Chapter 1

Introduction

Mean Field Particle Systems is about the study of particles which are represented by (stochastic) differential equations. This course in particular is concerned with the behaviour of the system as the size grows to infinity:

Definition 1.0.1 (Toy Mean Field Particle System)**.** Let $N \in \mathbb{N}$ then a Mean Field Particle System of first order is given by :

$$x_1(t), \ldots, x_n(t) \in \mathcal{C}^1([0,T];\mathbb{R}^d) \qquad x_i(0) = c_i.$$

Where each particle satisfies

$$dx_i = \frac{1}{N}\sum_{j=1}^{N} K(x_i, x_j)dt + \sigma dB_i(t).$$

Where B_i is a Brownian motion; For $\sigma = 0$ the system is called deterministic.

Example. Example choices for K are :

$$K(x_i, x_j) = \nabla(|x_i - x_j|^2).$$

or :

$$o\gamma = \frac{x_i - x_j}{|x_i - x_j|^d}.$$

Goal is to study what happens at $N \to \infty$, to do so we consider how the measure of a system converges

Definition 1.0.2 ((Empirical) Measure of a System)**.** Consider the point measure for every x_i : $\delta_{x_i(t)}$, then the measure of the System of order N is

:

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

Assumption 1.0.1. For initial data the empirical measure of a system converges $\mu_N(0) \to \mu(0)$ where μ is absolutely continuous with respect to the Lebesgue Measure

Corollary. By Radon Nikodym

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

It can be shown that μ solves a PDE, to do so we compute the derivative of μ using test functions

$$\forall\varphi \in C_0^\infty(\mathbb{R}^d).$$

$$\begin{aligned}
\frac{d}{dt}\langle\mu_N(t),\varphi\rangle = \frac{d}{dt}\int_{\mathbb{R}^d}\varphi(x)d\mu_N(t)(x) &= \frac{d}{dt}\int\frac{1}{N}\sum_{i=1}^{N}\varphi(x)d\delta_{x_i(t)} \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{d}{dt}\varphi(x_i(t)) \\
&\overset{\text{Chain.}}{=} \frac{1}{N}\sum_{i=1}^{N}\nabla\varphi(x_i(t))\frac{d}{dt}x_i(t) \\
&= \frac{1}{N}\sum_{i=1}^{N}\nabla_x\varphi(x_i(t))\cdot\underbrace{\frac{1}{N}\sum_{j=1}^{N}K(x_i(t),x_j(t))}_{\text{Def.}} \\
&= \frac{1}{N}\sum_{i=1}^{N}\nabla_x\varphi(x_i(t))\cdot\frac{1}{N}\sum_{j=1}^{N}\int_{\mathbb{R}^d}K(x_i(t),y)d\delta_{x_j(t)}(y) \\
&= \frac{1}{N}\sum_{i=1}^{N}\nabla_x\varphi(x_i(t))\cdot\int_{\mathbb{R}^d}K(x_i(t),y)d\mu_N(t) \\
&= \int_{\mathbb{R}^d}\nabla\varphi(x)\int_{\mathbb{R}^d}K(x,y)d\mu_N(t,y)d\mu_N(t,x)
\end{aligned}$$

.

Where the last line can be rewritten by using Integration by Parts (Divergence Theorem) :

$$\int_{\mathbb{R}^d} \nabla\varphi(x) \int_{\mathbb{R}^d} K(x,y)d\mu_N(t,y)d\mu_N(t,x) \overset{\text{Part.}}{=} -\langle \nabla \cdot (\mu_N \int_{\mathbb{R}^d} K(\cdot,y)d\mu_N(y)), \varphi\rangle$$

.

This means μ satisfies :

$$\partial_t \mu_N + \langle \nabla \cdot (\mu_N \int_{\mathbb{R}^d} K(\cdot,y)d\mu_N(y)), \varphi\rangle = 0 \xrightarrow{N\to\infty} \partial_t \mu + \langle \nabla \cdot (\mu \int_{\mathbb{R}^d} K(\cdot,y)d\mu(y)), \varphi\rangle = 0.$$

In practical applications (Theoretical Physics , Biology) systems that are considered are often of second order

Definition 1.0.3 (Toy Second Order System)**.** Given $N \in \mathbb{N}$ a Second Order System is given by

$$(x_i(t), v_i(t)), \ldots, (x_N(t), v_N(t)) \in \mathbb{R}^{2d}.$$

Such that :

$$\frac{d}{dt} x_i(t) = v_i(t)$$
$$\frac{d}{dt} v_i(t) = \frac{1}{N}\sum_{j=1}^{N} F(\underbrace{x_i(t), v_i(t)}_{\text{Position and Velocity of itself}} ; x_j(t), v_j(t)) + \sigma \frac{dB_t}{dt}$$

.

Example (Gravitational Force)**.** An example of F could be :

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

Definition 1.0.4 (Second Order Measure)**.** The Measure of a second order System is given by :

$$\mu_N(x,v) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_i(t), v_i(t))}.$$

Exercise. Show what PDE μ solves for $\sigma = 0$, *Hint :* Calculate $\frac{d}{dt} < \mu_N, \varphi >$ for some test function $\varphi \in C_0^\infty(\mathbb{R}^{2d})$

Chapter 2

Deterministic Mean Field Particle Systems

The goal for this chapter is to determine when a unique solution exists to the Mean-Field-Equation arising from our Particle Systems. Beginning by recapping standard ODE Theory on when a solution exists to an ODE IVP and continuing with the notion of Weak Solutions and Distributions which allows us the generalize the above ODE results.

Definition 2.0.1 (Deterministic Mean Field Particle System). For $N \in \mathbb{N}$ a deterministic mean field particle system is given by N particles :

$$x_1(t), \dots, x_n(t) \in \mathcal{C}^1([0,T]; \mathbb{R}^d) \qquad x_i(0) = c_i.$$

With initial points :

$$x_i(0) = x_{i,0} \in \mathbb{R}^d.$$

And the relation :

$$\frac{d}{dt} x_i = \frac{1}{N} \sum_{j=1}^{N} K(x_i, x_j).$$

The system is then given by :

$$X_N = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{pmatrix} \in \mathbb{R}^{dN}.$$

2.1 ODE Theory

Definition 2.1.1 (Initial Value Problem (standard)). For $\forall\ T > 0$ let the standard IVP be given by :

$$x' = f(t,x)$$
$$x|_{t=0} = x_0 \in \mathbb{R}^n.$$

with $t \in [0,T],\ x(t) \in \mathbb{R}^n$ and $f : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$

As opposed to Picard-Lindelöf where only need locally Lipschitz continuity since we first construct a solution on a small local subset and then extend this solution, here we require global Lipschitz continuity since we do not want to extend our solution.

Theorem 2.1.1 (Picard Iteration). Whenever $f : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$is globally Lipschitz continuous in the second component the standard IVP has a unique solution $x \in \mathcal{C}^1([0,T];\mathbb{R}^n)$

Proof. We begin by defining our Picard-Iteration by

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

The proof will be split into two steps

- **Step 1:** The first part of the proof consists of showing the above defined sequence is Cauchy and thus converges
- **Step 2:** The second part is then showing the limit is a solution to the IVP

We know that f is continuous such that $(x_n)_{n\in N} \subset \mathcal{C}^1([0,T];\mathbb{R}^n)$, by completeness we know that any sequence that is Cauchy must also converge against a limit in the space.

