Valentina Ros
PostDoc @ LPENS Paris

Youth in high dimensions *online*
ICTP- 3rd July 2020

Dynamics in high-dimensional non-convex landscapes: from slow descent to activation

VR, G. Biroli, C. Cammarota, arXiv:2006.08399
VR, J. Phys A: Math Theor 53 (2020)
VR, G.Biroli, C.Cammarota, EPL 126 (2019)

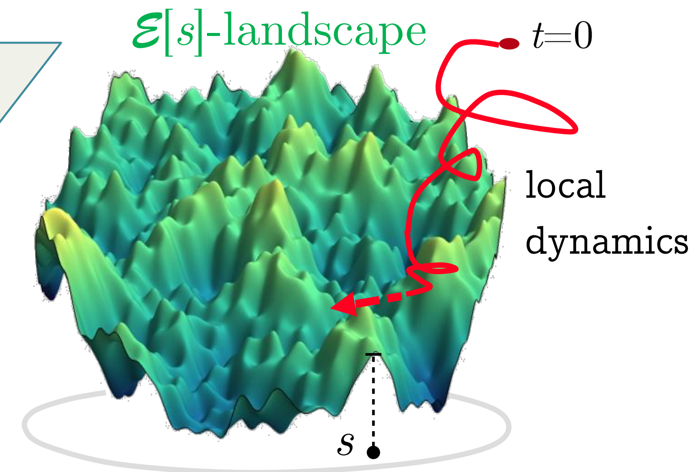
LPENS
LABORATOIRE DE PHYSIQUE DE L'ÉCOLE NORMALE SUPÉRIEURE

The problem

Stochastic dynamics in high-d landscapes

Goal: understand noisy gradient-descent dynamics in high-dimensional landscapes.

$$\frac{d\mathbf{s}_t}{dt} = -\nabla \mathcal{E}[\mathbf{s}_t] + \eta_t \qquad \begin{array}{l} \mathcal{E} = \text{landscape} \\ \eta_t = \text{weak noise} \end{array}$$



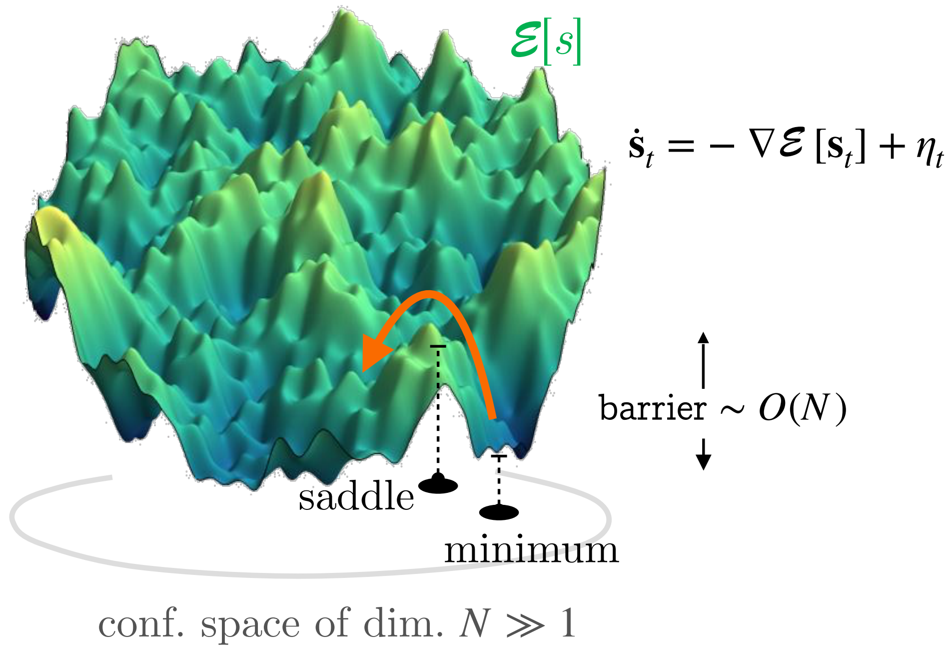
configuration space of dim. $N \gg 1$

Gain: optimization in high-d with local dynamics is ubiquitous.

- complex systems in physics —glasses, and biology —proteins, evolutionary biology
- inference and optimization —cost/loss landscapes
- quantum computing —fidelities in optimal control

Here: random high-d functionals with Gaussian statistics [→ e.g. E. Subag talk]

Challenging: why?

