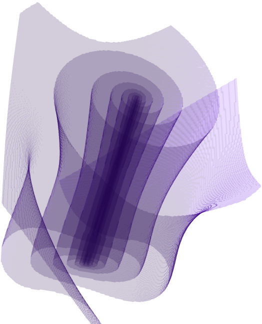
A deep-learning approach for nonequilibrium trajectories

or Geometrical Minimum Action Method revisited

Eric Simonnet

CNRS, INPHYNI, UMR 7010



Outline

- **Motivation**
- **Basics: instantons and the geometrical action**
- **Neural network formulation**
- **SDEs**
- **SPDEs**
- **Beyond the Freidlin-Wentzell regime: crisis, intermittency.**
- **Perspective**

Rare events: physical space versus phase space $\partial_t u = \mathcal{F}(u, \nabla_{\mathbf{x}} u, \Delta_{\mathbf{x}} u, \cdots) + \text{noise}, \; \mathbf{x} \in \mathbb{R}^d$.

Phase space, Galerking truncation: $\#$d.o.fs = $O(\text{precision}^{-d})$.

$$u(\mathbf{x}) = \sum_k u_k \mathbf{e}_k(\mathbf{x}),$$

- One is looking for (pointwise) estimates of the **Committor** = "Probability to first hit set *a* before another set *b*, given a present state". Solution of backward Kolmogorov operator in phase space: $\varepsilon \Delta_{u_1,\cdots,u_n} C + \mathbf{F} \cdot \nabla_{u_1,\cdots,u_n} C = 0, \; C(a) = 0, C(b) = 1, \; a \cap b = \emptyset$.

Strategies

- **AMS**. Selection-mutation on trajectories using many $\sim$ short-time integrations of the SDE system $du_k = \mathcal{F}_k(u_1, \cdots, u_n)\, dt + \sqrt{\varepsilon} dW_k$.
- **Dimension reduction/coarse-graining**, e.g. collective variables $O(100^d) \to \; O(10\text{-}100)$
  - Neural Network parametrisation *(Khoo et al. 2019; Q.Li et al. 2019; H.Li et al. 2021)*.
  - Simulation/Data-based approaches (analogues) $\to$ loop AMS *(D.Lucente et al. 2019, 2021)*.
- **Issues** Rare events: noise amplitude $\varepsilon \ll 1 \to$ AMS strong "apparent bias" due to nonoptimal reaction coordinate. Committor estimate very challenging (large deviations). Not enough data.

Motivation

Rare events: physical space versus phase space $\partial_t u = \mathcal{F}(u, \nabla_{\mathbf{x}} u, \Delta_{\mathbf{x}} u, \cdots) + \text{noise}, \ \mathbf{x} \in \mathbb{R}^d$.

Physical space, NN parametrisation: $\#$**d.o.fs =** $\text{Polynomial}(d, \text{precision}^{-1})$ **(A.Jentzen'theorems) .**

$$u(t,\mathbf{x};\theta) = \mathcal{N}_{\text{out}} \circ \mathcal{N}_L \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{N}_{\text{in}}((t,\mathbf{x})), \ \mathcal{N}_k(\mathbf{y}) = \sigma_k(W_k \mathbf{y} + \mathbf{b}_k), \theta = (W_k, \mathbf{b}_k).$$

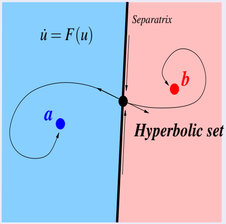
- **~~AMS~~**. Selection-mutation on trajectories in neural network space θ. NO PDE-integrators ! Physics-informed, Deep Galerking not good enough. Maybe with FNO/DeepONet approaches (AnandKumar, Karnadiakis) ? $\cdots$and there are still issues with the reaction coordinate.
- **Instanton**. Minimizing the Freidlin-Wentzell action $\mathcal{A}$ (additive noise) between two attractive sets a, b:

$$\lim_{T\to\infty} \min_{u\in V} \int_0^T ||\partial_t u - \mathcal{F}(u)||^2 \, dt, \ V = \{u(0) = a, u(T) = b\}.$$

- **Advantages:** rare events regime $\varepsilon \to 0$. Small dimension.
- **Issues:** one needs a relevant action (Freidlin-Wentzell regime). Cost functional might be difficult to minimize.

