Rearranged stochastic heat equation

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

Conference in Honor of Denis Talay

September 4-8, 2023

Joint works W. Hammersley (Nice, Université Côte d'Azur, France)

& Y. Ouknine (Marrakech, Morocco)

1. Motivation

Background

• 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 $(B_t)_{t\geq 0}$

$$dX_t = b_t(X_t)dt + dB_t$$

Background

• 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 $(B_t)_{t\geq 0}$

$$dX_t = b_t(X_t)dt + dB_t$$

• General objective

∘ similar smoothing properties but for

$$\partial_t \mu_t = -\mathrm{div}(b_t(\cdot, \mu_t)\mu_t)$$

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

∘ what is B here? Intuitively, should force μ_t to be random

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

Form of the noise

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

∘ throughout, dimension is 1 (work on $\mathcal{P}(\mathbb{R})$)

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

Form of the noise

- Intuitively, use kind of Brownian motion on the space of $\mathcal{P}_2(\mathbb{R})$
  - 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 $\varphi : \mathcal{P}(\mathbb{R}) \ni \mu \mapsto \varphi(\mu) \in \mathbb{R}$ 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 $L^2(\mathbb{S}, dx)$

Form of the noise

- Intuitively, use kind of Brownian motion on the space of $\mathcal{P}_2(\mathbb{R})$
  - 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 $\varphi : \mathcal{P}(\mathbb{R}) \ni \mu \mapsto \varphi(\mu) \in \mathbb{R}$ 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

$$dX_t(x) = \Delta X_t(x)dt + dW_t(x), \quad t > 0; \quad X_0(x) = X(x),$$

where $(W_t(x))_{t\geq 0, x\in\mathbb{R}}$ white noise with values in $L^2(\mathbb{S}, dx)$

  - but destroys the mean-field structure!

2. Rearranged Noise

General plan

- Throughout, $\boxed{d = 1}$
- Take as before SHE

$$dX_t(x) = \Delta X_t(x)dt + dW_t(x), \quad x \in \mathbb{S}, \ t \geq 0,$$

  - with

$$W_t(x) = \sum_{m \in \mathbb{Z}} W_t^m e_m(x)$$

where $((W_t^m)_{t \geq 0})_{m \in \mathbb{Z}}$ are independent Brownian motions and $(e_m)_{m \in \mathbb{Z}}$ shorter notation for Fourier basis

- Recall the shape of the solution

$$X_t(x) = \left[\exp(t\Delta)X_0 + \int_0^t \exp((t-s)\Delta)dW_s\right](x)$$

General plan

- Throughout, $\boxed{d = 1}$
- Take as before SHE

$$dX_t(x) = \Delta X_t(x)dt + dW_t(x), \quad x \in \mathbb{S}, \; t \geq 0,$$

  - with

$$W_t(x) = \sum_{m \in \mathbb{Z}} W_t^m e_m(x)$$

where $((W_t^m)_{t \geq 0})_{m \in \mathbb{Z}}$ are independent Brownian motions and $(e_m)_{m \in \mathbb{Z}}$ shorter notation for Fourier basis

- Recall the shape of the solution

$$X_t(x) = \left[\exp(t\Delta)X_0 + \int_0^t \exp((t-s)\Delta)dW_s\right](x)$$

- In order to make it intrinsic $\rightsquigarrow$ $\boxed{\text{RE-ARRANGE}}$
- Intuitively

$$X_t \rightsquigarrow \left[\exp(dt\Delta)X_t + \int_0^{dt} \exp((dt-s)\Delta)dW_{t+s}\right] \rightsquigarrow \text{re-arrangement} = X_{t+dt}$$

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 SHE 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 $a_{(1)} \leq a_{(2)} \leq ... \leq a_{(N)}$ 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

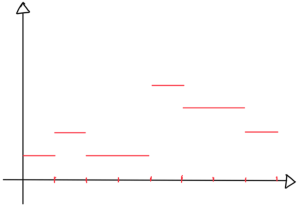
Re-arrangement in 1d – plots

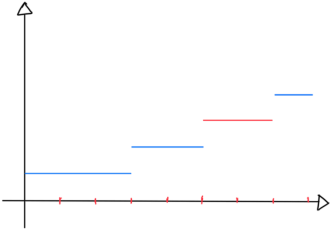
- Simplest example: $X(x) = \frac{1}{N}\sum_{i=0}^{N-1} a_i 1_{[i/N,(i+1)/N)}(x)$

