High-Dimensional Rough Landscapes from Physics to Machine Learning

Giulio Biroli

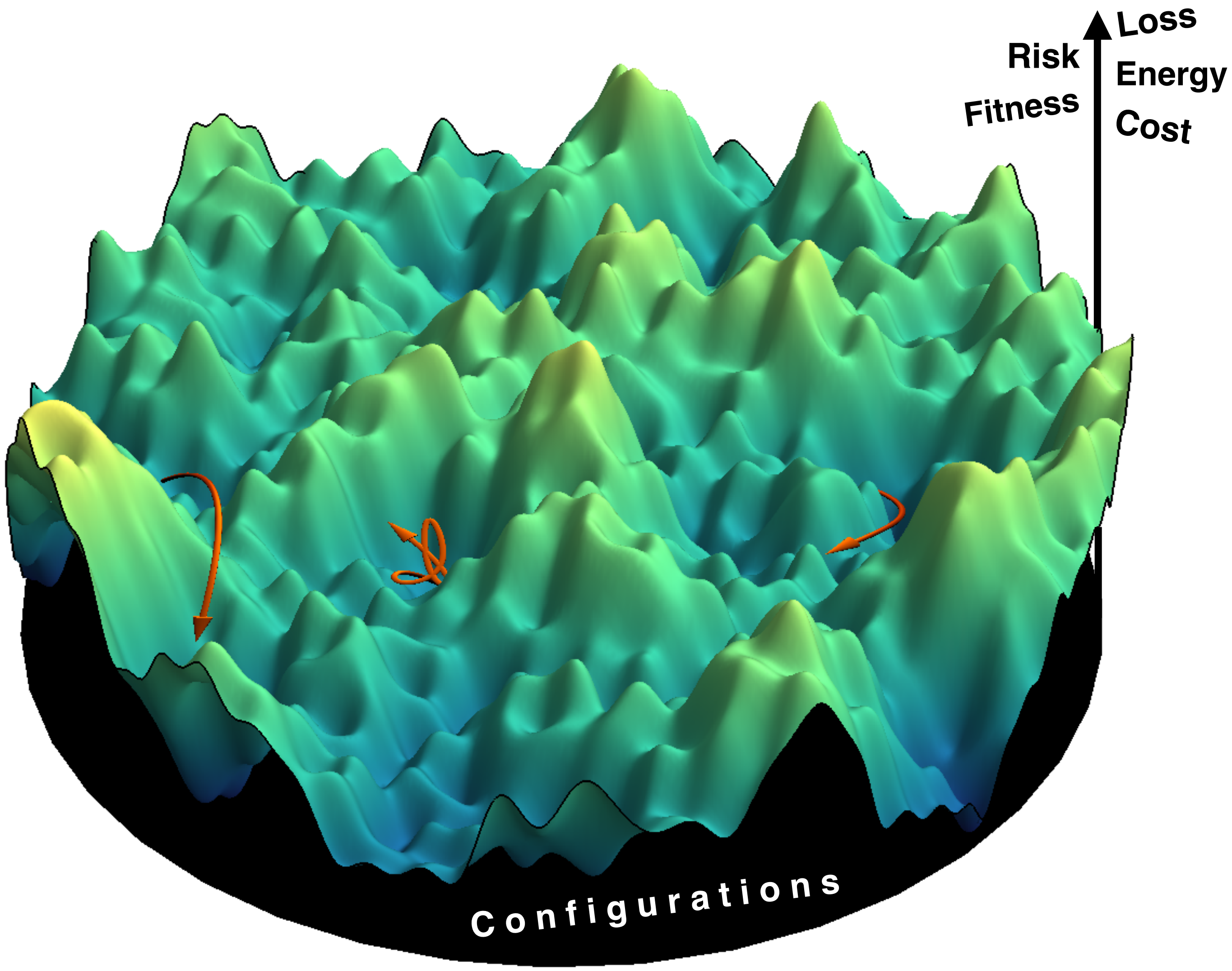
Ecole Normale Supérieure, Paris

Département de Physique
École normale supérieure

PR[AI]RIE
PaRis Artificial Intelligence Research InstitutE

SIMONS FOUNDATION

Rough Landscapes



Physics

Glasses, spin-glasses, amorphous materials,..

Biology

Evolution, Ecosystems,...

Information Theory

Optimization problems, inference,...

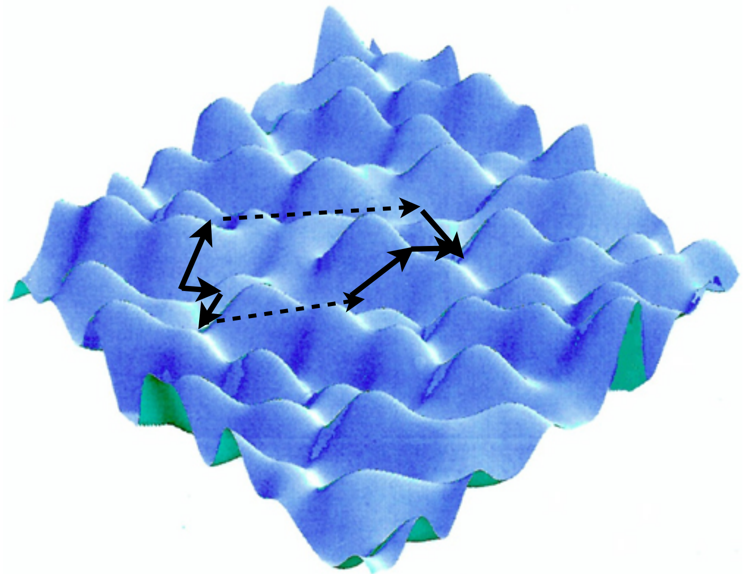
Statistics and Machine Learning

High-dimensional statistics,...

Glassy Systems

Glassy systems (super-cooled liquids, colloids,..): interacting particles with rough energy landscapes
Many systems are glassy: e.g., hard combinatorial optimization problems.

Glassy systems: diverging number of critical points and paths



Numerics example: 38 Lennard-Jones particles

$\mathcal{N}_{min} \sim 6000$

Wales et al. Science 1996

Statistics of the landscape and its influence on dynamics
Characterise minima (amorphous solids)

Cavagna, Giardina, Parisi 1998, …
Fyodorov 2004, … Auffinger, Cerny, Ben Arous 2013, …

Landscapes & Dynamics of Machine Learning Algorithms

Machine-Learning

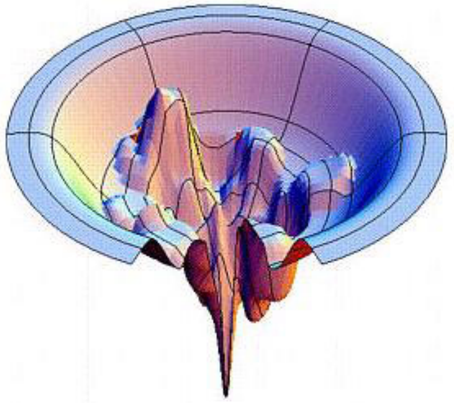
Minimizing Cost function with respect to many variables

$$\mathcal{L} = \frac{1}{M}\sum_{\alpha}^{M} \ell_\alpha(w)$$

$$\mathrm{Min}_{\{\mathrm{w}\}}\mathcal{L}(\{w\})$$

Dynamics: gradient descent & variants

Properties of the "landscape" $\mathcal{L}$?
properties of learning dynamics?