As such we show our sequence is Cauchy by first considering the distance between any two points

$$
\begin{aligned}
|x_2 - x_1| = \left|\int_0^t f(s, x_1(s))ds - \int_0^t f(s, x_0(s))ds\right| &= \left|\int_0^t f(s, x_{m-1}(s)) - f(s, x_{n-1}(s))ds\right| \\
&\leq \int_0^t |f(s, x_1(s)) - f(s, x_0(s))|ds \\
&\overset{\text{Lip.}}{\leq} L \int_0^t |x_1(s) - x_0(s)|ds \\
&= L \int_0^t \left|\int_0^{s_0} f(s, x_0)ds\right| ds_0 \\
&\leq L \cdot \int_0^t \int_0^{s_0} |f(s, x_0)|ds ds_0 \\
&\leq L \underbrace{M}_{=\max_{s\in[0,T]}|f(s,x_0)|} \frac{t^2}{2}.
\end{aligned}
$$

We can extend to arbitrary $m \in \mathbb{N}$ by using induction

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

(IS) : $m \to m+1$

$$
\begin{aligned}
|x_{m+1}(t) - x_m(t)| &\overset{\text{Lip.}}{\leq} L \int_0^t |x_m(s) - x_{m-1}(s)|ds \\
&\overset{\text{IA.}}{\leq} L \int_0^t \frac{ML^{m-1}s^m}{m!} ds = ML^m \frac{t^{m+1}}{(m+1)!}.
\end{aligned}
$$

Now for any $n, m \in \mathbb{N}$ and assuming without loss of generality that $n > m$ we can write $n = m + p$ for $p \in \mathbb{N}$:

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

This shows that $(x_m)_{m\in\mathbb{N}}$ is Cauchy and has a limit $x \in \mathcal{C}([0,T];\mathbb{R}^n)$ with

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

It remains to show that $x(t)$ is a solution to the IVP i.e :

$$x(t) = \lim_{m\to\infty} x_0 + \int_0^t f(s, x_{m-1}(s))ds \leftrightarrow x_0 + \int_0^t f(s, x(s))ds.$$

Consider the difference between both sides

$$\begin{aligned}
|\lim_{m\to\infty} \int_0^t f(s, x_{m-1}(s)) - f(s, x(s))ds| &\leq \lim_{m\to\infty} \int_0^t |f(s, x_{m-1}(s)) - f(s, x(s))|ds \\
&\leq \lim_{m\to\infty} L \int_0^t |x_{m-1}(s) - x(s)|ds \\
&\leq \lim_{m\to\infty} Lt \cdot \max_{s\in[0,t]} |x_{m-1}(s)x(s)| \\
&\leq \lim_{m\to\infty} Lt \cdot \max_{s\in[0,T]} |x_{m-1}(s)x(s)| \\
&= 0.
\end{aligned}$$

It remains to show that the solution is unique, for that assume $x, \hat{x} \in \mathcal{C}([0,T];\mathbb{R}^n)$ are both solutions to the IVP. Meaning that :

$$\begin{aligned}
x(t) &= x_0 + \int_0^t f(s, x(s))ds \\
\hat{x}(t) &= x_0 + \int_0^t f(s, \hat{x}(s))ds.
\end{aligned}$$

Then :

$$\begin{aligned}
|x - \hat{x}| &\leq \int_0^t |f(s, x(s)) - f(s, \hat{x}(s))|ds \leq L \cdot \int_0^t |x(s) - \hat{x}(s)|ds \\
&= L \int_0^t \underbrace{e^{-\alpha s}|x(s) - \hat{x}(s)|}_{=\rho(s)} e^{\alpha s} ds \\
&\leq L \max_{t\in[0,T]} \rho(t) \cdot \frac{1}{\alpha}(e^{\alpha t} - 1) \\
&\leq L \max_{t\in[0,T]} \rho(t) \cdot \frac{1}{\alpha} \cdot e^{\alpha t}.
\end{aligned}$$

By rearranging with the initial term :

$$\rho(t) = e^{-\alpha t}|x(t) - \hat{x}(t)| \leq \frac{L}{\alpha} \max_{t\in[0,T]} \rho(t)$$

$$\max_{t\in[0,T]} \rho(t) \leq \frac{L}{\alpha} \max t \in [0,T] \rho(t).$$

by choosing $\alpha = 2L$:

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

And the solutions must be equal for $\forall t \in [0, T]$. $\square$

The reason this proof deviates from the standard Picard-Lindelöf theorem, is that for our systems we require Global existence, doing so by requiring f to be globally Lipschitz continuous.

Theorem 2.1.2. The solution $x(t, t_0, x_0) \in \mathcal{C}$ is continuously dependent on (t_0, x_0)

Theorem 2.1.3 (Gronwalls inequality)**.** For $\alpha, \beta, \varphi \in \mathcal{C}([a, b]; \mathbb{R})$ $\beta \geq 0$ and

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

then :

$$\varphi(t) \leq \alpha(t) + \int_a^t \beta(s) \exp\left(\int_s^t \beta(\tau)d\tau\right) \alpha(s)ds.$$

Proof. Denote $\psi(t) = \int_a^t \beta(s)\varphi(s)ds$ then

$$\begin{aligned}\psi'(t) &= \beta(t)\varphi(t) \leq \beta(t)\alpha(t) + \beta(t)\psi(t)\\ &= \beta(t) \cdot (\alpha(t) + \psi(t))\end{aligned}$$

.

Recall $\dot{x} + a(t)x + b(t) = 0$

$$\begin{aligned}(\dot{\psi}(t) - \beta(t)\psi(t))e^{-\int_a^t \beta(s)ds} &\leq \beta(t)\alpha(t) \cdot e^{-\int_a^t \beta(s)ds}\\ (e^{-\int_a^t \beta(s)ds}\psi(t))' &\leq \beta(t)\alpha(t) \cdot e^{-\int_a^t \beta(s)ds}.\end{aligned}$$

Integrating gives :

$$(e^{-\int_a^t \beta(s)ds}\psi(t)) \overset{\psi(a)=0}{\leq} \int_a^t \beta(s)\alpha(s)e^{-\int_a^s \beta(r)dr}ds.$$

$\square$

Definition 2.1.2 (Regularity)**.** A function $K : \mathbb{R}^{2d} \to \mathbb{R}^d$ is called regular if :

1. $K \in \mathcal{C}^1(\mathbb{R}^{2d}; \mathbb{R}^d)$ (gives local lipschitz)
2. And $\exists L > 0$ s.t. :

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

Remark. We further assume K has the following properties :

$$\begin{aligned} K(x,y) &= -K(y,x) && \text{(antisymmetric)} \\ K(x,x) &= 0. \end{aligned}$$

Theorem 2.1.4. For regular K the MPS has a solution for all $T > 0$

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

has a unique solution by Picard-Iteration :

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

Definition 2.1.3 (Empirical Measure of a System)**.** Consider the point measure for every x_i : $\delta_{x_i(t)}$, then the measure of the System of order N is given by

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

As shown in the introduction μ_N is a (weak-) solution to the following PDE

$$\partial_t \mu_N + \nabla \cdot (\mu_N \cdot \int K(\cdot, y) d\mu_N(y)) = 0.$$

Intuition. Now for $N \to \infty$ if we have $\mu_N \xrightarrow{\text{in some sense}} \mu$ then μ is a (weak) solution to

$$\partial_t \mu + \nabla \cdot (\mu \cdot \int K(\cdot, y) d\mu(y)) = 0.$$

with

$$\mu_0 \leftarrow \mu_N(0).$$

2.2 Weak Solutions and Distributions

Distributions are a more general class of functions and can be seen as the dual space of the space of test functions

Definition 2.2.1 (Multi-Index)**.** A multi-index $\gamma \in \mathbb{N}_0^n$ of length $|\gamma| = \sum_i \gamma_i$ for example $\gamma = (0, 2, 1) \in \mathbb{N}_0^3$ can be used to denote partial derivatives of higher order as such :

$$\partial^\gamma = \prod_i (\frac{\partial}{\partial x_i})^{\gamma_i}.$$

Remark. Only sensible cause partial derivatives commute as otherwise the index would be ambiguous.

Definition 2.2.2 (Test Functions)**.** For $\Omega \subset \mathbb{R}^d$ the space of test functions $\mathcal{D}(\Omega) \supset \mathcal{C}_0^\infty(\Omega)$. We say a sequence of test functions $(\varphi_m)_{m\in\mathbb{N}} \subset \mathcal{C}_0^\infty(\Omega)$ converges against some limit $\varphi \in \mathcal{C}_0^\infty(\Omega)$ iff.