Shortcut: Freidlin-Wentzell is for noise-induced transitions.

- Exit time from a domain. Very few hypothesis to be valid: **a** must be isolated.



$$\begin{cases} d\mathbf{u}_t = F(\mathbf{u}_t)dt + \sqrt{\varepsilon}dW_t, \\ \Pr([\mathbf{u}_t];\mathbf{u}(0)=\mathbf{a},\mathbf{u}(T)=\mathbf{b}) \asymp \exp\left(-\frac{\mathcal{A}_T[\mathbf{u}]}{\varepsilon}\right) \\ \mathcal{A}_T[\mathbf{u}] = \frac{1}{2}\int_0^T ||\dot{\mathbf{u}} - F(\mathbf{u})||^2\, dt \end{cases}$$

- behavior for **large time intervals**: *most probable trajectory* or *instanton path* or *nonequilibrium transition* or *maximum likelihood path*: argmin of

$$\inf_T \inf_{\mathbf{u}\in V} \mathcal{A}_T[\mathbf{u}],\ V = \{\mathbf{u}(0)=\mathbf{a},\mathbf{u}(T)=\mathbf{b}\}.$$

- Pathological case: non-isolated attractors, e.g. Euler 2-D, AB model: F.Bouchet H.Touchette, J.Stat.Mech 2012.

M.Freidlin, A.Wentzell, *Random perturbations of Dynamical Systems*, Springer, 1984, 1998, 2012.

The geometrical action

- Relevant paths are for $T \to \infty$, one wishes to simplify the optimisation pb $\inf_T \inf_{\mathbf{u}\in V} \mathcal{A}_T[\mathbf{u}]$. The remarkable idea is to look at instanton arclength parametrisations.
- M.Heymann, E.Vanden-Eijnden, The geometric minimum action method: A least action principle on the space of curves, Comm.Pure Appl. Math, 61, 1052–1117 (2008); see also Olender, Elber, 1997 and FW, 1998.

$$\inf_T \inf_{\mathbf{u}\in V} \mathcal{A}_T[\mathbf{u}] = \inf_{\mathbf{u}\in W} \mathcal{A}_g[\mathbf{u}]$$

$$\mathcal{A}_g[\mathbf{u}] \equiv \int_0^1 \left(||\dot{\mathbf{u}}||\ ||F(\mathbf{u})|| - \langle \dot{\mathbf{u}}, F(\mathbf{u})\rangle\right) ds,\ W = \{\mathbf{u}(0) = \mathbf{a},\ \mathbf{u}(1) = \mathbf{b}\}.$$

- short proof: $|a|^2 + |b|^2 \geq 2|a|\ |b|$, gives $\frac{1}{2}\int_0^T ||\dot{\mathbf{u}} - F(\mathbf{u})||^2\, dt \geq \mathcal{A}_g$ and is equal for $\mathbf{u}$ such that $||\dot{\mathbf{u}}|| = ||F(\mathbf{u})||$. In fact all curves can be reparametrized such that $||\dot{\mathbf{u}}|| = ||F(\mathbf{u})||$ and due to the homogeneity of $\mathcal{A}_g$ in time, one can conclude.

Cost functional

- We replace u by some (deep) NN parametrisation: $\mathcal{N} : (s, \mathbf{x}) \in [0,1] \times \mathbb{R}^d \to \mathbb{R}^k,\ d \geq 0, k \geq 1$. Of course $\mathcal{N}$ depends also on parameters θ (weight/bias). We do not write them explicity.
- There is a continuum of possible parametrisations for $\mathcal{A}_g$, we fix one $||\dot{\mathcal{N}}|| \approx \mathrm{cst}$.

$$\mathcal{C}_{\mathrm{total}} = \gamma_g \mathcal{A}_g + \gamma_{\mathrm{bv}} \mathcal{C}_{\mathrm{BV}} + \gamma_{\mathrm{bc}} \mathcal{C}_{\mathrm{BC}} + \gamma_a \mathcal{C}_{\mathrm{arc}},\ \gamma_a \ll 1$$

$$\mathcal{C}_{\mathrm{arc}} = \int_0^1 ||\dot{\mathcal{N}}||^2\, ds - \left(\int_0^1 ||\dot{\mathcal{N}}||\, ds \right)^2$$