Statistics of the landscape and its influence on dynamics
Characterise minima (properties of solutions)

Krzakala, Montanari, Ricci-Tersenghi, Semejian, Zdeborová, …
Fyodorov 2018, Ros, Ben Arous, GB, Cammarota 2018, Baskerville, Keating, Mezzadri, Najudel 2020,…

Dynamics in Ecosystems and Biological Neural Networks

Quantitative Biology

Ecosystems formed by <u>many</u> species, e.g. bacteria, or recurrent neural networks

$$\frac{dN_i}{dt} = F_i(\{N_i\}) \neq -\frac{\partial V}{\partial N_i}$$

⟶

Properties of the equilibria?
properties of chaotic dynamics?

Lotka-Volterra equations $i = 1, \ldots, S \gg 1$

High-Dimensional Dynamics and Non-Conservative Forces

Statistics of the equilibria and their influence on dynamics
Characterise equilibria (ecosystems properties)

Wainrib, Touboul 2013 , Fyodorov, Khoruzhenko 2015, Ros, GB et al 2022…

Rough Energy Landscape Problem

- Complexity and/or Randomness

- Very High-Dimensional Landscape

10^3	Biology
10^8	Machine Learning & High-D statistics
10^{23}	Physics

- Rough: number of critical points (equilibria) is exponential in D

Methods for High-D Rough Landscapes

- Kac-Rice method Yan! (2004)

  It can be made fully rigorous (see Auffinger, Cerny, Ben Arous 2013)
  Many applications in recent years
  Put on a firm basis previous methods (see eg Cavagna, Giardina, Parisi 1998)

- Replica Method

  Indirect information on the landscape (see eg Monasson 1995)
  More widely applicable (discrete systems, non-Gaussian landscapes, ...)

Glass Transition & Gaussian Energy Landscapes

$E(s)$ Gaussian Random Function on the N-dimensional sphere $\sum_i s_i^2 = N$

$\langle E(s) \rangle = 0 \qquad \langle E(s) E(s') \rangle = f_R(s \cdot s'/N)$

Example: p-spin spherical model (paradigmatic model in the theory of glass transition)

$$E(s) = -\sum_{i_1,\ldots,i_p} J_{i_1,\ldots,i_p} s_{i_1} \ldots s_{i_p}$$

$J_{i_1,\cdots,i_p}$ iid Gaussian Random Variables

$$f_R(q) = \frac{N}{2} q^p$$

Kac-Rice Method for High-D Gaussian Landscapes

(from now on large D=N limit)

Density of critical points of a given energy

$$\mathcal{N}(E) = \int ds\, \delta(\nabla E(s))\, |det(\nabla^2 E(s))|\, \delta(E - E(s))$$

$$\ln\langle\mathcal{N}(E)\rangle = N\Sigma_A(E)$$

Annealed average

Easier, amenable to rigorous analysis

$$\langle\ln\mathcal{N}(E)\rangle = N\Sigma_Q(E)$$

Quenched average

More difficult, in general needs replica (except if Q=A)

Extensions: Density of critical points with a given index, at a given distance of another critical point

The method in 3 (rather) easy steps

Step 1: Isotropy

$$\langle \mathcal{N}(E) \rangle = \int ds \langle \, \delta(\nabla E(s)) \, |det(\nabla^2 E(s))| \, \delta(E - E(s)) \rangle$$

$$\downarrow$$

$$\langle \mathcal{N}(E) \rangle = S_N(\sqrt{N}) \langle \, \delta(\nabla E(1)) \, |det(\nabla^2 E(1))| \, \delta(E - E(1)) \rangle$$

The method in 3 (rather) easy steps

Step 1: Isotropy

Step 2: Covariances and density of gradients

$$\langle\, \delta(\nabla E(1))\, |det(\nabla^2 E(1))|\, \delta(E - E(1))\rangle$$

$$\langle E(1)E(1)\rangle = \frac{N}{2} f_R(1) \quad \langle \nabla_a E(1) \nabla_b E(1)\rangle = \frac{1}{2}\delta_{a,b} f'_R(1) \quad \nabla^2_{ab} E(1) = G_{ab} - \delta_{a,b} f'_R(1) \frac{E(1)}{N}$$

$$\downarrow$$

$$\left(\frac{1}{\sqrt{\pi f_R(1)}}\right)^{N-1} \frac{e^{-E^2/(N f_R(1))}}{\sqrt{\pi N f_R(1)}} \,\langle |det \left(G_{ij} - \delta_{i,j} f'_R(1) \frac{E}{N}\right)|\rangle$$

The method in 3 (rather) easy steps

Step 1: Isotropy

Step 2: Covariances and density of gradients

Step 3: RMT and absolute value of the determinant

$$\langle | det \left(G_{ij} - \delta_{i,j} f_R'(1) \frac{E}{N} \right) | \rangle$$

$$\downarrow$$

$$\langle \exp \left(N \int d\lambda \rho(\lambda) \ln |\lambda - f_R'(1) \frac{E}{N}| \right) \rangle$$

$$\downarrow$$

$$\exp \left(N \int d\lambda \langle \rho(\lambda) \rangle \ln |\lambda - f_R'(1) \frac{E}{N}| \right)$$