1. $\exists$ a compact set $K \subset \Omega$ s.t. $\operatorname{supp} \varphi_m \subset K$ for all $m \in \mathbb{N}$
2. $\forall$ multi-indexes $\alpha \in \mathbb{N}_0^n$:

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

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

Definition 2.2.3 (Distribution)**.** The space of distributions $\mathcal{D}(\Omega)'$ is the dual space of $\mathcal{D}(\Omega)$ i.e. $\mathcal{D}(\Omega)'$ contains all the continuous linear functionals T

$$T : \mathcal{D}(\Omega) \to \mathbb{K}.$$

Remark. Continuity refers to the notion that for a sequence $(\varphi_m)_{m\in\mathbb{N}} \subset \mathcal{D}(\Omega)$ with limit φ then :

$$\varphi_m \to \varphi \;\Rightarrow\; T(\varphi_m) \to T(\varphi).$$

linearity :

$$T(\alpha\varphi_1 + \beta\varphi_2) = \alpha T(\varphi_1) + \beta T(\varphi_2).$$

We sometimes write $\langle T, \varphi\rangle$ instead of $T(\varphi)$

Definition 2.2.4 (Convergence)**.** For a sequence of distributions $(T_m)_{m\in\mathbb{N}} \subset \mathcal{D}(\Omega)'$ we say it converges against a limit $T \in \mathcal{D}(\Omega)$ iff

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

Example. Every locally integrable function $f \in L^1_{\text{loc}}(\Omega) := \{f \mid \forall K \subset \Omega, \int_K f(x)dx < \infty\}$ defines a Distribution by :

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

i.e. $L^1_{\text{loc}}(\Omega) \subset \mathcal{D}'(\Omega)$

(Probability -) Measures $\mu \in \mathcal{M}(\Omega)$ define a distributions , by :

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

A prominent example is the δ distribution defined by :

$$\langle \delta, \varphi \rangle = \int_{\mathbb{R}^d} \varphi(x)d\delta = \varphi(0).$$

Remember for a measurable set E

$$\delta_x(E) = \begin{cases} 1, & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}.$$

We further introduce a class of operators called Mollifiers, which take non-smooth functions (like step functions) and smooth them out. In our case they prove useful as we can define test functions with desired properties easily, since the mollifier-kernel as defined below is required to be $\mathcal{C}_0^\infty(\mathbb{R}^d)$ the resulting function is a test function.

Definition 2.2.5 (Mollifier-Kernel)**.** A mollifier is given by a function $j \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ with the following properties

1. $j \geq 0$
2. $\operatorname{supp} j \subset \overline{B}_1(0)$
3. $\int_{\mathbb{R}^d} j(x)dx = 1$

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

Using the above example we can define a class of mollifiers called Standard Mollifier as follows

Definition 2.2.6 (Standard Mollifier)**.** For $\varepsilon > 0$ define the standard mollifier by

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

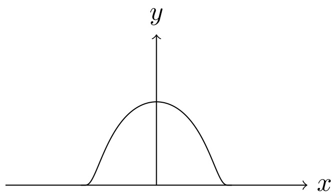


Figure 2.1: Example of a Standard Mollifier

Exercise. Proof that the standard mollifier converges to the delta distribution

$$j_\varepsilon \xrightarrow{\varepsilon \to 0} \delta_0.$$

Mollification is then defined as the convolution of a locally integrable function and our standard mollifier

Definition 2.2.7 (Mollification Operator)**.** $\forall u \in L^1_{\text{loc}}(\mathbb{R}^d)$ define

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

Lemma 2.2.1. If $u \in L^1(\mathbb{R}^d)$ and $\operatorname{supp} u$ is compact in $\mathbb{R}^d$ then for all fixed $\varepsilon > 0$

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

Furthermore if $u \in \mathcal{C}_0(\mathbb{R}^d)$ then :

$$J_\varepsilon(u) = j_\varepsilon \xrightarrow{\varepsilon \to 0} u \quad \text{uniformly in } \operatorname{supp} u.$$

Proof. Let $K = \operatorname{supp} u \subset \mathbb{R}^d$ compact then

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

$\forall i \in \{1, \dots d\}$ consider

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

Wether we can switch the derivative and the integral depends on wether we can bound the following

$$\frac{\partial}{\partial x_i} j_\varepsilon(x-y)|_K \leq \frac{C(j')}{\varepsilon^{d+1}} < \infty.$$

We only need to consider K, since the support is limited to K by property of u and j, then by DCT we can switch derivative and integral infinite many times

For $u \in \mathcal{C}_0(\mathbb{R}^d)$ we want to proof that

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

For any $x \in \operatorname{supp} u$ it suffices to show that the convergence does not depend on x

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

Note the first equality follows since the integral of any mollifier is equal to 1 and it acts as a "smart 1"

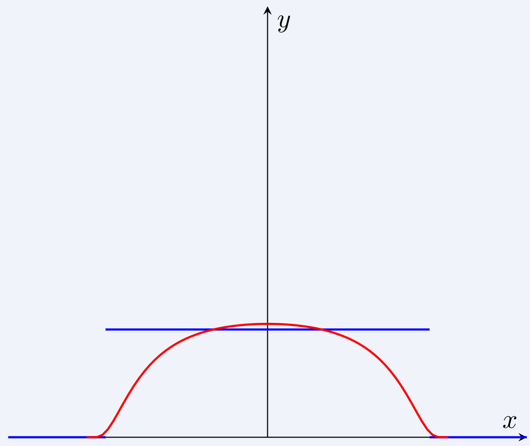
where we have used in the last step that $u \in C(\operatorname{supp} u)$ is uniformly continuous i.e $\forall \eta > 0$, $\exists \delta > 0$ s.t $\forall\ x, y \in \operatorname{supp} u$ and $|x - y| < \delta$ we have

$$|u(x) - u(y)| < \eta.$$

$\square$

An example of how Mollification can be useful consider the following example :

Example. Consider the following step function $\mathbb{1}_{[-1,1]}(\cdot)$, then the mollification $J_\varepsilon(u)$ will look like this



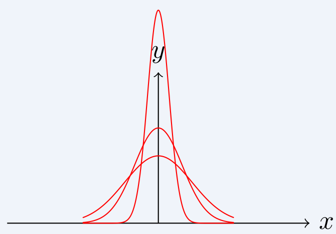
Exercise. Proof that the Mollification of the step function $\mathbb{1}_{[-1,1]}(\cdot)$ is a smooth function

Coming back to our empirical measure from 2.1.3, we can see the corresponding distribution is defined by :

$$\langle \mu_n, \varphi \rangle = \frac{1}{N} \sum_{i=1}^{N} \varphi(x_i).$$

Some examples in approximation of δ distribution

Example (Heat Kernel). The heat kernel $f_t(x) = \frac{1}{2\sqrt{\pi t}} e^{-\frac{x^2}{4t}}$ approximates the δ distribution



Proof.

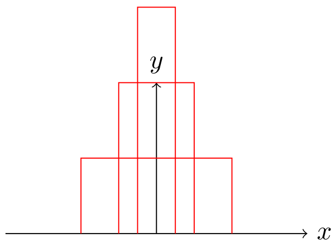
$$\begin{aligned}
\lim_{t\to 0+}\langle f_t,\varphi\rangle &= \lim_{t\to 0+}\int_{\mathbb{R}} f_t(x)\varphi(x)dx = \lim_{t\to 0+}\int_{-\infty}^{\infty}\frac{1}{2\sqrt{\pi t}}e^{-\frac{x^2}{4t}}\varphi(x)dx \\
&= \lim_{t\to 0+}\int_{-\infty}^{\infty}\frac{1}{\sqrt{\pi}}e^{-y^2}\varphi(2ty)dy \\
&\overset{*}{=}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

with the transformation $\frac{x}{\sqrt{2t}} = y$ □

Exercise. Show the change of integration and limit is valid in *

Further examples include :

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



And the dirichlet kernel

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

To define the notion of a distribution solving a PDE we need to first define the way we take the derivatives of distributions

Definition 2.2.8 (Weak derivative of Distributions)**.** $\forall T \in \mathcal{D}(\Omega)'$. $\partial_i T$ is given by

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

We first show this for all distributions that are defined by a $f \in L^1_{\text{loc}}$, every other distribution T also has to satisfy this property.

Exercise. Proof the above equality for distributions $f \in L^1_{\text{loc}}$ and show for arbitrary distribution T that :

$$-\langle T, \partial_i \varphi \rangle.$$

is continuous and linear.

