Regularising gradient descents on the space of probability measures with the rearranged stochastic heat equation

William Hammersley

Mean Field Models
12 – 16th June 2023
Rennes, France

Joint Work with François Delarue

Laboratoire J.A. Dieudonné, Université Côte d'Azur
Supported by the ERC Horizon program
(ELISA project, Grant agreement No. 101054746)

Motivation

We are interested in the minimisation problem,

$$\min_{\mu \in \mathcal{P}_2(\mathbb{R})} \{V(\mu)\}.$$

Following the mean field approach, one may consider the McKean-Vlasov dynamics:

$$\begin{aligned} dX_t(\omega) &= -\partial_\mu V(\mu_t, X_t(\omega))dt \\ \mu_t &:= \mathbb{P} \circ X_t^{-1} \\ X_0 &\sim \mu_0 = \nu \end{aligned}$$

Using Itô's formula along a flow of measures:

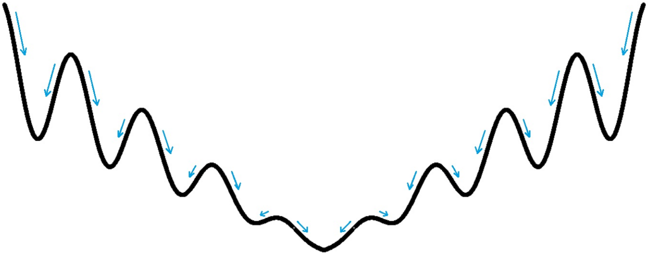
$$dV(\mu_t) = -\int_{\mathbb{R}} (\partial_\mu V(\mu_t, \cdot))^2 d\mu_t dt$$

Motivation

However, there are two issues to note:

Firstly, there are issues of solvability of this dynamics.

Secondly, without strong convexity assumptions, one does not expect that the gradient flow of the marginal laws will converge towards a solution of the minimisation problem.



To combat these issues, one can turn to two well-studied stochastic models: those of private and common noise.

Motivation

Introducing noise at the level of the particle system, we can examine what changes in the infinite particle limit.

$$dX_t^{i,N} = -\partial_\mu V(\mu_t^N, X_t^{i,N})dt + \sigma dM_t^{i,N} \overset{N\to\infty}{\to} dX_t = -\partial_\mu V(\mu_t, X_t)dt + \sigma dM_t$$

$$\mu_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{i,N}} \qquad \to \qquad \mu_t := \mathscr{L}(X_t|\mathcal{G}_t),$$

In the private noise model, $M^{i,N}$ are typically independent Brownian motions and the conditioning filtration $\{\mathcal{G}_t\}_t$ is trivial.

A non-trivial conditioning filtration $\{\mathcal{G}_t\}_t$ arises from correlation of the stochastic inputs $M^{i,N}$. For example, in the common noise case where $M^{i,N} := W^{i,N} + B$, one has $\mathcal{F}_t^B \subseteq \mathcal{G}_t$, $\forall t$.

Motivation

As is well-known, in the case with **private noise**, the corresponding Fokker-Planck-Kolmorogorov equation is modified with a Laplacian:

$$d\langle \varphi, \mu_t \rangle = -\langle \nabla\varphi \cdot \partial_\mu V(\mu_t, \cdot), \mu_t \rangle dt + \frac{1}{2}\sigma^2 \langle \Delta\varphi, \mu_t \rangle dt$$

Whilst, in many cases the issue of solvability is remedied, the evolution above is still deterministic.

In the **common noise** model, for example when $\sigma M^{i,N} := \sigma_1 W^{i,N} + \sigma_2 B$:

$$d\langle \varphi, \mu_t \rangle = -\langle \nabla\varphi \cdot \partial_\mu V(\mu_t, \cdot), \mu_t \rangle dt + \frac{1}{2}\sigma^2 \langle \Delta\varphi, \mu_t \rangle dt + \sigma_2 \langle \nabla\varphi, \mu_t \rangle dB_s.$$

The gradient flow is now random, however, the exploration of the resulting dynamics is limited by the dimension of the common noise.