The method in 3 (rather) easy steps

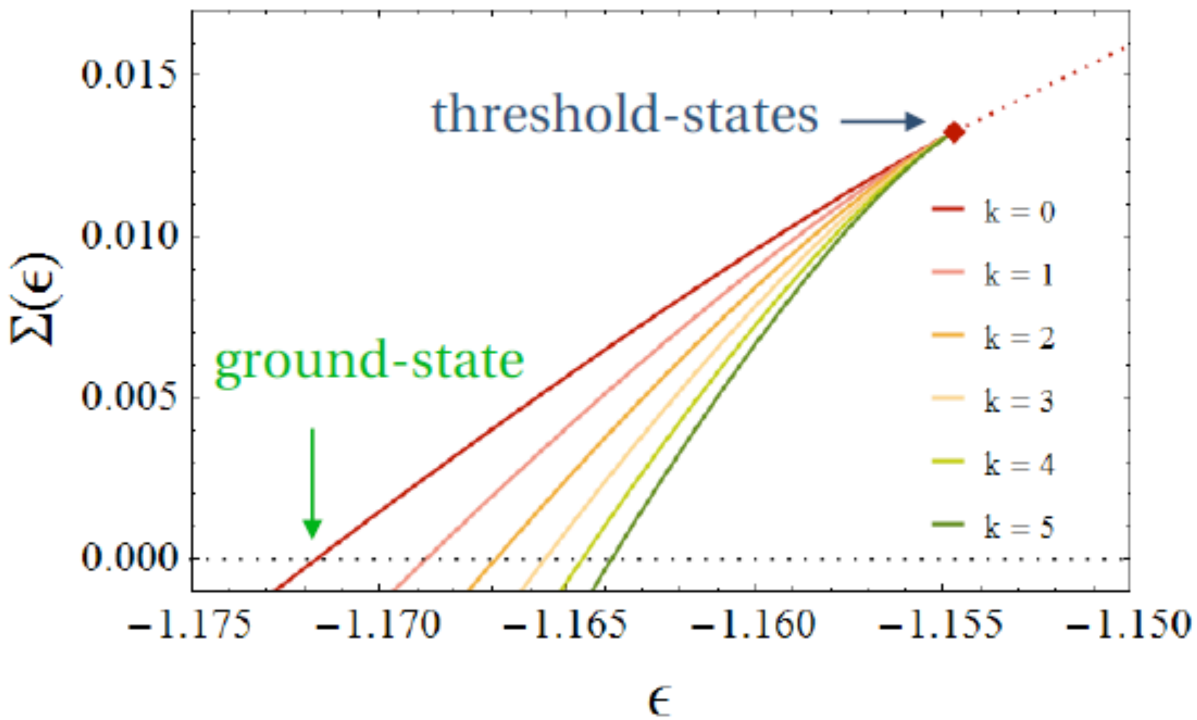
Step 1: Isotropy

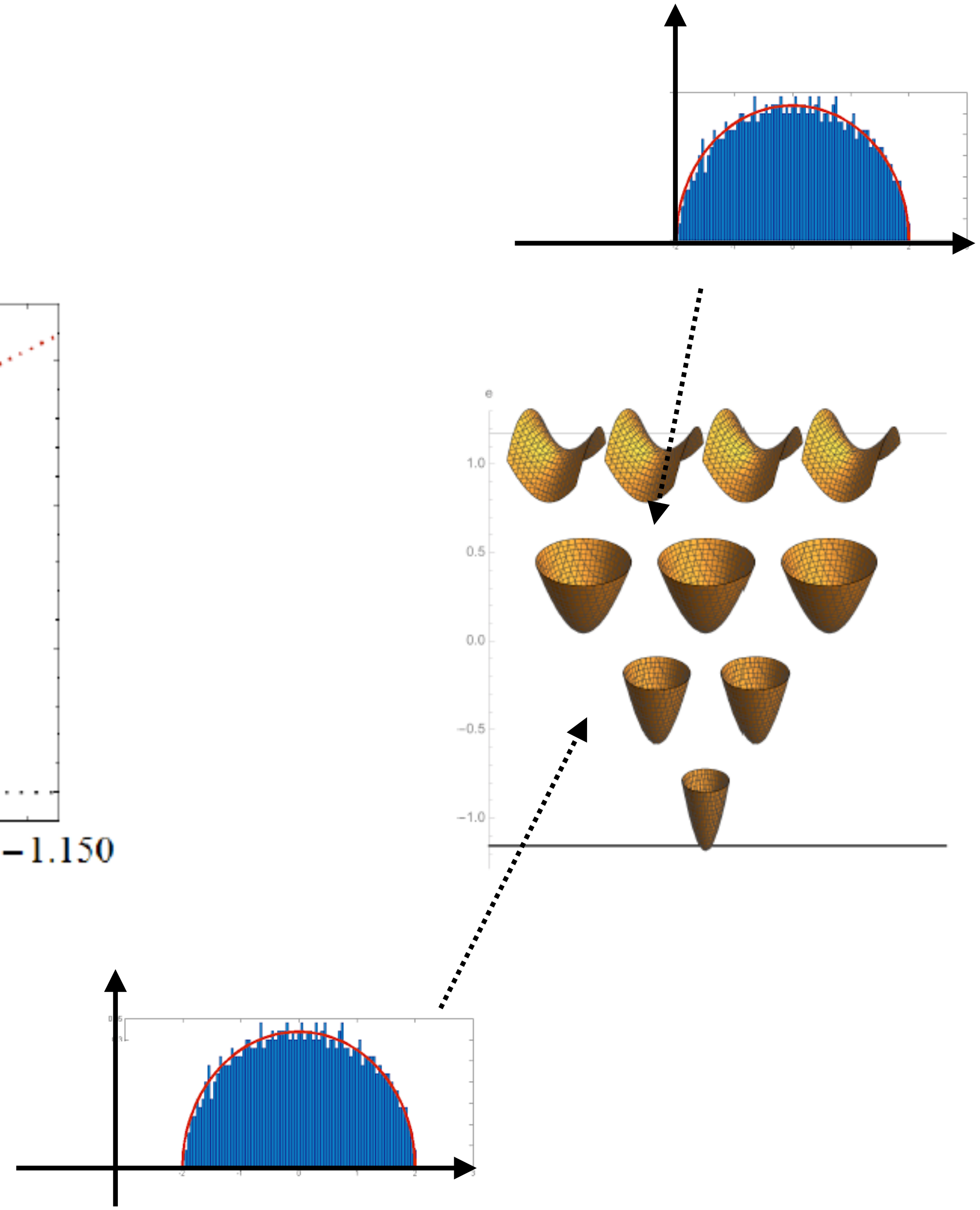
Step 2: Covariances and density of gradients

Step 3: RMT and absolute value of the determinant

$$\ln \langle \mathcal{N}(E) \rangle = N \left[\frac{1}{2} (\ln \pi + \ln 2 + 1) - \frac{1}{2} \ln(\pi f_R'(1)) - \frac{1}{f_R(1)} \frac{E^2}{N^2} + \int d\lambda \frac{\sqrt{2 f_R''(1) - \lambda^2}}{\pi f_R''(1)} \ln \left| \lambda - f_R'(1) \frac{E}{N} \right| \right]$$

Complexity for Mean-Field Glassy Systems (P-spin spherical model)





Cavagna, Giardina, Parisi 1998, …
Auffinger, Cerny, Ben Arous 2013, …
Subag, Zeitouni 2021

Universality class of 1RSB landscapes from replica theory
(spin systems, random graphs, interacting particles in large D)

Landscape & Dynamics

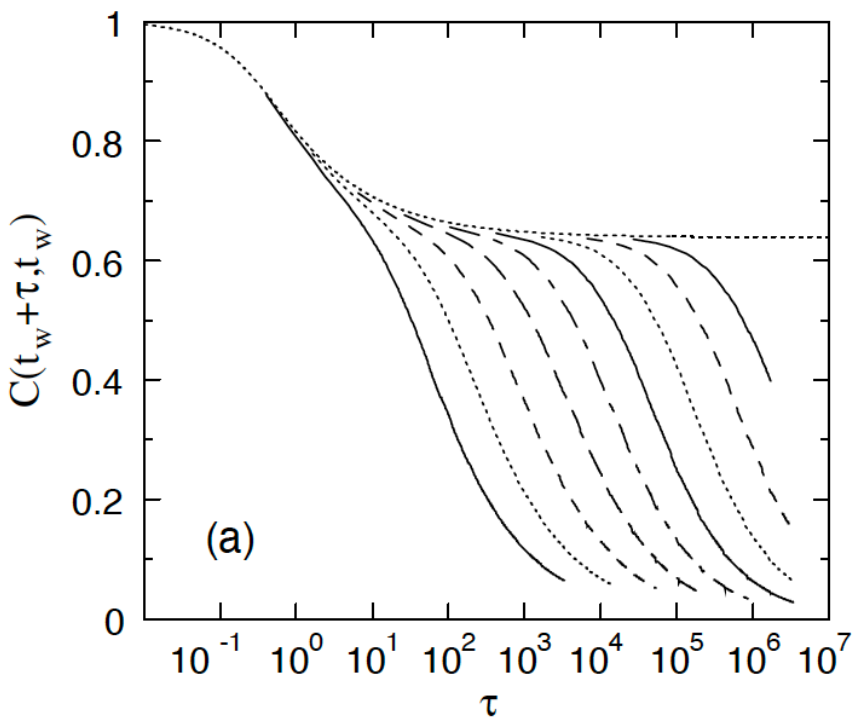
Langevin dynamics : $$\frac{dx}{dt} = -\nabla \mathcal{E} + \eta(t)$$