Hint : Integration by parts; Why does it vanish on the boundary ?

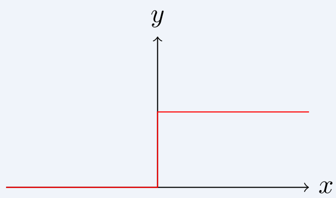
Example. For the δ distribution :

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

Example. Heaviside The Heaviside step function is defined by :

$$H = \begin{cases} 1, & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \in L^1_{\text{loc}}.$$

$$\begin{aligned} \langle H', \varphi \rangle &:= -\langle H, \varphi' \rangle = -\int_{-\infty}^{\infty} H(x)\varphi'(x)dx \\ &= -\int_0^{\infty} \varphi'(x)dx = \varphi(0) = \langle \delta, \varphi \rangle. \end{aligned}$$



Using all the above we can rewrite our many particle system (MPS) by using the empiric measure and distributions

$$\begin{cases} \frac{d}{dt} x_i &= \langle K(x_i, \cdot), \mu_N \rangle = \int K(x_i, y) d\mu_N(y) \\ x_i(0) &= x_{i,0} \end{cases}.$$

Definition 2.2.9 (Weak Solution of MFE)**.** We say μ is a weak solution of the Mean-Field-Equation (MFE) iff for $\forall t \in [0, T]$, $\mu_t \in \mathcal{M}(\mathbb{R}^d)$ satisfies for all test functions $\forall\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ the following equation

$$\langle \mu_t, \varphi\rangle - \langle \mu_0, \varphi\rangle = \int_0^t \langle \mu_s K \mu_s, \varphi\rangle ds.$$

Remark. If $\mu_0 = \mu_{N,0}$ then $\mu_{N,t}$ is a weak solution of the MFE

Theorem 2.2.1. Let the empirical measure 2.1.3 be denoted by μ_N then for "good" (regular ?) $K(x,y)$ we have

$$\frac{d}{dt}\langle \mu_N, \varphi\rangle = \langle \mu_N \underbrace{\int K(x,y) d\mu_N(y)}_{=K\mu_N}\rangle = -\langle \nabla \cdot \nabla(\mu_N K \mu_N), \varphi\rangle.$$

Such that μ_N is a weak solution of the MPDE

$$\partial_t \mu_N + \nabla \cdot \nabla(\mu_N K \mu_N) = 0.$$

Remark. Note when we talk about weak solution, it means the PDE is solved in the sense of distributions.

Exercise. Show $\mu_N K \mu_N$ as defined above is a distribution for regular/ good $K(x,y)$

Proof. □

Definition 2.2.10 (characteristic problem for MFE)**.** The corresponding characteristic is given by :

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

Notation (Push Forward)**.** For a measurable map $X : (\mathbb{R}^d, \mathcal{B}) \xrightarrow{X} (\mathbb{R}^d, \mathcal{B})$ and a measure $\mu_0 \in \mathcal{M}(\mathbb{R}^d)$ we have :

$$\forall B \in \mathcal{B}\ ,\ X\#\mu_0 = \mu_0(X^{-1}(B)).$$

Exercise. Show that if $x(t, x_0, \mu_0) \in \mathcal{C}^1(\mathbb{R}, \mathbb{R}^d)$ exists then $\mu_t = x(t, \cdot, \mu_0)\#\mu_0$ is a weak solution of MFE

Definition 2.2.11. Space of probability measures with bounded first moment

$$\mathcal{P}_1(\mathbb{R}) := \{\mu_0 \in \mathcal{M}_+(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x| d\mu_0(x) < \infty\}.$$

Theorem 2.2.2 (Uniqueness of Solution)**.** For regular K (2.1.2) and $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ then the characteristic problem 2.2.10 has a unique solution $x(t, x_0, \mu_0) \in \mathcal{C}^1(\mathbb{R}, \mathbb{R}^d)$ and $\mu_T \in \mathcal{P}_1(\mathbb{R}^d)$, for all $t > 0$

Proof. Existence

We consider

$$\begin{aligned} x(t, x_0, \mu_0) &= x_0 + \int_0^t \int_{\mathbb{R}^d} K(x(s, x_0, \mu_0), y) d\mu_s(y) ds \\ &\overset{\text{psh frwd.}}{=} x_0 + \int_0^t \int_{\mathbb{R}^d} K(x(s, x_0, \mu_0), x(s, \zeta, \mu_0)) d\mu_0(\zeta) ds. \end{aligned}$$

We define the following iteration for all $y \in \mathbb{R}^d$

$$\begin{aligned} x_0(t, y) &= y \\ x_1(t, y) &= y + \int_0^t \int_{\mathbb{R}^d} k(x_0(s, y), x_0(s, \zeta)) d\mu_0(\zeta) ds \\ &\vdots \\ x_n(t, y) &= y + \int_0^t \int_{\mathbb{R}^d} k(x_{n-1}(s, y), x_{n-1}(s, \zeta)) d\mu_0(\zeta) ds \\ &. \end{aligned}$$

Similar to our proof in we show the sequence is cauchy :

$$\begin{aligned} |x_n(t, y) - x_{n-1}(t, y)| &\leq \int_0^t \int_{\mathbb{R}^d} |K(x_{n-1}(s, y), x_{n-1}(s, \zeta)) - K(x_{n-2}(s, y), x_{n-2}(s, \zeta))| d\mu_0(\zeta) ds \\ &\leq L \int_0^t \int_{\mathbb{R}^d} |x_{n-1}(s, y) - x_{n-2}(s, y)| + |x_{n-1}(s, \zeta) - x_{n-2}(s, \zeta)| d\mu_0(\zeta) ds \\ &. \end{aligned}$$

To get rid of the ζ we define the following banach space $\mathcal{X} = \{v \in$

$\mathcal{C}(\mathbb{R}^d;\mathbb{R}^d) : \sup_x \frac{|v(x)|}{1+|x|} < \infty\}$ with norm

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

We can then further approximate by :

$$\begin{aligned} &L\int_0^t \int_{\mathbb{R}^d} |x_{n-1}(s,y) - x_{n-2}(s,y)| + |x_{n-1}(s,\zeta) - x_{n-2}(s,\zeta)| d\mu_0(\zeta) ds \\ &\le L\int_0^t |x_{n-1}(s,y) - x_{n-2}(s,y)| + \|x_{n-1}(s,\cdot) - x_{n-2}(s,\cdot)\|_{\mathcal{X}}(1+C_1) ds. \end{aligned}$$

Where $C_1 = \int |x| d\mu_0(x_0)$ is the first moment of our initial measure. Now we divide both sides of the inequality by $1+|y|$, and take the supremum in y

$$\|x_n(t,\cdot) - x_{n-1}(t,\cdot)\|_{\mathcal{X}} \le L(2+C_1)\int_0^{|t|} \|x_{n-1}(s,\cdot) - x_{n-2}(s,\cdot)\|_{\mathcal{X}} ds.$$

Then for $\forall n > m \gg 1$ we have

$$\|x_n(t,\cdot) - x_m(t,\cdot)\|_{\mathcal{X}} \le \sum_{i=m}^{n-1} \|x_{i+1}(t,\cdot) - x_i(t,\cdot)\|_{\mathcal{X}} \xrightarrow{m\to\infty} 0.$$

Therefore $(x_n(t,\cdot))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{X}$.

Now suppose $x_n(t,\cdot) \to x(t,\cdot)$:

$$\begin{aligned} x(t,y) &= y + \int_0^t \int_{\mathbb{R}^d} K(x(s,y), x(s,\zeta)) d\mu_0(\zeta) ds \\ &= y + \int_0^t \int_{\mathbb{R}^d} K(x(s,y), z) d\mu_0(z) ds \end{aligned}$$

.