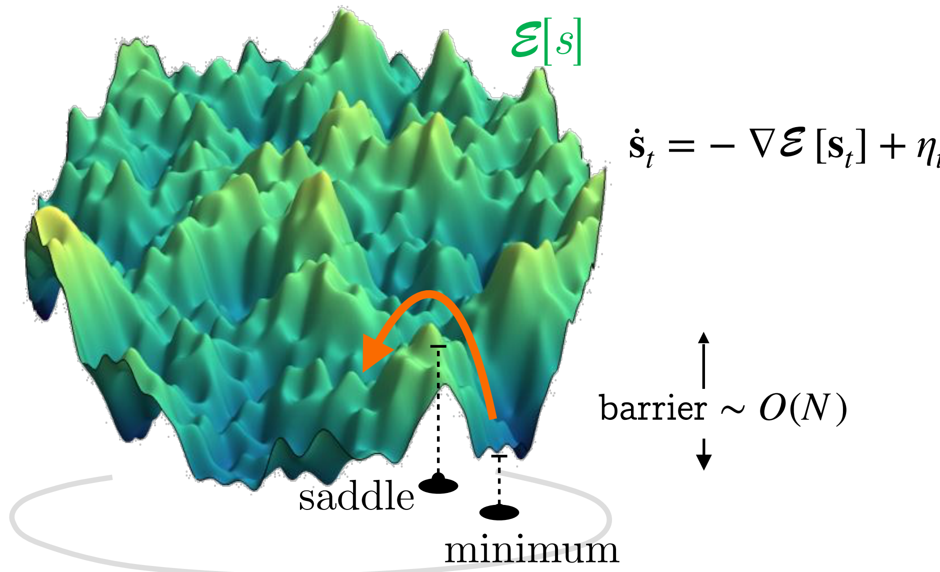


[1] Non-convexity & trapping

Proliferation of stationary points $\mathcal{N} \sim e^{N\Sigma}$ of different stability: [minima, maxima, saddles]

Need to classify exponential complexity

Challenging: why?



conf. space of dim. $N \gg 1$

[1] Non-convexity & trapping

Proliferation of stationary points $\mathcal{N} \sim e^{N\Sigma}$ of different stability: [minima, maxima, saddles]

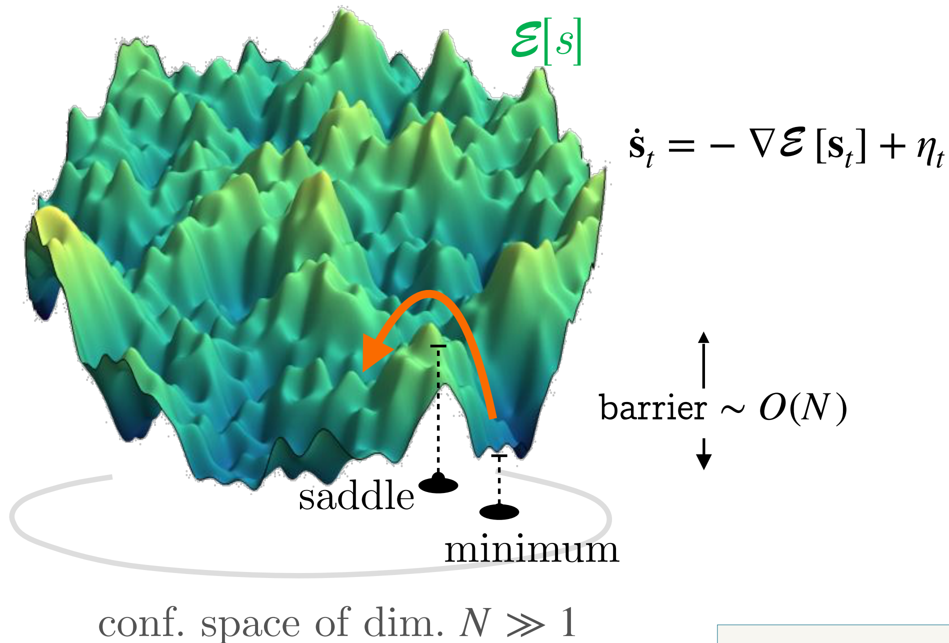
Need to classify exponential complexity

[2] Rare dynamical events

Barriers scale with N, jumps exponentially rare $\tau \sim e^{N\Delta\mathcal{E}}$ [suppressed when $N \to \infty$]

Need to go beyond standard mean field

Challenging: why?