Dynamics in mean-field glasses: closed self-consistent stochastic process

Crisanti, Sommers 1993, Cugliandolo Kurchan 1993,...,Ben Arous, Guionnet 1997

Ergodicity breaking (glass transition)
trapping in the threshold states

$$C(t_w + \tau, t_w) = s(t_w + \tau) \cdot s(t_w)/N$$



The glassy landscape plays a crucial role in the study of glassy dynamics within and beyond mean-field

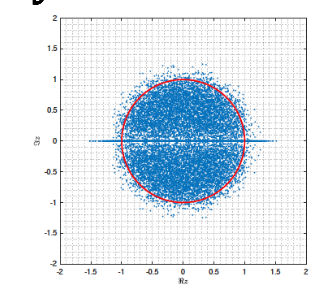
Extensions & Generalizations

- Replicated Kac-Rice to obtain quenched averages
  Ros, Ben Arous, GB, Cammarota 2018

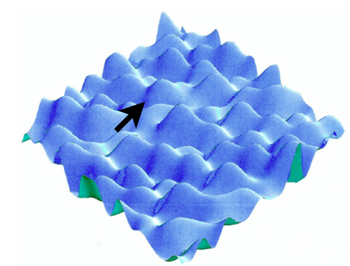
Parisi Matrix

$$\langle \ln \mathcal{N}(E) \rangle = \lim_{n \to 0} \frac{1}{n} \ln \langle \mathcal{N}(E)^n \rangle \longrightarrow Q_{ab}$$

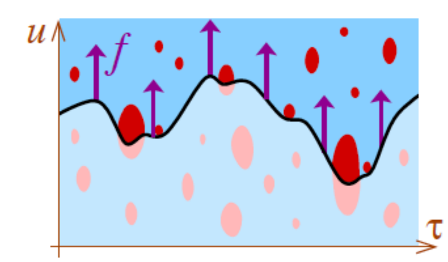
Non-symmetric matrices



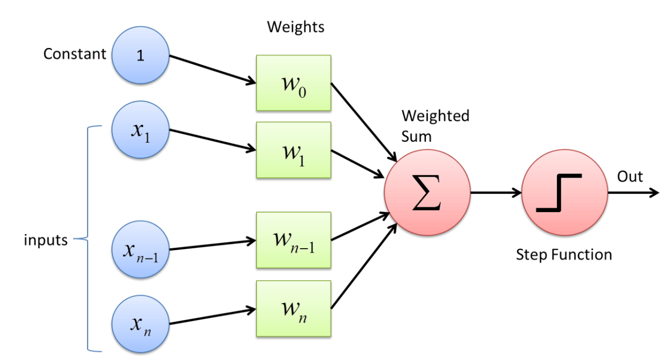
- Non-symmetric interactions and non-conservative systems
  Wainrib, Touboul 2013 , Fyodorov, Khoruzhenko 2015, Ros, GB et al 2022



- Barrier and paths between minima
  Ros, GB, Cammarota 2019-2021, Ros 2020



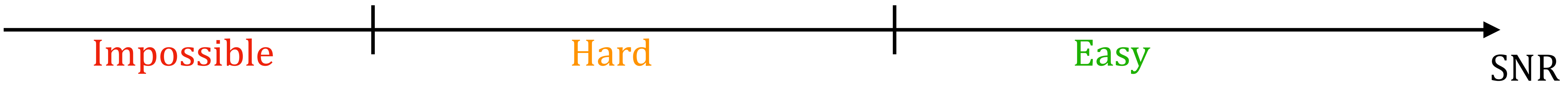
- Finite dimensional models
  Fyodorov, Le Doussal, Rosso, Texier 2017, Ben Arous, Bourgade, McKenna 2021



- Non-Gaussian energy (cost) functions
  Fyodorov 2018, Maillard, Ben Arous, GB 2019

Estimation problems in inference and machine learning

- Signal corrupted by noise
- Construct an energy function to minimise (e.g. MAP)
- Use a local optimisation algorithm to minimise it (eg gradient flow)



How does the landscape evolve with SNR?

Where is the trivialisation transition (just one global minimum) in this phase diagram?

How does the landscape influence the recovery of the signal?

Matrix & Tensor Principal Component Analysis

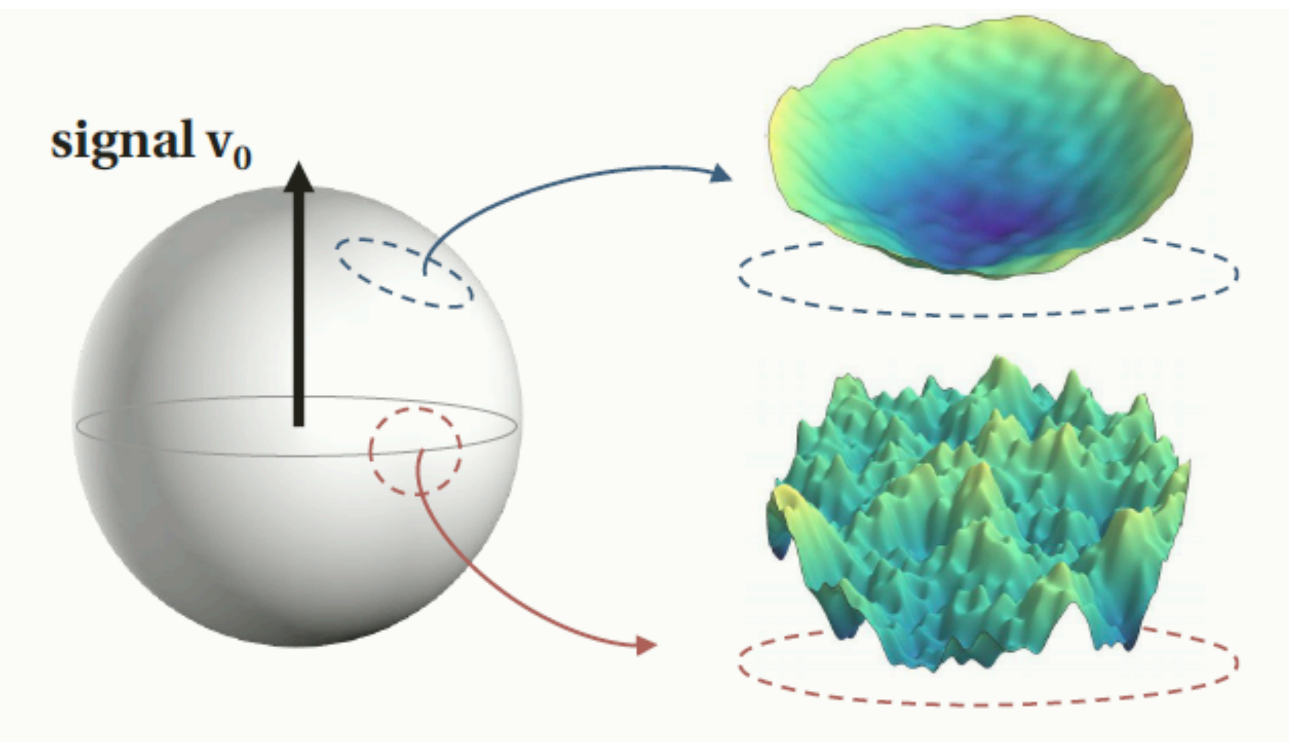
$$T_{i_1,\dots,i_p} = v_{i_1}\dots v_{i_p} + J_{i_i,\dots,i_p} \longrightarrow E(s) \sim \sum_{i_i,\dots,i_p} \left(T_{i_1,\dots,i_p} - s_{i_1}\dots s_{i_p}\right)^2$$