The minimisation problem for the potential V is set over the space of probability measures and so to develop a noise that would explore appropriately in this case, one would require an infinite dimensional noise.

Motivation

Rather than introducing noise at the finite particle regime, one may consider directly regularising by noise, the formal mean-field limit dynamics.

Then, to proceed with the mean field approach, one reverse-engineers the particle systems converging to this regularised mean-field limit.

This is the approach that we are following in this project.

Related to the study of diffusions in the space of measures, eg: Wasserstein Diffusion [4, 6], Fleming-Viot Process [8].

Setup

Consider that the particles live in $\mathbb{R}$ and that one expects solutions to have square integrable marginals. We do not begin with FPK, but we lift to the Hilbert-space L^2 and randomise in the following way:

We fix the underlying probability space of the McKean-Vlasov distribution dependent SDE and randomise the resulting PDE. Consider the unit circle equipped with Lebesgue measure $(\Omega, \mathbb{P}) := (\mathbb{S}, \text{Leb})$,

$$dX(\omega)_t = -\partial_\mu V(X(\omega)_t, \mu_t)dt \overset{rewrite}{\to} dX(x)_t = -\partial_\mu V(\text{Leb} \circ X_t^{-1}(x), X(x)_t)dt$$

$$\mu_t := \mathbb{P} \circ X_t^{-1}$$

Randomising in L^2

Consider introducing *stochastic heat* to the dynamics:

$$dX(x)_t = -\partial_\mu V(Leb \circ X_t^{-1}, X(x)_t)dt + \Delta_x X(x)_t dt + dW(x)_t$$

with periodic boundary conditions. W is cylindrical Brownian motion, expanded into the Fourier basis as:

$$W(t,x) := B_t^0 \cdot 1 + \sum_{m \in \mathbb{N}} \sqrt{2}\left(B_t^{m,+} cos(2\pi mx) + B_t^{m,-} sin(2\pi mx)\right) =: \sum_{m \in \mathbb{Z}} B_t^m e_m(x)$$

Expecting solutions to the above dynamics to be $L^2(\mathbb{S})$ valued; the Laplacian enables this.

Dynamics may change depending on the choice of representation of the initial distribution!

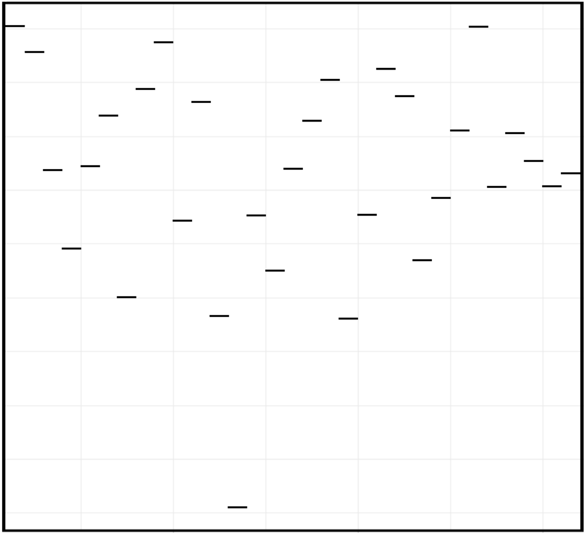
The idea is to constrain the above SPDE within a particular subset of $L^2(\mathbb{S})$ that correspond to the space of probability measures (with finite second moment).