[1] Non-convexity & trapping

Proliferation of stationary points $\mathcal{N} \sim e^{N\Sigma}$ of different stability: [minima, maxima, saddles]

Need to classify exponential complexity

[2] Rare dynamical events

Barriers scale with N, jumps exponentially rare $\tau \sim e^{N\Delta\mathcal{E}}$ [suppressed when $N \to \infty$]

Need to go beyond standard mean field

[3] Entropy of volumes

Given a reference minimum, exponential majority of other saddles/minima are far away and not locally connected

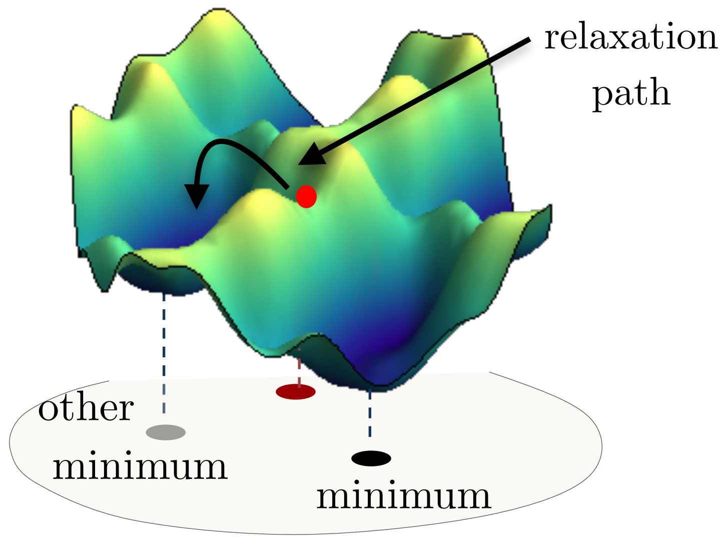
Need large deviations to look locally

Activation: the challenge, more precisely

In high N we know* how to describe "relaxation paths": DMFT [→ e.g. P. Urbani talk]

$$\mathscr{Z}_{\rm dyn} = \int \mathscr{D}Q \; \exp\left(-N\mathscr{A}[\,Q(t,t')\,]\right) \to \frac{\delta \mathscr{A}}{\delta Q} = 0$$

- Captures typical dynamics: $\mathscr{A} = 0$
- "Short" times $t \neq O(N)$



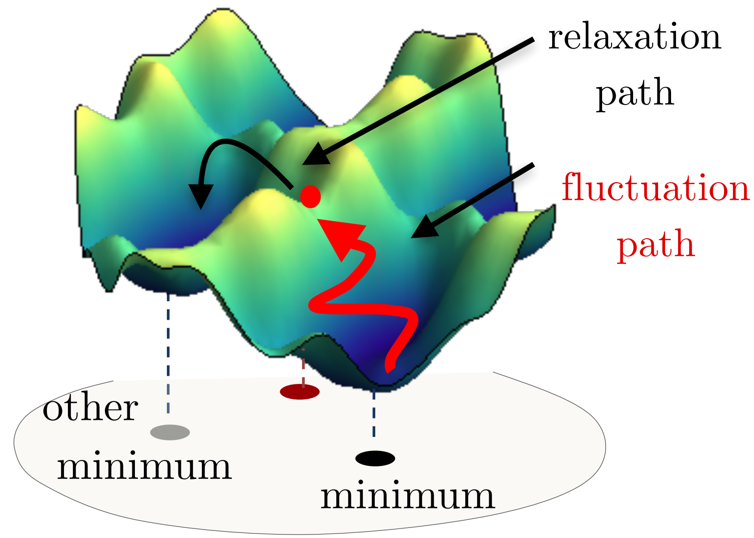
* Sompolinsky Zippelius '82, Cugliandolo Kurchan '93, … Langevin: $\dot{\mathbf{s}}_t = -\nabla \mathcal{E}\,[\mathbf{s}_t] + \eta_t$

Activation: the challenge, more precisely

In high N we know* how to describe "relaxation paths": DMFT [$\rightarrow$ e.g. P. Urbani talk]

$$\mathcal{Z}_{\rm dyn} = \int \mathcal{D}Q \; \exp\left(-N\mathcal{A}[\,Q(t,t')\,]\right) \rightarrow \frac{\delta\mathcal{A}}{\delta Q} = 0$$