$$E(s) \sim -\sum_{i_1,\dots,i_p} T_{i_1,\dots,i_p} s_{i_1}\dots s_{i_p} + cte$$

$J_{i_1,\cdots,i_p}$ Gaussian Random Variables

v and s are on the sphere

For p=2 Matrix PCA
For p>2 Tensor PCA (Richard, Montanari 2014)



$E(s)$ Gaussian Random Function on the N-dimensional sphere

$$\langle E(s)E(s')\rangle = f_R(s\cdot s'/N)$$

$$\langle E(s)\rangle = r f_D(s\cdot v/N)$$

Quenched Kac-Rice: Ros, Ben Arous, GB, Cammarota 2018
Annealed Kac-Rice (for tensor PCA): Ben Arous, Mei, Montanari, Nica 2018

Matrix and tensor PCA model

$$E(s) = -\frac{1}{\Delta_p} \sum_{i_1,\ldots,i_p} T_{i_1,\ldots,i_p} s_{i_1} \ldots s_{i_p} - \frac{1}{\Delta_2} \sum_{i_1,i_2} Y_{i_1,i_2} s_{i_1} s_{i_2}$$

$\frac{1}{\Delta_p}$ and $\frac{1}{\Delta_2}$ Signal to Noise ratio

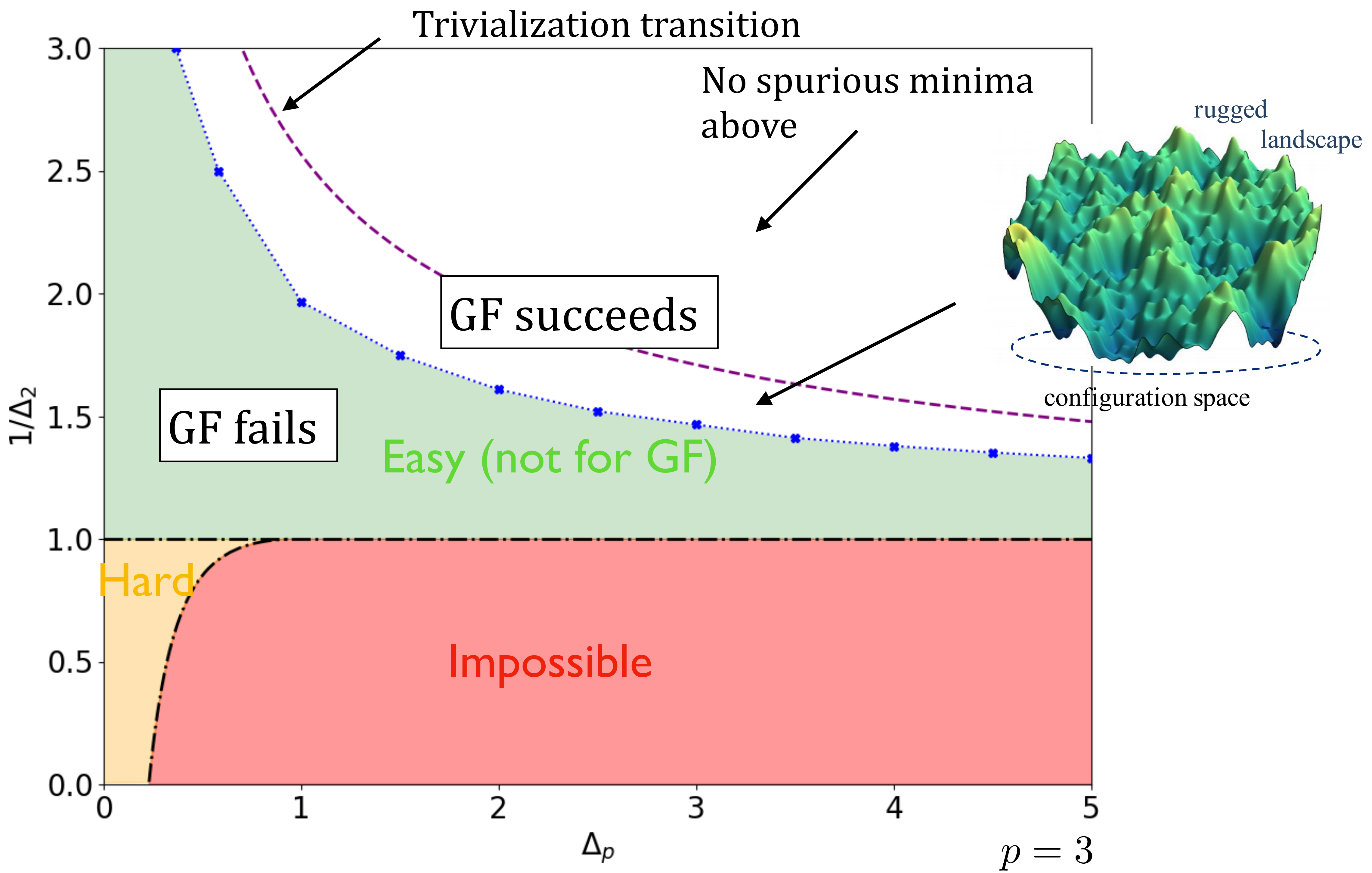
Optimisation Algorithm $\frac{ds}{dt} = -\nabla E$ Gradient Flow from a random initial condition

