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

Li Chen

October 30, 2023

Contents

Chapter 1

MEAN FIELD LIMIT FOR SDE SYSTEM

1.1 Basics On Probability Theory

This section is dedicated to a small review of basic concepts in probability theory in preparations of SDE's

1.1.1 Probability Spaces and Random Variables

Definition 1.1.1 (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 1.1.2 (Measure Space)**.** A tuple $(\Omega, \mathcal{F})$ is called a measurable space. The elements of $\mathcal{F}$ are called measurable sets

Definition 1.1.3 (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 1.1.4 (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 1.1.5 (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 1.1.6 (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 1.1.1 (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 1.1.7 (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 1.1.8 (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 1.1.9 (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 1.1.10 (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

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

Definition 1.1.11 (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

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

Definition 1.1.12 (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 1.1.13 (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 1.1.14 (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 1.1.1. $X_1, \dots, X_m : \Omega \to \mathbb{R}^d$ 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 1.1.2. 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 1.1.3. $X_1, \ldots, X_m : \Omega \to \mathbb{R}$ 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

1.1.2 Borel Cantelli

Definition 1.1.15. 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 1.1.2 (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

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

□

Definition 1.1.16 (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 1.1.4 (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.

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

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

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

□

1.1.3 Strong Law Of Large Numbers

Definition 1.1.17. A sequence of random variables $X_1, \dots, X_n$ is called identically distributed if

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

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

Theorem 1.1.5 (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

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

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

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

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

□

1.1.4 Conditional Expectation

Definition 1.1.18. 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 1.1.19. 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 1.1.6. 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 1.1.7 (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 1.1.3 (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

1.1.5 Stochastic Processes And Brownian Motion

Definition 1.1.20 (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 1.1.21 (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 1.1.22 (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 1.1.23 (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 1.1.4. 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 1.1.8 (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 □

1.1.6 Brownian Motion

Definition 1.1.24 (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 1.1.25. 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 1.1.9. 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 1.1.26 (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

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

1.1.7 Convergence of Measure and Random Variables

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

Definition 1.1.27 (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 1.1.28 (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 1.1.29 (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 1.1.10 (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 □

1.2 Itô Integral

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

Definition 1.2.1.

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 1.2.2 (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 1.2.3 (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 1.2.4.

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 1.2.5. 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 1.2.6 (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 1.2.7 ((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 1.2.1. Let $G, H \in \mathbb{L}^2([0,T])$ be two step processes, then for $\forall a, b \in \mathbb{R}$ it holds

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

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

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

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 1.2.1 ((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

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

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 1.2.2. 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

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

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 1.2.8 (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

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

   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

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

   it follows that

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

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

Theorem 1.2.1 (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 1.2.3 (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 1.2.8 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 $0 \le \tau \le T$ is a martingale

1.2.1 Itô's Formula

Definition 1.2.9 (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 $0 \le s \le r \le T$, then $X(\cdot)$ is called Itô process.

Furthermore we say $X(\cdot)$ has a stochastic differential.

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

Theorem 1.2.2 (Itô's Formula)**.** Let $X(\cdot)$ be an Itô process given by $dX = F dt + G dW_t$ for some $F \in \mathbb{L}^1([0,T])$ and $G \in \mathbb{L}^2([0,T])$. Assume $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ is continuous and $\frac{\partial u}{\partial t}, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}$ exists and are continuous. Then $Y(t) := u(X(t), t)$ satisfies

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

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

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

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

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

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

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

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

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

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

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

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

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

The summation of the above integrals yields

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

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

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

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

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

Notice that for each Itô process given by step processes

$$X^n_i(t) = X_i(0) + \int_0^t F^n_i ds + \int_0^t G^n_i dW_s.$$

the product rule holds, i.e.

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

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 ⇒ Step 5

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

1.2.2 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 1.2.10 (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 1.2.11 (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 1.2.3 (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 1.2.2. 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 1.2.12 (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 1.2.13 (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

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

1.3 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

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

1.4 Solving Stochastic Differential Equations

The setup of the following section will be the following

Definition 1.4.1 (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 1.4.2 (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

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

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 1.4.3 (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 1.4.4. 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 A.
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 A 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 1.4.1 (Existence and Uniqueness of Solution)**.** Let Assumption A 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

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

.

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 $d^n(t) \leq \frac{(Mt)^{n+1}}{(n+1)!}$ for some $M > 0$.
IA: For $n = 0$ we have

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

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

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

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

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

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 1.4.2 (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

□

1.5 Stochastic Mean Field Limit

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

Definition 1.5.1 (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

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

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. $\mu_n \rightharpoonup \mu$

1.5.1 Stochastic Particle System

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

Definition 1.5.2 (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 1.5.3 (Stochastic Particle System)**.** For N interacting particles $(X^1, \ldots, X^N)$ with i.i.d initial data $(X_i^N(0))_{i\in\{1,\ldots,N\}} \subset L^2(\Omega)$ 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 1.4.2

Remark. For our initial measure we already have

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

1.5.2 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 in the mean field limit this is only the case for our initial data all particles at $t > 0$ are not i.i.d.