This concludes the **Existence** proof

Uniqueness :

This proof closely mimics the one presented in by using the space $\mathcal{X}$ □

Remark. Showing the convergence of our Picard Iteration here is slightly more complicated, forcing us to use a different norm to get a simpler estimate to work with, remember similar trick as in functional analysis with

$$\|f\|_L = \sup_{t\in[0,1]} e^{-Lt}|f(t)|.$$

Exercise. Show that $\mathcal{X} = \{v \in \mathcal{C}(\mathbb{R}^d; \mathbb{R}^d) : \sup_x \frac{|v(x)|}{1+|x|} < \infty\}$ with norm

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

is a banach space

Hint : Compare to supremums norm

2.3 Wasserstein Distance

2.3.1 Goal

The goal of this section is to consider as $N \to \infty$ how the empirical measure $\mu_{N,\cdot}$ converges

$$\mu_{N,0} \overset{?}{\to} \mu_0$$
$$\mu_{N,t} \overset{?}{\to} \mu_t.$$

we have already shown that for arbitrary given measure μ_0 (on both sides of the arrows) the PDE problem is uniquely solved, the idea of the Mean Field Limit problem is to prove a stability result for the above convergence.

2.3.2 Weak Convergence of Measure (Wasserstein Distance)

Definition 2.3.1 (Weak Setting of PDE problem)**.** For all test functions $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$:

$$\int_{\mathbb{R}^d} \varphi(x) d\mu(t,x) - \int_{\mathbb{R}^d} \varphi(x) d\mu_0 = \int_0^t \int_{\mathbb{R}^d} K\mu(s,x) \nabla\varphi(x) d\mu(s,x).$$

Where

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

To give a small recap of what we have done so far :

1. If $\mu_0 = \mu_{N,0}$ then $\mu_{N,t}$ is a weak solution of the above PDE
2. Solve the PDE for given $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ for regular K then

$$\mu_t = x_t \# \mu_0.$$

   is a weak solution of the PDE

The next goal is to consider the problem

$$\text{if } \mu_{N,0} \to \mu_0 \quad \text{then } \mu_{N,t} \to \mu_t.$$

$\Leftrightarrow$ stability of PDE

Definition 2.3.2 (Wasserstein Distance). For all $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, $(p \geq 1)$ the Wasserstein Distance of μ and ν is given by

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

Where

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

Exercise. For two deterministic measures δ_x, δ_y prove

$$W^1(\delta_x, \delta_y) = |x-y|.$$

Remark. If $\varphi, \psi \in \mathcal{C}(\mathbb{R}^d)$ s.t :

$$\begin{aligned} \varphi(x) &\sim |x|^p \quad \forall\ |x| \gg 1 \\ \psi(x) &\sim |y|^p \quad \forall\ |y| gg1. \end{aligned}$$

Then

$$\int\int_{\mathbb{R}^{2d}} (\varphi(x) + \psi(y))\pi(dx,dy) = \int_{\mathbb{R}^d} \varphi(x) d\mu(x) + \int_{\mathbb{R}^d} \psi(y) d\nu(y).$$

Corollary (Kontonovich-Rubinstein duality).

$$\begin{aligned} \underset{MK,1}{\operatorname{dist}}(\mu,\nu) &= W^1(\mu,\nu) \\ &= \sup_{\varphi \in \operatorname{Lip}(\mathbb{R}^d)} \left| \int_{\mathbb{R}^d} \varphi(x) d\mu(x) - \int_{\mathbb{R}^d} \varphi(x) d\nu(x) \right|. \end{aligned}$$

$\operatorname{Lip}(\varphi) = 1$

The proof of the above can be found in the book xyz

Theorem 2.3.1 (Dobrushin's stability). Let $\mu_0, \overline{\mu}_0 \in \mathcal{P}_1(\mathbb{R}^d)$
Then let $(x(t,\cdot,\mu_0), \mu_t(\cdot)), (x(t,\cdot,\overline{\mu}_0), \overline{\mu}_t(\cdot))$ be solutions of the corresponding PDE problem. For arbitrary $\forall\ t > 0$ it holds that the distance

$$\underset{MK,1}{\operatorname{dist}}(\mu_t, \overline{\mu}_t) \leq e^{2|t|L} \underset{MK,1}{\operatorname{dist}}(\mu_0, \overline{\mu}_0).$$

Proof. For initial data $\mu_0, \overline{\mu}_0$ we want to compare the trajectories

$$\begin{aligned}
&x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0) \\
&= x_0 - \overline{x}_0 + \int_0^t \int_{\mathbb{R}^d} K(x(s, x_0, \mu 0), x(s, z, \mu_0)) d\mu_0(z) ds \\
&\qquad - \int_0^t \int_{\mathbb{R}^d} K(x(s, \overline{x}_0, \overline{\mu}_0), x(s, \overline{z}, \overline{\mu}_0)) d\overline{\mu}_0(\overline{z}) ds.
\end{aligned}$$

We need to combine the above two integrals together, while the time integrals are the same, but the space integral has to be converted into 2d dimensions by inserting $1 = \int_{\mathbb{R}^d} d\mu_0(z)$

$$\begin{aligned}
x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0) + \int_0^t \iint & K(x(s, x_0, \mu 0), x(s, z, \mu_0)) \\
& - K(x(s, \overline{x}_0, \overline{\mu}_0), x(s, \overline{z}, \overline{\mu}_0)) d(\mu_0 \times \overline{\mu}_0)(z, \overline{z}) ds.
\end{aligned}$$

Let $\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)$ then we can estimate point-wise for fixed $x_0, \overline{x}_0$:

$$\begin{aligned}
&\|x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0)\| \\
&\leq \|x_0 - \overline{x}_0\| + L \int_0^t \iint_{\mathbb{R}^{2d}} \|x(s, x_0, \mu_0) - x(s, \overline{x}_0, \overline{\mu}_0)\| \\
&\qquad + \|x(s, z, \mu_0) - x(s, \overline{z}, \overline{\mu}_0)\| d\pi_0(z, \overline{z}) ds.
\end{aligned}$$

Now taking the integral on both sides :

$$\begin{aligned}
&\iint_{\mathbb{R}^{2d}} \|x(t, x_0, \mu_0) - x(t, \overline{x}_0, \overline{\mu}_0)\| d\pi_0(x_0, \overline{x}_0) \\
&\leq \iint_{\mathbb{R}^{2d}} \|x_0 - \overline{x}_0\| d\pi_0(x_0, \overline{x}_0) \\
&+ L \int_0^t \iint_{\mathbb{R}^{2d}} \|x(s, x_0, \mu_0) - x(s, \overline{x}_0, \overline{\mu}_0)\| d\pi_0(z, \overline{z}) ds \\
&+ L \int_0^t \int \int_{\mathbb{R}^{2d}} \|x(s, z, \mu_0) - x(s, \overline{z}, \overline{\mu}_0)\| d\pi_0(z, \overline{z}) d\pi_0(z, \overline{z}) ds.
\end{aligned}$$

Now we define :

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

We obtain that the distance based on the measure π_0 at time t :

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

We can now use Gronwalls inequality

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

Where the above inequality holds for arbitrary $\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)$

$$\inf_{\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)} D[\pi_0](t) \leq \inf_{\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)} D[\pi_0](0) \cdot e^{2Lt} = \operatorname*{dist}_{MK,1} .$$

Now all we need to show is that

$$\operatorname*{dist}_{MK,1} = \inf_{\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)} D[\pi_0](t) = \inf_{\pi_t \in \Pi(\mu_t, \overline{\mu}_t)} \int \int |x - \overline{x}| d\pi_t(x, \overline{x}).$$

It remains to be prove that for

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

the push forward measure $\varphi_t \# \pi_0 \in \Pi(\mu_t, \overline{\mu}_t)$ for $\forall\ \pi_0 \in \Pi(\mu, \overline{\mu}_0)$ $\square$

Exercise. Prove that for arbitrary $\pi_0 \in \Pi(\mu_0, \overline{\mu}_0)$ the push forward measure $\varphi_t \# \pi_0 \exists \Pi(\mu_t, \overline{\mu}_t)$ for

$$\varphi_t(x_0, \overline{x}_0) : \mathbb{R}^{2d} \to \mathbb{R}^{2d} \ \varphi_t(x_0, \overline{x}_0) = (x(t, x_0, \mu_0), x(t, \overline{x}_0, \overline{\mu}_0)).$$