- Captures typical dynamics: $\mathcal{A} = 0$
- "Short" times $t \neq O(N)$



Because of trapping, dynamics when $N \gg 1$ is often activated:
"fluctuation paths" generated by rare noise fluctuations [$\mathcal{A} > 0$] that require $t \sim e^N$
-huge open problem in glasses and beyond-

* Sompolinsky Zippelius '82, Cugliandolo Kurchan '93, ... Langevin: $\dot{\mathbf{s}}_t = -\nabla\mathcal{E}\,[\mathbf{s}_t] + \eta_t$

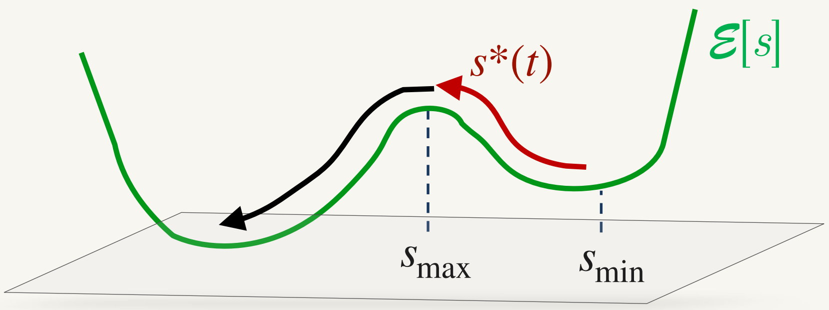
The strategy

[from slow descent to activation]

From relaxation to activation: building instantons

Instantons in low-d: double well, T small

$$\mathcal{Z}_{\rm dyn} = \int_{s(0)=s_0}^{s(\tau)=s_\tau} \mathcal{D}s \; \exp\left(-\frac{1}{T}\left[\dot{s}(t) + \mathcal{E}'(s)\right]^2\right)$$

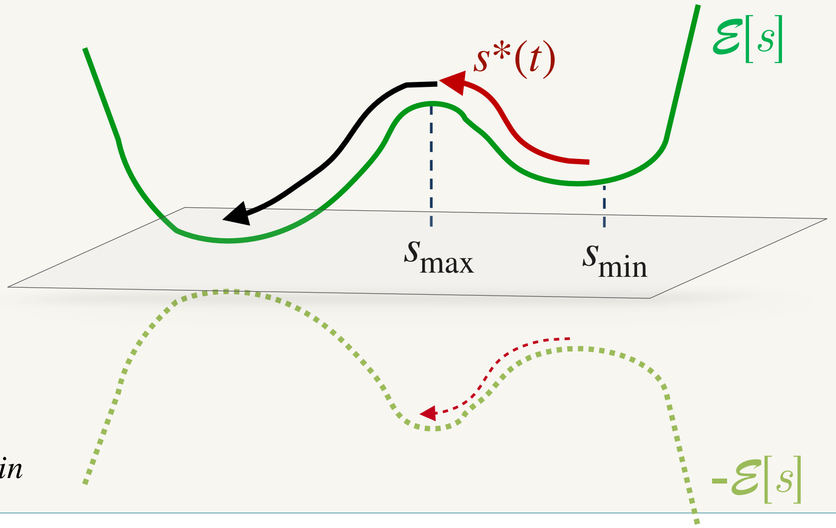


From relaxation to activation: building instantons

Instantons in low-d: double well, T small

$$\mathcal{Z}_{\text{dyn}} = \int_{s(0)=s_0}^{s(\tau)=s_\tau} \mathcal{D}s \; \exp\left(-\frac{1}{T}\left[\dot{s}(t) + \mathcal{E}'(s)\right]^2\right)$$

- Most probable fluctuation path $\mathbf{s}^*(t)$ from $\mathbf{s}_{\text{min}}$ to maximum $\mathbf{s}_{\text{max}}$ satisfies $\dot{\mathbf{s}}^* = \mathcal{E}'(\mathbf{s}^*)$
- Time reversed of relaxation path from $\mathbf{s}_{\text{max}}$ to $\mathbf{s}_{min}$