- **Lift or not Lift ?**. BV is for boundary values, we need $\mathcal{N}(0; \mathbf{x}) = \mathbf{a}(\mathbf{x})$ and $\mathcal{N}(1; \mathbf{x}) = \mathbf{b}(\mathbf{x})$. We choose the "naive" lift:

$$\mathbf{u}(s, \mathbf{x}) = (1-s)\mathbf{a}(\mathbf{x}) + s\mathbf{b}(\mathbf{x}) + s(1-s)\mathcal{N}(s, \mathbf{x}),\ s \in [0,1].$$

  Therefore $\gamma_{\mathrm{bv}} = 0$. *Do not lift if you are looking for gold.*
- $\mathcal{C}_{\mathrm{bc}}$ is for boundary conditions in the case of PDEs ($d \geq 1$) on $[0,1] \times \partial\mathcal{D}$. Note that in the case of periodic domains, periodicity of first-order derivatives is crucial.

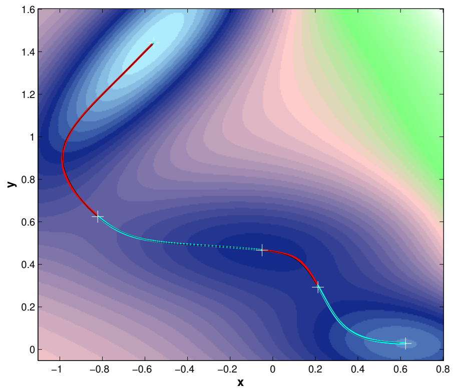
Comparison with the geometrical minimum action method

	gMAM	Deep NN
Discretisation	finite-diff, pseudo-spectral,...	NN parameter θ (exact)
Descent vector	$-\frac{\|\|F(\mathbf{u})\|\|}{\|\|\dot{\mathbf{u}}\|\|} \times$ EL, Euler-Lagrange $\equiv \frac{\delta \mathcal{A}_g}{\delta u}$	$-\nabla_\theta(\mathcal{A}_g + \mathcal{C}_{\mathrm{arc}})$
Descent algor.	**deterministic**: relaxation, CG, L-BFGS	**stochastic**: ADAM and variants
Renormalisation step	hard constraint $\|\|\dot{\mathbf{u}}\|\| = \mathrm{cst}$ (interpolation)	none (soft constraint)
Dim curse ($d \geq 3$)	Yes	No
Non-convexity	Moderate ($\sim$ continuous action)	Wild (NN par.)

gMAM: M.Heymann, E.VdE (Comm. Math., PRL, 2008), G.Poppe,T.Schäfer (J.Physics A, 2018), M.Tao (Physica D 2018). See also MAM, aMAM algorithms, e.g. X.Zhou, W.Ren, W.E (2008).

2-D Müller potential

$\mathcal{A}_g = 229.1$ (Arrhenius), $\langle \mathcal{A}_g \rangle = 229.3$

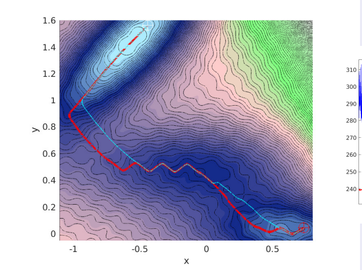


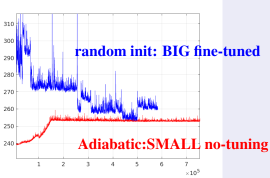
ANY: lifted $\mathcal{N} : [0,1] \to \mathbb{R}^2$, 4 hidden layers, 10 neurons/layer, swish, ADAM 10^{-3}, $N = 2048$, random init, $\gamma_{\rm arc} = 0.01$, $\gamma_{\rm geom} = 1$, $\#$ params $= 482$.

Rugged 2-D Müller potential $+\gamma \sin(10\pi x)\sin(10\pi y)$, $\gamma = 3$

How to bring NNs to relevant states ?

- **Brut force**: fine-tuning with deep NNs (penalty, capacity, depth, training rate, activation function, multi-scaling, etc...) NN converges to solution even with random initialisation (Xavier).
- **Adiabatic descent**: slow change of ad-hoc parameter during stochatic descent $\rightarrow$ **ANY**: cheap, light tuning, very efficient !



