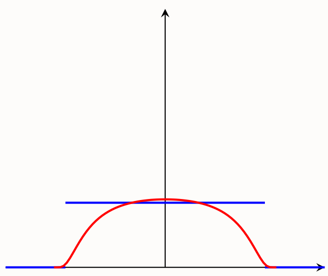
$$\int_{\mathbb{R}^d} \varphi(x) f(t, x) dx = \int_{\mathbb{R}^d} \varphi(x) f(\tilde{t}, x) dx + \int_{\tilde{t}}^t \iint_{\mathbb{R}^{2d}} f(s, x) K(x, y) f(s, y) \nabla \varphi(x) dx dy ds.$$

where the test function $\varphi \in \mathcal{C}_0^\infty$ is chosen arbitrarily. Now we take a sequence of test functions defined as follows.

For $\forall R > 0$

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

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

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

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

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

$\square$

2.7 Mean Field Limit

Theorem 2.7.1 (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

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

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

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

Then

$$\underset{\text{MK},1}{\text{dist}}(\mu^N(t), f(t,\cdot)) \le e^{2Lt} \underset{\text{MK},1}{\text{dist}}(\mu_0^N, f_0) \xrightarrow{N\to\infty} 0.$$

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 $\forall\varphi \in \text{Lip}(\mathbb{R}^d)$ we have

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

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 $\mathcal{C}_0$ 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 $\varphi \in \mathrm{Lip}(\mathbb{R}^d)$.

For $\forall \varphi \in \mathcal{C}_b^\infty$ 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 $\exists M >)$ 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 (MPS) and the associated Mean-Field equation (MFE) 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 small review of basic concepts in probability theory in preparations of SDE's

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 $A_1, A_2, \ldots \in \mathcal{F}$ countable, 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

$$\mathbb{P} \ : \ \mathcal{F} \to [0, 1].$$

s.t.

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 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 $\mathbb{P}(F) = 1$ the we say that F happens with probability 1 or almost surely (a.s.)

Remark. Let $\mathcal{H}$ be a family of subsets of Ω, then there exists a smallest $\sigma-$algebra of Ω called $\mathcal{U}_{\mathcal{H}}$ with

$$\mathcal{U}_{\mathcal{H}} = \bigcap_{\substack{\mathcal{H} \subset \mathcal{U} \\ \mathcal{H}\ \sigma-\text{alg.}}} \mathcal{H}.$$

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 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

From now on we denote $(\Omega, \mathcal{F}, \mathbb{P})$ as a given probability space.

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

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 there exists the simple relation

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

Remark. If $f : \mathbb{R}^d \to \mathbb{R}$ measurable and

$$\int_{\Omega} |f(X(\omega))| d\mathbb{P}(\Omega) < \infty.$$

then

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

Definition 3.1.9 (L^p spaces)**.** Let $X : \Omega \to \mathbb{R}^d$ be a random variable and $p \in [1, \infty)$. With the norm

$$\|X\|_p = \|X\|_{L^p(\mathbb{P})} = \left(\int_{\Omega} |X(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}}.$$

If $p = \infty$

$$\|X\|_\infty = \inf\{N \in \mathbb{R} : |X(\omega)| \leq N \text{ a.s.}\}.$$

the space $L^p(\mathbb{P}) = L^p(\Omega) = \{X \ : \ \Omega \to \mathbb{R}^d \mid \|X\|_p \leq \infty\}$ 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 3.1.10 (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 the ints 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 $X_1, \dots, X_m : \Omega \to \mathbb{R}^d$ are random variables, their joint distribution function is

$$F_{X_1,\dots,X_m} : (\mathbb{R}^d)^m \to [0,1]$$
$$F_{X_1,\dots,X_M} = \mathbb{P}(X_1 \le x_1, \dots, X_m \le x_m) \quad \forall x_i \in \mathbb{R}^d.$$

Definition 3.1.11 (Density Function Of X)**.** If there exists a non-negative function $f(x) \in L^1(\mathbb{R}^d; \mathbb{R})$ such that

$$F(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y)dy \quad y = (y_1, \dots, y_n).$$

then f is called 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 random variable with density function $x \in \mathbb{R}$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|x-m|^2}{2\sigma^2}}.$$

then we say that X has a Gaussian (or Normal) distribution with mean m and variance σ^2 and write

$$X \sim \mathcal{N}(m, \sigma^2).$$

Obviously

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

Definition 3.1.12 (Independent Events)**.** Events $A_1, \dots, A_n \in \mathcal{F}$ are called independent if $\forall 1 \le k_1 < \dots < k_m \le n$ it holds

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

Definition 3.1.13 (Independent $\sigma-$Algebra)**.** Let $\mathcal{F}_j \subset \mathcal{F}$ be $\sigma-$algebras for $j = 1, 2, \dots$. Then we say $\mathcal{F}_j$ are independent if for $\forall 1 \le k_1 < k_2 < \dots < k_m$ and $\forall A_{k_j} \in \mathcal{F}_{k_j}$ it holds

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

Definition 3.1.14 (Independent Random Variables)**.** We say random variables $X_1, \dots, X_m$: $\Omega \to \mathbb{R}^d$ are independent if for $\forall B_1, \dots, B_m \subset \mathcal{B}$ in $\mathbb{R}^d$ it holds

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

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

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

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

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

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

and

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

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

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

Exercise. Proof the above theorems

3.1.2 Borel Cantelli

Definition 3.1.15. Let $A_1, \dots, A_m \in \mathcal{F}$ then 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 $\mathbb{P}(A_{\text{ i.o. }}) = 0$

Proof. By definition we have

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

$\square$

Definition 3.1.16 (Convergence In Probability)**.** We say a sequence of random variables $(X_k)_{k=1}^{\infty}$ converges in probability to X if for $\forall \varepsilon > 0$

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

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

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

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

Proof. For $\forall j$ $\exists k_j$ with $k_j < k_{j+1} \to \infty$ s.t.

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

then

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

Let $A_j = \{\omega \ : \ |X_{k_j} - X| > \frac{1}{j}\}$ then by Borel Cantelli we have $\mathbb{P}(A_j \text{ i.o.}) = 0$ s.t.

$$\forall \omega \in \Omega \ \exists J \text{ s.t. } \forall j > J.$$

it holds

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

□

3.1.3 Strong Law Of Large Numbers

Definition 3.1.17. A sequence of random variables $X_1, \dots, X_n$ is called identically distributed if

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

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

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

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

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

Proof. Suppose for simplicity $\mathbb{E}[X^4] < \infty$ for $\forall i = 1, 2, \dots$. 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

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

Therefore for fixed $\varepsilon > 0$

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

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

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

□

3.1.4 Conditional Expectation

Definition 3.1.18. 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 3.1.19. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{U} \subset \mathcal{F}$ 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 $\forall A \in \mathcal{U}$

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

Exercise. Proof the following equalities

1. $\mathbb{E}[X|Y] = \mathbb{E}[X|\mathcal{U}]$
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

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

Note the equivalent notation $\chi_A \equiv \mathbb{1}_A$

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. $\mathbb{E}[aX + bY|\mathcal{U}] = a\mathbb{E}[X|\mathcal{U}] + b\mathbb{E}[Y|\mathcal{Y}]$
3. If X is $\mathcal{U}-$measurable and XY is integrable then
$$\mathbb{E}[XY|\mathcal{U}] = X\mathbb{E}[Y|\mathcal{Y}].$$
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 $X \leq Y$ a.s. then $\mathbb{E}[X|\mathcal{U}] \leq \mathbb{E}[Y\mathcal{U}]$ 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
$$\varphi(\mathbb{E}[X|\mathcal{U}]) \leq \mathbb{E}[\varphi(X)|\mathcal{U}].$$

Exercise. Proof the above Lemma

3.1.5 Stochastic Processes And Brownian Motion

Definition 3.1.20 (Stochastic Process)**.** A stochastic process is a parameterized collection of random variables
$$(X(t))_{t\in[0,T]} \ : [0,T] \times \Omega \ : \ (t,\omega) \mapsto X(t,\omega).$$
For $\forall \omega \in \Omega$ the map
$$X(\cdot,\omega) \ : \ [0,T] \to \mathbb{R}^d \ : \ t \mapsto X(t,\omega).$$
is called sample path

Definition 3.1.21 (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 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 3.1.22 (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 3.1.23 (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

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 3.1.24 (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

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

Furthermore based on the above remark for $t \geq s$

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

which means generally

$$\mathbb{E}[W(t)W(s)] = t \wedge s.$$

Definition 3.1.25. An $\mathbb{R}^d$ valued process $W(\cdot) = (W^1(\cdot), \dots, W^d(\cdot))$ is a $d-$dimensional Wiener process (or Brownian motion) if

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

Remark. If $W(\cdot)$ is a $d-$Dimensional Brownian motion, then $W(t) \sim \mathcal{N}(0,t)$ and for any Borel set $A \subset \mathbb{R}^2$

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

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 $\forall 0 < \gamma < \frac{\alpha}{\beta}$ and $T > 0$ 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

$$\mathbb{E}[|W(t)-W(s)|^{2m}] \leq C|t-s|^m.$$

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. $\omega, t \mapsto W(t,\omega)$ 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 3.1.26 (Markov Property)**.** An $\mathbb{R}^d-$valued process $X(\cdot)$ is said to have the Markov property, if $\forall 0 \leq s \leq t$ and $\forall B \subset \mathbb{R}^d$ Borel. , 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

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

3.1.7 Convergence of Measure and Random Variables

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

Definition 3.1.27 (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 $\forall B \in \mathcal{B}$
$$\mu_n(B) \to \mu(B).$$
4. For $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$ uniform continuous it holds
$$\int f d\mu_n \to \int f d\mu.$$

Definition 3.1.28 (Weak convergence of Random variable)**.** The following statements are equivalent

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

1. X_n converges to X in probability
2. For $\forall \varepsilon > 0$
$$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n\to\infty} 0.$$

Exercise. Prove that

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

Definition 3.1.29 (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

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

Theorem 3.1.10 (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

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

   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. $P = \{0 = t_0 < t_1 < \ldots < t_m = T\}$ s.t.

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

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

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

Let $0 \leq t_0 < t_1 < \ldots \leq t_n = T$ 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.