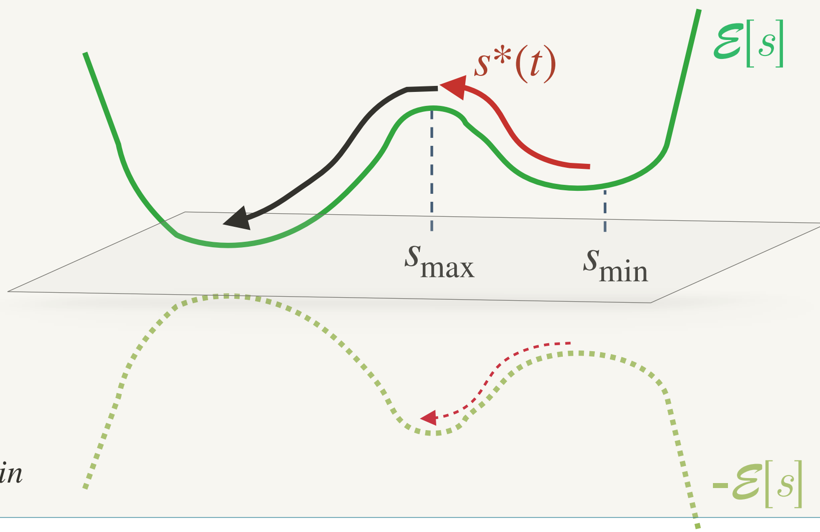


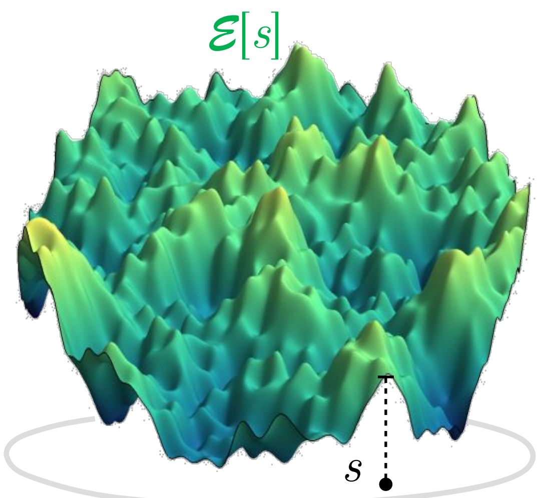
From relaxation to activation: building instantons

Instantons in low-d: double well, T small

$$\mathcal{Z}_{\text{dyn}} = \int_{s(0)=s_0}^{s(\tau)=s_\tau} \mathcal{D}s \; \exp\left(-\frac{1}{T}\left[\dot{s}(t) + \mathcal{E}'(s)\right]^2\right)$$

- Most probable fluctuation path $\mathbf{s}^*(t)$ from $\mathbf{s}_{\min}$ to maximum $\mathbf{s}_{\max}$ satisfies $\dot{\mathbf{s}}^* = \mathcal{E}'(\mathbf{s}^*)$
- **Time reversed** of relaxation path from $\mathbf{s}_{\max}$ to $\mathbf{s}_{min}$



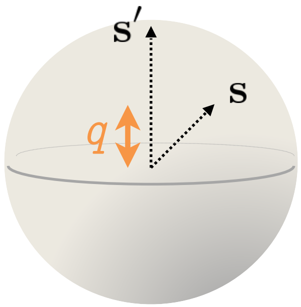


In high-dimension: — see also Lopatin, Ioffe '99

[1] **Non-convexity.** Exponentially many minima and saddles: how are they organized in dynamical paths?

[3] **Entropy.** What subset of the $\sim e^N$ saddles/minima represents escape states/final states?

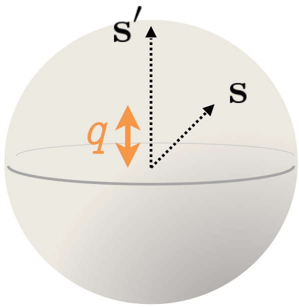
The program:



- Configuration space: sphere of radius $\sqrt{N} \gg 1$
- Functional: random Gaussian monomial with isotropic correlations