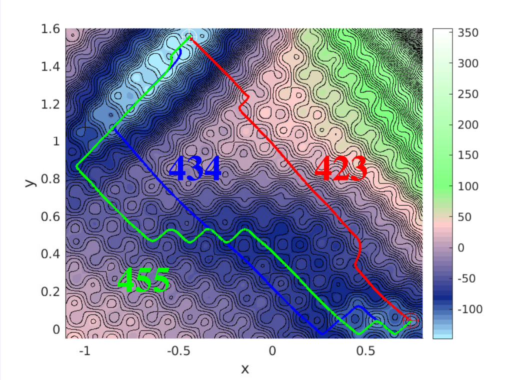


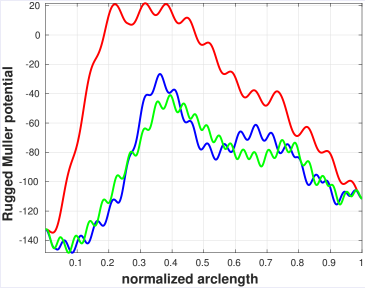
SMALL $=$ e.g. basic feedforward 8 hidd $\times$10 neur/l, $\#$ params = 922.

BIG $=$ Multiscaled 5 $\times$ 20 hidd $\times$10 neur/l. Skip connections, $N = 128$, ADAM 10^{-3}, $\#$ params = 11210

A challenging benchmark: many saddles

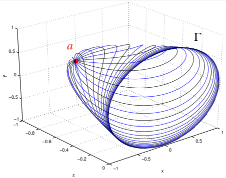
Rugged 2-D Müller potential $+\gamma \sin(10\pi x)\sin(10\pi y)$, $\gamma = 9$





A continuum set of instantons

- Separatrix can now contain more complicated *hyperbolic* sets Γ: limit cycles or attractors (e.g. Melancholia states, V.Lucarini).
- Consequence: continuum set of instantons which differ by their length [Tao, 2018]:



- To handle these situations, one must contraint the NN to look for instantons having a prescribed (arc)length L. New term in the cost functional:

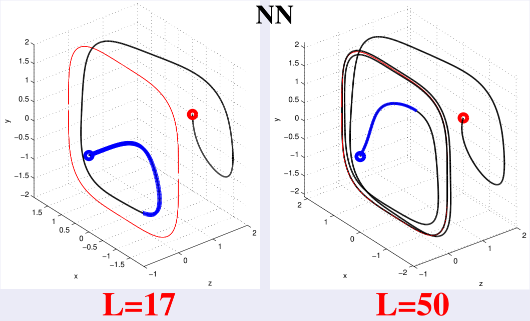
$$\mathcal{C}_L = \left(\int_0^1 ||\dot{\mathcal{N}}||\, ds - L \right)^2 .$$

M.Tao, Hyperbolic periodic orbits in nongradient systems and small-noise-induced metastable transitions, Physica D. 363, 2018.

A continuum set of instantons: example [Tao]

Two stable points $\mathbf{a} = (0,0,-1)$, $\mathbf{b} = (0,0,+2)$ and an hyperbolic limit cycle Γ on the separatrix $z = 0$

$$\left\{\begin{array}{lcl} \dot{x} & = & -(z+1)(z-2)\frac{x}{(x^4+y^4)^{1/4}} - x - y^3 \\ \dot{y} & = & -(z+1)(z-2)\frac{y}{(x^4+y^4)^{1/4}} + x^3 - y \\ \dot{z} & = & -(z+1)(z-2)z \end{array}\right. .$$



Ginzburg-Landau, Allen-Cahn PDE with shear

$$\partial_t u = \nu \Delta u + u - u^3 + \mathbf{c} \cdot \nabla u, \; \mathbf{x} \in \mathcal{D} \equiv [0,1]^d$$

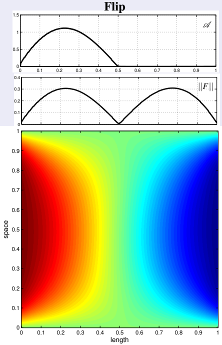
with boundary conditions, either Dirichlet or periodic. It contains a very large set of situations:

