# Groupe de travail du 24/01 : introduction aux systèmes de particules en interaction

## Introduction

$$\left(X_1(t), \dots, X_N(t)\right) \in \left(\mathbb{R}^d\right)^N \quad N \text{ grand}$$

Dynamique = ODE ou SDE couplées.

Idée : représenter le système par une mesure sur $\mathbb{R}^d$ :

$$\hat{\mu}_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta\left(x - X_i(t)\right) = \text{mesure empirique.}$$

On espère que $\hat{\mu}_t^N$ a une limite pour $N \to \infty$, par exemple $f(x,t)\,dx$.

A quoi ça sert ?

1. Modélisation (physique, bio, économie...)
On cherche une équation pour $f(x,t)$ (une EDP) et on l'étudie.

2. Méthode numérique : on a une EDP non linéaire ; on cherche un système de particules dont la mesure empirique est proche d'une solution de l'EDP.
$\longrightarrow$ facile à programmer.

Réseaux de neurones = différent.

$f : \Omega \subset \mathbb{R}^d \longrightarrow \mathbb{R}$ : on cherche à approximer $f$.

$$f^{(n)}(\theta, x) = \frac{1}{n} \sum_{i=1}^{n} c_i \, \sigma(w_i \cdot x + b_i)$$

$$\theta_i = (c_i, w_i, b_i)$$

On cherche à optimiser les $\theta_i$, en utilisant les données.

Apprentissage (stochastic gradient dynamics)

$\longrightarrow$ dynamique sur les $\theta_i$

$\longrightarrow$ système de particules en interaction.

# I Convergence de la mesure empirique.

## (0) A un instant donné

$\hat{\mu}^N$ = mesure aléatoire. donc la loi de $\hat{\mu}^N$ est un objet un peu compliqué...

<u>Lemme</u> : $\hat{\mu}^N \xrightarrow{\mathcal{L}} \mu$ $\qquad$ $\mu$ = mesure déterministe

$$\Leftrightarrow \forall \varphi \in \mathcal{C}_b(\mathbb{R}^d) \quad \mathbb{E}_{\mu^{(N)}}\left[\left|\int \varphi \, d\hat{\mu}^N - \int \varphi \, d\mu\right|\right] \xrightarrow[N\to\infty]{} 0$$

mesure empirique construite à partir de la mesure à N particules $\mu^{(N)}$.

<u>Ex</u> : si $\mu^{(N)} = \mu^{\otimes N}$ (les $X_i$ sont i.i.d.)

$$\int \varphi \, d\hat{\mu}_N = \frac{1}{N} \sum \varphi(X_i)$$

$$\mathbb{E}\left[ \left| \frac{1}{N} \sum_i \varphi(X_i) - \int \varphi \, d\mu \right| \right] \xrightarrow[N\to\infty]{} 0$$

<u>Def</u> : Soit $(\mu^{(N)})_{N\in\mathbb{N}}$ une suite de mesures sur $(\mathbb{R}^d)^N$.

$(\mu^{(N)})$ est <u>$\mu$-chaotique</u> si $\forall k \in \mathbb{N}$
$\forall \varphi_1, \dots, \varphi_k \in \mathcal{C}_b(\mathbb{R}^d)$

$$\int_{(\mathbb{R}^d)^N} \varphi_1(x_1) \dots \varphi_k(x_k) \, d\mu^{(N)}(x) \xrightarrow[N\to\infty]{} \left(\int \varphi_1 \, d\mu\right) \dots \left(\int \varphi_k \, d\mu\right)$$

ie : la loi $\mu^{(N)}$ "ressemble à une loi produit".

<u>Théo</u> : Soit $(\mu^{(N)})_{N\in\mathbb{N}}$ une suite de mesures sur $(\mathbb{R}^d)^N$ symétriques. Alors il y a équivalence entre :

i) $(\mu^{(N)})$ $\mu$-chaotique

ii) $\forall \varphi_1, \varphi_2 \in \mathcal{C}_b(\mathbb{R}^d)$ $\displaystyle\int_{(\mathbb{R}^d)^N} \varphi_1(x_1)\varphi_2(x_2)\,d\mu^{(N)}(x) \xrightarrow[N\to\infty]{} \left(\int_{\mathbb{R}^d} \varphi_1\,d\mu\right)\left(\int_{\mathbb{R}^d} \varphi_2\,d\mu\right)$

iii) $\hat{\mu}^N \xrightarrow{\mathcal{L}} \mu$.