Landscape by Kac-Rice method

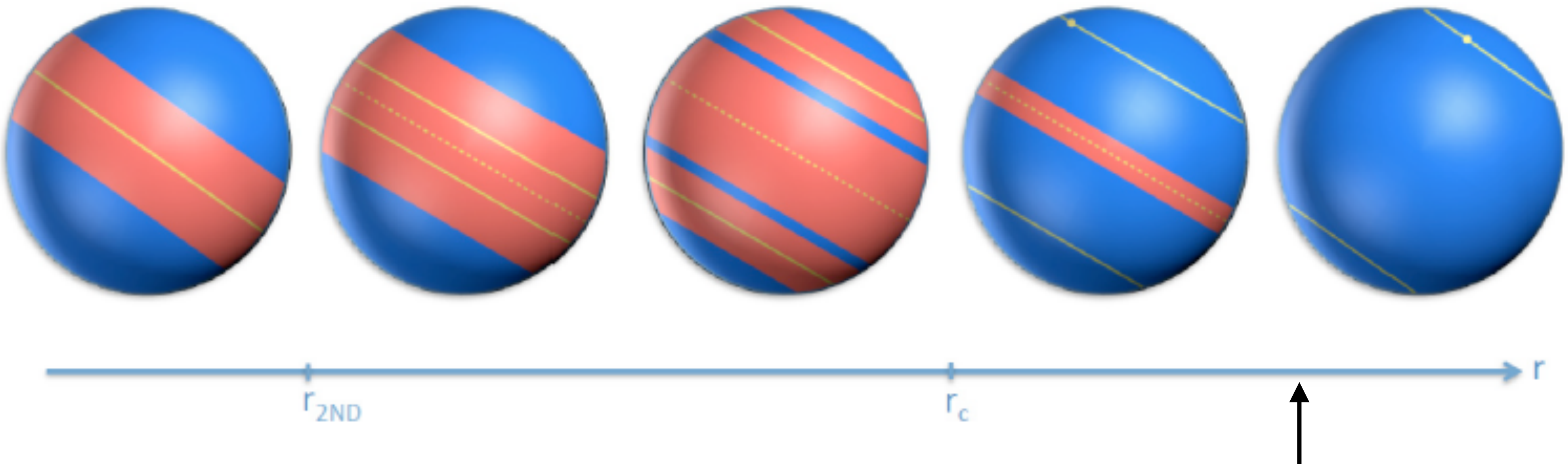
Dynamics by dynamical mean field theory

Sarao, GB, Cammarota, Krzakala, Urbani, Zdeborova 2020,…

Dynamical phase diagram



Evolution of the landscapes with SNR

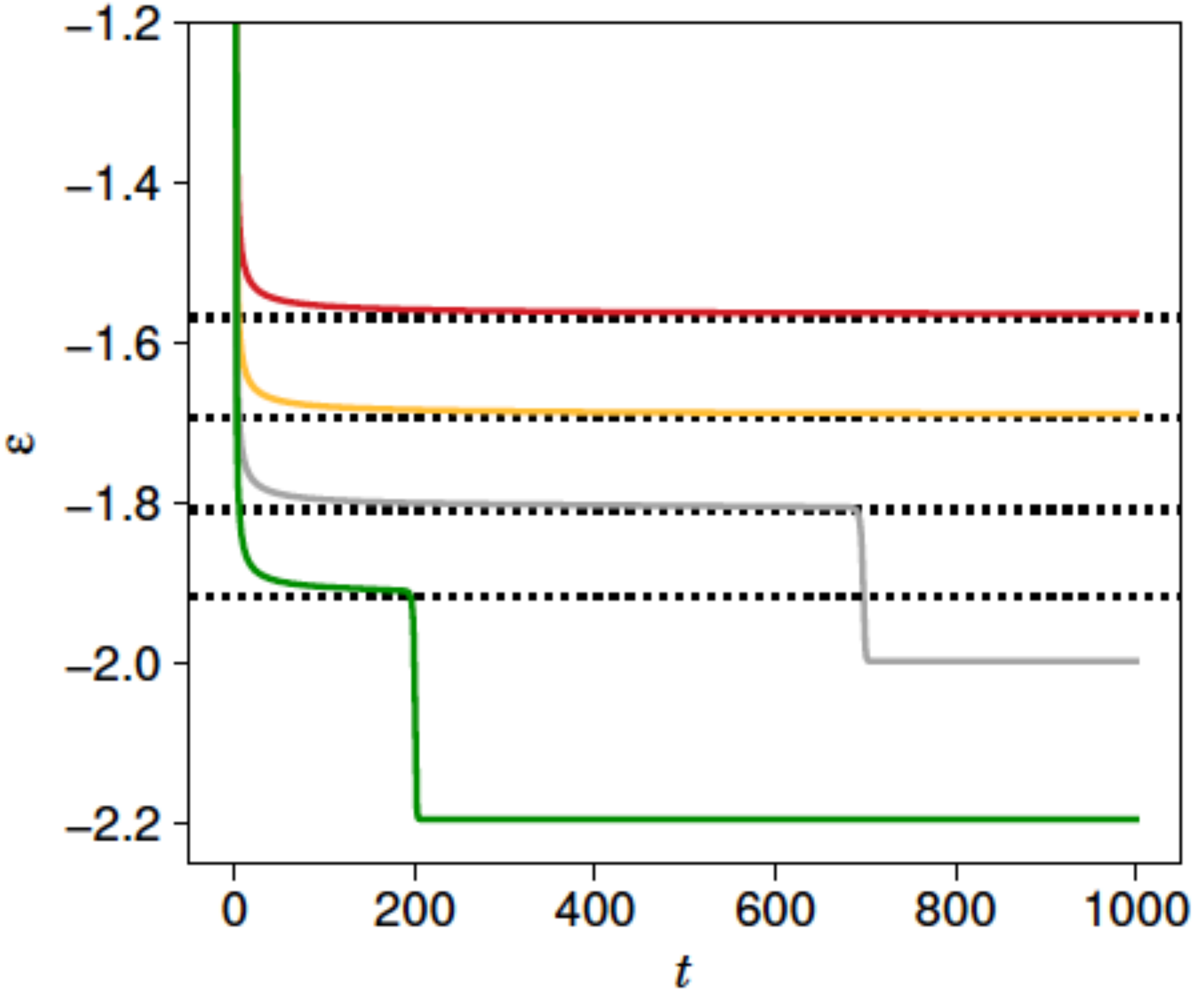


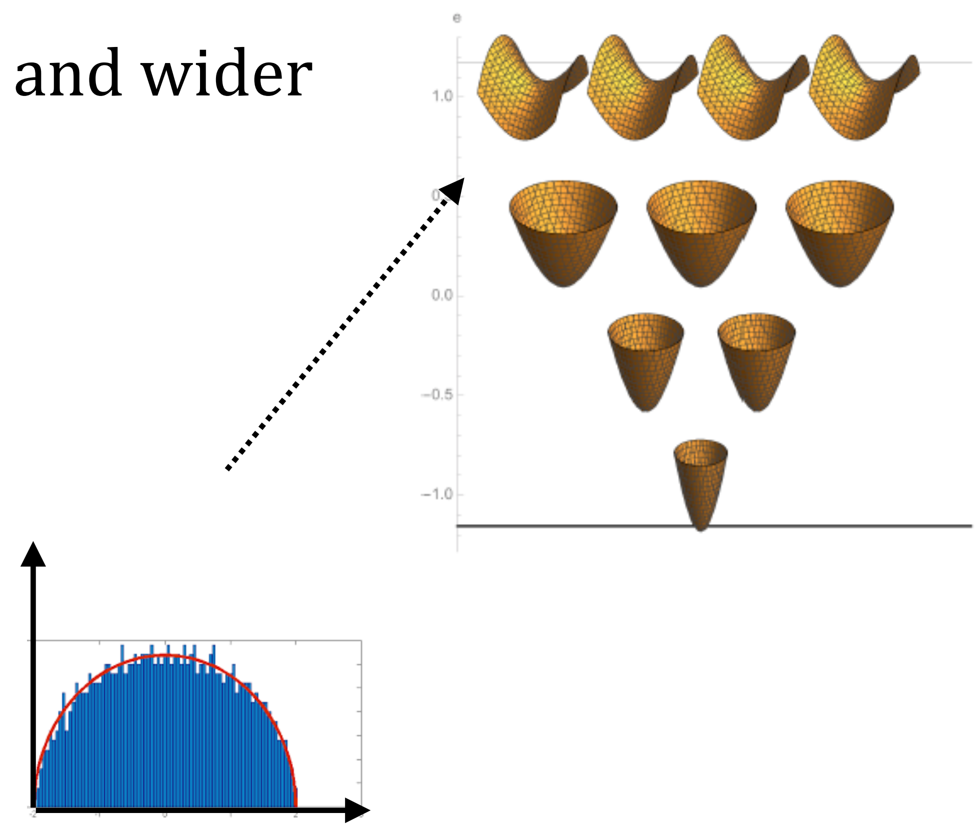
Ros, Ben Arous, GB, Cammarota, PRX 2018

What traps the dynamics?

