Physica A 631 (2024) 128152

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Introduction to the dynamics of disordered systems: Equilibrium and gradient descent

Giampaolo Folena [a,b], Alessandro Manacorda [c,a], Francesco Zamponi [a,*]

[a] Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France

[b] James Franck Institute and Department of Physics, University of Chicago, Chicago, IL 60637, USA

[c] Department of Physics and Materials Science, University of Luxembourg, Luxembourg, L-1511, Luxembourg

ARTICLE INFO

Article history:
Available online 5 September 2022

Keywords:
Complex landscapes
Gradient descent
Glass transition
Spin glasses
Disordered systems
Metastability

ABSTRACT

This paper contains the lecture notes of the short courses given by one of us (F.Z.) at the summer school *Fundamental Problems in Statistical Physics XV*, held in Brunico, Italy, in July 2021, and, just before that, at the summer school *Glassy Systems and Inter-Disciplinary Applications*, held in Cargese, France, in June 2021. The course was a short introductory overview of the dynamics of disordered systems, focused in particular on the equilibrium dynamics (with the associated glass transition), and on the simplest case of off-equilibrium dynamics, namely gradient descent. A few selected topics (and references) are chosen, based on the authors' own taste and competences, and on pedagogical reasons, without aiming at a complete review of the subject.

1. Motivations

In these notes, we will discuss the dynamics of two toy models of disordered systems, which provide a simple description for a broader class of more complex models. We will consider (1) the equilibrium dynamics, i.e. the dynamics that starts from the Boltzmann–Gibbs distribution at a fixed temperature and satisfies detailed balance at the same temperature, (2) the simplest case of off-equilibrium dynamics, i.e. a zero-temperature energy minimization via gradient descent from a random initial condition, and (3) a combination of the previous two, which provides a simple approximation to the simulated annealing protocol. The most important feature of the class of disordered systems we consider here is to have a *rough energy landscape*, i.e. an energy function featuring multiple local minima and saddle points that can trap the dynamics for long times; this *metastability* phenomenon will be one of the central theme of the notes.

While the physicists' interest in this problem has mostly been driven by the study of glasses and spin glasses, these phenomena play a fundamental role in many other fields of research, hence their study has benefited from a fruitful interaction between different disciplines such as statistical physics, probability theory, statistics, machine learning, and computer science. Given the vastity of the field, it is clearly impossible to give here a complete overview, so we will rather start the discussion by giving a few examples that will serve as motivations for the rest of the discussion.

* Corresponding author.
E-mail address: francesco.zamponi@ens.fr (F. Zamponi).

https://doi.org/10.1016/j.physa.2022.128152
0378-4371/

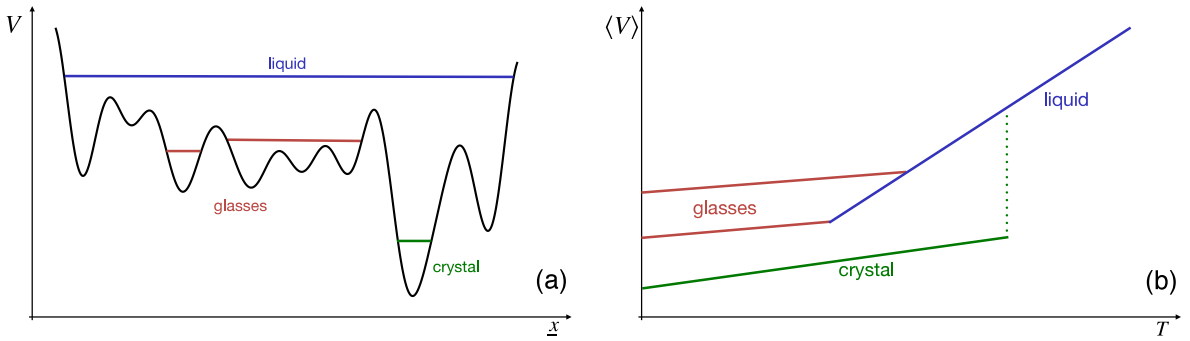

Fig. 1. A. Illustration of the potential energy landscape $V(\underline{x})$ of a structural glass model. B. Preparation of glassy states by cooling. The glass line with lower energy corresponds to slower cooling.

1.1. Glasses and spin-glasses: finding low-energy states

Structural glasses (such as common window glasses) are microscopically formed by a disordered collection of atoms or molecules. As a simple model, we can consider a system of N classical point particles in three dimensions, described by a set of positions $\underline{x} \in \mathbb{R}^{3N}$ with total potential energy $V(\underline{x})$. At low temperatures, the system visits low-energy states. The lowest-energy state is usually a crystal, corresponding to a periodic arrangement of the particles. In glassy systems, the potential energy also features disordered local minima, in which particles' configurations do not display any evident periodicity. Many different microscopic arrangements of the particles then give rise to similar potential energy values: at a given value of $V(\underline{x})$ there are many local minima with similar properties, as illustrated in Fig. 1A.

