Randomized Large Populations

François Delarue (Nice, Université Côte d'Azur, France)

Colloquium LJAD

November 13, 2023

1. Basics in Stochastic Analysis

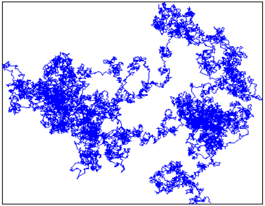
Brownian Motion [1920's-30's, Wiener, Lévy...]

- Continuous random walk in continuous time

- Brownian motion in dimension $d \Leftrightarrow d$ independent Brownian motions in dimension 1

$$W_t = (W_t^1, \cdots, W_t^d)$$

  - $(W_t^1)_{t\geq 0}, \cdots, (W_t^d)_{t\geq 0}$ independent
  - $W_{t+\mathrm{d}t}^i - W_t^i \perp\!\!\!\perp$ of the past before t and $\mathcal{N}(0, \mathrm{d}t)$ distributed
  - plot in $d = 2$



Brownian Motion [1920's-30's, Wiener, Lévy...]

- Continuous random walk in continuous time
- Brownian motion in dimension $d \Leftrightarrow d$ independent Brownian motions in dimension 1

$$W_t = (W_t^1, \cdots, W_t^d)$$

  - $(W_t^1)_{t\geq 0}, \cdots, (W_t^d)_{t\geq 0}$ independent
  - $W_{t+\mathrm{d}t}^i - W_t^i \perp\!\!\!\perp$ of the past before t and $\mathcal{N}(0, \mathrm{d}t)$ distributed

- Probabilistic interpretation of parabolic PDE

$$\partial_t u(t,x) + \frac{1}{2}\Delta_x u(t,x) + f(t,x) = 0, \quad (t,x) \in [0,T] \times \mathbb{R}^d$$
$$u(T,x) = g(x)$$

  - Kolmogorov equation (1930's)

$$u(0,x) = \mathbb{E}\left[g(x + W_T) + \int_0^T f(s, x + W_s)\, \mathrm{d}s \right]$$

Restoration of Uniqueness [1970's–..., Krylov, Flandoli...]

• Well-known illustration of smoothing properties of heat kernel:
ODE driven by bounded non-Lipschitz velocity field

$$\dot{X}_t = b_t(X_t)$$

∘ b continuous ⇒ existence but ~~uniqueness~~

∘ restore uniqueness by perturbing the dynamics by a Brownian motion $(B_t)_{t \geq 0}$

$$\underbrace{\mathrm{d}X_t}_{\text{infinitesimal variation of } X} = b_t(X_t)\,\mathrm{d}t + \underbrace{\mathrm{d}B_t}_{\text{infinitesimal variation of } B \sim \sqrt{\mathrm{d}t}}$$

Restoration of Uniqueness [1970's–..., Krylov, Flandoli...]

- Well-known illustration of smoothing properties of heat kernel:
ODE driven by bounded non-Lipschitz velocity field

$$\dot{X}_t = b_t(X_t)$$

  - b continuous $\Rightarrow$ existence but ~~uniqueness~~
  - restore uniqueness by perturbing the dynamics by a Brownian motion $(B_t)_{t\geq 0}$

$$\underbrace{\mathrm{d}X_t}_{\text{infinitesimal variation of } X} = b_t(X_t)\,\mathrm{d}t + \underbrace{\mathrm{d}B_t}_{\text{infinitesimal variation of } B \sim \sqrt{\mathrm{d}t}}$$

- Highly oscillating perturbation $\rightsquigarrow$ regularization effect
  - time averaging $\rightsquigarrow$ path by path $y \mapsto \int_0^t b_s(B_s + y)\,\mathrm{d}s$ (almost) Lipschitz
  - space averaging $\rightsquigarrow$ statistical behavior of solutions

Gradient Descent

- Minimization problem

$$\min_{x\in\mathbb{R}^d}\{V(x)\}, \qquad V : \mathbb{R}^d \to \mathbb{R}$$

- Gradient descent

$$\dot{x}_t = -\nabla_x V(x_t)$$

  - many issues with convergence without convexity of V

Gradient Descent

- Minimization problem

$$\min_{x\in\mathbb{R}^d}\{V(x)\}, \quad V:\mathbb{R}^d\to\mathbb{R}$$

- Gradient descent

$$\dot{x}_t = -\nabla_x V(x_t)$$

  ◦ many issues with convergence without convexity of V

- Stochastic gradient descent [Fokker, Planck, 1910's...]

$$dX_t = -\nabla_x V(X_t)\,dt + \sigma dB_t$$

◦ under confining properties of V (weaker than convexity [Bakry-Emery1980's...])

$$\mathrm{Law}(X_t) \underset{t\to\infty}{\longrightarrow} Z^{-1}\exp(-2\frac{V}{\sigma^2})$$

◦ in long time regime and for σ small, law is concentrated around the minimizers of V

2. Some Large Population Models

Prototype [1950-60's, Kac, McKean...]

- Typical example (see e.g. Sznitman [1990's])

$$dX_t^i = b\Big(X_t^i, \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^j}\Big)\,dt + dB_t^i, \quad i = 1, \cdots, N$$

  - velocity field at point x depends on the state of the population

Prototype [1950-60's, Kac, McKean...]