$\longrightarrow$ montrer que $\hat{\mu}^N$ a un comportement déterministe quand $N\to\infty$, c'est la même chose que montrer la chaoticité de $\mu^{(N)}$

$\longrightarrow$ importance de la "propagation du chaos".

Preuve :

$\boxed{\text{ii)} \Rightarrow \text{iii)}}$ $\varphi \in \mathcal{C}_b$.

$$\left[\int \varphi\, d\hat{\mu}_N - \int \varphi\, d\mu\right]^2 = \frac{1}{N^2}\sum_i \varphi(X_i)^2 + \frac{1}{N^2}\sum_{i\neq j} \varphi(X_i)\varphi(X_j)$$
$$+\left(\int \varphi\, d\mu\right)^2 - 2\left(\int \varphi\, d\mu\right)\left[\frac{1}{N}\sum_i \varphi(X_i)\right]$$

On prend l'espérance :

$$\mathbb{E}\Big(\qquad\Big)^2 = \underbrace{\frac{1}{N}\,\mathbb{E}\big(\varphi(X_1)^2\big)}_{\to 0} + \frac{N-1}{N}\,\mathbb{E}\big(\varphi(X_1)\varphi(X_2)\big)$$

$$+ \left(\int \varphi\, d\mu\right)^2 - 2\left(\int \varphi\, d\mu\right)\mathbb{E}\big(\varphi(X_1)\big)$$

Par l'hypothèse ii) $\mathbb{E}\big(\varphi(X_1)\big) \longrightarrow \int \varphi\, d\mu$

et $\mathbb{E}\big(\varphi(X_1)\varphi(X_2)\big) \longrightarrow \left(\int \varphi\, d\mu\right)^2$

Donc $\mathbb{E}\left(\left|\int \varphi\, d\hat{\mu}^N - \int \varphi\, d\mu\right|\right) \xrightarrow[N\to\infty]{} 0$

(Cauchy-Schwarz).

$\boxed{(iii) \to (i)}$ On fait le cas $k=2$.

$$\int_{(\mathbb{R}^d)^N} \varphi_1(x_1)\varphi_2(x_2)\, d\mu^{(N)} - \int \varphi_1\, d\mu \int \varphi_2\, d\mu =$$

$$\underbrace{\int_{(\mathbb{R}^d)^N} \varphi_1(x_1)\varphi_2(x_2)\, d\mu^{(N)} - \int_{(\mathbb{R}^d)^N} \left(\int \varphi_1\, d\hat{\mu}^N\right)\left(\int \varphi_2\, d\hat{\mu}^N\right) d\mu^{(N)}}_{1}$$

$$+ \underbrace{\int_{(\mathbb{R}^d)^N} \left(\int \varphi_1 \, d\hat{\mu}_N\right)\left(\int \varphi_2 \, d\hat{\mu}_N\right) d\mu^{(N)} - \left(\int \varphi_1 \, d\mu\right)\left(\int \varphi_2 \, d\mu\right)}_{2}$$

$2 \to 0$ par (iii) avec la fonction $\phi : \nu \to \left(\int \varphi_1 \, d\nu\right)\left(\int \varphi_2 \, d\nu\right)$

on a $\mathbb{E}\, \phi(\hat{\mu}^N) \to \mathbb{E}\, \phi(\mu)$

$$1 = \cdots + \int_{(\mathbb{R}^d)^N} \frac{1}{N^2} \sum_{i,j} \varphi_1(x_1)\, \varphi_2(x_2)\, d\mu^{(N)}(x)$$

$$= \cdots + \frac{1}{N} \int_{(\mathbb{R}^d)^N} \varphi_1(x_1)\, \varphi_2(x_1)\, d\mu^{(N)}(x)$$

$$+ \frac{N-1}{N} \int_{(\mathbb{R}^d)^N} \varphi_1(x_1)\, \varphi_2(x_2)\, d\mu^{(N)}(x)$$

On a utilisé la symétrie.