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

Proof. We roughly sketch the Idea here

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

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

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

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

$\square$

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

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

□

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 $\int_0^\tau GdW_t$ with $0 \le \tau \le T$ is a martingale

3.2.1 Itô's Formula

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

$$X(r) = X(s) + \int_s^r Fdt + \int_s^r GdW_t.$$

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

$$dX = Fdt + gdW_t \quad \forall 0 \le t \le T.$$

Theorem 3.2.2 (Itô's Formula)**.** Let $X(\cdot)$ be an Itô process given by $dX = Fdt + GdW_t$ 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

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

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

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

Proof. The proof is split into five steps

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

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

For (1) it is sufficient to prove $W_t^2 - W_0^2 = \int_0^t 2W_s dW_s + t$ a.s. By definition of Itô integral, for a.s. $\omega \in \Omega$ we have

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

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

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

holds true. Actually

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

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

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

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

For (2) : It is sufficient to prove $tW_t - 0W_0 = \int_0^t W_s ds + \int_0^t s dW_s$. Actually we have

$$\int_0^t s dW_s = \lim_{n\to\infty} \sum_{k=0}^{n-1} t_k^n \left(W(t_{k+1}^n - W(t_k^n))\right) \text{ a.s.}.$$

and for a.s. ω the standard Riemann sum

$$\int_0^t W_s ds = \lim_{n\to\infty}\sum_{k=0}^{n-1} W(t_{k+1}^n)(t_{k+1}^n - t_k^n).$$

The summation of the above integrals yields

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

Step 2. Now let us prove the Itô product rule. 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

$$X_i(t) = F_i t G_i W(t).$$

Then it holds a.s. that

$$\begin{aligned}&\int_0^t (X_2dX_1 + X_1dX_2 + G_1G_2ds)\\ &= \int_0^t (X_2F_1 + X_1F_2)ds + \int_0^t (X_2G_1 + X_1G_2)dW_s + \int_0^t G_1G_2ds\\ &= \int_0^t \left(F_1(F_2s + G_2W(s)) + F_2(F_1s + G_1W(s))\right) ds + G_1G_2t\\ &= \int_0^t \left(G_1(F_2s + G_2W(s)) + G_2(F_1s + G_1W(s))\right) dW_s\\ &= G_1G_2tF_1F_2t^2 + (F_1G_2 + F_2G_1)\left(\int_0^t W(s)ds + \int_0^t sdW_s\right)\\ &\quad + 2G_1G_2\int_0^t W(s)dW_s.\end{aligned}$$

using (1) and (2) from Step 1. It continues to hold that

$$G_1G_2(W(t))^2 + F_1F_2t^2 + (F_1G_2 + F_2G_1)tW(t) = X_1(t) + X_2(t).$$

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

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

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.

$$X_1^n(t)X_2^n(t) - X_1(0)X_2(0) = \int_0^t (X_1^n(s)dX_2^n(s) + X_2^n(s)dX_1^n(s) + G_1G_2ds)\,.$$

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 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 $m - 1 \to m$ then

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

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

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

Note 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 $K \subset \mathbb{R} \times [0, T]$. Since

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

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

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 $G \in \mathbb{L}^2_{n\cdot m}([0,T])$, $G_{ij} \in \mathbb{L}^2([0,T])$ $1 \leq i \leq n$, $1 \leq j \leq m$

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

With the Properties

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

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

$$X(t) = X(s) + \int_s^t F_{n\times 1}(r)dr + \int_0^t G_{n\times m}(r) dW_{m\times 1}(r)$$
$$dX^i = F^i dt + \sum_{j=1}^m G^{ij} dW_t^i \qquad 1 \leq i \leq n.$$

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

$$du(X(t),t) = \frac{\partial u}{\partial t}(X(t),t)dt + \nabla u(X(t),t) \cdot dX(t)$$
$$+ \frac{1}{2} \sum \frac{\partial^2 u}{\partial X_i \partial X_j}(X(t),t) \sum_{l=1}^m G^{il} G^{il} dt.$$

Proposition 3.2.2. For real valued processes X_1, X_2

$$\begin{cases} dX_1 & = F_1 dt + G_1 dW_1 \\ dX_2 & = F_2 dt + G_2 dW_2 \end{cases} \Rightarrow d(X_1, X_2) = X dX_2 + X_2 dX_1 + \sum_{k=1}^m G_1^k G_2^k dt.$$

Definition 3.2.12 (Multiplication Rules)**.** Formal multiplication rules for SDEs

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

.

Remark. Using the above we can simplify Itô's formula as follows

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

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

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

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

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

Definition 3.2.13 (Parabolic Operator)**.**

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

Example. If $F = 0$ $m = n$ and $G = \sqrt{2}I_{n\times n}$ then

$$dX = \sqrt{2}dW_t.$$

And the law μ of X fulfills the heat equation i.e

$$\dot{\mu}t - \Delta\mu = 0.$$

3.3 Relation To The Mean Field Limit

To find out how all this translates to our Mean field Limit we consider the particle system given by

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

And denote

$$\mathbb{X}_N = F(\mathbb{X}_N)dt + \sqrt{2}dW_t.$$

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

$$(X_1(t), \dots, X_N(t)) \sim u(X_1, \dots, X_n).$$

Where $u \in \mathcal{M}(\mathbb{R}^{dN})$, then by Itô's formula we get for arbitrary test function $\forall\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^{dN})$

$$\varphi(\mathbb{X}_N(t)) = \varphi(\mathbb{X}_N(0)) + \int_0^t \nabla\varphi\cdot\begin{pmatrix} \vdots \\ \frac{1}{n}\sum_{j=1}^{N} K(X_i, X_j) \\ \vdots \end{pmatrix}$$
$$+\int_0^t \Delta\mathbb{X}_N\varphi dt + \int_0^t \sqrt{2}\nabla\varphi dW_t^i.$$

Taking the expectation on both sides, then the last term disappears by definition of Itô processes

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

Now consider the Mean-Field-Limit, if the joint particle law can be rewritten as the tensor product of a single $\overline{u}$

$$u(X_1, \dots, X_N) = \overline{u}^{\otimes N}.$$

the equation simplifies

$$\partial_t - \sum_{i=1}^{N}\Delta_i u + \sum_{i=1}^{N}\nabla_{X_i}\left(\overline{u}^{\otimes N}k \star \overline{u}(X_i)\right) = 0.$$

3.4 Solving Stochastic Differential Equations

The setup of the following section will be the following

Definition 3.4.1 (Basic Setup). We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, With a m-D dimensional Brownian motion $W(\cdot)$. Let X_0 be an n-D dimensional random variable independent of $W(0)$, then our Filtration is given by

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

Definition 3.4.2 (SDE). Given the above basic setup we are trying to solve equations of the type

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

Where

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

Remark. The differential equation should always be understood as the Integral equation

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

Definition 3.4.3 (Solution). We say an $\mathbb{R}^n$-valued stochastic process $X(\cdot)$ is a solution of the SDE 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) $G := B(X_t, t) \in \mathbb{L}^2_{n\times m}([0, T]) \Leftrightarrow \ \int_0^t \mathbb{E}[|G_s|^2]ds < \infty$

Remark. (1) implies 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 solutions, for that we first define what it means for a solution to be unique

Definition 3.4.4. 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.

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 (in (t, x)) and Lipschitz continuous with respect to x for some $L > 0$. Furthermore assume they fulfill the linear

growth condition

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

Remark. Note the Lipschitz continuity from Assumption D implies that there $\exists L > 0$ such that

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

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

Proof. We begin with the uniqueness prove.

Suppose we have two solutions X and $\tilde{X}$ of the SDE then the goal is to show that they are indistinguishable, then by using the definition of a solution

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

If the diffusion term were 0 we could use a Grönwall type inequality and get the uniqueness.

Instead we consider the square of the above and apply Itôs isometry. Note that generally $|a+b|^2 \nleq (a^2+b^2)$ but $|a+b|^2 \leq 2(a^2+b^2)$

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

Now consider the following

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

.

Where the following Hoelders inequality was used

$$\begin{aligned}
\left(\int_0^t 1|f|ds\right)^2 &\leq \left(\int_0^t 1^2 ds\right)^{\frac{1}{2}\cdot 2} \cdot \left(\int_0^t |f|^2 ds\right)^{\frac{1}{2}\cdot 2} \\
&\leq t\int_0^t |f|^2 ds.
\end{aligned}$$

Now by Gronwalls inequality we have

$$\mathbb{E}[|X_t - \tilde{X}_t|^2] = 0.$$

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

Define

$$A_t = \{\omega \in \Omega \mid |X_t - \tilde{X}_t| > 0\} \qquad \mathbb{P}(A_t) = 0.$$

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

Now since $X_t(\omega)$ is continuous in t we can extend the maximum over the entire interval $[0, T]$

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

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 existence similar to the deterministic case we use Picard iteration.

First define the Picard iteration by

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

Let $d(t)^n = \mathbb{E}[|X_t^{n+1} - X_t^n|^2]$, then we claim by induction that $d^n(t) \le \frac{(Mt)^{n+1}}{(n+1)!}$ for some $M > 0$.
IA: For $n = 0$ we have

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

IV: suppose the assumption holds for $n - 1 \in \mathbb{N}$
IS: Take $n - 1 \to n$ then

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

Because of Ω we cannot use completeness to argue the convergence and instead are forced to use a similar argument as in the uniqueness proof.