Glasses are usually prepared by slow cooling from a high-temperature liquid melt. Suppose that the system is initialized in equilibrium at high temperature, in the liquid phase. There, it explores a large portion of the phase space, corresponding to the basin of attraction of many local minima, as illustrated in Fig. 1A. Upon a slow enough cooling, the system maintains thermodynamic equilibrium and a sharp first-order phase transition to the crystalline state is observed, as illustrated in Fig. 1B. However, in glass-forming systems, the nucleation time of the crystal can be astronomically large, and as a result, crystallization is not observed [1]. Upon slow cooling, instead, the system remains confined in the basin of attraction of one of the local minima that compose the original liquid state (Fig. 1A). Hence, during the cooling process, the average potential energy does not jump discontinuously to the crystal value, but instead displays a smooth crossover from the liquid regime to the glassy one (Fig. 1B). Because there are multiple glassy local minima within the liquid basin, different thermal histories (e.g. different temperature cooling rates) can bring the system to slightly different glassy states, hence leading to different low-temperature values of the potential energy. Typically, slower cooling leads to lower values of potential energy (Fig. 1B) [1].

Hence, the problem of describing structural glasses corresponds, mathematically, to the problem of characterizing low-energy local minima of the potential energy $V(\underline{x})$ [2]. Similar problems arise in other physical systems such as spin glasses, which are magnetic systems with impurities. In this case, the system is described by a N-spin configuration $\underline{\sigma}$ and an associated magnetic Hamiltonian $H(\underline{\sigma})$ [3]. Note that, both for glasses and spin glasses, the existence of disordered local minima is independent of the presence of explicit randomness in the potential energy function or Hamiltonian [4].

1.2. The jamming transition: colloids, emulsions, granular materials

A special case of interest is that of particles interacting via a finite-range pair potential, i.e. a potential such that there is no interaction when particles are sufficiently well separated. A specific common example is that of *harmonic soft spheres*,

$$V(\underline{x}) = \sum_{i<j} v(|x_i - x_j|) \,, \qquad v(r) = \varepsilon(1 - r/\ell)^2 \Theta(\ell - r) \,, \tag{1}$$

where $\Theta(x)$ is the Heaviside theta function, and therefore $v(r) = 0$ if particles do not overlap, i.e. $|x_i - x_j| > \ell$, being ℓ the particle diameter. Such a potential has multiple local disordered minima, as illustrated in Fig. 1A, but on top of that, a new phenomenon emerges due to its finite range. In fact, local minima of the potential (or $T = 0$ athermal states) can be separated in two classes. Those with $V = 0$ are such that all the overlaps between particles are removed (Fig. 2A). As a consequence, the structure is mechanically floppy, because of the absence of interactions. These states are called *unjammed*. On the contrary, minima with $V > 0$ have at least two particles overlapping, and in the generic case there is a full network of particle contacts (Fig. 2B) that provides mechanical stability (or rigidity) to the structure. These states are called *jammed*. Remarkably, a sharp *jamming phase transition* [5,6] separates a low-density region where unjammed minima are found with high probability (going to one in the thermodynamic limit) from a high-density region where jammed minima are found with high probability.

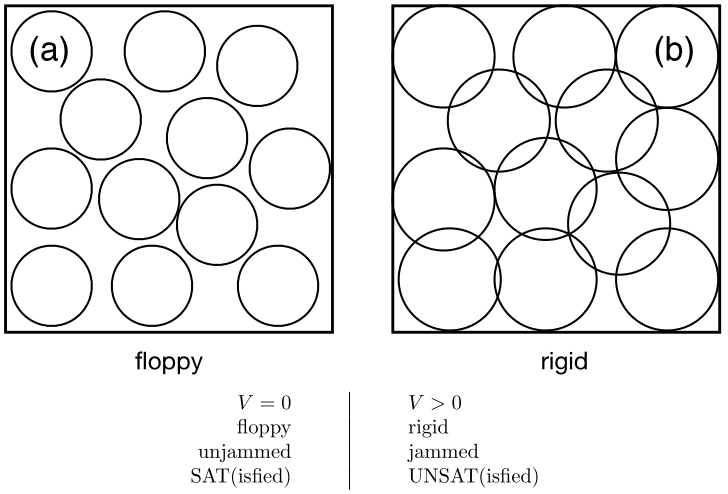

Fig. 2. The jamming transition from a floppy (A) to a rigid (B) state.

The potential in Eq. (1) has been used as a simple model in soft matter physics, to describe soft colloids, emulsions, and soft granular materials [7,8]. The jamming transition found application to these materials, because it describes the onset of rigidity upon athermal compression.

The problem of finding unjammed minima is also a *packing problem* [9,10] that can be formulated as follows: is it possible to find $\underline{x}$ such that $V(\underline{x}) = 0$, i.e. for which there is no overlap between particles? In other words, can a set of N hard spheres of diameter ℓ, which cannot overlap, be packed in a periodic box Ω of volume V? As such, it has a deep connection with many branches of mathematics and coding theory, because the problem of packing spheres (especially in high spatial dimension) is connected to the theory of error correcting codes [11,12].