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

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

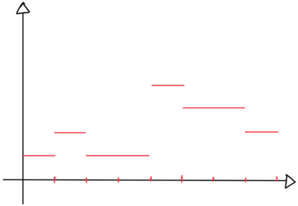
  - to get it on $\mathbb{S}$, use contraction of rate $1/2$ and symmetrize

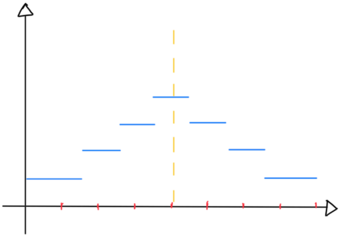




Re-arrangement in 1d – plots

- 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





Euler scheme with white noise

- Naive idea (from the general plan)

$$X_{n+1}^h = \left[e^{h\Delta} X_n^h + \int_0^h e^{(h-s)\Delta} dW_{nh+s} \right]^*$$

  - $h > 0$ is a time step

Euler scheme with white noise

- Naive idea (from the general plan)

$$X_{n+1}^h = \left[e^{h\Delta} X_n^h + \int_0^h e^{(h-s)\Delta} dW_{nh+s} \right]^*$$

  - $h > 0$ is a time step

- Not able to prove *tightness* (i.e., weak compactness)!

Euler scheme with white noise

- Naive idea (from the general plan)

$$X^h_{n+1} = \left[e^{h\Delta} X^h_n + \int_0^h e^{(h-s)\Delta} dW_{nh+s} \right]^*$$

  - $h > 0$ is a time step

- Not able to prove *tightness* (i.e., weak compactness)!

- Principle of the analysis taken from Brenier [09]

  - use non-expansion of the re-arrangement

$$\|u^* - v^*\|^2_{2,\mathbb{S}} = \int_{\mathbb{S}} |u^*(x) - v^*(x)|^2 dx \le \int_{\mathbb{S}} |u(x) - v(x)|^2 dx = \|u - v\|^2_{2,\mathbb{S}}$$

with $u^* = X^h_{n+1}$ and $\underbrace{v^* = e^{((n+1)-N)h\Delta} X^h_N}_{\text{sym. } \nearrow}$ for $N \le n$

and $u = e^{h\Delta} X^h_n + \int_0^h e^{(h-s)\Delta} dW_{nh+s}$ and $v = v^*$ for $N \le n$

Euler scheme with white noise

- Naive idea (from the general plan)

$$X^h_{n+1} = \left[e^{h\Delta} X^h_n + \int_0^h e^{(h-s)\Delta} dW_{nh+s} \right]^*$$

  - $h > 0$ is a time step

- Not able to prove *tightness* (i.e., weak compactness)!
- Principle of the analysis taken from Brenier [09]

$$\mathbb{E}\left[\left\| X^h_{n+1} - e^{((n+1)-N)h\Delta} X^h_N \right\|^2_{2,\mathbb{S}} \right]$$
$$\leq \mathbb{E}\left[\left\| e^{h\Delta}\left(X^h_n - e^{(n-N)h\Delta} X^h_N \right) \right\|^2_{2,\mathbb{S}} \right] + \underbrace{\mathbb{E}\left[\left\| \int_0^h e^{(h-s)\Delta} dW_{nh+s} \right\|^2_{2,\mathbb{S}} \right]}_{h^{1-\cdots}}$$

  - use contraction of $e^{h\Delta} \rightsquigarrow h^{-1} h^{1-\cdots} = h^{-\cdots} \rightsquigarrow$ BAD
  - need to combine $e^{h\Delta}$ and $* \rightsquigarrow$ NO SIMPLE WAY

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 $L^2(\mathbb{S}, \text{Leb})$

- May wonder why Δ is still needed in the equation
  - for the smoothing effect!! [Da Prato, ...]

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 $L^2(\mathbb{S}, \mathrm{Leb})$

- New scheme

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

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

3. Rearranged SHE

Equation satisfied by limit process

- Brenier's work $\rightsquigarrow$ infinitesimal impact of re-arrangement = reflection on symmetric non-decreasing functions
- Get a reflected SHE

$$dX_t = \Delta X_t dt + d\widetilde{W}_t + d\eta_t$$

  - recall that $X_t \in L^2(\mathbb{S}, \text{Leb})$ by symmetric non-decreasing

Equation satisfied by limit process

• Brenier's work $\rightsquigarrow$ infinitesimal impact of re-arrangement = reflection on symmetric non-decreasing functions