- Typical example (see e.g. Sznitman [1990's])

$$\mathrm{d}X_t^i = b\Big(X_t^i, \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^j}\Big)\,\mathrm{d}t + \mathrm{d}B_t^i, \qquad i = 1, \cdots, N$$

  - velocity field at point x depends on the state of the population

- Intuitively, particles become independent as N tends to ∞ and by exchangeability, weak limit should satisfy McKean-Vlasov equation

$$\mathrm{d}X_t = b(X_t, \mu_t)\,\mathrm{d}t + \mathrm{d}B_t$$

  - with $\mu_t = \mathrm{Law}(X_t)$

Prototype [1950-60's, Kac, McKean...]

- Typical example (see e.g. Sznitman [1990's])

$$\mathrm{d}X_t^i = b\Big(X_t^i, \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}\Big)\,\mathrm{d}t + \mathrm{d}B_t^i, \quad i = 1, \cdots, N$$

  - velocity field at point x depends on the state of the population

- Intuitively, particles become independent as N tends to ∞ and by exchangeability, weak limit should satisfy McKean-Vlasov equation

$$\mathrm{d}X_t = b(X_t, \mu_t)\,\mathrm{d}t + \mathrm{d}B_t$$

  - with $\mu_t = \mathrm{Law}(X_t)$

- Link with PDEs

  - nonlinear Fokker-Planck equation

$$\partial_t \mu_t - \mathrm{div}_x(b(\cdot, \mu_t)\mu_t) - \frac{1}{2}\Delta_x \mu_t = 0, \quad t \geq 0$$

Illustration 1: Gradient Flow on $\mathcal{P}(\mathbb{R}^d)$

- Minimization problem

$$\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \{V(\mu)\}, \qquad V : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

Illustration 1: Gradient Flow on $\mathcal{P}(\mathbb{R}^d)$

- Minimization problem

$$\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \{V(\mu)\}, \quad V : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Gradient descent [2000's, Otto...]

$$\dot{X}_t(\omega) = -\partial_\mu V(\mu_t)(X_t(\omega)), \quad \omega \in \Omega \,; \quad \mu_t := \mathrm{Law}(X_t)$$

◦ where $\partial_\mu V$ is Wasserstein derivative [Ambrosio-Gigli-Savaré, Lions...], i.e.

$$\partial_\mu V(\mathcal{L}(X))(X(\omega)) = D_{L^2(\Omega;\mathbb{R}^d)}[V(\mathcal{L}(X))](\omega), \quad \omega \in \Omega$$

◦ advection equation for $(\mu_t)_{t \geq 0}$

$$\partial_t \mu_t + \mathrm{div}_x\Big(\partial_\mu V(\mu_t)\mu_t\Big) = 0$$

Illustration 1: Gradient Flow on $\mathcal{P}(\mathbb{R}^d)$

- Minimization problem

$$\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ V(\mu) + \frac{\sigma^2}{2} \int_{\mathbb{R}^d} \ln\Big(\frac{\mathrm{d}\mu}{\mathrm{d}x}(x)\Big)\, \mathrm{d}\mu(x) \right\}, \qquad V : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Gradient descent [2000's, Otto...]

$$\mathrm{d}X_t(\omega) = -\partial_\mu V(\mu_t)(X_t(\omega))\, \mathrm{d}t + \sigma\, \mathrm{d}B_t, \quad \omega \in \Omega\,; \quad \mu_t := \mathrm{Law}(X_t)$$

  - where $\partial_\mu V$ is Wasserstein derivative, i.e.

$$\partial_\mu V(\mathcal{L}(X))(X(\omega)) = D_{L^2(\Omega;\mathbb{R}^d)}[V(\mathcal{L}(X))](\omega), \quad \omega \in \Omega$$

  - advection-diffusion equation for $(\mu_t)_{t \ge 0}$

$$\partial_t \mu_t + \mathrm{div}_x\Big(\partial_\mu V(\mu_t)\mu_t\Big) - \frac{\sigma^2}{2} \Delta_x \mu_t = 0$$

  - in general: $\sigma > 0$ doesn't suffice to get long-time convergence towards minimizers

Illustration 2: Mean Field Control (2010's, Lions...)

- Minimize cost to 'society'

$$J(\alpha) = \mathcal{G}(m_T) + \int_0^T \left(\int_{\mathbb{R}^d} \frac{1}{2} |\alpha(t,x)|^2 m_t(\mathrm{d}x) + \mathcal{F}(m_t) \right) \mathrm{d}t$$

  - infimum taken over Fokker-Planck equations

$$\partial_t m_t = \frac{1}{2} \Delta_x m_t - \mathrm{div}_x(m_t \alpha(t,\cdot)) \quad t \in [t_0, T]$$

    with fixed initial condition

  - $\mathcal{F} \equiv 0$, $\mathcal{G}(\mu) = \infty 1_{\{\mu \neq \nu_{\mathrm{target}}\}}$: entropic version of optimal transport

Illustration 2: Mean Field Control (2010's, Lions...)

