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

Li Chen

October 22, 2023

Contents

Abstract

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

Chapter 1

Model description and Introduction

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

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

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

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 $(x_1(t), x_2(t), \ldots, x_N(t)) \in \mathcal{C}^1([0,T];\mathbb{R}^d),\ i = 1,\ldots,N$ 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 $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ 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

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

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

Then our second order system is then governed by

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

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

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

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

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

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

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

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

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

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

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

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 $\square$

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

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

Proof. Define

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

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

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

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

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

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

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

Integrate the above inequality from a to t to get

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

Which implies

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

and

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

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

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

$\square$

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

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

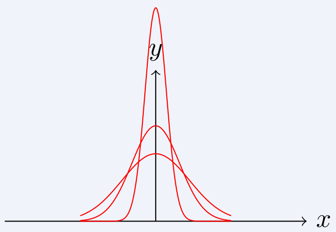


Figure 2.1: Heat Kernel for different t

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

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

Then

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

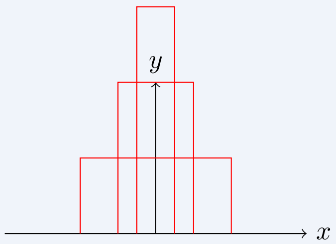


Figure 2.2: Rectangular functions for different n

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

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

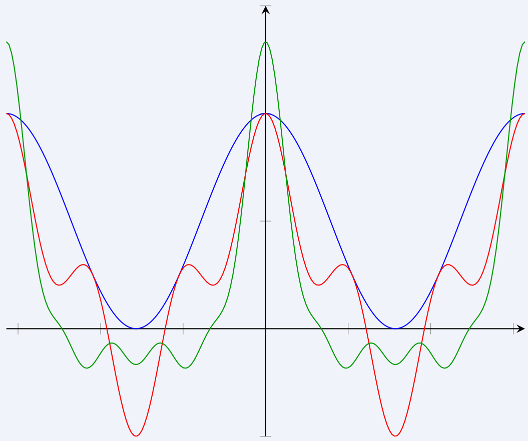


Figure 2.3: Dirichlet kernel for different n

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

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

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 $T \coloneqq T_f$ 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

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

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

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

There for by assumption on Regularity for K, we have

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

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

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

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

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

which implies by Gronwall's inequality that

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

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 $j(x) \in \mathcal{C}_0^\infty$ is called a mollification kernel if it satisfies the following properties

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

A typical example of a smooth kernel is given by

Example.

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

where k is given s.t the integral is 1

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

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

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

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

Definition 2.6.2 (Mollification Operator)**.** For $\forall u \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ 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

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

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 $\forall i = 1, 2, \dots, d$

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

where we have used the fact that

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

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

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

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

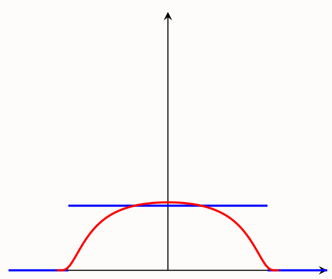
$$\int_{\mathbb{R}^d} \varphi(x) f(t,x) dx = \int_{\mathbb{R}^d} \varphi(x) f(\tilde{t},x) dx + \int_{\tilde{t}}^{t} \iint_{\mathbb{R}^{2d}} f(s,x) K(x,y) f(s,y) \nabla \varphi(x) dx dy ds.$$

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 $|\nabla\varphi_R(x)| \leq \frac{C}{R}$. Therefore with this test function, we obtain from the weak solution formula that

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

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

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

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

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

□

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

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

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

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

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

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

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

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

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

Since $\mathrm{Lip}(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$ and because the total mass is 1, the above also holds for test functions in $\mathcal{C}_b(\mathbb{R}^d)$. Hence the weak convergence in measure is true. The fact that $\mathrm{Lip}(\mathbb{R}^d)$ is dense in $\mathcal{C}_0$ 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 $p = \infty$

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

the space $L^p(\mathbb{P}) = L^p(\Omega) = \{X \ : \ \Omega \to \mathbb{R}^d \mid \|X\|_p \le \infty\}$ is a Banach space.