Finally, this problem can also be seen as a *constraint satisfaction problem*: a configuration space $\underline{x} \in \Omega^N$ is given together with a set of constraints, i.e. $|x_i - x_j| > \ell$, $\forall i, j$, which have to be satisfied. In this language, an unjammed configuration is also called SAT, because all constraints are satisfied, while a jammed configuration is called UNSAT, because some constraints are not satisfied; and the jamming transition corresponds to a SAT-UNSAT transition [13,14]. We will come back to this analogy in Section 1.4 below.

1.3. Optimization problems: finding low-energy configurations

The problem of finding low-energy configurations in a rough energy landscape, which is a central one in the physics of disordered systems (Section 1.1), is also very important in theoretical computer science. A typical example is the so-called *traveling salesman problem* [15,16], which is formulated as follows. One is given a set of cities labeled as $1, \dots, N$, and a cost w_{ij} to go from $i \to j$. One is then asked to find a path $i_1, i_2, \dots, i_N$ that goes once and only once through all the cities (Fig. 3A), and minimizes the total cost of the trip

$$E(i_1, \dots, i_N) = \sum_k w_{i_k i_{k+1}} \ . \tag{2}$$

This is called an *optimization problem*, i.e. one is given a cost function and has to minimize it over a large space of possibilities. Another version, called *decision problem*, is to find whether there exist a path such that the cost is below a given threshold, i.e. $E(i_1, \dots, i_N) < E_0$. Note that the cost w_{ij} can represent for example the distance between city i and j, the price of the train ticket, or any other positive number that is relevant to the problem.

This problem is one of the most intensively studied problems in optimization, and it is obviously relevant for many planning/scheduling problems, but it also find applications in biology and astronomy.[1] Many other optimization problems have been formulated in theoretical computer science, and in all cases the problem is to minimize a given cost function over a space of variables, with a number of configurations typically growing exponentially in the number of variables N, thus making a brute force search impossible.

[1] See e.g. https://en.wikipedia.org/wiki/Travelling_salesman_problem.

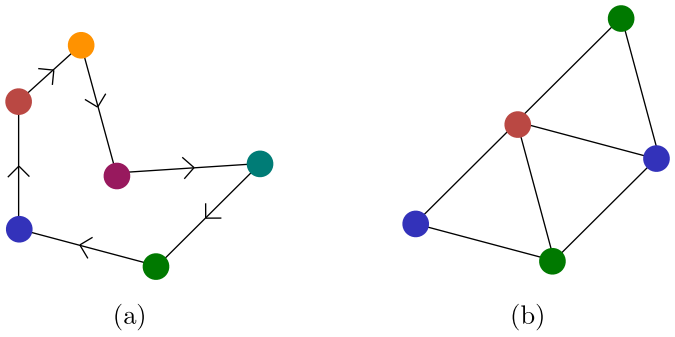

Fig. 3. A. Example of a traveling salesman problem. B. Example of a graph coloring problem.

1.4. Constraint satisfaction problems: finding solutions to a set of constraints

A special class of optimization problems, akin to the problem formulated in Section 1.2, is obtained when the cost function is a sum of local terms involving a finite number of variables, such that the local cost vanishes for certain assignments of the variables and is positive otherwise. An example is that of graph coloring [17] (Fig. 3B). One is given a graph with N nodes and M edges. Each node can be colored by q different colors (e.g. green, red, blue...), represented by a Potts spin $\sigma_i = 1, \dots, q$. The cost function is a sum over edges $\langle i, j \rangle$,

$$E(\underline{\sigma}) = \sum_{\langle i,j \rangle} \delta_{\sigma_i, \sigma_j} \ , \tag{3}$$

which counts the number of monochromatic edges connecting two nodes having the same color. Clearly, the cost function in Eq. (3) is a sum of local terms, each involving two variables, and vanishing whenever the two variables are different and being positive otherwise. Similarly, the potential in Eq. (1) is a sum of pair terms, vanishing when the pair of particles involved is well separated, and being positive otherwise.

One can then formulate two distinct problems. The first is a *satisfiability* problem: is there an assignment $\underline{\sigma}$ of the N variables such that $E(\underline{\sigma}) = 0$, or, in other words, is the problem satisfiable? If the answer is yes, the problem is deemed SAT, and UNSAT otherwise. Like in the setting of Section 1.2, for a broad class of cost functions involving disorder, a sharp SAT-UNSAT phase transition separates the two phases in the thermodynamic limit, upon varying the number of constraints per variable, i.e. the ratio $\alpha = M/N$. Furthermore, around the transition, there is a critical slowing down of search algorithms that look for solutions [13,18–22].

The second problem is an *optimization* (or *approximation*) problem, as in Section 1.3: if there is no solution to the set of constraints, is it possible to minimize the number of unsatisfied constraints, i.e. to find the ground state of $E(\underline{\sigma})$?