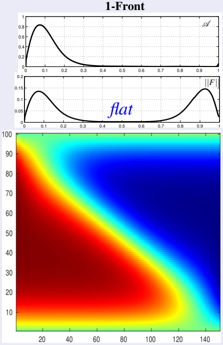
- Gradient PDE $\mathbf{c} = 0$: W.G.Faris, G.Jona-Lasinio, 1982 (LDP: $d = 1$).
- Nongradient with no transverse decomposition: $||\nabla_{\mathbf{x}} c|| \neq 0, d \geq 2$ [Tao]
- Separatrix contains many saddle fixed points and hyperbolic limit cycles.
- Many group symmetries and symmetry-breaking scenarii.
- $\sim$ exact estimates of minimum action values in the transverse case are available: Arrhenius law.
- $\nu \to 0$ singular limit
- Many exploratory studies, proof of concepts:
  W.E, W.Ren, E.VdE, 2004 (MAM), M.Heymann, E.VdE PRL2008 (gMAM), M.Tao 2018 (gMAM), J.Rolland et al. 2016 (AMS), H.Li et al., 2021 (NN, committor).

$d = 1$, Dirichlet boundary conditions, $c = 0$, $\nu = 5 \cdot 10^{-3}$.

Two stable points $\mathbf{a}(x) = +1$,$\mathbf{b}(x) = -1$ + *boundary-layer corrections* , many possible saddles. Periodic BCs: discrete set of orbits for the Hamiltonian $H(x,y) = y^2/2 - \frac{1}{4\nu}(x^4 - 2x^2)$.

$$\textbf{Arrhenius law}: \ \mathcal{A}^{(0)} \sim \frac{1}{2} - \frac{4}{3}\sqrt{2\nu},\ \mathcal{A}^{(n)} \sim \frac{4}{3} n\sqrt{2\nu},\ \nu \leq \frac{1}{(n+1)^2\pi^2}, n \geq 1.$$



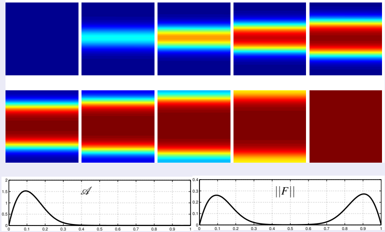


$\mathbf{L = 1.8}, \mathscr{A}_\mathbf{g} = \mathbf{0.36}$ $\mathbf{L = 3.1}, \mathscr{A}_\mathbf{g} = \mathbf{0.13}$

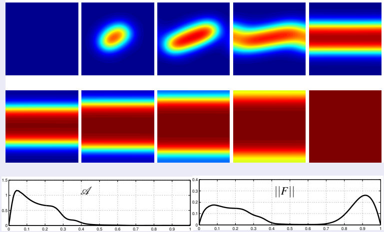
Nongradient 2-d PDE

$d = 2$, periodic boundary conditions, $0.1\sin(2\pi y)\frac{\partial u}{\partial x}$, $\nu = 5 \cdot 10^{-3}$.

Two stable points $\mathbf{a}(x) = -1$,$\mathbf{b}(x) = +1$, same saddles than for $c = 0$, in addition: presence of hyperbolic limit cycles.



$\mathbf{L = 2.9}, \mathscr{A} = \mathbf{0.26}$

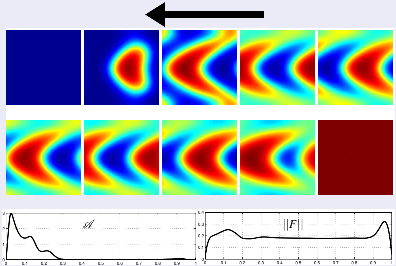


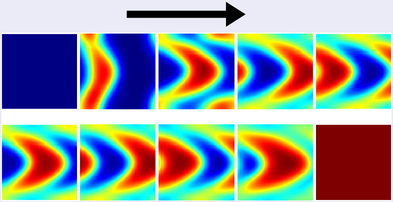
$\mathbf{L = 3.3}, \mathscr{A} = \mathbf{0.248}$

- shear-facilited nucleation: first obtained by M.Heynmann, E.VdE (PRL2008) using gMAM.
- goes through the same saddle ! transverse decomposition cannot hold [Tao, 2018].
- Here: NN initial conditions are *random* + **symmetry breaker**: $\left(\int_0^1 ||\partial_x \mathcal{N}||^2 ds - \mathrm{cst}\right)^2$.