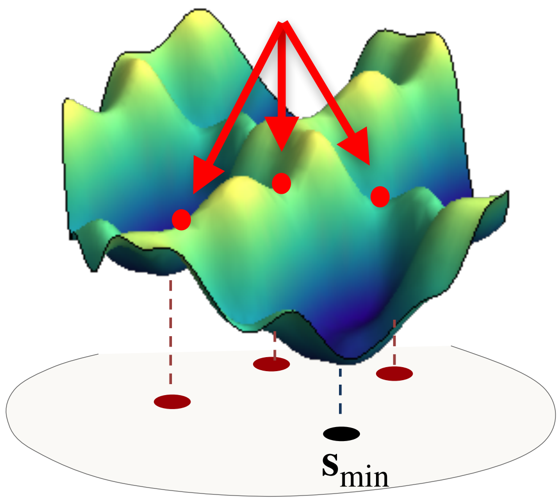
$$\mathcal{E}[\mathbf{s}] = \sum_{i_1 \leq i_2 \cdots \leq i_p} J_{i_1\, i_2 \cdots i_p} s_{i_1} s_{i_2} \cdots s_{i_p} \qquad \langle \mathcal{E}[\mathbf{s}]\mathcal{E}[\mathbf{s}'] \rangle = \frac{N}{2}\left(\frac{\mathbf{s}\cdot\mathbf{s}'}{N}\right)^p$$

The program:



- Configuration space: sphere of radius $\sqrt{N} \gg 1$
- Functional: random Gaussian monomial with isotropic correlations

$$\mathcal{E}[\mathbf{s}] = \sum_{i_1 \leq i_2 \cdots \leq i_p} J_{i_1\, i_2 \cdots i_p} s_{i_1} s_{i_2} \cdots s_{i_p} \qquad \langle \mathcal{E}[\mathbf{s}]\mathcal{E}[\mathbf{s}'] \rangle = \frac{N}{2}\left(\frac{\mathbf{s}\cdot\mathbf{s}'}{N}\right)^p$$



[1] Geometry: looking at the landscape locally

For any minimum, get connected saddles: how many, how high, how far.

VR, J. Phys A: Math Theor 53 (2020)
VR Biroli. Cammarota, EPL 126 (2019)

[2] Geometry-dependent dynamics

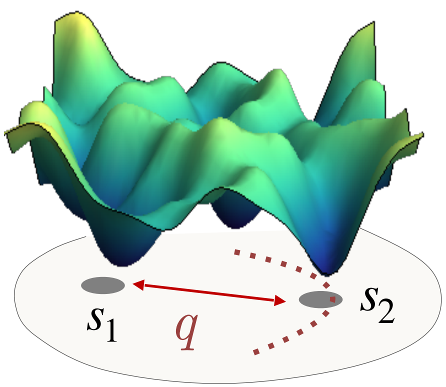
Dynamics initialized at saddles: where it ends up? How?