1.5. Machine learning: minimize a loss function

Similar problems arise in machine learning applications [23,24], for example in *supervised learning*. An example is data classification. Suppose one is given a set of M input data $\underline{x}_m$ (e.g. images), and a set of associated labels $y_m \in \{0, 1\}$ (e.g. "dog" or "cat"). The problem is to learn, from this set of *training* examples, the unknown function $y_m = f(\underline{x}_m)$ that relates the input to the output. One can then assume a class of functions $y_p = g(\underline{x}|\underline{\theta}) \in [0, 1]$ for $\underline{\theta} \in \mathbb{R}^p$, such that y_p is the probability of the input $\underline{x}$ being labeled as a "1", which depends on a set of unknown parameters $\underline{\theta}$, and define a *loss function* as:

$$L(\underline{\theta}) = \frac{1}{M} \sum_m e\big(y_m, g(\underline{x}_m|\underline{\theta})\big) \ , \tag{4}$$

where $e(y, y_p)$ is a measure of the error of the predicted probability y_p, e.g. the so-called *cross-entropy loss*:

$$e(y, y_p) = -y \log y_p - (1 - y) \log(1 - y_p) \ . \tag{5}$$

For a true label $y = 0$, minimizing the cross-entropy is equivalent to minimize $-\log(1 - y_p)$, i.e. to minimize y_p, and vice versa for $y = 1$. The aim is then to learn the best value of the parameters $\underline{\theta}$ by minimizing the loss function $L(\underline{\theta})$ calculated on the training examples. The training often consists in a gradient descent in the parameter space,

$$\dot{\underline{\theta}} = -\nabla_{\underline{\theta}} L(\underline{\theta}) \ , \tag{6}$$

possibly in the presence of a little additional noise.

Assuming that the function $f(\underline{x})$ is known, together with the statistical distribution of input data, the generalization capacity of the trained machine can be evaluated by computing the *generalization error*:

$$E_g(\underline{\theta}) = \mathbb{E}_{\underline{x}}\left[|f(\underline{x}) - \Theta[g(\underline{x}|\underline{\theta}) - 1/2]|\right] , \tag{7}$$

where $\Theta(x)$ is the Heaviside theta function. In words, we extract a new data point $\underline{x}$ at random, we compute its true label $y = f(\underline{x})$, and its predicted label $y' = \Theta[g(\underline{x}|\underline{\theta}) - 1/2]$ (i.e. we assign a label "1" if $y_p > 1/2$ and "0" otherwise), and we then compute the average of $|y - y'|$, which is one in case of a mistake and zero otherwise. This quantity then provides the probability that the trained machine commits a mistake when classifying a new data point, extracted from the same statistical distribution as the training data, but that was not used in the training.

In order to characterize the local minima of $L(\underline{\theta})$ that are reached by the gradient descent, one can ask several questions:

- Can we perfectly classify the training data, i.e. reach $L(\underline{\theta}) = 0$? The answer to this question defines a sharp SAT/UNSAT transition, as a function of the ratio $\alpha = M/P$ between the number of data and the number of parameters when both go to infinity, also called the *capacity transition* [25]. It separates a phase with $L = 0$ for low α from a phase with $L > 0$ for high α, as described in Sections 1.2 and 1.4.
- How does the generalization capacity of the machine depend on the dynamics (e.g. by the discretization of the gradient descent, by the initial condition, …) and on the ratio M/P [26,27]?
- What is the role of a little additional noise (e.g. in the so-called stochastic gradient descent) [28,29]?

The analytical and numerical study of the training dynamics, both using simple toy models of data (e.g. random uncorrelated data) and real datasets, has given insight on the functioning of the resulting machines.

1.6. Inference problems: maximize the likelihood of data

In inference problems, one is given a set of data, and has to deduce some information about the underlying statistical model from which the data were generated, see e.g. [30–34]. Many inference problems can be framed as *unsupervised learning* problems. The data $\{\underline{x}_m\}_{m=1,\dots,M}$ are independently and identically generated from an unknown probability distribution $P(\underline{x})$. One then assumes a model probability distribution $Q_{\underline{\theta}}(\underline{x})$, with unknown parameters $\underline{\theta}$. The *likelihood* of the data, i.e. the log-probability that the data were generated by $Q_{\underline{\theta}}(\underline{x})$, is defined as

$$\mathcal{L}(\underline{\theta}) = \frac{1}{M}\sum_m \log Q_{\underline{\theta}}(\underline{x}_m) . \tag{8}$$

Note that the likelihood also coincides with minus the cross-entropy $S_c[P_{\text{emp}}, Q_{\underline{\theta}}]$ of the empirical distribution $P_{\text{emp}}(\underline{x}) = M^{-1}\sum_m \delta_{\underline{x},\underline{x}_m}$ and the trial distribution $Q_{\underline{\theta}}(\underline{x})$. One can then seek for the value of $\underline{\theta}$ that maximizes the likelihood, or equivalently minimizes the cross-entropy, once again via a gradient descent:

$$\dot{\underline{\theta}} = \nabla_{\underline{\theta}}\mathcal{L}(\underline{\theta}) = -\nabla_{\underline{\theta}} S_c[P_{\text{emp}}, Q_{\underline{\theta}}] . \tag{9}$$

This intuitive procedure can also be justified more formally, either via the *maximum entropy principle* [30] or in a Bayesian setting with uniform prior [34].