The glassy landscape on the equator

Threshold states: most numerous and wider



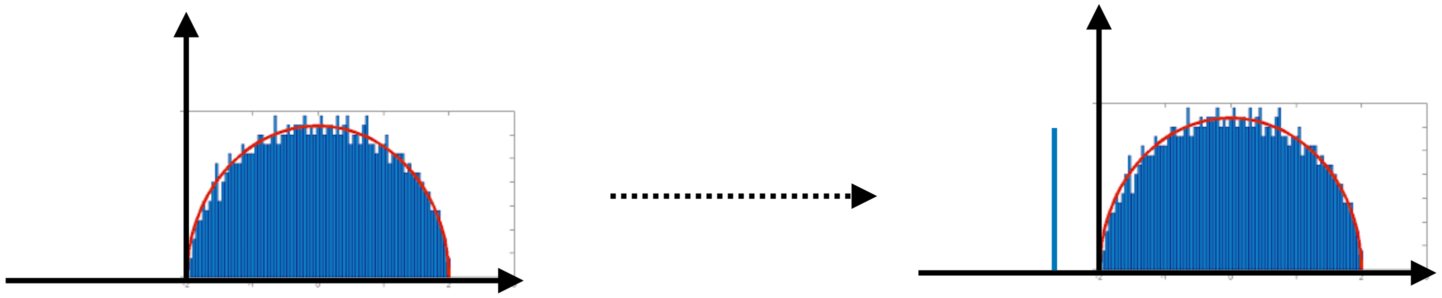


How does the algorithmic transition for GF work?

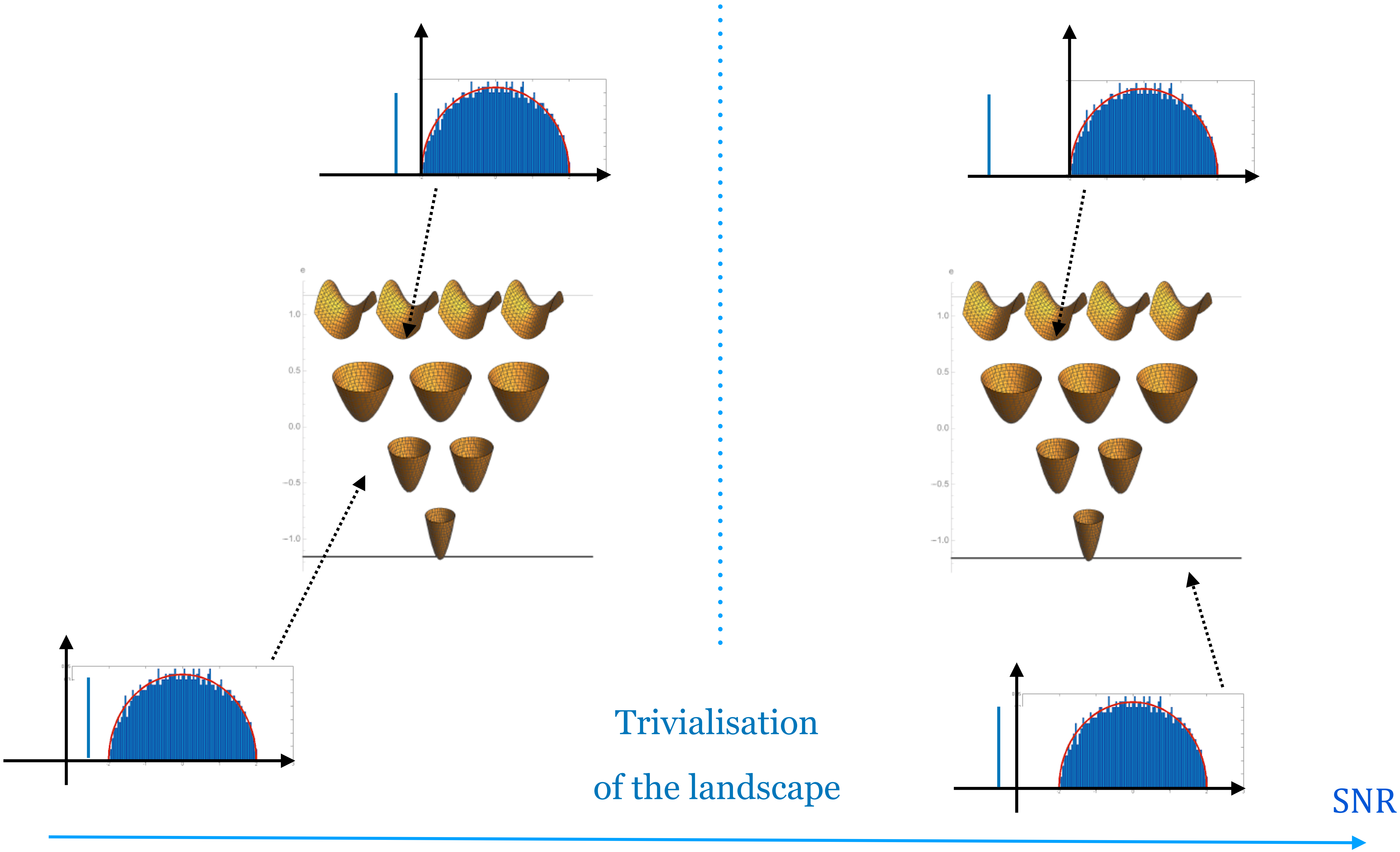
Their Hessian is a GOE matrix plus a rank-one perturbation in the direction of the signal

$$H_{ij} = G_{ij} - \frac{1}{2\Delta_2 N} v_i v_j$$

For large enough SNR: BBP transition —> a downward direction toward the signal emerges