- Minimize cost to 'society'

$$J(\alpha) = \mathcal{G}(m_T) + \int_0^T \Big(\int_{\mathbb{R}^d} \frac{1}{2} |\alpha(t,x)|^2 m_t(\mathrm{d}x) + \mathcal{F}(m_t) \Big) \mathrm{d}t$$

  - infimum taken over Fokker-Planck equations

$$\partial_t m_t = \frac{1}{2} \Delta_x m_t - \mathrm{div}_x\big(m_t \alpha(t,\cdot)\big) \quad t \in [t_0, T]$$

    with fixed initial condition

  - $\mathcal{F} \equiv 0$, $\mathcal{G}(\mu) = \infty 1_{\{\mu \neq \nu_{\text{target}}\}}$: entropic version of optimal transport

- Probabilistic formulation

$$J(\alpha) = \mathcal{G}\left(\mathcal{L}(X_T)\right) + \int_0^T \Big(\frac{1}{2} \mathbb{E}[|\alpha_t|^2] \, \mathrm{d}t + \mathcal{F}(\mathcal{L}(X_t)) \Big) \mathrm{d}t$$

subject to

$$\mathrm{d}X_t = \alpha_t \, \mathrm{d}t + \mathrm{d}B_t, \quad t \in [0,T]; \quad \mathcal{L}(X_0) = m_0$$

Illustration 2: Mean Field Control (2010's, Lions...)

- Minimize cost to 'society'

$$J(\alpha) = \mathcal{G}(m_T) + \int_0^T \Big(\int_{\mathbb{R}^d} \frac{1}{2} |\alpha(t,x)|^2 m_t(\mathrm{d}x) + \mathcal{F}(m_t) \Big) \mathrm{d}t$$

  - infimum taken over Fokker-Planck equations

$$\partial_t m_t = \frac{1}{2} \Delta_x m_t - \mathrm{div}_x \big(m_t \alpha(t,\cdot) \big) \quad t \in [t_0, T]$$

    with fixed initial condition

  - $\mathcal{F} \equiv 0$, $\mathcal{G}(\mu) = \infty 1_{\{\mu \neq \nu_{\text{target}}\}}$: entropic version of optimal transport

- 'Solution'

$$\alpha(t,x) = -\partial_\mu \mathcal{V}(t,\mu)(x)$$

  - $\mathcal{V}(t,\mu)$ optimal cost when population is initialized from μ at time t

Illustration 3: Mean Field Game (2010's, Lions...)

- Selfish individuals may evolve with time
  - focus on one typical player within the population

$$\mathrm{d}X_t = \alpha_t \,\mathrm{d}t + \mathrm{d}B_t, \quad t \in [0, T]$$

  - with α_t being the control (in $\mathbb{R}^d$)

Illustration 3: Mean Field Game (2010's, Lions...)

- Selfish individuals may evolve with time
  - focus on one typical player within the population

$$\mathrm{d}X_t = \alpha_t \,\mathrm{d}t + \mathrm{d}B_t, \quad t \in [0, T]$$

  - with α_t being the control (in $\mathbb{R}^d$)

- Compromise (or equilibrium) is described in terms of a flow

$$t \in [0, T] \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$$

  - equilibrium is unknown, but assuming that the population state $(\mu_t)_{t\in[0,T]}$ is given, typical player wants to minimize

$$\mathscr{J}\Big((X_t, \alpha_t, \mu_t)_{0\le t\le T}\Big)$$

Illustration 3: Mean Field Game (2010's, Lions...)

• Selfish individuals may evolve with time

$$\mathrm{d}X_t = \alpha_t \, \mathrm{d}t + \mathrm{d}B_t, \quad t \in [0, T]$$

• Compromise (or equilibrium) is described in terms of a flow

$$t \in [0, T] \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$$

∘ equilibrium is unknown, but assuming that the population state $(\mu_t)_{t \in [0,T]}$ is given, typical player wants to minimize

$$\mathscr{J}\Big((X_t, \alpha_t, \mu_t)_{0 \le t \le T}\Big)$$

∘ (Nash) equilibrium should be a fixed point

$$(\mu_t)_{t \in [0,T]} \mapsto \big(\mathcal{L}(X_t^{\mathrm{optim}})\big)_{t \in [0,T]}$$

Illustration 3: Mean Field Game (2010's, Lions...)

- Selfish individuals may evolve with time

$$dX_t = \alpha_t \, dt + dB_t, \quad t \in [0, T]$$

- Compromise (or equilibrium) is described in terms of a flow

$$t \in [0, T] \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$$

◦ equilibrium is unknown, but assuming that the population state $(\mu_t)_{t \in [0,T]}$ is given, typical player wants to minimize

$$\mathscr{J}\Big((X_t, \alpha_t, \mu_t)_{0 \le t \le T}\Big)$$

◦ (Nash) equilibrium should be a fixed point

$$(\mu_t)_{t \in [0,T]} \mapsto (\mathcal{L}(X_t^{\text{optim}}))_{t \in [0,T]}$$

- Example

$$\mathscr{J}\Big((X_t, \alpha_t, \mu_t)_{0 \le t \le T}\Big) = \mathbb{E}\Big[\int_0^T \Big(\frac{1}{2}|\alpha_t|^2 + f(X_t, \mu_t)\Big) dt + g(X_T, \mu_T)\Big]$$

◦ Guess: equilibrium described by $\alpha_t = -\partial_x \mathcal{U}(t, X_t, \mu_t)$

3. Randomisation

Philosophy

- General objective

  ◦ noisy version for

$$\partial_t \mu_t = -\mathrm{div}_x(b_t(\cdot, \mu_t)\mu_t) + \frac{\sigma^2}{2}\Delta_x \mu_t$$

where $\mu_t \in \mathcal{P}(\mathbb{R}^d)$

  ◦ what is noise here? Intuitively, should force μ_t to be random

  ◦ motivation: gradient descent on the space of probability measures, mean-field games

Philosophy

- General objective

  ◦ noisy version for

$$\partial_t \mu_t = -\mathrm{div}_x(b_t(\cdot, \mu_t)\mu_t) + \frac{\sigma^2}{2}\Delta_x \mu_t$$

where $\mu_t \in \mathcal{P}(\mathbb{R}^d)$

  ◦ what is noise here? Intuitively, should force μ_t to be random

  ◦ motivation: gradient descent on the space of probability measures, mean-field games

- Intuitively, use kind of Brownian motion on the space of $\mathcal{P}_2(\mathbb{R}^d)$

  ◦ challenging question, even in dimension 1

  ◦ no canonical definition: Stannat [02,06], Sturm and Von Renesse [09], Konarovskyi [15], Dello Schiavo [20]...