## 2°) Evolution temporelle.

### a) Déterministe, sans interaction

$$\begin{cases} \dot{X}_i = K(X_i) \\ X_i(t=0) \sim \mu_0 \quad (i.i.d) \end{cases}$$

$T_t : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ = flot associé

$$\text{loi de } X_i(t) \equiv \mu(t) = \underbrace{T_t \# \mu_0}_{\text{mesure } \mu_0 \text{ transportée par le flot.}}$$

Alors : * Pour toute fonction test $\varphi$

$$\int \varphi \, d\mu_t = \int \varphi \circ T_t \, d\mu_0$$

* $\mu_t$ solution de $\underbrace{\partial_t \mu_t + \nabla \cdot (K \cdot \mu_t) = 0}_{\text{équation de transport.}}$

* $\hat{\mu}_t^N$ proche de $\mu_t$ pour $N$ grand (LGN)

b) Déterministe, avec interaction

NB : premier cas important pour les réseaux de neurones.

$$\begin{cases} \dot{X}_i = \frac{1}{N} \sum_j K(X_j, X_i) \\ X_i(t=0) \sim \mu_0 \end{cases} \quad (1)$$

Idée : Si $\hat{\mu}_t^N$ proche de $\mu_t$, alors

$$\frac{1}{N} \sum_j K(X_j, X_i) = \int_{\mathbb{R}^d} K(x, X_i)\, d\hat{\mu}_t^N(x)$$

proche de $\int_{\mathbb{R}^d} K(x, X_i)\, d\mu_t(x) = \bar{K}[\mu_t](X_i)$

$\dot{y} = \bar{K}[\mu_t](y) \longrightarrow$ flot $T_t[\mu] : \mathbb{R}^d \longrightarrow \mathbb{R}^d$

On devine que $\mu_t$ solution de :

$$\partial_t \mu_t + \nabla \cdot \left( \bar{K}[\mu_t]\, \mu_t \right) = 0 \quad (2)$$

Remarque : $\hat{\mu}_t^N$ construite à partir de (1) est solution de (2).

Donc la question de savoir si $\hat{\mu}_t^N$ est proche de $\mu_t$ est une question sur la stabilité des solutions de 2.

<u>Théo</u> : $t_M > 0$ fixé. $\sup_{0 \leq t \leq t_M} W_1(\hat{\mu}_t^N, \mu_t) \xrightarrow[N \to \infty]{} 0$.

$W_1(\hat{\mu}_0^N, \mu_0) \xrightarrow[N \to \infty]{} 0$

$W_1$ = distance de Wasserstein d'ordre 1.

$$W_1(\mu, \nu) = \sup_{\mathrm{Lip}(\varphi) \leq 1} \left| \int \varphi \, d(\mu - \nu) \right|$$

$$= \inf_{\substack{(X,Y) \text{ t.q.} \\ X \sim \mu \\ Y \sim \nu}} \mathbb{E}|X - Y|$$

⚠ Vitesse de convergence pas donnée ici

⚠ Pas d'uniformité en temps. On y reviendra.

__Preuve__ : Soient $\mu_t, \nu_t$ 2 solutions de (e) avec CI $\mu_0, \nu_0$.

Alors $\left[ W_1(\mu_t, \nu_t) \leq e^{2L_K t} W_1(\mu_0, \nu_0) \right)$

$y \longmapsto K(y,x)$ $L_K$-lipschitzienne, $\forall x$.

$\longrightarrow$ pas uniforme en temps ; la convergence se dégrade rapidement en temps grand.

$$W_1(\mu_t, \nu_t) \leq W_1\left(T_t[\mu] \# \mu_0, T_t[\mu] \# \nu_0\right) \quad \text{(a)}$$
$$+ W_1\left(T_t[\mu] \# \nu_0, T_t[\nu] \# \nu_0\right) \quad \text{(b)}$$

(a) : $\leq \left| \int \varphi \circ T_t[\mu] \, d(\mu_0 - \nu_0) \right| \quad \forall \varphi, \ \mathrm{Lip}(\varphi) \leq 1$