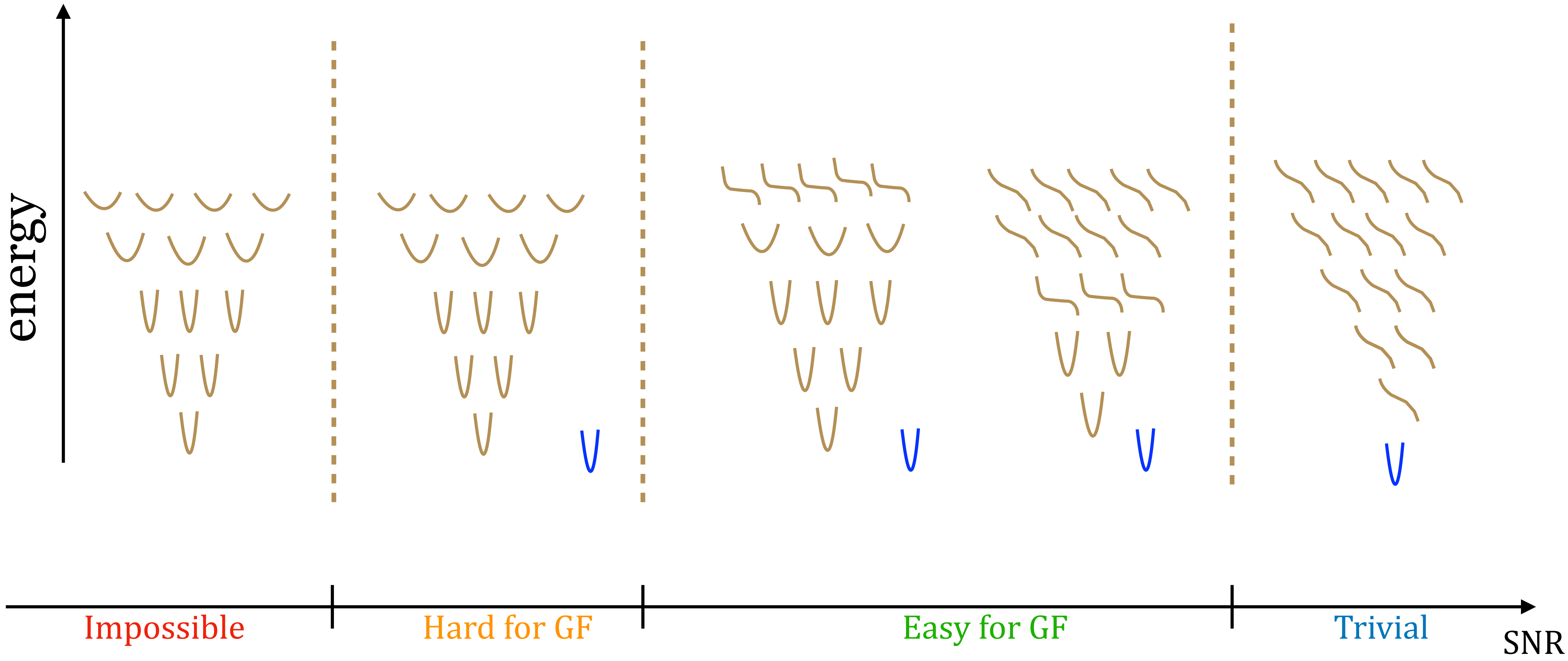


Sarao, GB, Cammarota, Krzakala, Zdeborova NeurIPS 2019



Cascade of instabilities from threshold to stable states

Sarao, GB, Cammarota, Krzakala, Zdeborova NeurIPS 2019
Ros, Ben Arous, GB, Cammarota, PRX 2018



•Gradient flow succeeds even in presence of an exponential number of spurious minima

•Algorithmic threshold is determined by the instability of the marginally stable spurious minima leading to glassy dynamics

Challenges & Perspectives

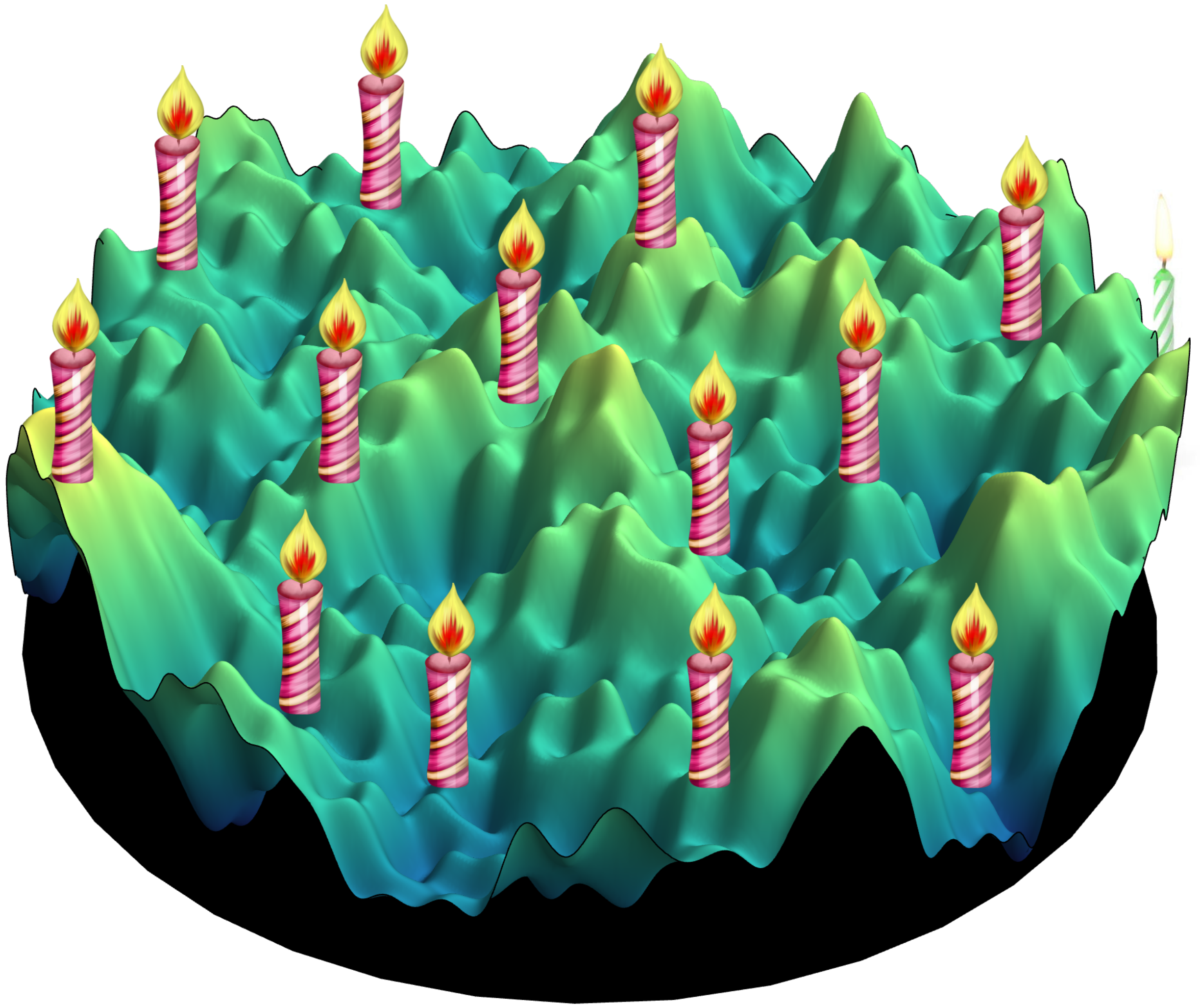
Applications

- Finite dimensional glassy models
- Non-conservative systems (ecology, biology)
- Landscapes of Neural Networks
- Connection “Landscape - Dynamics” in glassy systems and estimation problems

Methods

- Non-Gaussian random landscapes
- Rigorous approach of the replicated Kac-Rice method

The Rough Birthday Cake



Happy 60s Yan !