• Get a reflected SHE

$$dX_t = \Delta X_t dt + d\widetilde{W}_t + d\eta_t$$

∘ recall that $X_t \in L^2(\mathbb{S}, \text{Leb})$ by symmetric non-decreasing

∘ reflected SPDE $\rightsquigarrow$ Donati-Martin & Pardoux, Nualart & Pardoux, Zambotti (reflection to preserve positivity), Barbu & Da Prato & Tubaro, Röckner & Zhu and & Zhu (more general treatment)

Equation satisfied by limit process

- Brenier's work $\rightsquigarrow$ infinitesimal impact of re-arrangement = reflection on symmetric non-decreasing functions

- Get a reflected SHE

$$dX_t = \Delta X_t dt + d\widetilde{W}_t + d\eta_t$$

  - recall that $X_t \in L^2(\mathbb{S}, \text{Leb})$ by symmetric non-decreasing

- What is η_t?

$$d\eta_t = \left(e^{dt\Delta} X_t + \int_0^{dt} e^{(dt-s)\Delta} d\widetilde{W}_{t+s}\right)^* - \left(e^{dt\Delta} X_t + \int_0^{dt} e^{(dt-s)\Delta} d\widetilde{W}_{t+s}\right)$$

  - if u is smooth and symmetric non-decreasing

$$\langle u, d\eta_t \rangle_{2,\mathbb{S}} \geq 0$$

  - if $(z_t)_{t\geq 0}$ is smooth, symmetric $\nearrow$ and varies smoothly in time

$$\int_0^t \langle z_s, d\eta_s \rangle_{2,\mathbb{S}}$$

  makes sense (think of Stieltjes-integral) and ≥ 0

Equation satisfied by limit process

- Brenier's work $\rightsquigarrow$ infinitesimal impact of re-arrangement = reflection on symmetric non-decreasing functions

- Get a reflected SHE

$$dX_t = \Delta X_t dt + d\widetilde{W}_t + d\eta_t$$

  - recall that $X_t \in L^2(\mathbb{S}, \mathrm{Leb})$ by symmetric non-decreasing

- What is η_t?

$$d\eta_t = \left(e^{dt\Delta}\, X_t + \int_0^{dt} e^{(dt-s)\Delta} d\widetilde{W}_{t+s}\right)^* - \left(e^{dt\Delta}\, X_t + \int_0^{dt} e^{(dt-s)\Delta} d\widetilde{W}_{t+s}\right)$$

  - if u is smooth and symmetric non-decreasing

$$\langle u, d\eta_t \rangle_{2,\mathbb{S}} \geq 0$$

  - if $(z_t)_{t\geq 0}$ is smooth, symmetric $\nearrow$ and varies smoothly in time

$$\int_0^t \langle z_s, d\eta_s \rangle_{2,\mathbb{S}} = \sum_m \int_0^t \langle z_s, e_m \rangle_{2,\mathbb{S}} d\langle \eta_s, e_m \rangle_{2,\mathbb{S}}$$

makes sense (think of Stieltjes-integral) and ≥ 0

Definition of a solution (with W. Hammersley)

• For $(X_t)_{t\geq 0}$ a continuous process with values in $L^2(\mathbb{S}, \text{Leb})$ with each X_t symmetric non-decreasing

Definition of a solution (with W. Hammersley)

- For $(X_t)_{t\geq 0}$ a continuous process with values in $L^2(\mathbb{S}, \text{Leb})$ with each X_t symmetric non-decreasing

- We require the equation to be satisfied in a weak sense

$$\langle X_t - X_s, u\rangle_{2,\mathbb{S}} = \int_s^t \langle X_r, \Delta u\rangle_{2,\mathbb{S}} dr + \langle \widetilde{W}_t - \widetilde{W}_s, u\rangle_{2,\mathbb{S}} + \langle \eta_t - \eta_s, u\rangle_{2,\mathbb{S}}$$

  - for u smooth function on $\mathbb{S}$

Definition of a solution (with W. Hammersley)

- For $(X_t)_{t\geq 0}$ a continuous process with values in $L^2(\mathbb{S}, \mathrm{Leb})$ with each X_t symmetric non-decreasing

- We require the equation to be satisfied in a weak sense

$$\langle X_t - X_s, u\rangle_{2,\mathbb{S}} = \int_s^t \langle X_r, \Delta u\rangle_{2,\mathbb{S}} dr + \langle \widetilde{W}_t - \widetilde{W}_s, u\rangle_{2,\mathbb{S}} + \langle \eta_t - \eta_s, u\rangle_{2,\mathbb{S}}$$

  - for u smooth function on $\mathbb{S}$

- Non-decreasing property of the reflection term

$$\int_s^t \langle e^{\varepsilon\Delta} Z_r, d\eta_r\rangle_{2,\mathbb{S}} \geq 0,$$