$\leq \mathrm{Lip}\left(T_t[\mu]\right) W_1(\mu_0, \nu_0)$

$\leq e^{L_K t} W_1(\mu_0, \nu_0)$

(b) $\leq \left| \int \left(\varphi \circ T_t[\mu] - \varphi \circ T_t[\nu]\right) d\nu_0 \right| \leq \underbrace{\int \left| T_t[\mu] - T_t[\nu] \right| d\nu_0}_{= \lambda(t)}$

$$\frac{d\lambda}{dt} \leqslant \int \left| \bar{k}[\mu_t] \circ T_t[\mu] - \bar{k}[\nu_t] \circ T_t[\nu] \right| d\nu_0$$

$$\leqslant \int \left| \bar{k}[\mu_t] \circ T_t[\mu] - \bar{k}[\mu_t] \circ T_t[\nu] \right| d\nu_0 \quad b_1$$

$$+ \int \left| \bar{k}[\mu_t] \circ T_t[\nu] - \bar{k}[\nu_t] \circ T_t[\nu] \right| d\nu_0 \quad b_2$$

$$b_1 \leqslant \mathrm{Lip}(\bar{k}[\mu_t])\, \lambda(t) \leqslant L_K \lambda(t)$$

$$b_2 = \int \left| \bar{k}[\mu_t] - \bar{k}[\nu_t] \right| d\nu_t$$

$$= \int \left| \int k(y,x)\, d\mu_t(y) - \int k(y,x)\, d\nu_t(y) \right| d\nu_t(x)$$

$$\leqslant L_K \int W_1(\mu_t, \nu_t)\, d\nu_t(x) = L_K W_1(\mu_t, \nu_t)$$

Donc $\frac{d\lambda}{dt} \leqslant L_K \lambda + L_K W_1(\mu_t, \nu_t)$

(Gronwall) $\lambda(t) \leqslant L_K e^{L_K t} \int_0^t W_1(\mu_s, \nu_s) e^{-L_K s} ds$

$$\Rightarrow W_1(\mu_t, \nu_t) \leqslant e^{L_K t} W_1(\mu_0, \nu_0) + L_K e^{L_K t} \int_0^t W_1(\mu_s, \nu_s) e^{-L_K s} ds$$

(Gronwall) $W_1(\mu_t, \nu_t) \leqslant W_1(\mu_0, \nu_0) e^{2ct}$

c) Evolution aléatoire, sans interaction

$$dX_i = K(X_i^t)\,dt + \sqrt{2}\,dB_i^t$$

loi de $X_i^t = f(x,t)\,dx$

$$\partial_t f = \nabla\cdot\left(-K(x)f(x) + \nabla f(x)\right)$$

d) Evolution aléatoire, avec interaction

$$dX_i^t = \frac{1}{N}\sum_j K(X_j^t, X_i^t)\,dt + \sqrt{2}\,dB_i^t$$

Limite attendue : $\partial_t f = \nabla\cdot\left(-\bar{K}[f]f + \nabla f\right)$ (1)

avec $\bar{K}[f] = \int K(x,y)f(y)\,dy$

Problème : $\hat{\mu}_t^N$ n'est plus solution de (1).

Idée : Introduire les particules $\bar{X}_i$

$$d\bar{X}_i = \bar{K}[f_t](\bar{X}_i^t)\,dt + \sqrt{2}\,dB_i^t$$

le même que pour les $X_i$

densité $\bar{f}$

Alors $\partial_t \bar{f} = \nabla\cdot\left(-\bar{K}[f_t]\bar{f} + \nabla\bar{f}\right)$

Donc $\bar{f} = f$ ! (unicité de la solution de l'EDP).

<u>Remarque</u> : les $\bar{X}_i$ sont indépendants

$$\frac{d(X_i - \bar{X}_i)}{dt} = \frac{1}{N}\sum_j K(X_j, X_i) - \bar{K}[f_t](X_i)$$

$$= \frac{1}{N}\sum_j \left[K(X_j, X_i) - K(\bar{X}_j, \bar{X}_i)\right] \quad \text{contrôle par } |X - \bar{X}|$$

$$+ \underbrace{\frac{1}{N}\sum_j K(\bar{X}_j, \bar{X}_i) - \bar{K}[f_t](X_i)}_{\text{petit par LGN}}$$

puis Gronwall

$\Rightarrow$ à nouveau un contrôle pour tout horizon temporel fini, mais pas pour $t \to \infty$ ...