Corollary. If $(X_i)_{i\in\{1,\ldots,N\}}$ 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 $(X_i)_{i\in\{1,\ldots,N\}}$ 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 1.5.1 (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 means that

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

Lemma 1.5.2 (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 $\lim\limits_{x\to\infty} \frac{\varphi(x)}{x} = \infty$ and

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

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

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

Finally Lemma 1.5.2 implies the uniform integrability and we conclude by Lemma 1.5.1

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

□

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

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

1.5.3 Toy Example

Let us first consider a simple stochastic particle system given by

Assumption B. 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 1.5.3 with drift and diffusion for $\nabla V \in \mathrm{Lip}$

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

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

Proof. First we note that by Assumption B 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 $\mathbb{X} = (X_1^N, \ldots, X_N^N) \in \mathbb{R}^{dn}$ and $\mathbb{W} = (W^1, \ldots, W^N)$ then

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

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 A we get a solution by Theorem 1.4.1

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

For Σ the argument is analog where for the Wasserstein distance we used Then by Theorem 1.4.1 we get a solution $X \in L^2([0,T])$ for fixed N

$\square$

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

1.5.4 Makean-Vlasov

Definition 1.5.4 (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 1.5.5 (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

$$\|X\|_t = \sup_{0 \le t \le T} |X(t)|.$$

this norm $\|\cdot\|_T$ induces a σ-algebra on $\mathcal{C}^d$

Definition 1.5.6 (Random Variable). A random Variable on $\mathcal{C}^d$ is a map

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

Definition 1.5.7 (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

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

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

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

Definition 1.5.8 (Wasserstein Distance). And we can define for arbitrary measures $\mu, \tilde{\mu} \in \mathcal{P}^2(\mathcal{C}^d)$ the Wasserstein distance by

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

Where

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

Remark. Note that

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

Theorem 1.5.2 (Unique and Existence of Solution for Makean-Vlasov). If b and σ satisfy Assumption B 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

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

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

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

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

Now taking the expectation

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

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

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

All together

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

.

So by Grönwall inequality we get

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

using the inequality Definition 1.5.8
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 $\mathcal{P}^2(\mathcal{C}^d)$

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

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 1.5.3 (Mean-Field-Limit). Let b and σ fulfill Assumption B 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

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

Proof. The proof is similar to what we have done in the Theorem 1.5.2, the critical part is to work with our stochastic empirical measure, we do so by introducing an intermediate empirical measure. We compute

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

We get

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

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

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

We continue by taking the expectation

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

Goal is to get a Grönwall inequality for

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

.

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 □

1.6 PDE Approach To Solving the Makean-Vlasov Equation

Let us shortly explain the connection between the PDE approach and the SDE approach to solving the Makean-Vlasov equation, from now on assume $\sigma(Y(t), \mu(t)) = \sqrt{2}$constant, that means the

$$\begin{aligned}
Y(t) &= b(Y(t), \mu(t))dt + \sqrt{2}dW_t \\
Y(0) &= \xi \in L^2(\Omega).
\end{aligned}$$

and $\mu_0 = \mathcal{L}(\xi)$, applying Itôs formula for $\forall \varphi \in \mathcal{C}_0^\infty([0, T) \times \mathbb{R}^d)$

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

We take expectation on both sides (note the last term is 0)

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

Formally if μ is regular enough i.e it has density and the density is good enough, then μ should satisfy the weak PDE

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

Write out the entire thing with distributions as a middle step between the PDE and the Integral

Next goal is to solve this PDE

Definition 1.6.1 (Weak PDE)**.** If u is the density of μ we write

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

Remark. If we manage to find such u then we can plug it in to the (MVE) equation

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

if b is bounded and Lipschitz,then we have a solution, Y_t, let $\overline{u}$ be the Law of Y_t. Then we have by Itô formula for $\forall\varphi$

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

Which means $\overline{\mu}$ satisfies

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

This means in order to solve the Makean-Vlasov Equation e.g. we need to prove $\overline{u} = u$

The Laplacian gives us sufficient regularity, even for bad b i.e unbounded non Lipschitz, a standard example is

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

the regularity depends directly on the regularity of K or rather u by properties of convolution.

Definition 1.6.2 (Sobolev Spaces)**.** We define roughly

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

where $\forall\varphi\in\mathcal{C}_0^\infty$, with norm $\|u\|_{H_1} = \|u\|_2 + \|\nabla u\|_2$

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

And

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

Then

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

Definition 1.6.3 (Weak Solution)**.** We say that $u \in L^2([0,T]; H^1(\mathbb{R}^d) \cap L^\infty([0,T]; L^2(\mathbb{R}^d)))$ and $\partial_t u \in L^2([0,T]; H^{-1}(\mathbb{R}^d))$ is a weak solution of the (PDE) if for $\forall \varphi \in \mathcal{C}_0^\infty([0,T] \times \mathbb{R}^d)$ it holds

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

1.6.1 Heat Kernel

Definition 1.6.4 (Heat Equation)**.** The following PDE is called Heat Equation (last term is source term)

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

If $K(x,t)$ is the heat kernel, then we can formally rewrite the above equation into

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

This is called Duhamels formula.

For standard heat equation with source term f

$$(\text{HE}) := \partial_t u - \Delta u = f \quad u(0) = u_0.$$

$$u(x,t) = \int K(x-y,t) u_0(y) dy + \int_0^t \int_{\mathbb{R}^d} K(x-y,t-s) f(y,s) dy ds.$$

gives the solution formula of the Heat Equation (HE).

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

Definition 1.6.5 (Heat equation)**.** The following PDE is called Heat equation with source f

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

Using $x \in \mathbb{R}^d$ the Fourier transform

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

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

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

then we get the equivalent HE

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

With

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

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

inverse transform of the Fourier transformation

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

Then we get K as

Definition 1.6.6 (Heat Kernel)**.** The following is called Heat Kernel

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