Quantile Function Representation

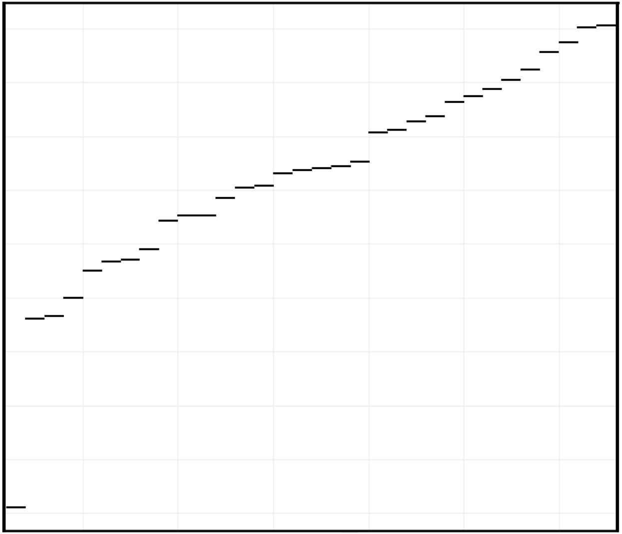
One could choose to represent a probability measure μ through its quantile, since there exists a unique non-decreasing function Q such that $\mu = \mathrm{Leb}_{\mathbb{S}} \circ Q^{-1}$.

$$Q(u) := \inf\left\{x \in \mathbb{R} : u \leq \mu_0\big((-\infty, x]\big)\right\}$$

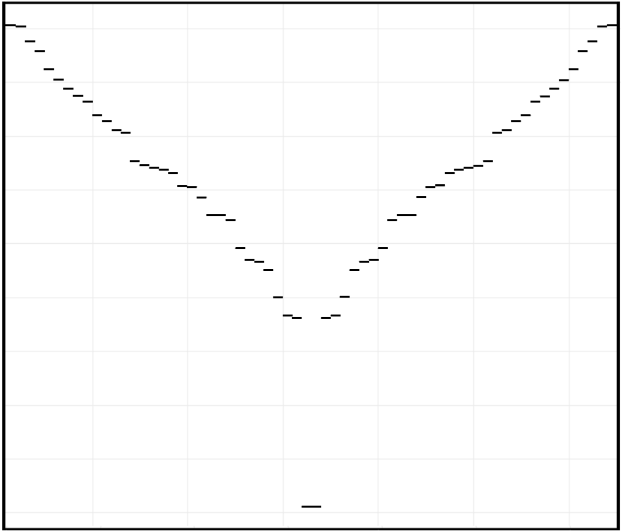
A Simple Random Variable



Its Quantile Function



Its Symmetric Non-increasing Rearrangement



Rearrangement

The rearrangement of f is denoted f^*.

Preservation of L_p norms

$$\|f^*\|_p = \|f\|_p$$

Non Expansive Property

$$\|f^* - g^*\|_p \leq \|f - g\|_p$$

NB: Non-expansion also holds for projections onto convex sets!

We will work with the subset of $L^2 = L^2(\mathbb{S})$ comprised of symmetric non-increasing functions, denoted $U^2 = U^2(\mathbb{S})$. It is a closed convex cone.

Also, we consider subsets of function spaces that are comprised of elements symmetric about zero.

For example $U^2 \subset L^2_{sym} \subset L^2$, and $H^2_{sym} \subset H^2$.

The Scheme

Consider the following scheme to construct a noise on U^2.

1. Initialise with the symmetric decreasing representative of the initial distribution:
$$X_0 = X_0^* \sim \mu_0.$$
2. Solve on some small time interval the stochastic heat equation:
$$\hat{X}_{\delta t} = SHE(X_0, \delta t) \in L^2(\mathbb{S}).$$
3. Rearrange the terminal state:
$$X_{\delta t} = \hat{X}_{\delta t}^* \in U^2(\mathbb{S}).$$
4. Iterate the above procedure.
5. Finally, we interpolate linearly between iterates.

The Scheme

For intervals of length $\delta t = h$, the scheme defined a family of continuous processes $X^h : \{X_t^h\}_{t\in[0,T]}$

We consider refining the time mesh, by decreasing $\delta t = h$.

However, we can only establish tightness if we penalise the amount of noise we add into the system.

As an informal intuition, the rearrangement operation preserves the L^2 norm, yet moves weight towards lower Fourier modes. Here the Laplacian has less of an effect and hence the need to penalise.