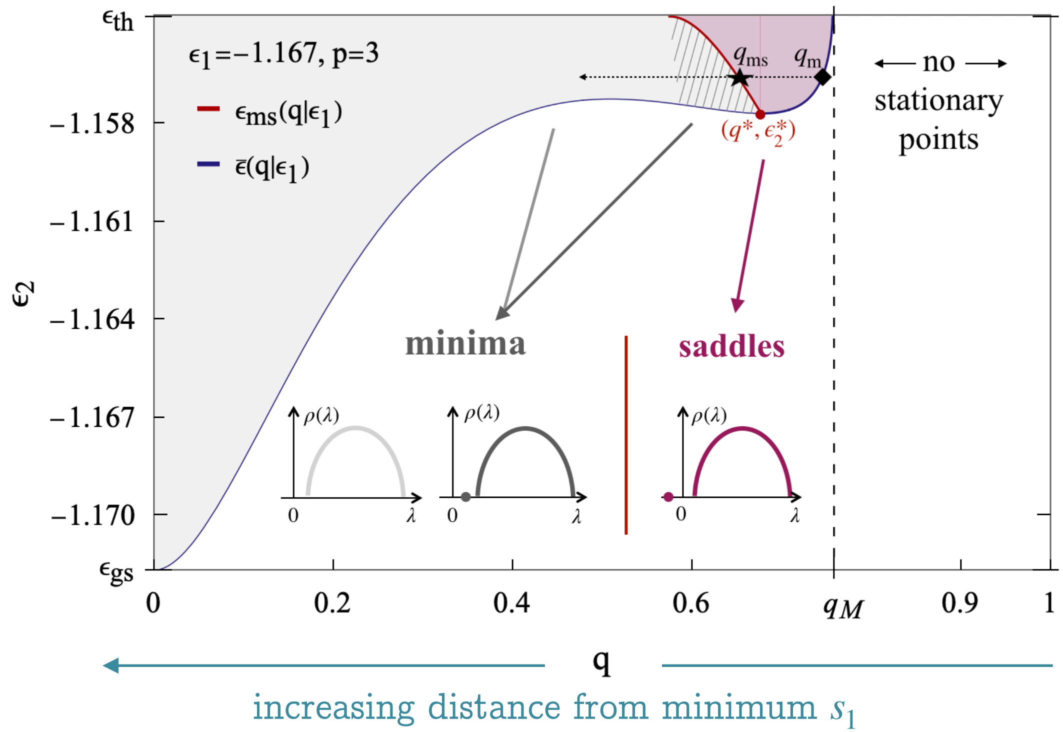
[3] Time reversal: building instantons

VR Biroli Cammarota, arXiv:2006.08399

Results

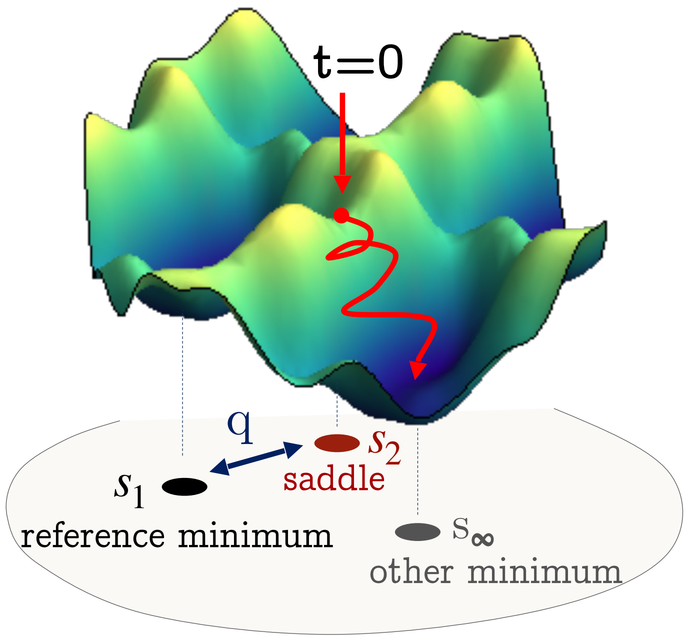
[1] Geometry: looking at the landscape locally





- Technically: a quenched Kac-Rice formalism to compute saddles complexity $\Sigma \sim \dfrac{\log \mathcal{N}}{N}$
- Stability of stationary points controlled with random matrix theory: Hessians are GOE shifted + deformed with finite-rank perturbations [$\rightarrow$ BBP transitions]

[2] Dynamics conditioned to the geometry



Dynamical mean-field equations with saddles as initial conditions*. Give typical values of:

$$c(t,t') = \frac{\mathbf{s}(t)\cdot\mathbf{s}(t')}{N}$$ correlations along dynamical path

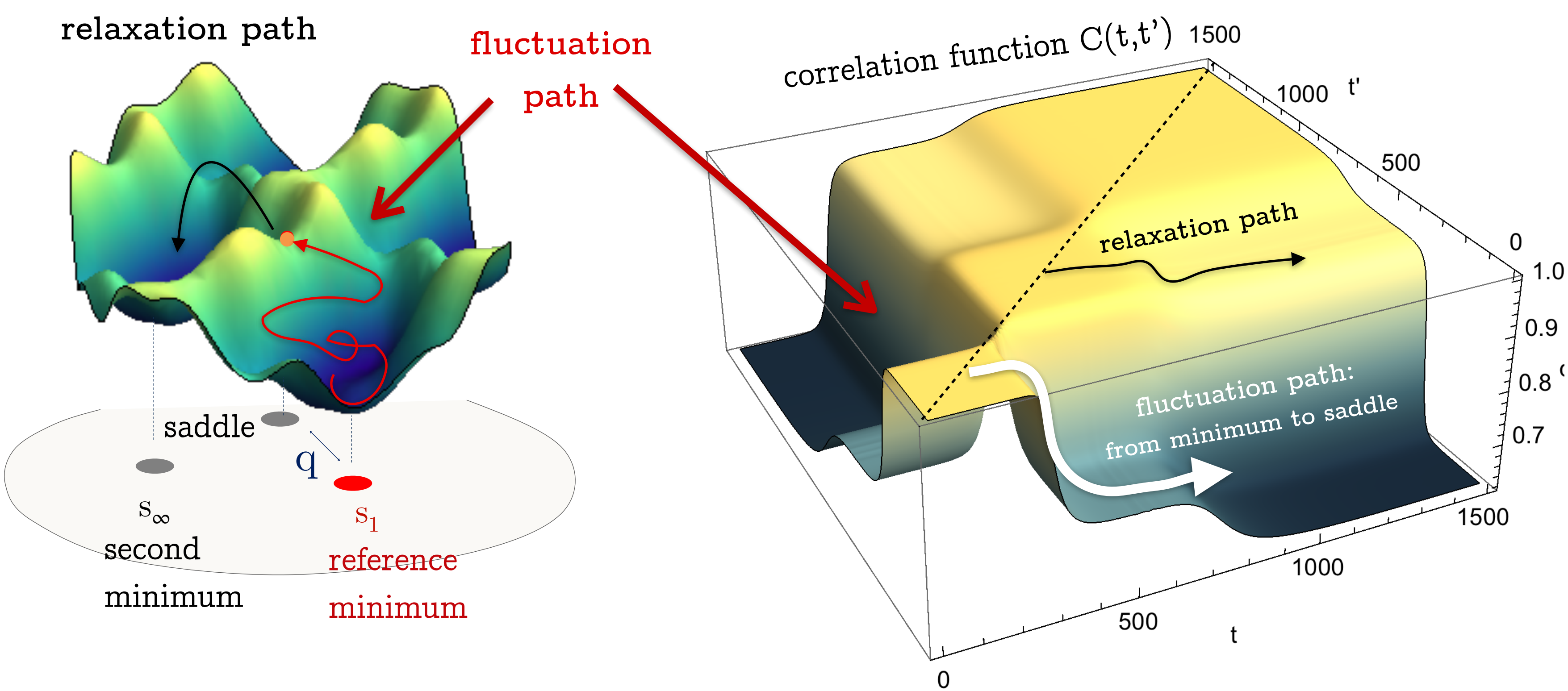
$$Q(t) = \frac{\mathbf{s}(t)\cdot\mathbf{s}_1}{N}$$ overlap with the fixed minimum