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

Where we used the following Doobs martingales L^p inequality

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

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

$$\begin{aligned}
C\cdot\mathbb{E}[\int_0^T |X_s^n-X_s^{n-1}|^2ds] &= C\cdot\int_0^T \mathbb{E}[|X_s^n-X_s^{n-1}|^2]ds \\
&\le \int_0^T \frac{(Mt)^n}{(n)!} \\
&= C\frac{M^nT^{n+1}}{(n+1)!}.
\end{aligned}$$

Further more we get with a Markovs inequality

$$\begin{aligned}
\mathbb{P}(\underbrace{\max_{0\le t\le T}|X_t^{n+1}-X_t^n|^2>\frac{1}{2^n})}_{A_n} &\le 2^{2n}\mathbb{E}[\max_{0\le t\le T}|X_t^{n+1}-X_t^n|^2] \\
&\le 2^{2n}\frac{CM^nT^{n+1}}{(n+1)!}.
\end{aligned}$$

Then by Borel-Cantelli

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

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

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

In fact we can give B directly by

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

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

$$\begin{aligned}\max_{0\le t\le T}|X_t^{n+k} - X_t^n| &\le \sum_{j=1}^{k} \max|X_t^{n+j} - X_t^{n+(j-1)}| \\ &\le \sum_{j=1}^{k} \frac{1}{2^{n+j-1}} \\ &< \frac{1}{2^{n-1}}.\end{aligned}$$

By the above we get

$$X_t^n(\omega) \to X_t(\omega) \qquad \text{uniform in } t \in [0,T].$$

Therefore for a.s. ω , 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

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

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

Using the above we conclude by Fatous's lemma

$$\mathbb{E}[|X_t|^2] = \mathbb{E}[\lim_{n\to\infty}|X_t^{n+1}|] \le \liminf_{n\to\infty}\mathbb{E}[|X_t^{n+1}|^2] \le C\cdot e^{Ct}.$$

Therefore

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

□

Remark. One should remember that if the diffusion term $B(X_t,t)$ is 0 then we get a unique solution iff $b(X_t,t)$ is Lipschitz

Theorem 3.4.2 (Higher Moments Estimate)**.** Assumptions for b , B and X_0 are the same as before, if in addition

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

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

$$\mathbb{E}[|X_t|^{2p}] \le C(1+\mathbb{E}[|X|_0^{2p}])e^{Ct}.$$

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

Proof. Left as an exercise

□

3.5 Stochastic Mean Field Limit

First recall the metric we use to talk about distance between two measures i.e the Wasserstein Distance

Definition 3.5.1 (Wasserstein Distance)**.** For all $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, $(p \geq 1)$ the Wasserstein Distance 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

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

Remark. Note that

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

follows naturally by Hölders inequality, in fact this holds for all $p > q$

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

Remark. 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 $|f(x)| \leq C(1 + |x|^p)$
$$\int f d\mu_n \to \int f d\mu.$$
3. $\mu_n \rightharpoonup \mu$

3.5.1 Stochastic Particle System

Let us begin by shortly defining the stochastic particle systems we study.

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

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

Then our stochastic particle system is given by,

Definition 3.5.3 (Stochastic Particle System)**.** For N interacting particles $(X^1,...,X^N)$ with i.i.d initial data $(X_i^N(0))_{i\in\{1,...,N\}} \subset L^2(\Omega)$ and law μ_0

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

Where μ_N is the stochastic empirical measure and note $\mathcal{L}(X_0) = \mu_0$

Remark. The dimensions for our Stochastic-Particle-System are the same as in Definition 3.4.2

Remark. For our initial measure we already have

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

3.5.2 I.I.D Case

Let us shortly consider the convergence of the empirical measure in the case where our random variables are i.i.d, note that in our mean field limit this is only the case for our initial data, since for $t > 0$ they are no longer i.i.d.

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

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

We can actually prove the stronger statement that the choice of $f \in \mathcal{C}_b$ does not matter for the convergence i.e. we can pull the function selection into the probability similarly to the difference between modification and indistinguishable.

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

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

i.e

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

Proof. Needs revision, this should only work for $\mathcal{C}_b(K)$ K compact in my opinion
The proof relies mainly on showing that $\mathcal{C}_b(\mathbb{R}^d)$ is separable for compact support we can use the density of the polynomials. Then we can go from arbitrary f to the union over a countable sequence of f and then argue through separability that this is equal to the entire space. □

Lemma 3.5.1 (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 $X \in L^p$

Proof. □

Remark. In general a sequence $(X_i)_{i\in\mathbb{N}}$ is called uniform integrability if

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

Lemma 3.5.2 (De la Vallèe Poussin Criterion)**.** A sequence of random variables (X_i) is uniformly integrable iff there $\exists\varphi$ convex with

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = \infty.$$

s.t.

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

Proof. As the construction of φ is heavily technical we refer to xyz □

Corollary. If $(X_i)_{i\in\{1,\dots,N\}}$ are i.i.d random variables with law μ_X and $\int |x|^p\mu < \infty$ i.e $\mu \in \mathcal{P}^p(\mathbb{R}^d)$

$$W_p(\mu_N, \mu) \to 0 \qquad \text{a.s..}$$

and

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

Where

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

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 $\lim_{n\to\infty} \sup_r \int_{|x|\geq r} |x|^p d\mu_N = 0$

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

$$\mathbb{E}[W^p(\mu_n, \mu)] \to 0.$$

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 $\lim_{x\to\infty}\frac{\varphi(x)}{x} = \infty$ and

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

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

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

Finally Lemma 3.5.2 implies the uniform integrability and we conclude by Lemma 3.5.1

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

□

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.

Remark. Formally our goal is to prove the convergence

$$\mathbb{E}[\sup_t W_2^2(\mu_N(t), \mu(t))] \to 0.$$

3.5.3 Toy Example

Let us first consider a simple stochastic particle system given by

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 Definition 3.5.3 with drift and diffusion for $\nabla V \in \text{Lip}$

$$\begin{aligned} b(X,\mu) &= \nabla V \star \mu(X) \\ \sigma(X,\mu) &= \sigma_0 > 0. \end{aligned}$$

Exercise. Think about what happens if the initial data is i.i.d but the diffusion term is 0, can you prove a convergence ?

Theorem 3.5.1 (Convergence Of Toy Model For Fixed N)**.** Let our (SDEN) be given with drift and diffusion as above and assume they fulfill Assumption E, then for fixed N we get a unique strong solution in $\mathbb{L}^2_{dN}([0,T])$

Proof. First we note that by Assumption E we get

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

Let use the notation $\mathbb{X} = (X_1^N,\dots,X_N^N) \in \mathbb{R}^{dn}$ and $\mathbb{W} = (W^1,\dots,W^N)$ then

$$\begin{aligned} B(\mathbb{X}) &= \begin{pmatrix} \vdots \\ b(X_i^N, \frac{1}{N}\sum_{k=1}^N \delta_{X_k}) \end{pmatrix}_{dN} \\ \Sigma(\mathbb{X})_{dN\times mN} &: \ \text{diag}(\Sigma(\mathbb{X})) = \left(\delta(X_1, \tfrac{1}{N}\textstyle\sum_{k=1}^N \delta_{X_k}),\dots\delta(X_N, \tfrac{1}{N}\textstyle\sum_{k=1}^N \delta_{X_k})\right). \end{aligned}$$

Then our SDE is given by

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

Now if B and Σ satisfy Assumption D we get a solution by Theorem 3.4.1

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

For Σ the argument is analog where for the Wasserstein distance we used Then by Theorem 3.4.1 we get a solution $X \in L^2([0,T])$ for fixed N

□

Remark. To get a bound on the Wasserstein Distance we used the following

$$\pi = \frac{1}{N}\sum_{k=1}^{N}\delta_{(X_k,Y_k)} \in \Pi.$$

then the Wasserstein distance is given by

$$\frac{1}{N}\sum_{k=1}^{N}|X_k - Y_k|^2.$$

and one can further simplify to get the bound used.

Remark. As $N \to \infty$ we expect to get the following

$$\begin{cases} dY^i(t) & = b(Y^i(t), \mu(t))dt + \sigma(Y^i(t), \mu(t))dW_t^i \\ Y^i(0) & = X_{i,0}^N \in L^2(\Omega) \text{ i.i.d} \end{cases}.$$

In fact since the above system beyond the initial data is independent of N, we may consider the simplified equation

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

this equation is called Makean-Vlasov equation which is a non-linear non-local SDE

3.5.4 Makean-Vlasov

Definition 3.5.4 (Makean-Vlasov Equation)**.** The following non-linear and non-local SDE is called Makean-Vlasov Equation

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

Add Space of Y and dimensions

Definition 3.5.5 (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 $\mathcal{C}^d$

Definition 3.5.6 (Random Variable)**.** A random Variable on $\mathcal{C}^d$ is a map

$$X : \Omega_{\text{a.s.}} \to \mathcal{C}^d.$$

Definition 3.5.7 (Measure)**.** Since the norm $\|\cdot\|_T$ induces a σ-algebra on $\mathcal{C}^d$ we can define measures $\mu \in \mathcal{P}^2(\mathcal{C}^d)$ by

$$\mu := (\mu(t))_{t\in[0,T]} \qquad \mu(t).$$

and by using the function

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

then we get a measure on $\mathbb{R}^d$ by using the pushforward

$$\mu_t := \mathcal{B} \to \mathbb{R}^d \; A \mapsto \mu(I_t^{-1}(A)).$$

Definition 3.5.8 (Wasserstein Distance)**.** And we can define for arbitrary measures $\mu, \tilde{\mu} \in \mathcal{P}^2(\mathcal{C}^d)$ the Wasserstein distance by

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

Where

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

Corollary. Let us prove the inequality

We choose concrete $\pi_t = I_t \# \pi$ for $\pi \in \Pi_{\mathcal{C}^d}(\mu, \tilde{\mu})$

Proof.

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

It remains to check that $p_t \in \Pi(\mu_t,\tilde{\mu}_t)$, let $A\in\mathcal{B}(\mathbb{R}^d)$

$$
\begin{aligned}
\pi_t(A\times\mathbb{R}^d) &= \pi(I_t^{-1}(A\times\mathbb{R}^d)) \\
&= \pi(\{(x,y)\in\mathcal{C}^d\times\mathcal{C}^d \;:\; I_t(x)\in A, I_t(y)\in\mathbb{R}^d\}) \\
&= \pi(\{(x,y)\in\mathcal{C}^d\times\mathcal{C}^d \;:\; I_t(x)\in A, y\in\mathcal{C}^d\}) \\
&= \pi(I_t^{-1}(A)\times\mathcal{C}^d) \\
&= \mu(I_t^{-1}(A)) \\
&= I_t\#\mu(A) \\
&= \mu_t(A).
\end{aligned}
$$

□

Remark. Note that

$$
\int_{\mathcal{C}^d} f(x)d\mu(x) = \int_{\mathbb{R}^d} f(x(t))d\mu_t.
$$

Theorem 3.5.2 (Unique and Existence of Solution for Makean-Vlasov)**.** If b and σ satisfy Assumption E then MVE has a unique and strong solution $Y\in\mathbb{L}^2([0,T])$ and $\mu\in\mathcal{L}(Y)$

Proof. We use the notation

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

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

$$
|b(x,\mu(t))-b(\tilde{x},\mu(t))| + |\sigma(x,\mu(t))-\sigma(\tilde{x},\mu(t))| \le L|x-\tilde{x}|.
$$

Since it is the same measure the Wasserstein distance is 0 and the above is true by Assumption E.