One can then ask several questions to characterize the quality of the inference, such as:

- If $P(\underline{x}) = Q_{\underline{\theta}^*}(\underline{x})$, can one reconstruct the true $\underline{\theta}^*$ via gradient ascent on the likelihood [35–38]? If noise is added to the measurements, what kind of information can one obtain on $\underline{\theta}^*$ [39]?
- Whatever the form of the unknown $P(\underline{x})$, are samples from the inferred $Q_{\underline{\theta}}(\underline{x})$ statistically identical to samples from $P(\underline{x})$? If the answer if yes, then $Q_{\underline{\theta}}(\underline{x})$ is called a *generative model* [40].
- Can one use the learned model $Q_{\underline{\theta}}(\underline{x})$ to infer hidden structure in the data, e.g. to cluster the data or to project them on lower-dimensional manifolds (i.e. lower-dimensional *representations* or *features*)? Can one use this knowledge to generate data with desired properties [41]?

1.7. Simulated annealing: a general algorithm for optimization

We have seen by this sequence of examples that the optimization of multi-dimensional functions is a highly interdisciplinary problem, which finds application in very diverse fields, ranging from soft matter to inference. A very powerful algorithm to search for a global minimum is simulated annealing [42], that was introduced by statistical physicists in the context of optimization problems (Section 1.3). In this algorithm, a random initial configuration is chosen (corresponding, in statistical physics, to an initial state at temperature $T = \infty$) and the system is then cooled, with different cooling protocols, to zero temperature (Fig. 4). This process mimics in the computer the physical cooling process used to created structural glasses (Fig. 1B). Simulated annealing is one of the most widely used algorithms for optimization.

Yet, in many of the applications described above, the function to be optimized is *rough*, i.e. it displays multiple local minima at different levels, separated by saddle points and local maxima, as in Fig. 1A. Because local minima can trap

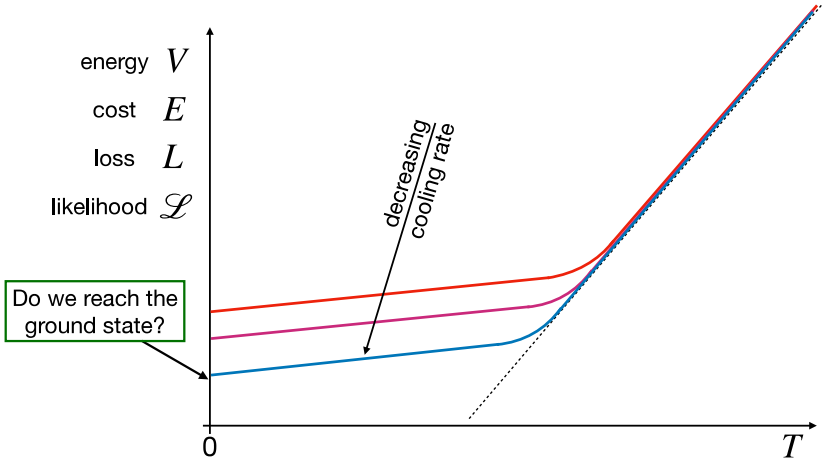

Fig. 4. Simulated annealing, i.e. exploring the energy landscape by slowly decreasing the temperature.

simulated annealing for extremely long times (infinite for gradient descent), and high-energy minima are likely to be more numerous than low-energy ones, one might wonder what are the performances of this algorithm in these complex situations. Does the algorithm get stuck in high-energy, bad minima? Is it able to reach the ground state? Or is it able to find a good compromise, i.e. low enough states that provide good approximation to the ground state?

Statistical physics (together with other disciplines) has made much progress on this question, see e.g. [26,28,29,35,43–47]. In these introductory notes we will consider the two simplest toy models of rough function, that give some insight on the problem:

1. the mixed p-spin model;
2. the perceptron model.

Moreover, while in simulated annealing the cooling protocol can be rather arbitrary, we will focus on some limit cases that can be more easily described analytically:

1. full equilibrium (very slow cooling);
2. gradient descent (very fast cooling);
3. equilibrium down to some temperature T, followed by gradient descent (mixed dynamics).

We will see how the performances of these annealing protocols can be studied in the two toy models. The central theme of the notes will be the relation between the function geometry (i.e., how the local minima and the saddle points are organized) and the long-time properties of the dynamics.

2. Definition and properties

In the following, we will denote by $\underline{\sigma} \in \mathbb{R}^N$ the system's configuration, because we will be dealing with continuous spin systems, and by $H(\underline{\sigma})$ the energy, or cost, or loss, or minus likelihood, function, which we will call *Hamiltonian* in the language of physics. The equilibrium Boltzmann–Gibbs distribution of a system in contact with a heat bath at temperature $T = \frac{1}{\beta}$ is

$$P_{eq}(\underline{\sigma}) = \frac{e^{-\beta H(\underline{\sigma})}}{Z} , \qquad Z = \int d\underline{\sigma}\, e^{-\beta H(\underline{\sigma})} , \tag{10}$$

being Z the partition function.