$$\begin{aligned}
(\partial_t + z(t))\, c(t,t') &= \alpha r(t',t) + \frac{p(p-1)}{2}\int_0^\infty ds r(t,s)[c(t,s)]^{p-2}c(t',s) + \frac{p}{2}\int_0^\infty ds [c(t,s)]^{p-1} r(t',s) \\
&- \frac{p(p-1)}{2}\frac{q^2}{q^2-q^{2p}} c(t')\int_0^\infty ds r(t,s)\left\{c^{p-2}(t)c^{p-1}(s) - \frac{q^{p-1}}{2}\left(c^{p-2}(t)Q^{p-1}(s) + Q^{p-2}(t)Q(s)c^{p-2}(s)\right)\right\} \\
&- \frac{p(p-1)}{2}\frac{q^2}{q^2-q^{2p}} Q(t')\int_0^\infty ds r(t,s)\left\{Q^{p-2}(t)Q^{p-1}(s) - \frac{q^{p-1}}{2}\left(Q^{p-2}(t)c^{p-1}(s) + c^{p-2}(t)Q^{p-2}(s)c(s)\right)\right\} \\
&- \frac{p}{2}\frac{q^2}{q^2-q^{2p}}\int_0^\infty ds r(t',s)\left\{[Q(t)Q(s)]^{p-1} + [c(t)c(s)]^{p-1} - q^{p-1}\left([Q(t)c(s)]^{p-1} + [c(t)Q(s)]^{p-1}\right)\right\} \\
&+ \mathcal{F}_{\mathcal{E},q}[c,Q]
\end{aligned}$$

boundary terms— initial conditions

* Similar conditioning: Dembo Subag '19, Barrat Burioni Mezard '96, Barrat Franz '98

[3] The shape of high-d instantons



First example of instanton for high-d Langevin: correlation function along escape paths from minimum to saddle to another minimum

ϵ_1=-1.167
ϵ_{saddle}=-1.1555
q=.75
overlap wt saddle= 0.957351
overlap wt min= 0.619584
ϵ_∞=-1.15595

In the talk:

- Dynamics in high-d non-convex landscapes with extensive barriers: sharp separation of timescales — slow descent vs activated jumps
- Can build instantons for 2-point functions — via DMFT & time reversal
- Crucial ingredient: control on local landscape geometry — via Kac-Rice

Some perspectives:

- Contribution from saddles entropy: more general escape processes
- Activated processes from family of rare saddles
- Non-Gaussian random landscapes

Details:

arXiv:2006.08399 J. Phys A: Math Theor 53 (12) 125002 EPL 126 (2) 20003