If φ has a fixpoint $\overline{\mu}$, then $\overline{\mu}$ is the solution of MVE. We prove this by first bounding the difference between two measures, let $\mu, \tilde{\mu}$ be arbitrary given measure i n $\mathcal{P}^2(\mathcal{C}^d)$, first note

$$Y^\mu(t) - \xi = \int_0^t b(Y^\mu(s), \mu(s))ds + \int_0^t \sigma(Y^\mu(s), \mu(s))dW_s \qquad \mu = \mu, \tilde{\mu}.$$

then by taking the difference

$$\begin{aligned}
&\sup_{0\le t\le \tau} |Y^\mu(t) - Y^{\tilde{\mu}}(t)|^2 \\
&= \sup_{0\le t\le s} \left| \int_0^t b(Y^\mu(s), \mu(s)) - b(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))ds + \int_0^t \sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))dW_s \right|^2 \\
&\le \sup_{0\le t\le \tau} 2t \int_0^t |b(Y^\mu(s), \mu(s)) - b(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))|^2 ds \\
&+ \sup_{0\le t\le \tau} 2 \left| \int_0^t \sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))dW_s \right|^2
\end{aligned}$$

.

Now taking the expectation

$$\begin{aligned}
&\mathbb{E}[\sup_{0\le t\le \tau} |Y^\mu(t) - Y^{\tilde{\mu}}(t)|^2] \\
&\le 4\tau L^2 \mathbb{E}\left[\int_0^\tau |Y^\mu(s) - Y^{\tilde{\mu}}(s)|^2 + W_2^2(\mu(s), \tilde{\mu}(s))ds\right] \\
&+ 16L^2 \mathbb{E}[\int_0^\tau |Y^\mu(s) - Y^{\tilde{\mu}}(s)|^2 + W_2^2(\mu(s), \tilde{\mu}(s))ds].
\end{aligned}$$

Where we used Doobs-L^p inequality for the second term.

$$\begin{aligned}
&\mathbb{E}[\sup_{0\le t\le \tau} \left| \int_0^t \sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))dW_s \right|^2] \\
&\le 8\mathbb{E}[\int_0^\tau |\sigma(Y^\mu(s), \mu(s)) - \sigma(Y^{\tilde{\mu}}(s), \tilde{\mu}(s))|^2 ds] \\
&\le 8\mathbb{E}[\int_0^\tau |Y^\mu(s) - Y^{\tilde{\mu}}(s)|^2 + W_2^2(\mu(s), \tilde{\mu}(s))ds].
\end{aligned}$$

All together

$$\mathbb{E}[\|Y^\mu - Y^{\tilde{\mu}}\|_\tau^2] \le C\int_0^\tau \mathbb{E}[\|Y^\mu - Y^{\tilde{\mu}}\|_s^2]ds + C\int_0^\tau \mathbb{E}[W_2^2(\mu(s), \tilde{\mu}(s))]ds$$

.

So by Grönwall inequality we get

$$\begin{aligned}
\mathbb{E}[\|Y^\mu - Y^{\tilde{\mu}}\|_\tau^2] &\le C(\tau) \cdot \int_0^\tau W_2^2(\mu(s), \tilde{\mu}(s))ds \\
&\le C(\tau) \cdot \int_0^\tau \sup_{0\le t\le s} W_2^2(\mu(t), \tilde{\mu}(t))ds \\
&\le C(\tau) \int_0^\tau d_s(\mu, \tilde{\mu})ds.
\end{aligned}$$

using the inequality Definition 3.5.8

remember that $\varphi(\mu) = \mathcal{L}(Y^\mu)$ and $\varphi(\tilde{\mu}) = \mathcal{L}(Y^{\tilde{\mu}})$, then

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

now if we take joint distribution of Y^μ and $Y^{\tilde{\mu}}$. π_1 we can write

$$\begin{aligned}\mathbb{E}[\|Y^\mu - Y^{\tilde{\mu}}\|_\tau^2] &= \int_{\mathcal{C}^d,\mathcal{C}^d} \|x-y\|_\tau^2 d\pi_1(x,y) \\ &\leq C(\tau)\int_0^\tau d_s(\mu,\tilde{\mu})ds.\end{aligned}$$

Lets summarize, for $\forall \mu, \tilde{\mu}\mathcal{P}^2(\mathcal{C}^d)$ we obtained

$$d_t(\varphi(\mu),\varphi(\tilde{\mu})) \leq C(t)\int_0^t d_s(\mu,\tilde{\mu})ds. \tag{*}$$

To prove the uniqueness of solutions. If we have two solutions $\mu, \tilde{\mu}$ i.e.

$$\begin{aligned}\varphi(\mu) &= \mu \\ \varphi(\tilde{\mu}) &= \tilde{\mu}.\end{aligned}$$

then the above estimate (*) says

$$d(\mu,\tilde{\mu}) \leq C(t)\int_0^t ds(\mu,\tilde{\mu})ds \Rightarrow d_t(\mu,\tilde{\mu}) = 0.$$

To prove the existence. Take arbitrary $\mu_0 \in \mathcal{P}^2(\mathcal{C}^d)$, (for example $\mu_0 = \mathcal{L}(\xi)$)

$$\begin{aligned}\varphi(\mu_0) &= \mu_1 \\ \varphi(\mu_1) &= \mu_2 \\ &\vdots \\ \varphi(\mu_k) &= \mu_{k+1}.\end{aligned}$$

the estimate means that (μ_k) is Cauchy in $\mathcal{P}^2(\mathcal{C}^d)$

$$d_t(\mu_{k+m},\mu_m) \leq \sum \ldots.$$

Then there exists a $\mu \in \mathcal{P}^2(\mathcal{C}^d)$ such that

$$W_2^2(\mu_k,\mu) \to 0.$$

□

Remark. That in our case the empirical measure μ_N is not exactly the law of X^N and is stochastic, such that the above proof does not exactly holds for our (SDEN)
For our initial data we already know that

$$\mathbb{E}[W_2^2(\mu_N(0),\mu_0)] \xrightarrow{N\to\infty} 0.$$

and we expect for any $t > 0$

$$\mathbb{E}[W_{\mathcal{C}^d,2}^2(\mu_N(t),\mu)] \to 0.$$

Theorem 3.5.3 (Mean-Field-Limit)**.** Let b and σ fulfill Assumption E and use μ_N the empirical measure, then there exists a measure $\mu \in \mathcal{P}^2(\mathcal{C}^d)$ s.t.

$$\lim_{N\to\infty} \mathbb{E}]W_{\mathcal{C}^d,2}^2(\mu_N,\mu) = 0.$$

and for any fixed $k \in \mathbb{N}$ it holds

$$(X_1^N, \dots, X_k^N) \xrightarrow{(D)} (Y_1, \dots, Y_k)\,.$$

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

$$\begin{aligned}|X_i^N(t) - Y_i(t)|^2 \le 2t \int_0^t &|b(X_i^N(s), \mu_N(s)) - b(Y_i(s), \mu(s))| \\ &+ 2\left|\int_0^t \sigma(X_i^N(s), \mu_N(s)) - \sigma(Y_i(s), \mu(s)) dW_s^i\right|^2.\end{aligned}$$

We get

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N \mathbb{E}[\sup_{0\le r\le t} |X_i^N(r) - Y_i(r)|^2] &\le C\mathbb{E}\int_0^t W_2^2(\mu_N(s), \mu(s))ds \\ &\le C \cdot \mathbb{E}[\int_0^t d_r^2(\mu_N, \mu)dr].\end{aligned}$$

Let $\overline{\mu}_N$ be the empirical measure of Y_i

$$\overline{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{Y_i}.$$

And let $\mu \sim \mathcal{L}(Y_i)$ for $\forall t > 0$ then

$$\mathbb{E}[W_2^2(\overline{\mu}_N, \mu)] \to 0.$$

Now we consider for $\forall$ a.s. $\omega \in \Omega$

$$d_t^2(\overline{\mu}_N, \mu_N) = \inf_{\pi\in\Pi(\mu_N,\overline{\mu}_N)} \int_{\mathcal{C}^d\times\mathcal{C}^d} \|x-y\|_t^2 d\pi(x,y).$$

By taking $\pi = \mu_N \otimes \overline{\mu}_N$ we can write the above integral explicitly

$$\le \frac{1}{N}\sum_{i=1}^N \|X_i^N - Y_i\|_t^2.$$

We continue by taking the expectation

$$\begin{aligned}\mathbb{E}[d_t^2(\mu_N, \overline{\mu}_N)] &\le \frac{1}{N}\sum_{i=1}^N \mathbb{E}[\sup_{0\le s\le t} \|X_i(s)^N - Y_i(s)\|_t^2] \\ &\le 2C\int_0^t \mathbb{E}[d_r^2(\mu_N, \mu)]dr.\end{aligned}$$

Goal is to get a Grönwall inequality for

$$\begin{aligned}\mathbb{E}[d_t^2(\mu_N, \mu)] &\le 2\mathbb{E}[d_t^2(\mu_N, \overline{\mu}_N)] + 2\mathbb{E}[d_t^2(\mu_N, \mu)] \\ &\le C\int_0^t \mathbb{E}[d_r^2(\mu_N, \mu)]dr + C\mathbb{E}[d_t^2(\overline{\mu}_N, \mu)]\end{aligned}$$

.