Degenerate Model [2010's, Lasry-Lions...]

- Mean field game with common noise W

  ∘ asymptotic formulation for a finite player game with

$$dX_t^i = b(W_t, X_t^i, \bar{\mu}_t^N)\,dt + \sigma\,dB_t^i$$

  ∘ uncontrolled version ⇝ asymptotic SDE with $\bar{\mu}_t^N$ replaced by $\mathcal{L}(X_t|(W_s)_{0\le s\le T}) = \mathcal{L}(X_t|(W_s)_{0\le s\le t})$

  ∘ particles become independent conditional on W and converge to the solution

$$dX_t = b(W_t, X_t, \mathcal{L}(X_t|W))\,dt + \sigma\,dB_t$$

Degenerate Model [2010's, Lasry-Lions...]

- Mean field game with common noise W
  - asymptotic formulation for a finite player game with

$$\mathrm{d}X_t^i = b(X_t^i, \bar{\mu}_t^N)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t^i + \eta\,\mathrm{d}W_t$$

  - particles become independent conditional on W and converge to the solution

$$\mathrm{d}X_t = b\big(X_t, \mathcal{L}(X_t|W)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t + \eta\,\mathrm{d}W_t$$

Degenerate Model [2010's, Lasry-Lions...]

- Mean field game with common noise W
  - asymptotic formulation for a finite player game with

$$dX_t^i = b(X_t^i, \bar{\mu}_t^N)\,dt + \sigma\,dB_t^i + \eta\,dW_t$$

  - particles become independent conditional on W and converge to the solution

$$dX_t = b(X_t, \mathcal{L}(X_t|W))\,dt + \sigma\,dB_t + \eta\,dW_t$$

- Stochastic Fokker-Planck equation

$$\partial_t \mu_t - \mathrm{div}_x(b(\cdot,\mu_t)\mu_t) - \Big(\frac{1}{2} + \frac{1}{2}\Big)\Delta_x \mu_t - \underbrace{\mathrm{div}_x(\mu_t \dot{B}_t)}_{\nabla_x \mu_t \cdot \dot{B}_t} = 0$$

  - completely degenerate model: noise is d-dimensional whilst state variable is infinite dimensional (locally, translation of the mean)

Ex. 1: Time Asymptotic Dictated by the Mean

(with Maillet & Tanré)

- Stationary regimes to

$$\partial_t \mu_t - \mathrm{div}_x\Big(\big[\nabla V + \nabla W * \mu_t\big]\mu_t\Big) + \frac{\sigma^2}{2}\Delta_x \mu_t = 0$$

  ◦ case of interest

V symmetric non-convex (e.g., $V(x) = \frac{x^4}{4} - \frac{x^2}{2}$, 1d), $W(x) = \frac{\alpha}{2}|x|^2$

  ◦ probabilistic interpretation

$$\mathrm{d}X_t = -\nabla_x V(X_t)\,\mathrm{d}t - \alpha\big(X_t - \mathbb{E}(X_t)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

  ◦ competition between α and σ: σ small $\Rightarrow$ non-uniqueness of stationary measures including one with 0 mean

Ex. 1: Time Asymptotic Dictated by the Mean

(with Maillet & Tanré)

- Stationary regimes to

$$\partial_t \mu_t - \mathrm{div}_x\Big(\big[\nabla V + \nabla W * \mu_t\big]\mu_t\Big) + \frac{\sigma^2}{2}\Delta_x \mu_t = 0$$

  ◦ case of interest

V symmetric non-convex (e.g., $V(x) = \frac{x^4}{4} - \frac{x^2}{2}$, 1d), $W(x) = \frac{\alpha}{2}|x|^2$

  ◦ probabilistic interpretation

$$\mathrm{d}X_t = -\nabla_x V(X_t)\,\mathrm{d}t - \alpha\big(X_t - \mathbb{E}(X_t)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

  ◦ competition between α and σ: σ small $\Rightarrow$ non-uniqueness of stationary measures including one with 0 mean

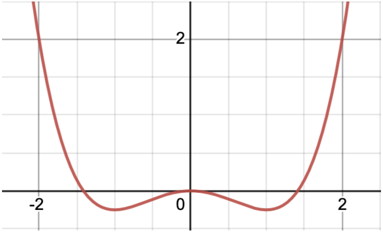
- When $\sigma = 0$, minimize

$$\int_{\mathbb{R}^d} V(x)\,\mathrm{d}\mu(x) + \frac{\alpha}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} W(x-y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$$

Ex. 1: Time Asymptotic Dictated by the Mean

(with Maillet & Tanré)

- When $\sigma = 0$, minimize
$$\int_{\mathbb{R}^d} V(x)\, \mathrm{d}\mu(x) + \frac{\alpha}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} W(x-y)\, \mathrm{d}\mu(x)\, \mathrm{d}\mu(y)$$
  - W quadratic $\Rightarrow$ variance
  - get δ masses at minima of V