We are also interested in what happpens when the initial data is good

Corollary. If μ_0 has a density $f_0 \in L^1(\mathbb{R}^d)$a with finite first moment :

$$\int_{\mathbb{R}^d} |x| f_0(x) dx < \infty.$$

Then the Cauchy problem

$$\begin{aligned} \partial_t f + \nabla \cdot (fkf) &= 0 \\ f_0|_{t=0} &= f_0. \end{aligned}$$

has a unique weak solution $f \in \mathcal{C}(\mathbb{R}; L^1(\mathbb{R}^d))$ i.e $\forall\ \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ it holds :

$$\int_{\mathbb{R}^d} \varphi(x) f(x,t) dx - \int_{\mathbb{R}^d} \varphi(x) f_0(x) dx = \int_0^t \int_{\mathbb{R}^d} f(x,t) K f(x,t) \nabla \varphi(x) dx ds.$$

and for all $t \in \mathbb{R}$

$$\|f(\cdot, t)\|_{L^1(\mathbb{R}^d)} \in \mathcal{C}(\mathbb{R}).$$

Proof. $\forall t \in \mathbb{R}$ proving $f_t \in L^1(\mathbb{R}^d)$ is left as an exercise

The weak formulation of the problem is :

$$\int_{\mathbb{R}^d} \varphi(x) f(t,x) dx = \int_{\mathbb{R}^d} \varphi(x) f(t,x) dx + \int_0^t \int_{\mathbb{R}^d} f(s,x) K f(s,x) \nabla \varphi(x) dx ds.$$

The idea is to choose a special test function φ as follows

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

to get the above desired function it suffices to take the mollification of the indicator function $\mathbb{1}_{[-R,R]}$ with $\frac{R}{2}$ we then get

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

i.e for $F_R(t) = \int_{\mathbb{R}^d} f(t,x)\varphi_R(x)dx$ we get

$$|F_R(t) - F_R(\hat{t})| \leq C \cdot |t-\hat{t}| \xrightarrow{t\to\hat{t}} 0.$$

on the other hand

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

This limit is a result of how we defined $\varphi_R(x)$, we choose $\varphi_R(x)$ such that it converges against 1 as $R \to \infty$ pointwise □

Exercise. We already know the weak-solution μ_t exists such that it remains to show that for $\forall\ t \in \mathbb{R}$, $\mu_t \in \mathcal{P}_1(\mathbb{R}^d)$ is absolute continuous with respect to the Lebesgue measure, prove this statement

Exercise. Show that $|\nabla\varphi_R(x)| \leq \frac{C}{R}$

Exercise. Proof that $\varphi_R(x) \to 1$ as $R \to \infty$

Theorem 2.3.2 (mean field limit). For arbitrary initial data $\forall\ f_0 \in L^1(\mathbb{R}^d)$

, let $\mu_{N,0} = \frac{1}{n}\sum_{i=1}^{N}\delta_{x_{i,0}}$ s.t the distance

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

and $x_N(t)$ be the solution of many particle system wiht ID $x_{i,0}$ then the corresponding empirical measure $\mu_{N,t} = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$ it holds :

$$\underset{MK,1}{\text{dist}}(\mu_{N,t}, f_t) \leq e^{2Lt}\underset{MK,1}{\text{dist}}(\mu_{N,0}, f_0).$$

Where f_t is the corresponding density resulting from the previous corollary of the PDE weak solution.
and furthermore $\mu_{N,t} \xrightarrow{N\to\infty} f(\cdot, t)$ weakly in measure i.e

$$\forall \varphi \in \mathcal{C}_b(\mathbb{R}^d) \ : \ \int \varphi d\mu_{N,t} \to \int \varphi(x) f_t(x) dx.$$

Proof. We have to show that $\forall\ \varphi \in C_0(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \varphi d\mu_{N,t} \to \int_{\mathbb{R}^d} \varphi f_t dx.$$

means $\forall\ \varphi \in Lip(\mathbb{R}^d)$ it holds

$$\int \varphi d\mu_{N,t} \to \int \varphi f_t dx.$$

The proof relies on using ε-δ language and combining that with our mollification results
$\forall\ \varepsilon > 0$ set $R \gg 1$ s.t $\frac{2C\|\varphi\|_\infty}{2R} \leq \frac{\varepsilon}{2}$, and let $\varphi_m \in \mathcal{C}_0^\infty(B_{2R})$ s.t $\exists\ M > 0\ \forall\ m > M$

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

For $\varphi_{M+1} \in \mathcal{C}_0^\infty(B_{2R})$ there $\exists\ N_1$ s.t $\forall\ N > N_1$

$$\left|\int \varphi_{M+1}(d\mu_{N,t} - f_t dx)\right| < \frac{\varepsilon}{4}.$$

Using the above

$$
\begin{aligned}
&\left|\int \varphi d\mu_{N,t} - \int \varphi f_t dx\right| \\
&\leq \left|\int_{B_R} \varphi(x)(d\mu_{N,t} f_t dx)\right| + \left|\int_{B_R^C} \varphi(x)(d\mu_{N,t} - f_t d(x))\right| \\
&\overset{\text{Tri.}}{\leq} \left|\int_{B_R} \varphi(x)_{M+1}(d\mu_{N,t} - f_t dx)\right| + \left|\int_{B_R} (\varphi_{M+1} - \varphi)(d\mu_{N,t} - f_t dx)\right| \\
&+ \left|\int_{B_R^c} \varphi(x)\frac{|x|}{R}(d\mu_{N,t} - f_t dx)\right| \\
&\leq \frac{C}{R} < \frac{\varepsilon}{2}.
\end{aligned}
$$

□

Exercise. Find a sequence of empirical measures that converges against a density.

Chapter 3

Stochastic Mean Field Particle Systems

From now on let the underlying probability space be given by $(\Omega, \mathcal{F}, \mathbb{P})$.

3.1 Basics of probability

Definition 3.1.1 (Brownian Motion)**.** Real valued stochastic process $W(\cdot)$ is called a Brownian motion (Wiener process) if

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

Remark (Properties)**.** 1. $\mathbb{E}[W(t)] = 0$, $\mathbb{E}[W(t)^2] = t$, for all $t > 0$

2. $\mathbb{E}[W(t)W(s)] = t \wedge s$ a.s
3. $W(t) \in \mathcal{C}^\gamma[0, T]$, $\forall 0 < \gamma < \frac{1}{2}$.
4. $W(t)$ is nowhere differentiable a.s

additionally Brownian motions are martingales and satisfy the Markov property

Definition 3.1.2 (Progressively measurable)**.** In addition to adaptation of a Stochastic process X_t we say it is progressively measurable w.r.t $\mathcal{F}_t$ if $X(s, \omega) : [0, t] \times \Omega \to \mathbb{R}$ is $\mathcal{B}[0, t] \times \mathcal{F}_t$ measurable, i.e the t is included

Definition 3.1.3 (Simple functions)**.** Instead of $\mathcal{H}^2$ she uses $\mathbb{L}^2(0,T)$ is the space of all real-valued progressively measurable proceses $G(\cdot)$ s.t

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

define $\mathbb{L}$ analog

Definition 3.1.4 (Step Process)**.** $G \in \mathbb{L}^2(0,T)$ is called a step process when Partition of $[0,T]$ exists s.t $G(t) = G_k$ for all $t_k \le t \le t_{k+1}$, $k = 0, \dots, m-1$ note G_k is $\mathcal{F}_{t_k}$ measurable R.V.

For step process we define the ito integral as a simple sum

Definition 3.1.5 (Ito integral for step process)**.** Let $G \in \mathbb{L}^2(0,T)$ be a step process is given by

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

We take the left value of the process such that we converge against the right integral later

Remark. For two step processes $G, H \in \mathbb{L}^2(0,T)$ for all $a, b \in \mathbb{R}$, we have linearity (note they may have two different partitions, so we need to make a bigger (finer) one to include both,)

1. $\int_0^T (aG + bH)dW_t = a \int G + b \int H$
2. $\mathbb{E}[\int_0^T GdW_t] = 0$, because the Brownian motion has EV of 0
3. $\mathbb{E}[(\int_0^T GdW_t)^2] = \mathbb{E}[\int_0^T G^2 dt]$ Ito isometry

Proof. First property is just defining a new partition that includes both process. Second property, the Idea of the proof is that

$$\begin{aligned}\mathbb{E}[\int_0^t GdW_t] &= \mathbb{E}[\sum_{k=0}^{m-1} G_k(W_{t_{k+1}} - W_{t_k})] \\ &= \sum_{k=0}^{m-1} \mathbb{E}[G_k(W(t_{k+1}) - W(t_k))]\end{aligned}$$

.