2.1. Langevin dynamics

We consider the simplest model for the dynamics of a system in contact with a heat bath, the *over-damped Langevin dynamics* [48,49]:

$$\frac{d\sigma_i}{dt} = -\frac{\partial H}{\partial \sigma_i} + \xi_i(t) , \qquad i = 1, \dots, N , \tag{11}$$

where $\xi_i(t)$ is a white Gaussian noise that models the thermal bath, with mean and variance given by

$$\langle \xi_i(t) \rangle = 0 \ , \qquad \langle \xi_i(t)\, \xi_j(t') \rangle = 2T\delta_{ij}\delta(t-t') \ . \tag{12}$$

The evolution of the probability distribution is described by the Fokker–Planck equation,

$$\frac{dP(\underline{\sigma},t)}{dt} = \sum_i \frac{\partial}{\partial \sigma_i}\Big(\frac{\partial H}{\partial \sigma_i} + T\frac{\partial}{\partial \sigma_i}\Big)P(\underline{\sigma},t) = -\mathcal{L}_{\rm FP}P(\underline{\sigma},t) \ . \tag{13}$$

This equation conserves the total probability, i.e. $\frac{d}{dt}\int d\underline{\sigma} P(\underline{\sigma},t) = 0$, and admits the equilibrium distribution $P_{eq}(\underline{\sigma})$ as fixed point. It is easy to check that

$$0 = -\mathcal{L}_{\rm FP}P_{eq}(\underline{\sigma}) \ , \tag{14}$$

which implies that if at any time the system is in $P(\underline{\sigma},t) = P_{eq}(\underline{\sigma})$, then it stays there at all times, because the time derivative of $P(\underline{\sigma},t)$ then vanishes.

Because $\mathcal{L}_{\rm FP}$ is not Hermitian, its left and right eigenmodes associated to an eigenvalue λ_α, respectively Q_α and P_α, differ. For finite N and under a few assumptions, physically needed to ensure that the Hamiltonian is confining at infinity and to exclude the existence of disconnected regions of phase space (see Ref. [49] for details), the operator $\mathcal{L}_{\rm FP}$ has a unique ground state corresponding to the equilibrium distribution,

$$\lambda_0 = 0 \ , \qquad P_0 = P_{eq} \ , \qquad Q_0 = 1 \ , \tag{15}$$

and presents a discrete spectrum of strictly positive excited states with eigenvalues $\lambda_\alpha > 0$. The dynamics starting from a generic initial distribution $P_{in}(\underline{\sigma})$ can then be decomposed on eigenmodes of $\mathcal{L}_{\rm FP}$,

$$P(\underline{\sigma},t) = e^{-\mathcal{L}_{\rm FP}t}P_{in}(\underline{\sigma}) = \sum_\alpha e^{-\lambda_\alpha t}P_\alpha(\underline{\sigma})\langle Q_\alpha | P_{in}\rangle \ . \tag{16}$$

The λ_α then physically correspond to decay rates, and the second eigenvalue λ_1 defines the rate of convergence to equilibrium with a corresponding time scale $\tau_{rel} = 1/\lambda_1$.

2.2. Observables

Given generic observables $A(\underline{\sigma})$ and $B(\underline{\sigma})$, we define the dynamical average as

$$\langle A(t) \rangle \equiv \langle A(\underline{\sigma}(t)) \rangle = \int \mathrm{d}\underline{\sigma} P(\underline{\sigma},t) A(\underline{\sigma}) \ , \tag{17}$$

and the correlation between observables at two different times as

$$C_{AB}(t_w + t, t_w) \equiv \langle A(t_w + t) B(t_w) \rangle \ . \tag{18}$$

Note that the dynamical average encodes an average over both the initial condition P_{in} and over different stochastic paths generated by the realizations of the noise $\xi_i(t)$ in the Langevin Eq. (11). In these notes, we do not need to consider the two averages separately.

If we perturb the Hamiltonian with an external time-dependent field $h(t)$ conjugated to the observable $B(\underline{\sigma})$,

$$H(\underline{\sigma}) \longrightarrow H(\underline{\sigma}) - h(t)B(\underline{\sigma}) \ , \tag{19}$$

we can define the linear response:

$$\langle A(t) \rangle_h = \langle A(t) \rangle_0 + \int_0^t \mathrm{d}s R_{AB}(t,s) h(s) \ . \tag{20}$$

If the perturbation is a pulse at time t_w of amplitude δh, i.e. $h(t) = \delta h\, \delta(t - t_w)$, then the response of the system precisely corresponds to the response function $R_{AB}(t, t_w) = \frac{\delta \langle A(t) \rangle_h}{\delta h}\Theta(t - t_w)$, as illustrated in Fig. 5.