Then by Grönwall

$$\mathbb{E}[d_t^2(\mu_N, \mu)] \leq e^{CT}\mathbb{E}[\mu_{N,0}] + e^{CT}\mathbb{E}[d_t^2(\overline{\mu}_N, \mu)] \xrightarrow{N\to\infty} 0.$$

and then for $\forall 1 \leq k < \infty$.

$$\begin{aligned}\mathbb{E}[\max_{1\leq i\leq k} \sup_{0\leq r\leq t} \|X_i^N(r) - Y_i(r)\|^2] &\leq \max_{1\leq i\leq k} \frac{1}{N}\sum_{i=1}^{k} \mathbb{E}[\|X_i^N - Y_i\|_t^2] \\ &\leq C \cdot k\mathbb{E}[d_t^2(\mu_N, \mu)] \\ &\xrightarrow{N\to\infty} 0.\end{aligned}$$

This concludes the proof. Add small summary □

Chapter 4

PDE Approach To Solving the Makean-Vlasov Equation

This entire chapter needs alot of reworking, formulating Theorems etc. so there is more structure

4.1 Motivation

Above we saw an SDE approach to solving the Makean-Vlasov Equation, in this section we instead focus on a PDE based approach. From now on we assume $\sigma(Y(t), \mu(t)) = \sqrt{2}$ is a constant, then the (MVE) can be rewritten as

$$(\text{MVE*}) \begin{cases} Y(t) & = b(Y(t), \mu(t))dt + \sqrt{2}dW_t \\ Y(0) & = \xi \in L^2(\Omega) \\ \mu_0 & = \mathcal{L}(\xi) \end{cases}.$$

by applying Itôs formula for $\forall \varphi \in \mathcal{C}_0^\infty([0, T) \times \mathbb{R}^d)$

$$\begin{aligned} \varphi(Y(t), t) - \varphi(Y(0), 0) = & \int_0^t \frac{\partial \varphi}{\partial t}(Y(s), s) + \nabla \varphi(Y(s), s) \cdot b(Y(s), \mu(s)) \\ & + \frac{1}{2} \underbrace{\sqrt{2} \cdot \sqrt{2})}_{tr(\sigma \cdot \sigma^T)} \cdot \Delta \varphi(Y(s), s) ds \\ & + \int_0^t \nabla \varphi(Y(s), s) \sqrt{2} dW_s. \end{aligned}$$

and taking the expectation on both sides, such that the last term disappears

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi(x, t) d\mu(t) - \int_{\mathbb{R}^d} \varphi(x, 0) d\mu_0 \\ & = \int_0^t \int_{\mathbb{R}^d} \frac{\partial \varphi}{\partial t}(x, s) + \nabla \varphi(x, s) \cdot b(x, \mu(s)) \cdot \Delta \varphi(x, s) d\mu(s) ds. \end{aligned}$$

This leads us to formulating the following weak PDE, if μ is regular enough i.e it has density and the density has enough regularity, then μ should satisfy

$$\begin{cases} \partial_t \mu - \Delta \mu + \nabla \cdot (b(x, \mu) \cdot \mu) = 0 \\ \mu(0) = \mu_0 \end{cases}.$$

> **Remark.** Compare this weak PDE to the one we got in the discrete case, what do you notice ?

Exercise. Show that the integral equation and the weak formulation are equal.

Remark. Now suppose we find μ with density u satisfying the weak PDE, then we can plug it in to the (MVE) equation to get

$$\begin{cases} dY_t = b(Y_t, u)dt + \sqrt{2}dW_t \\ Y(t) = \xi \in L^2(\Omega) \quad \mathcal{L}(\xi) = u \end{cases}.$$

Now if b is bounded and Lipschitz continuous, then we get a solution Y_t. Now if $\overline{u}$ is the Law of Y_t. Then by Itô formula we have for $\forall \varphi \in \mathcal{C}_0^\infty$

$$\int_{\mathbb{R}^d} \varphi(x,t) d\overline{\mu(t)} - \int_{\mathbb{R}^d} \varphi(x,0) u_0(x) dx$$
$$= \int_0^t \int_{\mathbb{R}^d} \left(\frac{\partial \varphi}{\partial t}(x,s) + \nabla\varphi(x,s) \cdot b(x,u) - \Delta\varphi(x,s) \right) \overline{u}(x,t) dx ds.$$

Which means $\overline{\mu}$ satisfies

$$\begin{cases} \partial_t \overline{\mu} - \Delta\overline{\mu} + \nabla \cdot (b(x,u) \cdot \overline{\mu}) = 0 \\ \overline{\mu}|_{t=0} = u_0 \end{cases}.$$

If we can prove $\overline{u} = u$, then we get a solution to the Makean-Vlasov Equation.

Example. A common choice of b is the following for some kernel K

$$b(Y_t, u) = \int K(Y_t - y) u(y) dy = \int K(y) u(Y_t - y) dy.$$

then the regularity of b by convolution depends on either K or u

4.2 Problem Definition

Definition 4.2.1 (Weak PDE)**.** Let μ have density u, then we write

$$\text{(PDE)} \begin{cases} \partial_t u - \Delta u + \nabla \cdot (b(x,u) \cdot u) = 0 \\ u(0) = u_0 \end{cases}.$$

Formalize by adding the relevant spaces

Definition 4.2.2 (Sobolev Spaces)**.** We define roughly

$$H^1(\mathbb{R}^d) = \{u \in L^2(\mathbb{R}^d) : \nabla u \in L^2(\mathbb{R}^d)\}$$
$$\|u\|_{H_1} = \|u\|_2 + \|\nabla u\|_2.$$

where the gradient is defined for $\forall \varphi \in \mathcal{C}_0^\infty$

$$\nabla u = \langle \nabla u, \varphi \rangle = -\langle u, \nabla \varphi \rangle.$$

And the dual space

$$H^{-1}(\mathbb{R}^d) = (H^1(\mathbb{R}))' = \{l : l \text{ is bounded linear functional of } H^1(\mathbb{R}^d)\}.$$

Then

$$L^2([0,T];H^1(\mathbb{R}^d)) = \{u : \int_0^T \|u(t)\|_{H^1} dt < \infty\}.$$

Remark. The Sobolev space H^1 is a separable Hilbert space

Definition 4.2.3 (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 (PDE) if for $\forall \varphi \in \mathcal{C}_0^\infty([0,T]\times\mathbb{R}^d)$ it holds

$$\begin{aligned}\int_0^T \langle \partial_t u, \varphi\rangle_{(H^{-1},H^1)} dt = &\int_0^T \int_{\mathbb{R}^d} \nabla\varphi \cdot (b(x,u)\cdot u) dx dt \\ &- \int_0^T \int_{\mathbb{R}^d} \nabla u \cdot \nabla\varphi dx dt.\end{aligned}$$

4.3 Heat Equation and the Heat Kernel

4.3.1 Motivation

Definition 4.3.1 (Heat equation)**.** The following PDE is called the inhomogenes Heat equation with source term f

$$\text{(HE)} \begin{cases} \partial_t u(x,t) - \Delta u(x,t) & = f(x,t) \\ u|_{t=0} & = u_0 \end{cases}.$$

Remark. Compare this to our PDE which looks similar, but is in fact non-linear

$$\partial_t u - \Delta u + \nabla \cdot (b(x,u)\cdot u) = 0.$$

Remark. Let us suppose $K(x,t)$ is a heat kernel, then

$$\begin{aligned}u(x,t) &= \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy - \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s)\nabla\cdot(b(y,u(y,s)u(y,s))dyds \\ &= u_1(x,t) + u_2(x,t).\end{aligned}$$

is a solution to the inhomogenous Heat-Equation, this is called Duhamel's principle i.e. we can "add" up solutions to homogeneous problems and get the solution to the inhomogenous.

Remark. We say the heat kernel is the density of the Brownian Motion.

4.3.2 Derivation by Fourier Transform

Definition 4.3.2 (Fourier Transform)**.** For $x \in \mathbb{R}^d$ the Fourier transform is defined as

$$\mathcal{F} : L^2 \to L^2 \ u \mapsto \hat{u}.$$

where

$$\hat{u}(k) = \int_{\mathbb{R}^d} u(x) e^{ix\cdot k} dx.$$

Exercise. Proof

$$-\widehat{\Delta u} = |k|^2 \hat{u}(k).$$

Hint

$$\widehat{\nabla u} = \frac{k}{i} \hat{u}(k).$$

Remark. Using the Fourier transformation we can transform our PDE into an ODE

$$\begin{cases} \partial_t \hat{u} - \widehat{\Delta u} & = \hat{f} \\ \hat{u}|_{t=0} & = \hat{u}_0 \end{cases}.$$

that is

$$\begin{cases} \partial_t \hat{u}(k) + |k|^2 \hat{u}(k) = \hat{f}(k) \\ \hat{u}_0(k) = \hat{u}_0 \end{cases}.$$

where

$$\hat{u}(k,t) = e^{-|k|^2 t} \hat{u}_0(k) + \int_0^t e^{-|k|^2(t-\tau)} \hat{f}(k,\tau) d\tau.$$

Lemma 4.3.1 (Inverse transformation of the Fourier transformation)**.**

$$u(x,t) = \frac{1}{(4\pi t)^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-\frac{|x-y|^2}{4t}} u_0(y) dy + \int_0^t \int_{\mathbb{R}^d} \frac{1}{(4\pi(t-\tau))^{\frac{d}{2}}} e^{\frac{-|x-y|^2}{4(t-\tau)}} f(y,\tau) dy d\tau.$$

Definition 4.3.3 (Heat Kernel)**.** The following is called Heat Kernel

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

and for $\forall t > 0$ it is a solution to the homogeneous heat equation

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

And

$$K \xrightarrow{t \to 0^+} \delta.$$

In the sense of distributions.

Theorem 4.3.1 (Solution To Heat Equation)**.** Let K be the Heat kernel and 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 (schwarz function would work as well, since they lie dense in compact)

$$\begin{aligned}u(x,t) &= \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy - \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s)\nabla\cdot(b(y,u(y,s)u(y,s))dyds \\ &= u_1(x,t) + u_2(x,t).\end{aligned}$$

is a solution to the heat equation, in fact u_1 and u_2 are solutions to

$$(\text{P1})\begin{cases}\partial_t u_1 - \Delta u_1 = 0 \\ u_1(0) = u_0\end{cases} \qquad (\text{P2})\begin{cases}\partial_t u_2 - \Delta u_1 = f \\ u_2(0) = u_0\end{cases}.$$

respectively

Proof. We begin by showing that u_1 is a solution to (P1) by showing

$$\lim_{t\to 0^+} u_1(x,t) = \lim_{t\to 0^+} \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy \overset{!}{=} u_0.$$

$$\begin{aligned}\lim_{t\to 0^+} \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy &= \lim_{t\to 0^+} \int_{\mathbb{R}^d} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{4t}} u_0(y)dy \\ &= \lim_{t\to 0^+} \int_{\mathbb{R}^d} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-|z|^2} u_0(x+2\sqrt{t}z)dz \\ &= \int_{\mathbb{R}^d} \lim_{t\to 0^+} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-|z|^2} u_0(x+2\sqrt{t}z)dz \\ &= u_0(x).\end{aligned}$$

where we used the change of variables

$$\frac{x-y}{2\sqrt{t}} = -z.$$

Then for $\forall t > 0$

$$\begin{aligned}\partial_t u_1 - \Delta u_1 &= (\partial_t - \Delta)\int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy \\ &= \int_{\mathbb{R}^d} (\partial_t - \Delta)K(x-y,t)u_0(y)dy \\ &= 0.\end{aligned}$$

by properties of the Heat-Kernel.