Coloured Noise

Instead of the cyclindrical Brownian motion,

$$W(t,x) = \sum_{m \in \mathbb{Z}} B_t^m e_m(x),$$

we redefine W as a coloured noise (Q Brownian motion):

$$W := B^0 e_0 + \sum_{m \in \mathbb{N},\, n \neq 0} m^{-\lambda} B^m e_m \equiv \sum_{m \in \mathbb{N}_0} \lambda_m B^m e_m.$$

It is assumed that λ_m defines a square-summable sequence, i.e. $\lambda > \frac{1}{2}$.
Whilst the noise is now in L^2, it is useful to retain the Laplacian for its smoothing effect.

Limiting Dynamics

We expect that the limit process X should satisfy a reflected equation of the type considered in Röckner, Zhu and Zhu [7],

$$dX_t = -\partial_\mu V(\mathrm{Leb} \circ X_t^{-1}, X_t)dt + \Delta X_t dt + dW_t + d\eta_t.$$

Such an equation was studied for stochastic heat constrained to live above some fixed/static boundary in a series of papers between Donati-Martin, Nualart, Pardoux and Zambotti, [3, 5, 9].

In analogy to previous studied models with reflection, one imagines that η may be decomposed into a directed normal force n and a local time L:

$$\eta = n \cdot L.$$

So our setting, we expect the stochastic heat equation to be altered with a forcing term that reflects the process X into U^2. However, in the absence of a corresponding integration by parts formula, the dynamics will be studied via a smaller class of test functions, sufficient to demonstrate the well-posedness and regularising effect of the limiting dynamics and ergodicity for a class of drifts.

The (driftless) rearranged stochastic heat equation

Definition (of a solution)

On a given probability space $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$ (filtration $\mathbb{F}$ satisfying the usual conditions) equipped with a Q-Brownian motion $(W_t)_{t \in I}$ (with respect to $\mathbb{F}$) and with an $\mathcal{F}_0$-measurable initial condition X_0, valued in $U^2(\mathbb{S})$ with finite moments of any order (allowing for random initial distribution), say that a pair of processes $(X_t, \eta_t)_{t \in I}$ solves the drifted rearranged stochastic heat equation driven by $(W_t)_{t \in I}$ and started from X_0 up to time T if

1. $(X_t)_{t \in [0,T]}$ is a continuous $\mathbb{F}$-adapted process with values in $U^2(\mathbb{S})$;
2. $\mathbb{P}$-a.s., for any $u \in H^2(\mathbb{S}) \subset L^2$, $\forall t \in [0, T]$,

$$\langle X_t, u \rangle = \int_0^t \langle X_r, \Delta u \rangle dr + \langle W_t, u \rangle + \langle \eta_t, u \rangle.$$

3. $(\eta_t)_{t \in [0,T]}$ is a continuous $\mathbb{F}$-adapted process with values in $H^{-2}_{\mathrm{sym}}(\mathbb{S})$, started from 0, such that, $\mathbb{P}$-almost surely, for any $u \in H^2_{\mathrm{sym}}(\mathbb{S})$ that is non-increasing, the path $(\langle \eta_t, u \rangle)_{t \in [0,T]}$ is non-decreasing;
4. for any $t \in [0, T]$,

$$\lim_{\varepsilon \searrow 0} \mathbb{E}\left[\int_0^t e^{\varepsilon \Delta} X_r \cdot d\eta_r \right] = 0.$$

Well-posedness

For the dynamics to be well posed (in the strong probabilistic sense with 1-Lipschitz dependence on initial condition), we have the following assumption on the initial condition:

• The initial distribution μ_0 on $U^2(\mathbb{S})$, is such that there exists a constant $c_0 > 0$ such that

$$\int_{U^2(\mathbb{S})} \exp\{c_0 \| \cdot \|_2^2\} d\mu_0 < \infty. \tag{A1}$$

Regularisation Effect

As a first regularisation effect, we studied the smoothing effect of the semigroup $\{P_t\}_{t\in I}$ defined by $P_t f(x) := \mathbb{E}[f(X_t^x)]$ for $f \in B_b(U^2)$

Considering the difference, for $u, v \in U^2$,

$$P_T f(u+\delta v) - P_T f(u) = \mathbb{E}[f(X_T^{u+\delta v})] - \mathbb{E}[f(X_T^u)].$$

We reproduce the law of the process started from $u+\delta v$ by the law of the *shifted* process $X_t^{u^M+\delta\frac{T-t}{T}v^M}$ under a change of measure. We do not have a well defined change of measure without the Laplacian!