2.3. Special properties of the equilibrium dynamics

Let us consider an initial equilibrium distribution $P_{in}(\underline{\sigma}) = P_{eq}(\underline{\sigma})$, which is then left invariant by the $\mathcal{L}_{\rm FP}$ operator. Physically, if the system is initialized in equilibrium, then it remains in equilibrium at all subsequent times. The invariance of P_{eq} with time has several important implications [48,49]:

- Time-Translation Invariance (TTI): because the origin of time can be shifted arbitrarily leaving the dynamics statistically invariant, two-times observables depend only on the time differences, i.e.

$$C_{AB}(t + t_w, t_w) = C_{AB}(t) \ , \qquad R_{AB}(t + t_w, t_w) = R_{AB}(t) \ . \tag{21}$$

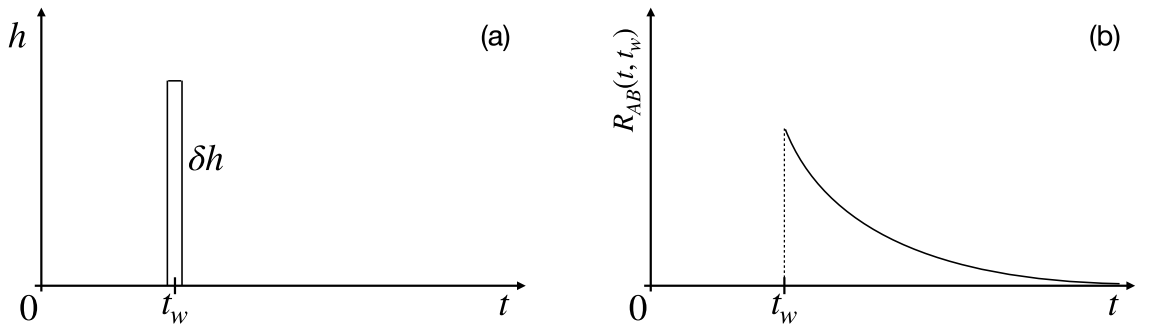

Fig. 5. A. Pulse δh at time t_w. B. Response of the system to the pulse δh at time t_w.

- Onsager Reciprocity: combining the time reversal symmetry, which implies $C_{AB}(t) = C_{AB}(-t)$ (because the Langevin dynamics is statistically reversible), and TTI, which implies $C_{AB}(t) = \langle A(t)B(0)\rangle = \langle A(0)B(-t)\rangle = C_{BA}(-t)$, we have

$$C_{AB}(t) = C_{BA}(t) \ . \tag{22}$$

  This symmetry has important implications for transport coefficients.
- Fluctuation–Dissipation Theorem (FDT): the response and correlation functions are related by

$$R_{AB}(t) = -\frac{1}{T}\Theta(t)\frac{d}{dt}C_{AB}(t) \ , \tag{23}$$

  where $\Theta(t)$ is the Heaviside function.
- Decorrelation (at finite N): because of the discreteness of the spectrum of the evolution operator (Section 2.1), the connected correlation and response functions decay exponentially at long times,

$$\langle A(t)B(0)\rangle_{eq} - \langle A\rangle_{eq}\langle B\rangle_{eq} \sim e^{-t/\tau_{rel}} \ , \qquad R_{AB}(t) \propto \frac{d}{dt}C_{AB}(t) \sim e^{-t/\tau_{rel}} \ , \tag{24}$$

  as illustrated in Fig. 5B for the response function.

As we will see in Section 3, the relaxation time τ_{rel} can however diverge in the thermodynamic limit, leading to ergodicity breaking at a dynamical phase transition.

3. Equilibrium dynamics in the thermodynamic limit: the *p*-spin spherical model

A major breakthrough in glass physics was the recognition by Kirkpatrick, Thirumalai and Wolynes that simple mean field spin glass models share the basic phenomenology of real structural glasses [50–53]. In particular, they identified the presence and role of two distinct phase transitions: a thermodynamic phase transition happening at T_K (already identified in Refs. [54–57]) and a new dynamical phase transition at a higher temperature $T_{MCT} > T_K$, below which equilibrium dynamics is fully arrested in mean-field theory and activated in finite dimensional systems [58]. These results, originally derived using Ising and Potts spin glass models, can be derived in a technically much simpler form using the continuous spherical p-spin model [59,60]. We give an overview in this section, see Ref. [3] for a more complete introduction.

The pure p-spin spherical model is defined by the Hamiltonian:

$$H_p(\underline{\sigma}) = -\sum_{i_1<i_2<\cdots<i_p} J^{(p)}_{i_1\cdots i_p}\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_p} \ , \tag{25}$$

where the spins $\sigma_i \in \mathbb{R}$ with $i = 1, \ldots, N$ are constrained on a sphere $\sum_i \sigma_i^2 = N$. This Hamiltonian contains two kind of objects. The σ_i are dynamical variables that evolve via the Langevin equation. The J parameters are instead fixed at the beginning, and characterize a given *instance* or realization of the Hamiltonian, which is then a random function. Because they are fixed in time and disordered, they go under the name of *quenched disorder*. A simple choice of couplings $J^{(p)}$ is to consider i.i.d. Gaussian variables with mean $\overline{J} = 0$ and variance $\overline{J^2} = \frac{p!}{2N^{p-1}}$; from now on, an overline denotes the quenched average over the random couplings. A mixed p-spin is given by a mixture of pure p-spins with independent couplings:

$$H(\underline{\sigma}) = \sum_p \sqrt{\alpha_p}H_p(\underline{\sigma}) \ . \tag{26}$$