For $u_2(x,t)$, we further assume f has compact support

$$u_2(x,t) = \int_0^t \int_{\mathbb{R}^d} K(y,s)f(x-y,t-s)dyds.$$

First note that

$$\lim_{t\to 0^+} u_2(x,t) = 0.$$

Then by applying

$$\begin{aligned}(\partial_t - \Delta)u_2 &= \int_0^t \int_{\mathbb{R}^d} K(y,s)(\partial_t - \Delta_x) f(x-y,t-s)\,dyds \\ &+ \int_{\mathbb{R}^d} K(y,t) f(x-y,0)\,dy \\ &= \int_0^\varepsilon \int_{\mathbb{R}^d} K(y,s)\left(-\partial_s - \Delta_y\right) f(x-y,t-s)\,dyds + \int_\varepsilon^t \int_{\mathbb{R}^d} \dots \\ &+ \int_{\mathbb{R}^d} K(y,t) f(x-y,0)\,dy \\ &= I_\varepsilon + J_\varepsilon + L.\end{aligned}$$

We are allowed to exchange the order because of the Heat-Kernel since it decays very fast this gives uniform integrability the further away we get from the origin
We have, since $f \in \mathcal{C}_b^{2,1}$ we get that its term is bounded and the integral for the Heat kernel is just 1

$$\begin{aligned}|I_\varepsilon| &\leq C \cdot \varepsilon \\ J_\varepsilon &= \int_\varepsilon^t \int_{\mathbb{R}^d} K(y,s)(-\partial_t - \Delta_y) f(x-y,t-s)\,dyds \\ &= \int_\varepsilon^t \int_{\mathbb{R}^d} \underbrace{(-\partial_t - \Delta_y)K(y,s)}_{=0} f(x-y,t-s)\,dyds \\ &+ \int_{\mathbb{R}^d} K(y,\varepsilon) - f(x-y,t-\varepsilon)\,dy \\ &\underbrace{- \int_{\mathbb{R}^d} K(y,t) f(x-y,0)\,dy}_{=Li}.\end{aligned}$$

□

Together we have

$$\begin{aligned}\partial_t u_2 - \Delta u_2 &= \lim_{\varepsilon \to 0} \left(\int_{\mathbb{R}^d} \underbrace{K(y,\varepsilon)}_{\to \delta} f(x-y,t-\varepsilon)\,dy + \underbrace{C\varepsilon}_{\to 0} \right) \\ &= f(x,t).\end{aligned}$$

This shows that

$$u(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)\,dy - \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s)\nabla \cdot (b(y,u(y,s)u(y,s))\,dyds$$
.

is a solution to the inhomogenous heat equation.

Remark. When changing order or variable of derivative one has to be aware of the boundary terms appearing. For example when changing the order for the ∂_t derivative one gets two boundary terms , where one is not well behaved $t = 0$

$$\begin{cases} \partial_t u - \Delta u + \nabla \cdot (b(x,u) \cdot u) = 0 \\ u|_{t=0} = u_0 \in L^1 \quad \int (1+|x|^2) u_0 < \infty \end{cases}.$$

Formally

$$u(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)\,dy - \int_0^t \int_{\mathbb{R}^d} K(x-y,t-\tau)\nabla \cdot (b(y,u(y,\tau)) \cdot u(y,\tau))\,dyd\tau.$$

Now we start with bounded drift term for a linear equation.

Definition 4.3.4 (LDE). For bounded drift term $\bar{b} \in L^\infty$ we define

$$\text{(LDE)} \begin{cases} \partial_t - \Delta u + \nabla \cdot (\bar{b}(x,t)u) = 0 \\ u|_{t=0} = u_0 \end{cases}.$$

Remark. By first proving the existence of a solution to this simpler equation we can then construct an iteration that will yield a solution to the more complex non-linear , non-local one.

Theorem 4.3.2 (Uniqueness and Existence of LDE Solution). If $b \in L^\infty([0,T] \times \mathbb{R}^d)$ and $u_0 \in L^1(\mathbb{R}^d)$, then the (LDE) has a unique solution $u \in L^\infty([0,T]; L^1(\mathbb{R}^d))$

$$u(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy + \int_0^t \int_{\mathbb{R}^d} \nabla K(x-y,t-\tau) \cdot (\bar{b}(y,\tau)u(y,\tau))dyd\tau.$$

Proof. We prove again by Iteration, and fix point argument, consider a map

$$\begin{aligned} &\mathcal{T} : L^\infty([0,T]; L^1(\mathbb{R}^d)) \to L^\infty([0,T]; L^1(\mathbb{R}^d)) \\ &u \mapsto \mathcal{T}(u) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy + \int_0^t \int_{\mathbb{R}^d} \nabla K(x-y,t-\tau) \cdot (\bar{b}(y,\tau)u(y,\tau))dyd\tau. \end{aligned}$$

We need to check $\mathcal{T}(u) \in L^\infty([0,T]; L^1(\mathbb{R}^d))$, for $\forall t > 0$

$$\begin{aligned} \int_{\mathbb{R}^d} |\mathcal{T}(u)(x,t)|dx \leq &\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(x-y,t)|u_0(y)|dydx \\ &+ \int_0^t d\tau \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} dy |\nabla K(x-y,t-\tau)\bar{b}(y,\tau)u(y,\tau)| \\ &= I + II. \end{aligned}$$

Since we have fixed $t > 0$ we use Fubini

$$\begin{aligned} I &\leq \int_{\mathbb{R}^d} \underbrace{\int_{\mathbb{R}^d} K(x-y,t)dx}_{=1} |u_0(y)|dy \\ &\leq \|u_0\|_{L^1(\mathbb{R}^d)}. \end{aligned}$$

First consider the gradient of K

$$\begin{aligned} \int_{\mathbb{R}^d} \nabla K(x,s)dx &= \frac{1}{(4\pi s)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \frac{1}{\sqrt{s}} |\frac{4}{2\sqrt{s}}| e^{-\frac{|x|^2}{4s}} dx \\ &\leq \frac{1}{\sqrt{s}} C. \end{aligned}$$

Then for the second term we get

$$\begin{aligned} II &\leq \|\bar{b}\|_{L^\infty} \int_0^t d\tau \int_{\mathbb{R}^d} |\nabla K(x,t-\tau)|dx \int_{\mathbb{R}^d} u(y,\tau)dy \\ &\leq \|\bar{b}\|_{L^\infty} \|u\|_{L^\infty(L^1)} C \cdot \int_0^t \frac{1}{\sqrt{s}} dy \\ &= C \cdot \sqrt{t-\tau}. \end{aligned}$$

Note we 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 our map $\mathcal{T}$ is indeed well defined, next we proof $\mathcal{T}(u)$ is a contraction, for

$\forall u_1, u_2 \in L^\infty(L^1)$, for $t^\star$ s.t. $C\|\bar{b}\|_\infty\sqrt{t^\star} < \frac{1}{2}$

$$\begin{aligned}
&\|\mathcal{T}(u_1) - \mathcal{T}(u_2)\|_{L^\infty(L^1)} \\
&= \operatorname*{ess\,sup}_{0\le t\le t^\star} \int_{\mathbb{R}^d} |\mathcal{T}(u_1) - \mathcal{T}(u_2)|(x,t)dx \\
&\le \operatorname*{ess\,sup}_{0\le t\le t^\star} \int_0^t d\tau \int_{\mathbb{R}^d} dy |\nabla K(x-y, t-\tau)(\bar{b}(y,\tau)(u_1-u_2))(y,\tau)| \\
&\le \operatorname*{ess\,sup}_{0\le t\le t^\star} \|\bar{b}\|_{L^\infty} \|u_1 - u_2\|_{L^\infty(L^1)} \int_0^t \frac{1}{\sqrt{t-\tau}} d\tau \\
&\le \operatorname*{ess\,sup}_{0\le t\le t^\star} C\|\bar{b}\|_{L^\infty}\sqrt{t}\|u_1 - u_2\|_{L^\infty(L^1)} \\
&\le C\|\bar{b}\|_\infty\sqrt{t^\star}\|u_1 - u_2\|_{L^\infty(L^1)}.
\end{aligned}$$

Then $\mathcal{T}$ is a contraction. Since $t^\star$ only depends on $\|\bar{b}\|_\infty$ and dimension d. Then for any given $T > 0$ we can repeat the above argument finite many time and obtain

$$u \in L^\infty([0,T]; L^1(\mathbb{R}^d)).$$

□

Exercise. Think about wether you can proof it for b satisfying linear growth condition

$$|\bar{b}| \le C(1 + |x|).$$

Let us discuss how a solution to the LDE leads back to a solution to the more complex

$$\partial_t u - \Delta u + \nabla \cdot (b(x,u)u) = 0.$$

where $u \in L^\infty(L^1)$ under the assumption on $b(x,u)$

$$|b(x,u) - b(\tilde{x},\tilde{u})| \le L(|x - \tilde{x}| + W_2(u,\tilde{u})).$$

By fixing $\tilde{x} = 0$

$$|b(x,u) - b(0,\delta_0)| \le L(|x| + W_2(u,\delta_0)) \le L(1 + |x|).$$

where

$$W_2(u,\delta_0) \le \left(\int_{\mathbb{R}^d} |x|^2 u(x) dx\right)^{\frac{1}{2}}.$$

when b is unbounded we consider , the cutoff

$$\bar{b}(x, v(x,t)) = \min\{b(x, v(x,t)), M\}.$$

Assumption F. Assume $b(x,u)$ is given by the convolution

$$b(x,u) = \nabla V \star u(x).$$

Remark. In the (SDE) case we would assume

$$\nabla V \in \mathrm{Lip}. \tag{SDE}$$

since our (SDE) is given by

$$dX_i = \nabla V \star u(x_i)dt + \sqrt{2}dW_t.$$

and we know a solution to the linear SDE exists for Lipschitz continuous coefficients