Ultimately, we are able to show that the semigroup maps bounded functions f into Lipschitz functions with constant

$$c_{\|f\|_\infty} \cdot T^{-(\frac{1}{2}+\frac{\lambda}{2})}$$

Notably for $\lambda < 1$ this is integrable.

The results up to this point are contained within an arXiv preprint, [Delarue, H.] [2].

The drifted rearranged stochastic heat equation

Definition (of a solution)

On a given probability space $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$ (filtration $\mathbb{F}$ satisfying the usual conditions) equipped with a Q-Brownian motion $(W_t)_{t\in I}$ (with respect to $\mathbb{F}$) and with an $\mathcal{F}_0$-measurable initial condition X_0, valued in $U^2(\mathbb{S})$ with finite moments of any order (allowing for random initial distribution), say that a pair of processes $(X_t, \eta_t)_{t\in I}$ solves the drifted rearranged stochastic heat equation driven by $(W_t)_{t\in I}$ and started from X_0 up to time T if

1. $(X_t)_{t\in[0,T]}$ is a continuous $\mathbb{F}$-adapted process with values in $U^2(\mathbb{S})$;
2. $\mathbb{P}$-a.s., for any $u \in H^2(\mathbb{S}) \subset L^2$, $\forall t \in [0, T]$,

$$\langle X_t, u\rangle = \int_0^t \langle F(X_r), u\rangle + \langle X_r, \Delta u\rangle dr + \langle W_t, u\rangle + \langle \eta_t, u\rangle.$$

3. $(\eta_t)_{t\in[0,T]}$ is a continuous $\mathbb{F}$-adapted process with values in $H^{-2}_{\mathrm{sym}}(\mathbb{S})$, started from 0, such that, $\mathbb{P}$-almost surely, for any $u \in H^2_{\mathrm{sym}}(\mathbb{S})$ that is non-increasing, the path $(\langle \eta_t, u\rangle)_{t\in[0,T]}$ is non-decreasing;
4. for any $t \in [0, T]$,

$$\lim_{\varepsilon \searrow 0} \mathbb{E}\left[\int_0^t e^{\varepsilon\Delta} X_r \cdot d\eta_r\right] = 0.$$

Assumptions

For weak existence (via Girsanov Transformation):

- Let the drift F be such that $L^2_{sym} \ni u \mapsto F(u) \in L^2_{sym}$, and
- The drift, F, satisfies the following growth assumptions:

$$\|F(u)\|_2^2 \leq c_F(1 + \|u\|_2^2) \tag{G1}$$

$$\sum_m m^2 \langle F(u), e_m \rangle^2 \leq C_F \left(1 + \sum_m m^2 \langle u, e_m \rangle^2 \right) \tag{G2}$$

Note that the condition (G2) holds true if F is derived from a mean field potential with bounded derivatives, i.e. $F(u)(x) = -\partial_\mu V(u \circ \mathrm{Leb}_{\mathbb{S}}^{-1}, u(x))$ and $\partial_y \partial_\mu V(u \circ \mathrm{Leb}_{\mathbb{S}}^{-1}, y)$ is bounded in $y \in \mathbb{R}$, uniformly in u.