Ex. 1: Time Asymptotic Dictated by the Mean

(with Maillet & Tanré)

- Stationary regimes to

$$\partial_t\mu_t - \mathrm{div}_x\Big(\big[\nabla V + \nabla W * \mu_t\big]\mu_t\Big) + \Big(\frac{\sigma^2}{2} + \frac{\sigma_0^2}{2}\Big)\Delta_x\mu_t + \mu_t \cdot \dot{W}_t = 0$$

  - e.g.,

V symmetric non-convex (e.g., $V(x) = \frac{x^4}{4} - \frac{x^2}{2}$, 1d), $W(x) = \frac{\alpha}{2}|x|^2$

  - probabilistic interpretation

$$\mathrm{d}X_t = -\nabla_x V(X_t)\,\mathrm{d}t - \alpha\big(X_t - \mathbb{E}(X_t | W)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t + \sigma_0\,\mathrm{d}W_t$$

- α large, $\sigma_0 > 0$ and σ small, unique invariant regime (law on $\mathcal{P}(\mathbb{R}^d)$)
  - α large and σ small $\Rightarrow X_t \sim \mathbb{E}(X_t | W)$ and

$$\mathrm{d}\mathbb{E}(X_t | W) \sim -\nabla_x V\big(\mathbb{E}(X_t | W)\big)\,\mathrm{d}t + \sigma_0\,\mathrm{d}W_t$$

Ex. 2: Game with Gaussian Equilibria (with Foguen)

- Mean Field Game with dynamics of the form

$$dX_t = \Big[(c_b X_t + b(\mu_t)) + \alpha_t\Big]\,dt + \sigma\,dW_t$$

  - cost functional of the form

$$J(\alpha) = \mathbb{E}\Big[\frac{1}{2}(c_g X_T + g(\mu_T))^2 + \int_0^T \big[\frac{1}{2}(c_f X_t + f(\mu_t))^2 + \frac{1}{2}\alpha_t^2\big]\,dt\Big]$$

  - coefficients c_b, c_f, c_g may be arbitrarily chosen (say 1)
  - $(\mu_t)_{0 \le t \le T}$ flow of probability measures

- Look for fixed point $(\mu_t)_{0 \le t \le T}$ such that

$$\mu_t = \mathrm{Law}(X_t^*), \qquad t \in [0, T]$$

  - in general, no uniqueness (even smooth coefficients)

Ex. 2: Game with Gaussian Equilibria (with Foguen)

- Mean Field Game with dynamics of the form

$$dX_t = \Big[(c_b X_t + b(\mu_t(B))) + \alpha_t\Big]\,dt + \sigma\, dW_t + \sigma_0\, dB_t$$

  - cost functional of the form

$$J(\alpha) = \mathbb{E}\Big[\frac{1}{2}(c_g X_T + g(\mu_T(B)))^2 + \int_0^T [\frac{1}{2}(c_f X_t + f(\mu_t(B)))^2 + \frac{1}{2}\alpha_t^2]\,dt\Big]$$

  - $(\mu_t(B))_{0\le t\le T}$ random flow of probability measures

- Look for fixed point $(\mu_t(B))_{0\le t\le T}$ such that

$$\mu_t(B) = \mathrm{Law}(X_t^*|B), \quad t \in [0,T]$$

  - uniqueness (smooth coefficients...)

Ex. 2: Game with Gaussian Equilibria (with Foguen)

- Mean Field Game with dynamics of the form

$$dX_t = \Big[(c_b X_t + b(\mu_t(B))) + \alpha_t\Big]\,dt + \sigma\, dW_t + \sigma_0\, dB_t$$

  - cost functional of the form

$$J(\alpha) = \mathbb{E}\Big[\frac{1}{2}(c_g X_T + g(\mu_T(B)))^2 + \int_0^T \big[\frac{1}{2}(c_f X_t + f(\mu_t(B)))^2 + \frac{1}{2}\alpha_t^2\big]\,dt\Big]$$

  - $(\mu_t(B))_{0\le t\le T}$ random flow of probability measures

- Look for fixed point $(\mu_t(B))_{0\le t\le T}$ such that

$$\mu_t(B) = \text{Law}(X_t^*|B), \quad t \in [0,T]$$

  - uniqueness (smooth coefficients...)

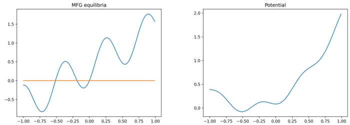
- General form of the optimizer over α when μ is fixed

$$\alpha_t = -\eta_t X_t - h_t$$

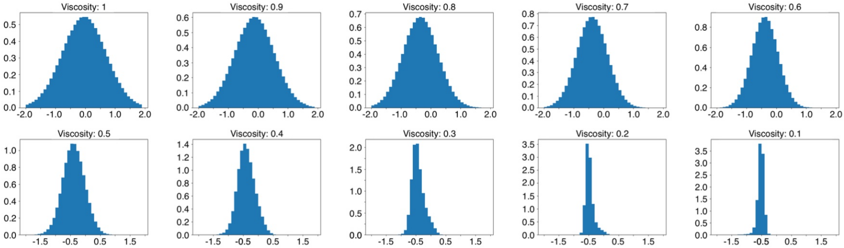
  - η and $h \rightsquigarrow$ deterministic and η independent of μ!
  - X^* is Gaussian with fixed variance $\rightsquigarrow$ fixed point on the mean only!