Definition 4.3.5. Assume $b(x,u)$ is given by the convolution

$$b(x,u) = \nabla V \star u(x).$$

Then let our (PDE) be given by

$$(\text{PDE})\begin{cases} u_t - \Delta u + \nabla \cdot (\nabla V \star u\ u) = v \\ u|_{t=0} = u_0 \in L^1((1+|x|^2)dx) \end{cases}.$$

Definition 4.3.6 (Epsilon Problem)**.** For $v \in L^\infty(L^1)$ consider $j_\varepsilon \star \nabla V \star v \leftarrow \nabla V \star v$

$$(\text{PDE})_\varepsilon \begin{cases} u^\varepsilon_t - \Delta u^\varepsilon + \nabla \cdot (j_\varepsilon \star \nabla V \star v\ (j_\varepsilon \star \mathbb{1}_{|x|\le \frac{1}{\varepsilon}}\ u^\varepsilon)) = v \\ u^\varepsilon|_{t=0} = \tilde{j}_\varepsilon \star u_0 \end{cases}.$$

where j_ε is the mollification kernel in x, t

$$\begin{aligned} \nabla V \star u &\leftarrow j_\varepsilon \star \nabla V \star u \\ u &\leftarrow j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}u}) \\ u_0 &\leftarrow \tilde{j}_\varepsilon \star u_0. \end{aligned}$$

where $\tilde{j}_\varepsilon$ is mollification just in x

Remark. We first solve this problem for v, and then construct a fixpoint argument to get a solution for u, in order to solve PDE , for bounded $\overline{b}$ we already have a solution

$$v \in L^\infty(L^1) \xrightarrow{\text{sol for bounded } \overline{b}} u \in L^\infty(L^1).$$

Where we know our solution is given by

$$u(x,t) = \int_{\mathbb{R}^d} K(x-y,t)u_0(y)dy + \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s)\nabla \cdot (\nabla V \star v\ u)(s,y)dyds.$$

We further want that

$$\lim_{|x|\to\infty} u(x,t) = 0.$$

Remark. For the linear PDE with term

$$j_\varepsilon \star \nabla V \star v\ (j_\varepsilon \star \mathbb{1}_{|x|\le\frac{1}{\varepsilon}}\ u^\varepsilon).$$

We get

$$u^\varepsilon(x,t) = \int_{\mathbb{R}^d} K(x-y,t) j_\varepsilon \star u_0(y)dy + \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s)\nabla \cdot (j_\varepsilon \star \nabla V \star v\ j_\varepsilon(\mathbb{1}_{|x|\le\frac{1}{\varepsilon}})u^\varepsilon))(s,y)dyds.$$

then u^ε satisfies the eq. In the classical sense

$$\begin{cases} \partial_t u^\varepsilon - \Delta u^\varepsilon + \underbrace{\nabla \cdot (j_\varepsilon \star \nabla V \star v\ (j_\varepsilon \star \mathbb{1}_{|x|\le\frac{1}{\varepsilon}}\ u^\varepsilon))}_{\in C^{2,1}_0} = 0 \\ u^\varepsilon|_{t=0} = \tilde{j}_\varepsilon \star u_0 \end{cases}.$$

Lemma 4.3.2 (Estimates)**.** We have $u \in L^\infty(L^1) = L^\infty([0,T]; L^1(\mathbb{R}^d))$

Proof.

$$\begin{aligned}
\|u^\varepsilon\|_{L^\infty(L^1)} &= \sup_t \int_{\mathbb{R}^d} |I(x,t) + II(x,t)| dx \\
&\le \sup_t \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} K(x-y,t) j_\varepsilon \star u_0(y) \right| dy dx + II \\
&\le \|K(t,\cdot)\|_{L^1} \cdot \|j_\varepsilon \star u_0\|_{L^1} + \|II\| \\
&\le \|u_0\|_{L^1} + \|II\| \\
&\le C_0 + C\sqrt{t}\|u^\varepsilon\|_{L^\infty(L^1)}.
\end{aligned}$$

Then for $C\sqrt{t^\star} \le \frac{1}{2}$ we have $\|u^\varepsilon\|_{L^\infty(L^1)} \le C$ since $\sqrt{t^\star}$ does not depend on ε we have for $\forall T > 0$ the bound

$$\|u^\varepsilon\|_{L^\infty(L^1)} \le C.$$

where

$$\begin{aligned}
I &= \int_{\mathbb{R}^d} K(x-y,t) j_\varepsilon \star u_0(y) dy \\
II &= \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s) \nabla \cdot (j_\varepsilon \star \nabla V \star v \, j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}} u^\varepsilon))(s,y) dy ds.
\end{aligned}$$

where

$$\begin{aligned}
\|II\| &\le \sup_t \int_0^t ds \int_{\mathbb{R}^d} dx \left| \int_{\mathbb{R}^d} dy K(x-y,t-s) \nabla \cdot (j_\varepsilon \star \nabla V \star v \cdot j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}} u^\varepsilon))(s,y) \right| \\
&\le \sup_t \int_0^t ds \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} dy \left| \nabla K(x-y,t-s) \cdot (j_\varepsilon \star \nabla V \star v \cdot j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}} u^\varepsilon))(s,y) \right| \\
&\le \sup_t \int_0^t ds \|\nabla V\|_\infty \|\nabla K(\cdot, t-s)\|_{L^1} \cdot \|j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}} u^\varepsilon\| \\
&\le C \sup_t \int_0^t \frac{1}{\sqrt{t-s}} ds \cdot \|u^\varepsilon\|_{L^\infty(L^1)} \\
&= C \cdot \sqrt{t} \cdot \|u^\varepsilon\|_{L^\infty(L^1)}.
\end{aligned}$$

□

Lemma 4.3.3 (Second Moment Bound)**.** The second moment of $u^\varepsilon \in L^\infty(L^1)$

$$\int |x|^2 u^\varepsilon dx < \infty.$$

is bounded

Proof.

$$\begin{aligned}
\int_{\mathbb{R}^d} |x|^2 u^\varepsilon(x,t) dx &= \int_{\mathbb{R}^d} |x|^2 \int_{\mathbb{R}^d} K(x-y,t) j_\varepsilon \star u_0(y) dy dx \\
&\quad + \int_0^t \int_{\mathbb{R}^d} dx |x|^2 \int_{\mathbb{R}^d} \nabla K(x-y,t-s) \cdot (j_\varepsilon \star \nabla V \star v \, j_\varepsilon \star (\mathbb{1}_{|x|\le\frac{1}{\varepsilon}} u^\varepsilon))(s,y) dy ds \\
&= I + II.
\end{aligned}$$

We bound again individually

$$
\begin{aligned}
I &\leq \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} |x-y|^2 K(x-y,t) j_\varepsilon \star u_0(y) dy dx + \int_0^t ds \iint_{\mathbb{R}^d\times\mathbb{R}^d} dx dy |y|^2 K(x-y,t) j_\varepsilon \star u_0(y) \\
&\leq \int_{\mathbb{R}^d} |x|^2 K(x,t) dx \cdot \int j_\varepsilon \star u_0(y) dy + \int_0^t ds \int_{\mathbb{R}^d} |y|^2 u_0(y) dy \\
&\leq C \cdot t \left(\|u_0\|^{L^1} + \int |y|^2 u_0(y) dy \right).
\end{aligned}
$$

Where

$$
\begin{aligned}
\int |x|^2 K(x,t) dx &= 4t \int \frac{|x|^2}{4t} \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} dx \\
&= C \cdot t.
\end{aligned}
$$

For

$$
\begin{aligned}
II &= \int_0^t \int_{\mathbb{R}^d} dx \underbrace{2x}_{|x|\leq|x-y|+|y|} \int_{\mathbb{R}^d} K(x-y,t-s) \cdot (j_\varepsilon \star \nabla V \star v\ (j_\varepsilon \star (\mathbb{1}_{|x|\leq \frac{1}{\varepsilon})} u^\varepsilon))(s,y) dy ds \\
&\leq C \cdot \int_0^t \sqrt{t-s} ds + C \cdot \int_0^t \int_{\mathbb{R}^d} K(x,t-s) dx \cdot \underbrace{\int |y| (j_\varepsilon \star \mathbb{1}_{|x|\leq \frac{1}{\varepsilon}} u^\varepsilon)(y) dy}_{III} \\
&\leq C(t) + C \int_0^t \int_{\mathbb{R}^d} |y| u^\varepsilon(y) dy ds \\
&\leq C(t) + C \int_0^t \int |x|^2 u^\varepsilon(x) dx ds.
\end{aligned}
$$

where in the above estimate we used for III that for arbitrary $u_0 \geq 0$ (just use the absolute value)

$$
\begin{aligned}
\int_{\mathbb{R}^d} |y|^2 \cdot j_\varepsilon \star u_0(y) dy &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |y|^2 j_\varepsilon j_\varepsilon(z) u_0(y-z) dz dy \\
&\leq \iint |y-z|^2 j_\varepsilon(z) u_0(y-z) dz dy + \iint |z|^2 j_\varepsilon(z) u_0(y-z) dy dz \\
&\leq \|j_\varepsilon\|_{L^1} \int |y|^2 u_0(y) dy + \varepsilon^2 \iint \frac{|z|^2}{\varepsilon^2} \frac{1}{\varepsilon^d} j(\frac{z}{\varepsilon}) u_0(y-z) dy dz \\
&\leq \int |y|^2 u_0(y) dy + C\varepsilon^2 \int u_0(y) dy.
\end{aligned}
$$

□

Remark. The proof above main trick is to use the moment estimates we have for the individual parts

Lemma 4.3.4. We have for $u^\varepsilon(x,t)$ in

$$
\begin{cases}
\partial_t u^\varepsilon - \Delta u^\varepsilon + \underbrace{\nabla \cdot (j_\varepsilon \star \nabla V \star v\ (j_\varepsilon \star \mathbb{1}_{|x|\leq \frac{1}{\varepsilon}}\ u^\varepsilon))}_{\in C_0^{2,1}} = 0 \\
u^\varepsilon|_{t=0} = \tilde{j}_\varepsilon \star u_0
\end{cases}
.
$$

that

$$
\lim_{|x|\to\infty} u^\varepsilon(x,t) = 0.
$$

Proof. we multiply the eq. by u^ε and integrate on $\mathbb{R}^d$

$$\int_{\mathbb{R}^d} \partial_t u^\varepsilon \cdot u_\varepsilon - \int_{\mathbb{R}^d} \Delta u^\varepsilon \cdot u^\varepsilon = -\int \nabla \cdot (\dots) \cdot u^\varepsilon$$