$d = 2$, periodic boundary conditions, $0.1\sin(2\pi y)\frac{\partial u}{\partial x}$, $\nu = 5\cdot 10^{-3}$.

Arrow$^+$, Arrow$^-$ hyperbolic limit cycles



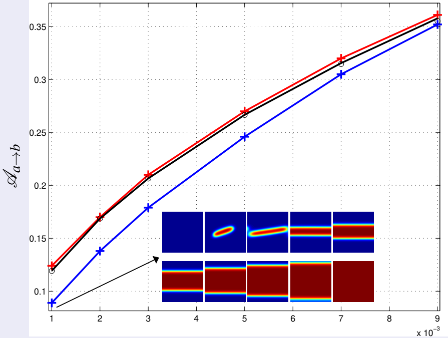


$$\mathbf{L = 10, \mathscr{A} = 0.35}$$

- undocumented: symmetry-breaking of Tao's snake?

$d = 2$, periodic boundary conditions, $0.1\sin(2\pi y)\frac{\partial u}{\partial x}$

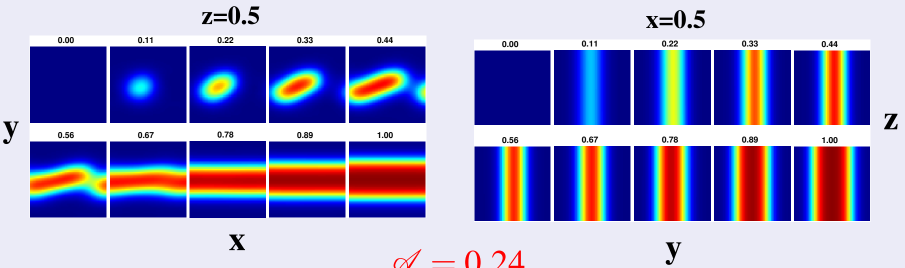
$\nu \to 0$ using adiabatic descent



ν

$d = 3$, periodic boundary conditions, $\nu = 5 \cdot 10^{-3}$.

$$\partial_t u = \nu \Delta u + u - u^3 + 0.1 \sin 2\pi y \, (\partial_x u + \partial_z u)$$



$$\mathscr{A} = 0.24$$

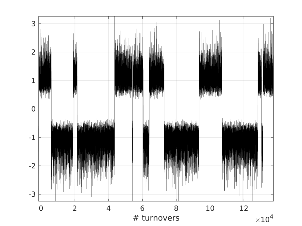
7 hidden layers, 10 neurons/layer, swish, tensorial batch:16 $\times$ 32 using two symmetry breakers in x and z.

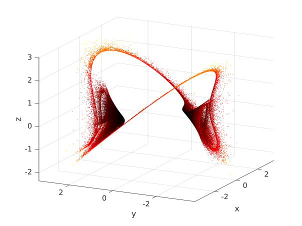
- **converge with a very small $\#$ of parameters** ! Here 1662, to be compared to O(10M) with classical methods.
- no codim-2 nucleations ?

Bistability in deterministic systems

- One does not need noise to observe bistability: **Crisis-induced intermittency**: Grebogi, Ott, York (1982). Basin of attractions can touch/overlap $\longrightarrow$ Chaos is doing the job...
- Even more generic with group symmetry: road to crisis. P.Chossat, M.Golubitsky, *Symmetry-increasing bifurcation of chaotic attractors*, Physica D, 1988 . Double-gyre circulation, Kuroshio bimodality, atmospheric jet transitions in Phillips models.
- Rare event is "intermittency with a timescale separation" !
- Example Dynamo reversal: C.Gissinger, E.Dormy, S.Fauve EPL 2010