Assumptions

For long-time weak existence-uniqueness (also via Girsanov Transformation):

Assumption (L^2 norm square is Lyapunov with negative exponent)

There exist non-negative constants C_L,c_1 and c_2 such that, for $u \in U^2$

$$\langle F(u), u \rangle \leq C_L + c_1 \sum_{m \neq 0} \hat{u}_m^2 - c_2 \hat{u}_0^2 \equiv C_L + c_1 \| u - \bar{u} \|_2^2 - c_2 \bar{u}^2. \tag{1}$$

Where $c_1 < \frac{2}{c_P}$, c_P the Poincaré constant for $\mathbb{S}$.

Under the above assumptions, there is exponential convergence to equilibrium for the regularised descent. This follows from an adaptation of arguments of Debussche, Hu and Tessitore [1].

This assumption is fairly strong and limits the class of potentials one can currently consider, however it should be possible to extend with use of the following Itô formula

Itô Formula

Proposition

Under the assumptions of the previous slide, let $(X_t)_{t\geq 0}$ *solve the drifted rearranged SHE. Then, for any smooth mean-field function* $\varphi : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ *with bounded and jointly continuous derivatives* $\partial_\mu \varphi : \mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$, $\partial_x \partial_\mu \varphi : \mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ *and* $\partial_\mu^2 \varphi : \mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, *with* $\mathcal{P}_2(\mathbb{S})$ *being equipped with the 2-Wasserstein distance, the following formula holds, with probability 1, for any* $t \geq 0$, *with* $\mathcal{L}(X_t) := Leb \circ X_t^{-1}$,

$$
\begin{aligned}
\varphi\big(\mathcal{L}(X_t)\big) = \varphi\big(\mathcal{L}(X_0)\big) &- \int_0^t \int_{\mathbb{S}} \partial_x \partial_\mu \varphi\big(\mathcal{L}(X_s)\big)\big(X_s(x)\big)\big[DX_s(x)\big]^2 dx\, ds \\
&- \int_0^t \int_{\mathbb{S}} \partial_\mu \varphi\big(\mathcal{L}(X_s)\big)\big(X_s(x)\big) \partial_\mu V\big(\mathcal{L}(X_s)\big)\big(X_s(x)\big) dx\, ds \\
&+ \int_0^t \int_{\mathbb{S}} \partial_\mu \varphi\big(\mathcal{L}(X_s)\big)\big(X_s(x)\big) dW_s(x) \\
&+ \frac{1}{2} \int_0^t \int_{\mathbb{S}} \partial_x \partial_\mu \varphi\big(\mathcal{L}(X_s)\big)\big(X_s(x)\big) \sum_{k \in \mathbb{Z}} \lambda_k^2 e_k(x) dx\, ds \\
&+ \frac{1}{2} \int_0^t \int_{\mathbb{S}} \int_{\mathbb{S}} \partial_\mu^2 \varphi\big(\mathcal{L}(X_s)\big)\big(X_s(x), X_s(y)\big) \sum_{k \in \mathbb{Z}} \lambda_k^2 e_k(x) e_k(y) dx\, dy\, ds.
\end{aligned}
$$

Returning to the minimisation problem

Convergence to equilibrium is all well and good, but it doesn't tell us about the form of the invariant measure, and hence about the minimisation problem.

Recalling that the drifted rearranged stochastic heat equation is envisioned as a *regularised* gradient descent for the minimisation problem, $\min_{\mu\in\mathcal{P}_2(\mathbb{R})}\{V(\mu)\}$, we consider the small viscosity version of the equation:

$$dX^\varepsilon_t(x) = -\partial_\mu V\big(\mathcal{L}(X^\varepsilon_t)\big)\big(X^\varepsilon_t(x)\big)dt + \varepsilon^{2\alpha}\Delta X^\varepsilon_t(x)dt + \varepsilon dW^\varepsilon_t(x) + d\eta^\varepsilon_t(x),$$

where,

$$W^\varepsilon_t(\cdot) = \sum_{k\in\mathbb{N}} \lambda^\varepsilon_k B^k_t, \quad \textit{with} \quad \lambda^\varepsilon_k := \begin{cases} 1 & \text{if} \quad k \leq \varepsilon^{-\alpha} \\ k^{-\lambda} & \text{if} \quad k > \varepsilon^{-\alpha} \end{cases}.$$

This may be viewed as a stochastic gradient descent for the modified potential:

$$V\big(\mathcal{L}(X)\big) + \frac{\varepsilon^{2\alpha}}{2}\|\nabla X\|_2^2, \quad X \in L^2(\mathbb{S}),$$

Unfortunately, we are unable to establish a Gibbs type representation for the invariant measures of these dynamics from which to draw conclusions about the minimisation problem.