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^d} |u^\varepsilon|^2 dx + \int_{\mathbb{R}} |\nabla u^\varepsilon|^2 = \int (\dots) \cdot \nabla u^\varepsilon$$

$$\leq \frac{1}{2}\int_{\mathbb{R}^d} |\nabla u^\varepsilon|^2 dx + \frac{1}{2}\int |\dots|^2 dx$$

$$\frac{d}{dt}\int_{\mathbb{R}^d} |u^\varepsilon|^2 dx + \int_{\mathbb{R}^d} |\nabla u^\varepsilon|^2 dx \leq \|\nabla V\|_{L^\infty} \cdot \int_{\mathbb{R}} |j_\varepsilon \star (\mathbb{1}_{|x|\leq \frac{1}{\varepsilon}})|^2 dx$$

$$\leq C\|\nabla V\|_{L^\infty}^2 \cdot \int_{\mathbb{R}^d} |u^\varepsilon|^2 dx.$$

After applying Grönwall Above needs some restructuring

$$\sup_{0\leq t\leq T} \|u^\varepsilon\|_{L^2}^2 + \int_0^t \int_{\mathbb{R}^d} |\nabla u^\varepsilon|^2 \leq C(\|u_0\|_{L^2}).$$

□

Now we get for $\forall \varphi \in \mathcal{C}_0^\infty([0,T];\mathbb{R}^d)$

$$\begin{aligned}&\langle \partial_t u^\varepsilon, \varphi\rangle \\ &= \langle \Delta u^\varepsilon - \nabla \cdot (j_\varepsilon \star \nabla V \star v \cdot j_\varepsilon \star (\mathbb{1}_{|x|\leq \frac{1}{\varepsilon}} u^\varepsilon)), \varphi\rangle \\ &\leq \|\nabla u^\varepsilon\|_{L^2(L^2)} \cdot \|\nabla \varphi\|_{L^2(L^2)} + \|u^\varepsilon\|_{L^2(L^2)} \cdot \|\nabla \varphi\|_{L^2(L^2)}.\end{aligned}$$

It means

$$\|\partial_t u^\varepsilon\|_{(L^2(H_1))'} \leq C.$$

If we take $\varepsilon \to 0$ we obtain that there $\exists u \in L^\infty(L^2) \cap L^2(H^1)$ s.t.

$$u_\varepsilon \overset{\star}{\rightharpoonup} u \quad \text{in } L^\infty(L^2) \cap L^2(H^1).$$

and u satisfies the weak version of the PDE, for $\forall \varphi \in L^2(H^1)$

$$\int_0^T \langle \partial_t u, \varphi\rangle_{(H^1,H^1)} dt = -\int_0^T \int_{\mathbb{R}^d} (\nabla u - \nabla V \star v \cdot u) \cdot \nabla \varphi dx dt.$$

Lemma 4.3.5. Let $u \in L^\infty(L^2) \cap L^2(H^1)$ s.t. for $\forall \varphi \in L^2(H^1)$ it is a solution to

$$\int_0^T \langle \partial_t u, \varphi\rangle_{(H^1,H^1)} dt = -\int_0^T \int_{\mathbb{R}^d} (\nabla u - \nabla V \star v \cdot u) \cdot \nabla \varphi dx dt.$$

with $u_0 \geq 0$, then it is non negative $u \geq 0$ a.e. and unique

Proof. Choose $\varphi = u_- = \min\{0, -u\}$ then we know that $u \in L^2(H^1) \Rightarrow u_- \in L^2(H^1)$ then

That u_- lies in the space is in fact non trivial and is part of the PDE lecture

$$\int_0^t \partial_t u \cdot u_- ds = \int_0^t \int_{\mathbb{R}^d} -\nabla u \cdot \nabla u_- dx ds - \int_0^t \int_{\mathbb{R}^d} \nabla V \star v \cdot u \nabla u_- dx ds$$

.

And

$$\frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \partial_t |u_-|^2 dx ds + \int_0^t \int_{\mathbb{R}^d} |\nabla u_-|^2 dx ds \leq \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} |\nabla u_-|^2 dx ds + \frac{C}{2}\int_0^t \int_{\mathbb{R}^d} |u_-|^2 dx ds.$$

By Grönwall

$$\int_{\mathbb{R}^d} |u_-|^2 dx \leq e^{Ct} \int_{\mathbb{R}^d} |u_{0,-}|^2 dx = 0.$$

Since $u_0 \geq$ we get $u_{0,-} = 0$

The uniqueness of the solution follows also from the L^2 estimate, □

Theorem 4.3.3 (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

$$\|u\|_U \leq C.$$

Then the map $T(\cdot, 1)$ has a fixed point

Lemma 4.3.6 (Aubin-Lions Lemma (extended version))**.** First note that

$$H^1 \subset L^2 \subset H^{-1}.$$

If $(f_n)_n \subset L^2([0, T]; H^1(\mathbb{R}^d))$ satisfying

1. $\|f_n\|_{L^2(H^1)} \leq C$
2. $\|\partial_t f_n\|_{L^2(H^1)} \leq C$
3. $\sup_t \int |x|^2 |f_n(t, x)| dx \leq C$

Then f_n is relatively compact in $L^2([0, T]; L^2(\mathbb{R}^d))$ i.e. there $\exists f \in L^2(L^2)$ s.t.

$$\|f_{n_j} - f\|_{L^2([0,T];L^2(\mathbb{R}^d))} \to 0.$$

Theorem 4.3.4 (Existence of Solution)**.** The nonlinear non-local (PDE) has a solution,

Proof. We show by fix point argument using Theorem 4.3.3 on the map

$$M : (v, \sigma) \in L^\infty(L^1) \to X.$$

where

$$X = \{u | u \in L^\infty(L^1(1 + |x|^2)dx) \cap L^2 \cap L^2(H^1), \partial_t u \in L^2(H^{-1})\}.$$

where u is a weak Solution of the (PDE), it is then sufficient to show that M has fixed point.

$$(\text{PDE})(\delta) \begin{cases} u_t - \Delta u + \sigma \nabla \cdot (\nabla V \star u \; u) = v \\ u|_{t=0} = \sigma u_0 \geq 0 \end{cases}.$$

Using Leray-Schauder Fixed Point Theorem we obtain a fixed point of $M(\cdot, 1)$ which means

$$(\text{PDE})(1) \begin{cases} u_t - \Delta u + \nabla \cdot (\nabla V \star u \; u) = v \quad \nabla V \in L^\infty \\ u|_{t=0} = \geq 0 \quad u_0 \in L^1((1 + |x|^2)dx) \cap L^2(dx) \end{cases}.$$

Rewatch video and add the arguments why we can apply schauder □

4.3.3 Back to the Makean-Vlasov Equation

We revisit the MVE

$$\text{(MVE)} \begin{cases} dY(t) = (\nabla V \star u)(Y(t))dt + \sqrt{2}dW_t \\ Y(0) = \xi \in L^2(\Omega) \\ \mathcal{L}(\xi) = u_0 \in L^1((1+|x|^2)dx) \cap L^2(dx) \\ \mathcal{L}(Y) = u \end{cases}.$$

Suppose now that u is the solution of the (PDE) we obtained above, then consider for $\nabla V \in \mathrm{Lip}$

$$\text{(SDE)} \begin{cases} dY(t) = (\nabla V \star u)(Y(t))dt + \sqrt{2}dW_t \\ Y(0) = \xi \in L^2(\Omega) \end{cases}.$$

then by convolution $(\nabla V \star u)$ is also Lipschitz, and by SDE theory there $\exists! Y \in \mathbb{L}^2([0,T])$ solving (SDE). Then for $\forall \varphi \exists C_0^\infty([0,T] \times \mathbb{R}^d)$, by Itô's formula

$$\begin{aligned} \varphi(Y(t),t) - \varphi(Y(0),0) = &\int_0^t \partial_t \varphi(Y(s),s) + \nabla\varphi(Y(s),s) \cdot \nabla V \star u(Y(s),s) + \Delta\varphi(Y(s),s)ds \\ &+ \int_0^t \dots dW_s. \end{aligned}$$

Then after taking the expectation the stochastic term disappears, suppose $\mu^Y = \mathcal{L}(Y)$ then we get

$$\begin{aligned} &\int_{\mathbb{R}^d} \varphi(x,t)d\mu^Y(x,t) - \int_{\mathbb{R}^d} \varphi(x,0)u_0(x)dx \\ &= \int_0^t \int_{\mathbb{R}^d} \partial_t\varphi(x,s) + \nabla\varphi(x,s) \cdot \nabla V \star u(x,s) + \Delta\varphi(x,s)d\mu^Y(x,s)ds. \end{aligned}$$

Now we want to prove $\mu^Y = u$ i.e. $\forall \varphi \in C_b$ a.e. in $t \in [0,T]$

$$\int_{\mathbb{R}^d} \varphi d\mu^Y = \int_{\mathbb{R}^d} \varphi u dx.$$

We can rewrite the integral version in the weak sense as follows

$$\partial_t \mu^Y = \Delta\mu^Y - \nabla \cdot (\nabla V \star u \cdot \mu^Y).$$

How to prove : since u as a solution of (PDE) it is also a solution of the (LDPE) the proof $u = \mu^Y$ can be reduced to show that (LPDE) has a unique measure valued solution.
Let $\tilde{\mu} \in L^\infty(L^1)$ be a solution for $\forall \varphi \in \mathcal{C}_0^\infty([0,T);\mathbb{R}^d)$

$$\int_0^T \int_{\mathbb{R}^d} \tilde{u}(\partial_t\varphi + \Delta\varphi - \nabla\varphi \cdot \nabla V \star u)dxdt = 0.$$

then $\tilde{u} = 0$ implies the uniqueness of (LPDE), which means we need to show for $\forall g \in \mathcal{C}_c^\infty(\mathbb{R}^d \times [0,T])$ it holds

$$\int_0^T \int_{\mathbb{R}^d} \tilde{u} g dxdt = 0.$$

if $\exists \varphi$ s.t. $g = \partial_t\varphi + \Delta\varphi - \nabla\varphi \cdot \nabla V \star u$