  - if $(Z_r)_{r\geq 0}$ continuous process with values in $L^2(\mathbb{S}, \mathrm{Leb})$ such that Z_r is symmetric decreasing

  - $\varepsilon > 0$ is an arbitrarily small regularization parameter

Definition of a solution (with W. Hammersley)

- For $(X_t)_{t\geq 0}$ a continuous process with values in $L^2(\mathbb{S}, \mathrm{Leb})$ with each X_t symmetric non-decreasing
- We require the equation to be satisfied in a weak sense

$$\langle X_t - X_s, u\rangle_{2,\mathbb{S}} = \int_s^t \langle X_r, \Delta u\rangle_{2,\mathbb{S}} dr + \langle \widetilde{W}_t - \widetilde{W}_s, u\rangle_{2,\mathbb{S}} + \langle \eta_t - \eta_s, u\rangle_{2,\mathbb{S}}$$

  - for u smooth function on $\mathbb{S}$

- Non-decreasing property of the reflection term

$$\int_s^t \langle e^{\varepsilon\Delta} Z_r, d\eta_r\rangle_{2,\mathbb{S}} \geq 0,$$

  - if $(Z_r)_{r\geq 0}$ continuous process with values in $L^2(\mathbb{S}, \mathrm{Leb})$ such that Z_r is symmetric decreasing
  - $\varepsilon > 0$ is an arbitrarily small regularization parameter

- Orthogonality principle

$$\lim_{\varepsilon\searrow 0} \mathbb{E} \int_s^t \langle e^{\varepsilon\Delta} X_r, d\eta_r\rangle_{2,\mathbb{S}} = 0$$

Definition of a solution (with W. Hammersley)

- We require the equation to be satisfied in a weak sense

$$\langle X_t - X_s, u\rangle_{2,\mathbb{S}} = \int_s^t \langle X_r, \Delta u\rangle_{2,\mathbb{S}} dr + \langle \widetilde{W}_t - \widetilde{W}_s, u\rangle_{2,\mathbb{S}} + \langle \eta_t - \eta_s, u\rangle_{2,\mathbb{S}}$$

  - for u smooth function on $\mathbb{S}$

- Non-decreasing property of the reflection term

$$\int_s^t \langle e^{\varepsilon\Delta} Z_r, d\eta_r\rangle_{2,\mathbb{S}} \geq 0,$$

  - if $(Z_r)_{r\geq 0}$ continuous process with values in $L^2(\mathbb{S}, \text{Leb})$ such that Z_r is symmetric decreasing
  - $\varepsilon > 0$ is an arbitrarily small regularization parameter

- Orthogonality principle

$$\lim_{\varepsilon \searrow 0} \mathbb{E} \int_s^t \langle e^{\varepsilon\Delta} X_r, d\eta_r\rangle_{2,\mathbb{S}} = 0$$

- Implies uniqueness as in finite dimension

4. Smoothing Effect

Result (with W. Hammersley)

- Smoothing effect of the semi-group is standard folklore of SPDEs

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

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

Result (with W. Hammersley)

- Smoothing effect of the semi-group is standard folklore of SPDEs

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

  - for $\varphi : L^2(\mathbb{S}, \text{Leb}) \to \mathbb{R}$ 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 $t \in (0, T]$

Result (with W. Hammersley)

- Smoothing effect of the semi-group is standard folklore of SPDEs

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

  - 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. Application to Stochastic Gradient Descent on $\mathcal{P}(\mathbb{R})$

Gradient Descent

- Minimization problem

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

Gradient Descent

- Minimization problem

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

- Gradient descent

$$dX_t(\omega) = -\partial_\mu V(\mu_t, X_t(\omega))dt, \quad \mu_t := \mathcal{L}(X_t)$$

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

$$\partial_\mu V(\mathcal{L}(X))(X(x)) = D_{L^2(\mathbb{S},dx)}[V(\mathcal{L}(X))](x), \quad x \in \mathbb{S}$$

Gradient Descent

- Minimization problem

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

- Gradient descent

$$dX_t(\omega) = -\partial_\mu V(\mu_t, X_t(\omega))dt, \quad \mu_t := \mathcal{L}(X_t)$$

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

$$\partial_\mu V(\mathcal{L}(X))(X(x)) = D_{L^2(\mathbb{S},dx)}[V(\mathcal{L}(X))](x), \quad x \in \mathbb{S}$$