$$\dot{x} = \mu x - yz, \dot{y} = -\nu y + xz, \; \dot{z} = -z + xy + \Gamma$$





$\mu = 0.1192805, \nu = 0.1, \Gamma = 0.9$

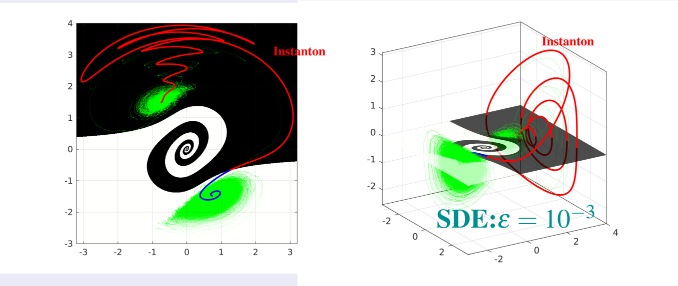
Freidlin-Wentzell action

- **No large deviations** for $\mu > \mu_c$! The *mean waiting time* follows a power law: $\tau \sim (\mu_c - \mu)^{-\gamma}$, $\gamma \sim \frac{1}{2}$ in the dynamo model. Here $\mu_c \approx 0.1192805$. The exponent depends on the type of crisis tangency.
- **FW minimizers have zero action !**

Which strategy to use if one does not know (stochastic versus deterministic) ?

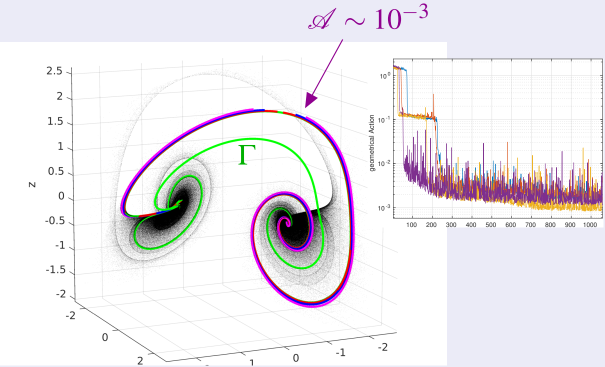
- **Direct simulation**: in the rare event regime (e.g. $|\mu - \mu_c| \ll 1$), one must integrate for a very long time some uninteresting dynamics. This is what we want to avoid.
- **AMS**: solve this problem by selections, mutations in the space of trajectories. Still the issue of reaction coordinate when ε is too small: poor estimates due to $\exists$ of large tails (bias on probability *and* ensemble).
- **Geometrical action**: it looks like direct simulation with reparametrisation of time...

FW regime



- One saddle $O = (0,0,\Gamma)$, two stable (damped oscillations) points: $(\pm\sqrt{\nu+\Gamma\sqrt{\frac{\nu}{\mu}}}, -\pm\sqrt{\mu+\Gamma\sqrt{\frac{\mu}{\nu}}}, -\sqrt{\mu\nu})$
- dim $W^u(O) = 2$ (oscillatory).
- noisy precursor of crisis ? Silchenko et al., Fluctuational transitions across different kinds of fractal basin boundaries, PRE 2005. See also Maier, Stein, Luchinsky, Kraut, Grebogi, Demaeyer...
- Very different from Lorenz63: X.Zhou, W.E, Study of noise-induced transitions in the Lorenz system using the minimum action method, Comm.Math.Sci, 2010.

A waiting time killer ?



- Bad guy: initial conditions *a* are important. Do like in AMS: DNS sampling of $\mu_{\text{eq}}^{\text{hit}}\{\partial C\}$ where C contains *a*. Note that the transition path Γ is still correct up to a scaling factor in phase space (i.e $\exists \gamma > 0, \gamma\Gamma$ close to crisis path).

CONCLUSION

Advantages

- Natural stochastic extension of classical gMAM in the context of deep learning.
- Simple to implement, possibility to add algorithmic constraints without extra work
- NNs are most often able to find ad-hoc solutions without help (stochastic gradient descent).
- Possibility to work as well in the context of deterministic chaotic systems: **"waiting time filtering"**.
- Hard to compete with in dimension $3+1$ and more (curse dimensionality).

Drawbacks

- Hyperparameters tuning: mostly penalty parameters, sometimes tricky...
- Chaotic systems (e.g. eddy-resolving models): one needs to consider more powerfull NN structures : *how far can we go ?*
- Pathological situations: barotropic QG with additive noise (Eulerian limit).
- Cannot be used for understanding finite-noise effects (see triple-well potential).

Perspectives

- intermediate-complexity deterministic models having crisis (ODEs, 1-D PDEs ?)
- challenge: baroclinic QG 2-layer model (Phillips) with atmospheric jet transitions.
- Room for algorithmic improvments: e.g. coupling with time integrators, multiscaled architectures...