Remember $G_k \sim \mathcal{F}_{t_k}$ m.b. and $W(t_{k+1}) - W(t_k)$ is mb. wrt to $W^t(t_k)$ future sigma algebra and it is independent of $\mathcal{F}_{t_k}$ s.t the expectation decomposes

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

For the variance decompose into square and non square terms , the non square terms dissapear by property 2 the rest follows by the variance of Brownian motion, be careful of which terms are actually independent , at leas one will always be independent of the other 3 □

Definition 3.1.6 (Ito Formula)**.** If $u \in \mathcal{C}^{2,1}(\mathbb{R} \times [0,T]; R)$ then

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

For $dX = Fdt + GdW_t$ for $F \in L^1([0,T])$, $G \in L^2([0,T])$

Proof. The proof is split into the steps

1. $$\begin{aligned} d(W_t^2) &= 2W_t dW_t + dt \\ d(tW_t) &= W_t dt + tdW_t. \end{aligned}$$

2. $$\begin{aligned} dX_i &= F_i dt + G_i dW_t \\ d(X_1, X_2) &= X_2 dX_1 + X_1 dX_2 + G_1 G_2 dt \end{aligned}$$
.

3. $$u(x) = x^m \quad m \geq 2.$$

4. Itos formula for $u(x,t) = f(x)g(t)$ where f is a polynomial

I.e we prove the Ito formula for functions of the form $u(x) = x^m$ and then Step 1 :

1. $d(W_t^2) = 2W_t dW_t + dt$ which is equivalent to $W^2(t) = W_0^2 + \int_0^t 2W_s dW_t + \int_0^t ds$

2. $d(tW_t) = W_t dt + t dW_t$ which is equivalent to $tW(t) - sW(0) = \int_0^t W_s ds + \int_0^t s dW_s$

Actually $\forall$ a.e $\omega \in \Omega$:

$$2\int_0^t W_s dW_s = 2 \lim_{n\to\infty} .$$

Now we prove (2) $tW_t - 0W_0 = \int_0^t W_s ds + \int_0^t s dW_s$

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

We choose the right value for the second integral

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

Its product rule

$$\begin{aligned} dX_1 &= F_1 dt + G_1 dW_t \\ dX_2 &= F_2 dt + G_2 dW_t. \end{aligned}$$

This can be written as

$$d(X_1, X_2) = X_2 dX_1 + X_1 dX_2.$$

this shorthand notation actually means

$$\begin{aligned} X_1(t)X_2(t) - X_1(0)X_2(0) = &\int_0^t X_2 F_1 ds + \int_0^t X_2 G_1 dW_s \\ &+ \int_0^t X_1 F_2 ds + \int_0^t X_1 G_2 dW_s \\ &+ \int_0^t G_1 G_2. \end{aligned}$$

We prove for F_1, F_2, G_1, G_2 are time independent

$$
\begin{aligned}
&\int_0^t (X_2 dX_1 + X_1 dX_2 + G_1 G_2 ds) \\
&= \int_0^t (X_2 F_1 + X_1 F_2 + G_1 G_2) ds + \int_0^t (X_2 G_1 + X_1 G_2) dW_s \\
&= \int_0^t (\underbrace{F_2 F_1 s + F_1 G_2 W_s}_{=X_2} + \underbrace{F_1 F_2 s + F_2 G_1 W_s}_{=X_1} + G_1 G_2) ds \\
&+ \int_0^t (F_2 G_1 s + G_2 G_1 W_s + F_1 G_2 s + G_1 G_2 W_s) dW_s \\
&= G_1 G_2 t + F_1 F_2 t^2 + (F_1 G_2 + F_2 G_1) \underbrace{\left(\int_0^t W_s ds + \int_0^t s dW_s \right)}_{tW_t} + 2 G_1 G_2 \underbrace{\int_0^t W_s dW_s}_{W_t^2 - t} \\
&= G_1 G_2 t + F_1 F_2 t^2 + (F_1 G_2 + F_2 G_1) t W_t + G_1 G_2 W_t^2 - G_1 G_2 t \\
&= X_1(t) \cdot X_2(t).
\end{aligned}
$$

Where $X_2(t) = \int_0^t F_2 ds + \int_0^t G_2 dW_s \overset{\text{Cons.}}{=} F_2 t + G_2 W_t$

Extend the above idea by considering step processes (F_1, F_2, G_1, G_2)instead of time independent. Step processes are constant (related to time) and we can use the above prove for every time step t and just consider a summation over it.

For general $F_1, F_2 \in L^1(0,T), G_1, G_2 \in L^2(0,T)$ then we take step processes to approximate them

$$
\begin{aligned}
&\mathbb{E}[\int_0^T |F_i^n - F_i| dt] \to 0 \\
&\mathbb{E}[\int_0^T |G_i^n - G_i|^2 dt] \to 0
\end{aligned}
$$

.

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

It holds

$$
\begin{aligned}
X_1^n(t) X_2(t)^n - X_1(0) X_2(0) = &\int_0^t X_2(s)^n F_1^n(s) ds + \int_0^t X_2(s) G_1(s)^n dW_s \\
&+ \int_0^t X_1^n(s) F_2^n(s) ds + \int_0^t X_1(s)^n G_2^n(s) dW_s + \int_0^t G_1(s)^n G_2^n(s) ds.
\end{aligned}
$$

Only thing left is a convergence result (i.e DCT) sinc the processes are bounded or smth like that.

Step 3 if $u(x) = x^m$, $\forall m = 0, \dots$ to prove

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

For $m = 2$ the result is obtained by the product rule, By induction we prove for arbitrary m

(IV) Suppose the statement hold for $m - 1$
(IS) $m - 1 \to m$

$$\begin{aligned} d(X^m) &= d(X \cdot X^{m-1}) = XdX^{m-1} + X^{m-1}dx + (m-1)X^{m-2}G^2dt \\ &\overset{\text{IS}}{=} X(m-1)X^{m-2}dx + X \cdot \frac{1}{2}(m-1)(m-2)X^{m-3}G^2dt + X^{m-1}dx + (m-1)X^{m-2}G^2dt \\ &= mX^{m-1}dx + (m-1)(\frac{m}{2} - 1 + 1)X^{m-2}G^2dt \\ &= \underbrace{mX^{m-1}}_{\partial_x u}dx + \frac{1}{2}\underbrace{m(m-1)X^{m-2}}_{\partial_x^2 u}G^2dt. \end{aligned}$$

Now $u(x) = x^m$

$$dX = Fdt + GdW_t.$$

Step 4 If $u(x,t) = f(x)g(t)$ where f is a polynomial

$$\begin{aligned} d(u(x,t)) &= d(f(x)g(t)) = f(x)dg + gdf(x) + G \cdot 0dt \\ &\overset{\text{S3}}{=} f(x)g'(t)dt + gf'(x)dx + \frac{1}{2}gf''(x)G^2dt. \end{aligned}$$

Itos formula is true for $f(x)g(t)$, it should thus also be true for functions $u(x,t) = \sum_{i=1}^m g^i(t)f^i(x)$

Step 5: if $u \in \mathcal{C}^{2,1}$ then we know there exists a sequence of polynomials $f^i(x)$ s.t

$$u_n(x,t) = \sum_{i=1}^n f^i(x)g^i(t).$$

Then $u_n \to u$ uniformly for any compact set $K \subset \mathbb{R} \times [0,T]$, we can thus apply Itos formula for each of the u_n and take the limit term wise $\square$

Remark. One can get the existence of the polynomial sequence by using Hermetian polynomials

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

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

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

Remark (Multi Dimensional Brownian Motion)**.** Multi dimensional Brownian motion

$$W(t) = (W^1(t), \ldots, W^m(t)) \in \mathbb{R}^m$$

.