Selection (in Learning): illustration (with Vasileiadis)

- Potential mean field game: $d = 1, f \equiv 0$, equilibria vs. minimizers



- Selection by randomisation of the population



  - evolution of the random mean state at terminal time

4. Infinite Dimensional Noise

(with Hammersley)

Form of the noise

- Throughout, dimension is 1 (work on $\mathcal{P}_2(\mathbb{R})$)
- Here, follow P.L. Lions' approach to differential calculus on $\mathcal{P}_2(\mathbb{R})$
  - see function $\varphi : \mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto \varphi(\mu) \in \mathbb{R}$ as

$$L^2(\mathbb{S} = \mathbb{R}/\mathbb{Z}, dx) \ni X \mapsto \varphi(\mathcal{L}(X))$$

Form of the noise

- Throughout, dimension is 1 (work on $\mathcal{P}_2(\mathbb{R})$)
- Here, follow P.L. Lions' approach to differential calculus on $\mathcal{P}_2(\mathbb{R})$
  - see function $\varphi : \mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto \varphi(\mu) \in \mathbb{R}$ as

$$L^2(\mathbb{S} = \mathbb{R}/\mathbb{Z}, dx) \ni X \mapsto \varphi(\mathcal{L}(X))$$

... and then define derivative as Fréchet derivative in $L^2(\mathbb{S}, dx)$

- Proceed here in the same way for smoothing out φ:

$$L^2(\mathbb{S}, dx) \ni X \mapsto \varphi(\mathcal{L}(X_t)), \quad t > 0,$$

with $(X_t(x))_{t \geq 0, x \in \mathbb{S}}$ Gaussian process with values in $L^2(\mathbb{S}, dx)$

  - but destroys the mean-field structure!

Form of the noise

- Throughout, dimension is 1 (work on $\mathcal{P}_2(\mathbb{R})$)
- Here, follow P.L. Lions' approach to differential calculus on $\mathcal{P}_2(\mathbb{R})$
  - see function $\varphi : \mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto \varphi(\mu) \in \mathbb{R}$ as

$$L^2(\mathbb{S} = \mathbb{R}/\mathbb{Z}, dx) \ni X \mapsto \varphi(\mathcal{L}(X))$$

- Proceed here in the same way for smoothing out φ:

$$L^2(\mathbb{S}, dx) \ni X \mapsto \varphi(\mathcal{L}(X_t)), \quad t > 0,$$

with $(X_t(x))_{t \geq 0, x \in \mathbb{S}}$ Gaussian process with values in $L^2(\mathbb{S}, dx)$

  - but destroys the mean-field structure!

- In order to make it intrinsic $\rightsquigarrow$ RE-ARRANGE
  - intuitively

$$X_t \rightsquigarrow \text{Gaussian step} \rightsquigarrow \text{re-arrangement} = X_{t+\mathrm{d}t}$$

Re-arrangement in 1d

• Take a probability measure μ on $\mathbb{R}$

$$\mu \leftrightarrow \text{quantile function } F_\mu^{-1}$$

◦ where $x \in (0,1) \mapsto F_\mu^{-1}(x)$ is the quantile function

◦ $x \in (0,1) \mapsto F_\mu^{-1}(x)$ is the canonical random variable for representing μ, i.e.

$$\mathrm{Leb}_{(0,1)} \circ \left(x \in (0,1) \mapsto F_\mu^{-1}(x)\right)^{-1} = \mu$$

Re-arrangement in 1d

- Take a probability measure μ on $\mathbb{R}$

$$\mu \leftrightarrow \text{quantile function } F_{\mu}^{-1}$$

  - where $x \in (0,1) \mapsto F_{\mu}^{-1}(x)$ is the quantile function
  - $x \in (0,1) \mapsto F_{\mu}^{-1}(x)$ is the canonical random variable for representing μ, i.e.

$$\mathrm{Leb}_{(0,1)} \circ \left(x \in (0,1) \mapsto F_{\mu}^{-1}(x)\right)^{-1} = \mu$$

- Conversely, re-arranging $X_t(x)$ in Gaussian dynamics is choosing canonical representative of

$$\mathrm{Leb} \circ (x \in \mathbb{S} \mapsto X_t(x))^{-1}$$

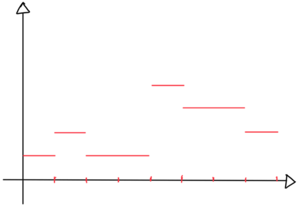
  - on $[0,1)$, choose quantile function of law of $x \mapsto X_t(x)$
  - on $\mathbb{S} \simeq [0,1]$, choose non-decreasing on $[0,1/2]$ and reflect w.r.t. $1/2$ to get it periodic

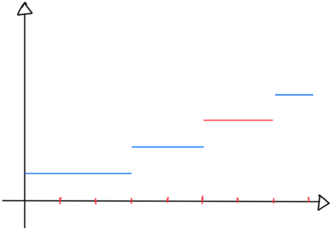
Re-arrangement in 1d – plots

- Simplest example: $X(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_i 1_{[i/N,(i+1)/N)}(x)$
  - rearrangement on [0, 1): $X^*(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_{(i)} 1_{[i/N,(i+1)/N)}(x)$
  - where $a_{(1)} \le a_{(2)} \le ... \le a_{(N)}$ is the non-decreasing rearrangement of $a_1, \cdots, a_N$
  - to get it on $\mathbb{S}$, use contraction of rate $1/2$ and symmetrize