Theorem

Assume that for some $\kappa > 0$ *and some* $X_0 \in U^2(\mathbb{S}) \cap H^1_{\mathrm{sym}}(\mathbb{S})$*, there exists a convex neighbourhood* $\mathcal{O} \subset U^2(\mathbb{S})$ *of* X_0 *such that (the lift of)* V *is* κ*-convex on* $\mathcal{O}$ *with* X_0 *as the minimizer. By* κ*-convex, that is to say that the function*

$$U^2(\mathbb{S}) \ni X \mapsto V\big(\mathcal{L}(X)\big) - \frac{\kappa}{2}\|X\|_2^2 = V\big(\mathcal{L}(X)\big) - \frac{\kappa}{2}\int_{\mathbb{S}} X^2(x)\mathrm{d}x$$

is convex. Then, there exists a constant C *such that, for any* $a > 0$ *satisfying* $\inf_{X \in \partial O} V(X) - V(X_0) > a$*,*

$$\exp\Big(\frac{a}{C\varepsilon^2}\Big) \le \mathbb{E}\big(\tau_\varepsilon\big) \le \exp\Big(\frac{Ca}{\varepsilon^2}\Big),$$

where

$$\tau_\varepsilon := \inf\Big\{t \ge 0 : V\big(\mathcal{L}(X_t^\varepsilon)\big) \ge V\big(\mathcal{L}(X_0)\big) + a\Big\}.$$

Note that κ convexity on U^2 is weaker than κ-convexity on L^2 which helps to justify the model over randomising in L^2.

Comments

There are evidently many questions that remain for this model, to name a few:

- Propagation of chaos for corresponding particle systems.
- Large Deviations.
- Regularisation of stochastic McKean-Vlasov SDE. I.e. private noise model.
- Higher dimension.
- Application to mean field control.

Thank you for listening!

Selected References I

[1] A. Debussche, Y. Hu, and G. Tessitore. "Ergodic BSDEs under weak dissipative assumptions". In: *Stochastic Processes and their Applications* 121.3 (2011), pp. 407–426.

[2] F. Delarue and W. R. P. Hammersley. "Rearranged Stochastic Heat Equation". In: *arXiv:2210.01239* (2022).

[3] C. Donati-Martin and E. Pardoux. "White noise driven SPDEs with reflection". In: *Probability Theory and Related Fields*. 95, 1-24 (1993).

[4] V. Marx. "A Bismut-Elworthy inequality for a Wasserstein diffusion on the circle". In: *arXiv* 2005.04972 (2020).

[5] D. Nualart and E. Pardoux. "White noise driven quasilinear SPDEs with reflection.". In: *Probab. Th. Rel. Fields*. 93, 77–89 (1992).

[6] M.-K. von Renesse and K.-T. Sturm. "Entropic measure and Wasserstein diffusion". In: *Ann. Probab.* (2009).

Selected References II

[7] M. Röckner, R.-C. Zhu, and X.-C. Zhu. "The stochastic reflection problem on an infinite dimensional convex set and BV functions in a Gelfand triple". In: *The Annals of Probability* 40.4 (2012), pp. 1759 –1794.

[8] W. Stannat. "Long-time behaviour and regularity properties of transition semigroups of Fleming-Viot processes". In: *Probab. Theory Related Fields* 122.3 (2002), pp. 431–469.

[9] L. Zambotti. "A Reflected Stochastic Heat Equation as Symmetric Dynamics with Respect to the 3-d Bessel Bridge". In: *Journal of Functional Analysis* 180 (2001), pp. 195–209.