In each direction we should have a 1 dimensional Brownian motion and any two directions should be independent. We use the natural filtration $\mathcal{F}_t = \sigma(W(s); 0 \leq s \leq t)$

Definition 3.1.7 (Multi-Dimensional Ito's Integral)**.** We the define the n dimensional integral for $G \in L^2_{n \cdot m}([0,T])$, $G_{ij} \in L^2([0,T])$ $1 \leq i \leq n$, $1 \leq j \leq m$

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

With the Properties

$$\mathbb{E}[\int_0^T G dW_t] = 0$$

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

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

Definition 3.1.8 (Multi-Dimensional Ito process)**.** We define the n dimensional Ito process as

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

$$dX^i = F^i dt + \sum_{j=1}^m G^{ij} dW_t^i \qquad 1 \leq i \leq n.$$

Theorem 3.1.1 (Multi Dimensional Ito's formula)**.** We define the n dimensional Ito's formula as $u \in \mathcal{C}^{2,1}(\mathbb{R}^n \times [0,T], \mathbb{R})$

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

Proposition 3.1.1. For real valued processes X_1, X_2

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

Working with SDEs relies on a lot of notational rules as seen in the differential notation is just shorthand for the Integral form

Definition 3.1.9. Formal multiplication rules for SDEs

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

Using this notation we can simply itos 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

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

Typical example

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

Example. If F and G are deterministic

$$dX_{n\times 1}F(t)_{n\times 1}dt + G_{n\times m}dW_t m \times 1.$$

Then for arbitrary test function $u \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ then by Ito'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 we take the expectation of the above integral

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

Definition 3.1.10 (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 of X , μ fulfills the heat equation

$$\mu_t = \triangle\mu = 0.$$

How does this all translate to our Mean field Limit, consider a particle system given by

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

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

$$(x_1(t), \ldots, x_N(t)) \sim u(x_1, \ldots, x_n).$$

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

$$\varphi(X_N) = \varphi(X_N(0)) + \int_0^t \nabla_{dN}\varphi \cdot \begin{pmatrix} \vdots \\ \frac{1}{n}\sum_{j=1}^N K(x_i, x_j) \\ \vdots \end{pmatrix} X_N$$
$$+ \int_0^t \triangle_{X_N}\varphi dt + \int_0^t \sqrt{2}\nabla\varphi dW_t^i.$$

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

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

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

$$u(x_1, \ldots, x_N) = \overline{u}^{\otimes N}.$$

the equation simplifies

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

3.2 Solving Stochastic Differential Equations

The setup of the following section will be the following

Definition 3.2.1 (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

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

Note for better understanding the dimensions will be included in the following definition, but we generally leave them out.

Definition 3.2.2 (SDE). Given the above basic setup we are trying to solve equations of the type

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

Where

$$b : (x, t) \in \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$$
$$B : (x, t) \in \mathbb{R}^n \times [0, T] \to M^{nxm}.$$

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

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

Definition 3.2.3 (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) $F := b(X_t, t) \in L^1_n([0, T]) \Leftrightarrow \int_0^t \mathbb{E}[F_s]ds < \infty$
3. (diffusion) $G := B(X_t, t) \in L^2_{n\times m}([0, T]) \Leftrightarrow \int_0^t \mathbb{E}[|G_s|^2]ds < \infty$

Reminder that (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, we formulate the following theorem, one should remember that if the diffusion term $B(X_t, t)$ is 0 then we get a unique solution iff $b(X_t, t)$ is Lipschitz

Theorem 3.2.1 (Existence and Solution)**.** Suppose $b : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ and $B : \mathbb{R}^n \times [0, T] \to M^{n\times m}$, then we get the necessary condition that they are continuous and (globally) Lipschitz continuous with respect to x i.e $\exists\ L > 0$ such that for arbitrary $\forall\ x, \tilde{x} \in \mathbb{R}^n$ and $t \in [0, T]$ it holds

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

and the linear growth condition

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

The initial data X_0 should be square integrable $x_0 \in L^2_n(\Omega)$ and that X_0 is independent of $W^t(0)$

Whenever the above conditions hold then there exists a unique solution $X \in L^2_n([0, T])$ of the SDE.

Proof. We begin by proving the uniqueness of solution.

Suppose we have two solutions X and $\tilde{X}$ to the SDE then we need 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 to be 0 we could use a Grönwall type inequality and get the uniqueness. To work with the diffusion term we consider the square of the above and apply Itos isometry. Note that generally $|a+b|^2 \nleq (a^2+b^2)$ which is why we need the extra 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 as in the deterministic case we use Picard iteration.

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

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

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

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

Issue now is that because of Ω we cannot use completeness to argue the convergence, instead we use a similar argument to the uniqueness proof.

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

Where we used the following Doobs martingales Lp inequality

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

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

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

Further more we get with a Markovs inequality

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

Then by Borel-Cantelli we know

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

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

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

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 $\omega\in B$ we can make a Cauchy sequence argument

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

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

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

This means X_t is just a stochastic process and by properties of W_s it is already progressively measurable. It remains to show that $X_t\in\mathbb{L}^2([0,T])$

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

This follows by $\mathbb{E}[X_0]=0$, linear growth condition for the first integral and Ito isometry for second and then linear growth condition, noting that the $\mathbb{L}^2$ norm of the initial value X_0 is bounded

By Dominated Convergence we obtain that the expectation of X_t^2 is bounded, whats the dominating rv ?

$$\mathbb{E}[|X_t|^2]\le C(t)\Rightarrow\int_0^T\mathbb{E}[|X_t|^2]<\infty.$$

We can bound it by

$$\mathbb{E}[|X_t|^2]=\mathbb{E}[|X_t+X_0-X_0|^2]\le 2\mathbb{E}[|X_t-X_0|^2]+2\mathbb{E}[|X_0|^2].$$

□

Remark. Uniqueness in a stochastic sense means that for two solution $X, \tilde{X}$ we have

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

I.e they are indistinguishable

As a small side note we consider this example to distinguish modifications and indistinguishable.

Example. First note that for any $t \in [0,T]$ we have the following inclusion

$$A \coloneqq \{X(t) = \tilde{X}(t),\ \forall\, t \in [0,T]\} \subset \{X(t) = \tilde{X}(t)\} \coloneqq A_t.$$

i.e

$$\mathbb{P}(A) \leq P(A_t).$$

Such that indistinguishability implies modification where modification means

$$\forall\, t \in [0,T]\ :\ \mathbb{P}(A_t) = 1.$$

Theorem 3.2.2 (Higher moments estimate)**.** Assumptions for b , B and X_0 are the same as before, if in addiotion

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

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

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

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

Proof. Left as an exercise
Hint: Expand the definition use the same growth condition stuff , holder and then Ito isometry ☐

3.2.1 General Convergence Results

Definition 3.2.4 (Weak convergence of measures)**.** The following statements are equivalent

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

Definition 3.2.5 (Weak convergence of Random variable)**.** The following statements are equivalent

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

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

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

Definition 3.2.6 (Tightness)**.** A set of probability measures $S \subset \mathcal{P}(\mathbb{R}^d)$ is called tight, if for $\forall\ \varepsilon > 0$ there exists $\exists\ K \subset \mathbb{R}^d$ compact such that

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

Theorem 3.2.3 (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 □

We can now define the Stochastic Empirical measure

Definition 3.2.7 (Empirical Measure (Stochastic version))**.** For a set of random variables $(X_i)_{i\leq N}$ we define the (random) empirical measure

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

Point wise this definition coincides with the deterministic case.

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

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

Actually one can prove that the choice of $f \in \mathcal{C}_b$ does not matter for the convergence We get the stronger corollary

Corollary. If X_i are i.i.d random variables with law μ_X then it holds that

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

i.e

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

Proof. The proof relies mainly on proving 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. □

Remembering the definition of the Wasserstein distance

Definition 3.2.8 (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

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

then the following convergences 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. $\mu_n \rightharpoonup \mu$

Corollary. If X_i are i.i.d random variables with law μ_X and $\int |x|^p \mu < \infty$ then

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

and

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

Where

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