## II Structure de flot gradient

Outil important pour aborder le comportement au temps grand.

Dimension finie : $\dot{x} = -\nabla U(x)$ $U : \mathbb{R}^d \to \mathbb{R}$

alors $\frac{dU(x_t)}{dt} = -|\nabla U(x_t)|^2 \leqslant 0$

$= 0$ seulement si $\nabla U|_{x_t} = 0$

$\to$ très utile pour la convergence en temps grand, le "paysage" de $U$ contient toute la dynamique.

Remarque : discrétisation en temps

$$x(t+\Delta t) = \underset{y}{\operatorname{argmin}} \left( \frac{1}{2\Delta t} \|y - x_t\|^2 + U(y) \right)$$

2 ingrédients : fonctionnelle $U$ et métrique, ici norme euclidienne.

Pour des mesures : distance de Wasserstein $W_2$

$$W_2^2(\mu,\nu) = \inf_{\substack{X\sim\mu \\ Y\sim\nu}} \mathbb{E}\left(|X-Y|^2\right)$$

$$\mu_{t+\Delta t} = \operatorname*{argmin}_{\mu} \left( \frac{1}{2\Delta t} W_2^2(\mu_t,\mu) + U(\mu) \right)$$

Ex : $U(\mu) = \int V \, d\mu + \frac{1}{2}\int W(x,y)\, d\mu(x)\, d\mu(y)$

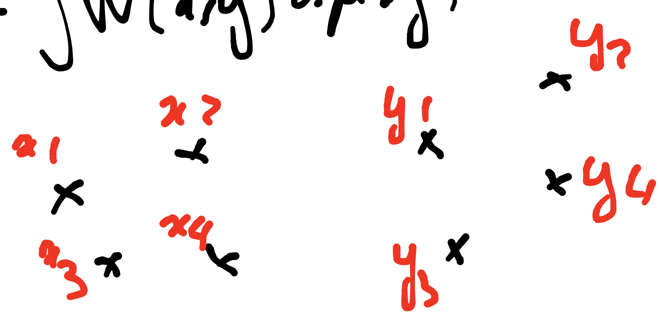
Alors $\mu_t$ satisfait l'équation :

$$\partial_t \mu_t = \nabla\cdot\left(\mu_t \nabla \frac{\delta U}{\delta\mu}\right)$$

équation de transport

$$\frac{\delta U}{\delta \mu} = V + \int W(x,y)\, d\mu(y)$$

Nuage de points :




$$W_2^2\left(\hat{\mu}_t^N, \hat{\mu}_y^N\right) = \frac{1}{N} \sum_i (y_i - x_i)^2$$

(couplage trivial, OK si $\Delta t$ petit)

$$U(\hat{\mu}_y^N) = \frac{1}{N}\sum_i V(y_i) + \frac{1}{2N^2}\sum_{i,j} W(y_i, y_j)$$

argmin

$\to \forall i \quad \frac{y_i - x_i}{\Delta t} = -\nabla V(y_i) - \frac{1}{N}\sum_j \nabla_1 W(y_i, y_j)$

hyp W symétrique.

$\Rightarrow y_i = x_i - \Delta t \left[\nabla V(y_i) + \frac{1}{N}\sum_j \nabla_1 W(y_i, y_j)\right]$

$$\dot{X}_i = -\nabla V(X_i) - \frac{1}{N}\sum_j \nabla W(X_i, X_j)$$

$$= -\nabla_i U^N(X_1, \dots, X_N)$$

$$U^N(X_1, \dots, X_N) = \frac{1}{N}\sum_i U(X_i) + \frac{1}{2N^2}\sum_{i,j} W(X_i, X_j)$$

<u>Remarque</u> : si $U(\mu) = \int f \ln f \, dx$ si $\mu = f dx$

flot gradient : $\partial_t f = +\nabla(f \nabla \ln f) = \Delta f$

diffusion !