- Stochastic gradient descent

$$dX_t(x) = -\partial_\mu V(\mathcal{L}(X_t))(X_t(x))dt + \Delta X_t(x)dt + d\widetilde{W}_t(x) + d\eta_t(x)$$

for $x \in \mathbb{S}$ and $t \geq 0$

  - same interpretation as before

Results (with W. 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)

  - solution to SGD and unique invariant measure

Results (with W. 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)

- No explicit shape of the invariant measure but metastability for rescaled forcing

$$dX_t^\varepsilon(x) = -\partial_\mu V(\mathcal{L}(X_t^\varepsilon))(X_t^\varepsilon(x))dt + \varepsilon^2 \Delta_x X_t^\varepsilon(x)dt + \varepsilon dW_t^\varepsilon(x) + d\eta_t^\varepsilon(x)$$

  - where

$$\widetilde{W}_t^\varepsilon(x) = \sum_{|m|\leq \varepsilon^{-1}} B_t^m e_m(x) + \sum_{|m|>\varepsilon^{-1}} m^{-\lambda} B_t^m e_m(x)$$

  - same result

Results (with W. 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)

- metastability for rescaled forcing

$$\begin{aligned} dX_t^\varepsilon(x) &= -\partial_\mu V(\mathcal{L}(X_t^\varepsilon))(X_t^\varepsilon(x))dt \\ &\quad + \varepsilon^2 \Delta_x X_t^\varepsilon(x)dt + \varepsilon dW_t^\varepsilon(x) + d\eta_t^\varepsilon(x) \end{aligned}$$

  - where

$$\widetilde{W}_t^\varepsilon(x) = \sum_{|m|\leq\varepsilon^{-1}} B_t^m e_m(x) + \sum_{|m|>\varepsilon^{-1}} m^{-\lambda} B_t^m e_m(x)$$

  - same result and mean time to exit from convex well is of order $\exp(a/\varepsilon^2)$ for a the height of the well

6. Application to mean field games

Application to MFG (with Y. Oukníne)

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

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

Application to MFG (with Y. Ouknine)

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

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

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

$$\begin{aligned} dX_t(x) &= -Y_t(x)dt \\ dY_t(x) &= -\partial_x f(X_t(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1})dt \\ Y_T(x) &= \partial_x g(X_T(x), \mathrm{Leb}_{\mathbb{S}} \circ X_T^{-1}), \quad x \in \mathbb{S} \end{aligned}$$

Application to MFG (with Y. Ouknine)

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

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

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

$$\begin{aligned}
dX_t(x) &= -Y_t(x)dt \ + \Delta X_t(x)dt + dW_t(x) + d\eta_(x) \\
dY_t(x) &= -\partial_x f(X_t(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1})dt \\
Y_T(x) &= \partial_x g(X_T(x), \mathrm{Leb}_{\mathbb{S}} \circ X_T^{-1}), \quad x \in \mathbb{S}
\end{aligned}$$

  - $\partial_x f$ 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 Y. Ouknine)

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

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

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

$$\begin{aligned} dX_t(x) &= -Y_t(x)dt \ + \Delta X_t(x)dt + dW_t(x) + d\eta_(x) \\ dY_t(x) &= -\partial_x f(X_t(x), \mathrm{Leb}_{\mathbb{S}} \circ X_t^{-1})dt \ + \ dM_t(x) \\ Y_T(x) &= \partial_x g(X_T(x), \mathrm{Leb}_{\mathbb{S}} \circ X_T^{-1}), \quad x \in \mathbb{S} \end{aligned}$$

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

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

Combining with Idiosyncratic Noise

• Consider $(B_t)_{t\geq 0}$ another Brownian motion constructed on some Ω, whilst $\widetilde{W}$ is constructed on some Ω^0

∘ with g_t Gaussian kernel, let

$$\mu_t(\omega_0) = \mathcal{L}_{x,\omega}\Big(X_t(x,\omega_0) + B_t(\omega)\Big) = g_t \star \mathcal{L}_x\Big(X_t(x,\omega_0)\Big)$$

• Trotter-Kato?

$$X_0 \rightsquigarrow X_{dt}^{X_0}(x,\omega_0) \rightsquigarrow g_{dt} \star \mathcal{L}_x\Big(X_{dt}^{X_0}(x,\omega_0)\Big)$$

∘ generator is $\varphi \mapsto \frac{1}{2}\partial_x\partial_\mu\varphi + \mathbf{L}_\mu\varphi$

∘ new stochastic differential inclusion for $X_t(x)$?

• Higher dimension?

∘ replace re-arrangement by optimal transport: $X \sim \nabla_x\varphi(U)$ for U d-dimensional with a density