Re-arrangement in 1d – plots

- Simplest example: $X(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_i 1_{[i/N,(i+1)/N)}(x)$
  - rearrangement on [0, 1): $X^*(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_{(i)} 1_{[i/N,(i+1)/N)}(x)$
  - where $a_{(1)} \leq a_{(2)} \leq \ldots \leq a_{(N)}$ is the non-decreasing rearrangement of $a_1, \cdots, a_N$
  - to get it on $\mathbb{S}$, use contraction of rate 1/2 and symmetrize





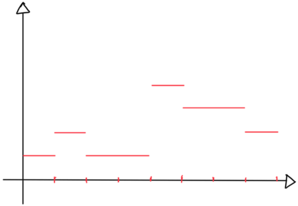
Re-arrangement in 1d – plots

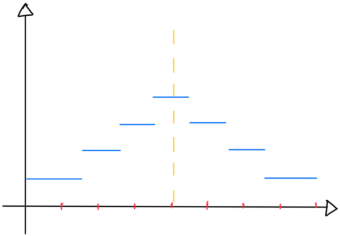
- Simplest example: $X(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_i 1_{[i/N,(i+1)/N)}(x)$

    - rearrangement on [0, 1): $X^*(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_{(i)} 1_{[i/N,(i+1)/N)}(x)$

    - where $a_{(1)} \leq a_{(2)} \leq ... \leq a_{(N)}$ is the non-decreasing rearrangement of $a_1, \cdots, a_N$

    - to get it on $\mathbb{S}$, use contraction of rate 1/2 and symmetrize





Euler scheme with colored noise

- Replace white noise by colored noise

$$\widetilde{W}_t(x) = \sum_{m\in\mathbb{Z}} m^{-\lambda} W_t^m e_m(x)$$

where $\lambda \in (1/2, 1]$ and $((W_t^m)_{t\geq 0})_{m\in\mathbb{Z}}$ are independent Brownian motions

- $\mathbb{E}[\|\widetilde{W}_t(\cdot)\|_2^2] = ct < \infty$
- the noise takes values in $L^2(\mathbb{S}, \text{Leb})$

Euler scheme with colored noise

- Replace white noise by colored noise

$$\widetilde{W}_t(x) = \sum_{m \in \mathbb{Z}} m^{-\lambda} W_t^m e_m(x)$$

where $\lambda \in (1/2, 1]$ and $((W_t^m)_{t \geq 0})_{m \in \mathbb{Z}}$ are independent Brownian motions

  - $\mathbb{E}[\|\widetilde{W}_t(\cdot)\|_2^2] = ct < \infty$
  - the noise takes values in $L^2(\mathbb{S}, \mathrm{Leb})$

- Scheme

$$X_{n+1}^h = \left[e^{h\Delta} X_n^h + \int_0^h e^{(h-s)\Delta} \, \mathrm{d}\widetilde{W}_{nh+s} \right]^*$$

  - $h > 0$ is a time step
  - get tightness in any $C([0, T]; L^2(\mathbb{S}, dx))$

Smoothing effect

- Semi-group of limiting dynamics

$$\mathcal{P}_t : X_0 \in L^2(\mathbb{S}, \mathrm{Leb}) \mapsto \mathbb{E}\left[\varphi\left(X_t^{X_0^*}\right)\right]$$

  - for $\varphi : L^2(\mathbb{S}, \mathrm{Leb}) \to \mathbb{R}$ bounded and measurable

Smoothing effect

- Semi-group of limiting dynamics

$$\mathcal{P}_t : X_0 \in L^2(\mathbb{S}, \mathrm{Leb}) \mapsto \mathbb{E}\Big[\varphi\Big(X_t^{X_0^*}\Big)\Big]$$

  - for $\varphi : L^2(\mathbb{S}, \mathrm{Leb}) \to \mathbb{R}$ bounded and measurable

- Bound on the Lipschitz constant

$$\Big|\mathcal{P}_t\varphi((X_0 + z)^*) - \mathcal{P}_t\varphi(X_0^*)\Big| \leq \frac{C_T}{t^{(1+\lambda)/2}} ||\varphi||_\infty ||z||_{L^2}$$

  - for $t \in (0, T]$

Smoothing effect

- Semi-group of limiting dynamics

$$\mathcal{P}_t : X_0 \in L^2(\mathbb{S}, \mathrm{Leb}) \mapsto \mathbb{E}\Big[\varphi\Big(X_t^{X_0^*}\Big)\Big]$$

  - for $\varphi : L^2(\mathbb{S}, \mathrm{Leb}) \to \mathbb{R}$ bounded and measurable

- Bound on the Lipschitz constant

$$\Big|\mathcal{P}_t\varphi((X_0 + z)^*) - \mathcal{P}_t\varphi(X_0^*)\Big| \leq \frac{C_T}{t^{(1+\lambda)/2}} \|\varphi\|_\infty \|z\|_{L^2}$$

  - for $t \in (0, T]$

- Discussion on the rate
  - blow-up exponent $(1 + \lambda)/2 \in (3/4, 1)$, close to $3/4$ for $\lambda \sim 1/2$
  - NOT AS GOOD as in finite dimension (blow up like $t^{-1/2}$)
  - but INTEGRABLE in small time, which is crucial for nonlinear models

5. Applications

Stochastic Gradient Descent on $\mathcal{P}_2(\mathbb{R})$ (with Hammersley)