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

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

Definition 3.1.10 (Distribution Functions)**.** Note for $x, y \in \mathbb{R}^d$ we write $x \le y$ if $x_i \le y_i$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

Exercise. Proof the above theorems

3.1.2 Borel Cantelli

Definition 3.1.15. Let $A_1, \ldots, A_m \in \mathcal{F}$ 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

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

□

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 $X_1, \ldots, X_n$ is called identically distributed if

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

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

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

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

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

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

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

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

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

Then

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

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 (History). Let $X(t)$ be a real valued process. The $\sigma-$algebra
$$\mathcal{U}(t) := \mathcal{U}(X(s) \mid 0 \leq s \leq t).$$
is called the history of X until time $t \geq 0$

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

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

2. If $X(s) \leqq_{\geq} \mathbb{E}[X(t)|\mathcal{U}(s)]$ 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 3.1.23 (Brownian Motion). A real valued stochastic process $W(\cdot)$ is called a Brownian motion or Wiener process if

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

Remark. One can derive directly that

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

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

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

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

Borel set $A \subset \mathbb{R}^2$

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

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 3.1.25 (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

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

3.2 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

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

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

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

(2). By definition we have

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

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

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

□

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

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

Proof. By definition we can write

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

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

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

Then we have

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

□

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

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

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

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

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

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

□

Definition 3.2.8 (Itþ 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

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

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 Itøisometry we have

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

   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

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

it follows that

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

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. $\mathbb{E}[\int_0^T G dW_t \cdot \int_0^T H dW_t] = \mathbb{E}[\int_0^T G H dt]$

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

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

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

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

□

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 $\int_0^\tau G dW_t$ with $0 \leq \tau \leq T$ 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

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

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

Furthermore we way $X(\cdot)$ 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

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

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

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

holds true. Actually

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

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

$$dX_1=F_1dt+G_1dW_t\quad\text{ and }\quad dX_2=F_2dt+G_2dW_t.$$

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

$$d(X_1X_2)=X_2dX_1+X_1dX_2+G_1G_2dt\quad=(X_2F_1+X_1F_2+G_1G_2)dt+(X_2G_1+X_1G_2)dW_t.$$

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

$$X_i(t)=F_itG_iW(t).$$

Then it holds a.s. that

$$\begin{aligned}
&\int_0^t (X_2 dX_1 + X_1 dX_2 + G_1 G_2 ds) \\
&= \int_0^t (X_2 F_1 + X_1 F_2) ds + \int_0^t (X_2 G_1 + X_1 G_2) dW_s + \int_0^t G_1 G_2 ds \\
&= \int_0^t \left(F_1(F_2 s + G_2 W(s)) + F_2(F_1 s + G_1 W(s))\right) ds + G_1 G_2 t \\
&= \int_0^t \left(G_1(F_2 s + G_2 W(s)) + G_2(F_1 s + G_1 W(s))\right) dW_s \\
&= G_1 G_2 t F_1 F_2 t^2 + (F_1 G_2 + F_2 G_1) \left(\int_0^t W(s) ds + \int_0^t s dW_s \right) \\
&\quad + 2 G_1 G_2 \int_0^t W(s) dW_s.
\end{aligned}$$

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

$$G_1 G_2 (W(t))^2 + F_1 F_2 t^2 + (F_1 G_2 + F_2 G_1) t W(t) = X_1(t) + X_2(t).$$

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.

$$X_1^n(t) X_2^n(t) - X_1(0) X_2(0) = \int_0^t \left(X_1^n(s) dX_2^n(s) + X_2^n(s) dX_1^n(s) + G_1 G_2 ds\right).$$

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

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

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

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

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 $\square$