- Assume V is smooth potential that confines the mean, typically

$$V(\mu) = V_0(\mu) + \lambda\Big(\int_{\mathbb{R}} x d\mu(x)\Big)^2,$$

for V_0 smooth (with bounded derivatives)

$$X_t \rightsquigarrow -\text{Gradient step} + \text{Gaussian move} \rightsquigarrow \text{re-arrangement} = X_{t+\mathrm{d}t}$$

  - solution to SGD, unique invariant measure and convergence

Stochastic Gradient Descent on $\mathcal{P}_2(\mathbb{R})$ (with Hammersley)

- Assume V is smooth potential that confines the mean, typically

$$V(\mu) = V_0(\mu) + \lambda\Big(\int_{\mathbb{R}} x d\mu(x)\Big)^2,$$

for V_0 smooth (with bounded derivatives)

$X_t \rightsquigarrow -$Gradient step $+$ Gaussian move $\rightsquigarrow$ re-arrangement $= X_{t+\mathrm{d}t}$

  - solution to SGD, unique invariant measure and convergence

- No explicit shape of the invariant measure but metastability for rescaled Gaussian move
  - same result
  - mean time to exit from convex well is of order $\exp(a/\varepsilon^2)$ for a the height of the well

Application to MFG (with Ouknine)

- Back to the first section $\rightsquigarrow$ MFG without idiosyncratic noise
  - 1d representative player $\rightsquigarrow$ $\mathrm{d}X_t = \alpha_t \,\mathrm{d}t$
  - cost functional with f, g convex in x

$$J(\alpha) = \mathbb{E}\Big[g(X_T, \mu_T) + \int_0^T \Big(f(X_t, \mu_t) + \tfrac{1}{2}|\alpha_t|^2 \Big) \,\mathrm{d}t \Big]$$

Application to MFG (with Ouknine)

- Back to the first section $\rightsquigarrow$ MFG without idiosyncratic noise
  - 1d representative player $\rightsquigarrow$ $\mathrm{d}X_t = \alpha_t \, \mathrm{d}t$
  - cost functional with f, g convex in x

$$J(\alpha) = \mathbb{E}\Big[g(X_T, \mu_T) + \int_0^T \Big(f(X_t, \mu_t) + \frac{1}{2}|\alpha_t|^2\Big)\, \mathrm{d}t\Big]$$

- Optimal trajectories with $\mu_t = \mathcal{L}(X_t)$ (on $L^2(\mathbb{S}, dx)$)

$$\mathrm{d}X_t(x) = -Y_t(x)\, \mathrm{d}t$$
$$\mathrm{d}Y_t(x) = -\partial_x f\big(X_t(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1}\big)$$
$$Y_T(x) = \partial_x g\big(X_T(x), \mathrm{Leb}_{\mathbb{S}} \circ X_T^{-1}\big), \quad x \in \mathbb{S}$$

Application to MFG (with Ouknine)

- Back to the first section $\rightsquigarrow$ MFG without idiosyncratic noise
  - 1d representative player $\rightsquigarrow$ $\mathrm{d}X_t = \alpha_t \, \mathrm{d}t$
  - cost functional with f, g convex in x

$$J(\alpha) = \mathbb{E}\Big[g(X_T, \mu_T) + \int_0^T \Big(f(X_t, \mu_t) + \tfrac{1}{2}|\alpha_t|^2 \Big) \, \mathrm{d}t \Big]$$

- Optimal trajectories with $\mu_t = \mathcal{L}(X_t)$ (on $L^2(\mathbb{S}, dx)$)

$$\begin{aligned} \mathrm{d}X_t(x) &= -Y_t(x) \, \mathrm{d}t \ + \text{noise} \\ \mathrm{d}Y_t(x) &= -\partial_x f\big(X_t(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1}\big) \\ Y_T(x) &= \partial_x g\big(X_T(x), \mathrm{Leb}_{\mathbb{S}} \circ X_T^{-1}\big), \quad x \in \mathbb{S} \end{aligned}$$

  - $\partial_x f$ and $\partial_x g$ smooth, then existence and uniqueness hold for stochastic system! Solution is distributed:

$$Y_t(x) = v(t, X_(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1})$$

Application to MFG (with Ouknine)

- Back to the first section $\rightsquigarrow$ MFG without idiosyncratic noise
  - 1d representative player $\rightsquigarrow$ $\mathrm{d}X_t = \alpha_t \, \mathrm{d}t$
  - cost functional with f, g convex in x

$$J(\alpha) = \mathbb{E}\Big[g(X_T, \mu_T) + \int_0^T \Big(f(X_t, \mu_t) + \tfrac{1}{2}|\alpha_t|^2 \Big) \, \mathrm{d}t \Big]$$

- Optimal trajectories with $\mu_t = \mathcal{L}(X_t)$ (on $L^2(\mathbb{S}, dx)$)

$$\mathrm{d}X_t(x) = -Y_t(x) \, \mathrm{d}t \;\; + \text{noise}$$

$$\mathrm{d}Y_t(x) = -\partial_x f\big(X_t(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1}\big) \;\; + \ldots$$

$$Y_T(x) = \partial_x g\big(X_T(x), \mathrm{Leb}_{\mathbb{S}} \circ X_T^{-1}\big), \quad x \in \mathbb{S}$$

  - $\partial_x f$ and $\partial_x g$ smooth, then existence and uniqueness hold for stochastic system! Solution is distributed:

$$Y_t(x) = v(t, X_(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1})$$