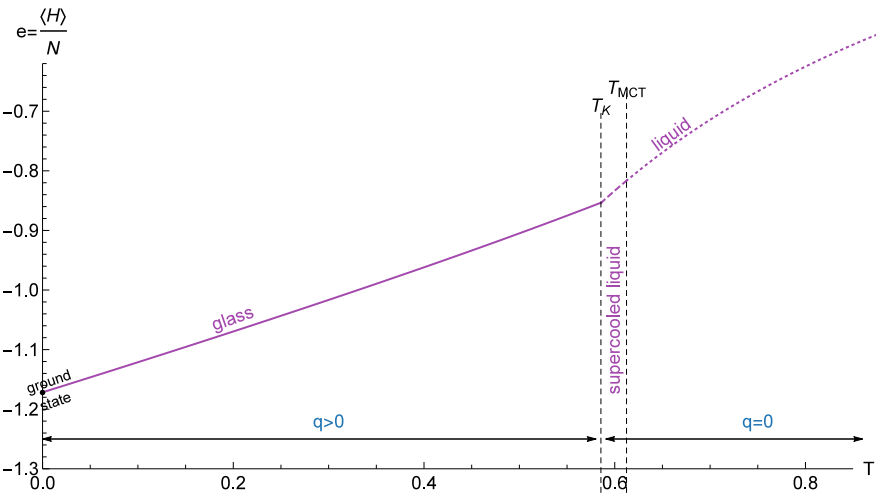

Fig. 6. Equilibrium energy in the 3-spin model.

For fixed $\underline{\sigma}$, $H[\underline{\sigma}]$ is a random Gaussian variable with average and covariance given by

$$\overline{H[\underline{\sigma}]} = 0 \ , \qquad \overline{H[\underline{\sigma}]H[\underline{\tau}]} = Nf(q_{\underline{\sigma}\,\underline{\tau}}) \qquad \text{with} \quad f(q) = \frac{1}{2}\sum_p \alpha_p q^p \ , \tag{27}$$

where $q_{\underline{\sigma}\,\underline{\tau}} = \underline{\sigma}\cdot\underline{\tau}/N = \hat{\sigma}\cdot\hat{\tau}$ is the overlap (or scalar product) between different configurations (Fig. 8A). Energies for different configurations $\underline{\sigma}, \underline{\tau}$ are thus correlated, but their correlation depends only on the overlap between the configurations. The function $f(q)$ is the polynomial that uniquely defines each specific mixed p-spin model. By means of $f(q)$ it is possible to define different classes of models, which correspond to different kinds of rough landscapes with ergodicity breaking at low temperature (see Ref. [61] for more details).

As in Section 2, we consider the simplest equilibrium dynamics, the overdamped Langevin equation

$$\partial_t \sigma_i = -\mu\sigma_i - \frac{\partial H}{\partial \sigma_i} + \xi_i \ , \tag{28}$$

where ξ_i is the thermal noise with zero mean and white correlation $\langle \xi_i(t)\xi_j(t')\rangle = 2T\delta_{ij}\delta(t-t')$, and the term $-\mu\sigma_i$ is added to enforce the spherical constraint on the spins [48]. We will discuss later on how the parameter μ is determined. This mixed p-spin spherical model with Langevin dynamics is the simplest toy model of rough energy landscape, and it is exactly solvable in the thermodynamic limit $N\to\infty$, both for the thermodynamics (via replicas) and for the dynamics (via dynamical mean field theory, or DMFT) [43,48,52,59,60,62,63]. A very pedagogical review on its solution is Ref. [3]. Several results obtained via these methods for the spherical p-spin have been rigorously confirmed [64–69].

3.1. Equilibrium phase diagram

A rather peculiar case is given by the pure $p = 2$ spherical model, for which the Hamiltonian $H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j$ is quadratic in the spins $\underline{\sigma}$. This model presents a phenomenology that differs, both thermodynamically and dynamically, from models with $p > 2$. This is a result of the fact the GOE matrix J_{ij} can be diagonalized, leading to a decoupling of the degrees of freedom in the diagonal basis, $H = -\sum_\alpha \lambda_\alpha \sigma_\alpha^2$. The thermodynamics of the system is characterized by a condensation transition at a critical temperature T_c, below which a finite component of the Gibbs measure is concentrated in the lowest eigenvalue λ_0, and the system enters a "disordered ferromagnetic" phase, where a spontaneous magnetization along the lowest eigenvalue of the matrix J_{ij} is present [70]. The Newtonian dynamics of the system is fully integrable, with N integrals of motion [71]. Because the energy landscape is convex, hence not rough, and the model is integrable, we do not consider it further in these notes.

We will thus be mostly interested in the equilibrium behavior that is generically observed in all pure p-spin models with $p \geq 3$, and in most cases of mixed models, as illustrated in Fig. 6. A critical temperature T_K (Kauzmann temperature) then separates two different thermodynamic phases.

- **Paramagnetic phase** ($T > T_K$): The system is paramagnetic, hence the local magnetization is $\langle\sigma_i\rangle = 0$ for all spins. Furthermore, the average overlap of two configurations that are drawn independently from the same Boltzmann–Gibbs measure (i.e. with the same couplings, corresponding to the same physical system) is

$$q = \overline{\langle\underline{\sigma}\cdot\underline{\tau}\rangle}/N = \frac{1}{N}\sum_i \overline{\langle\sigma_i\tau_i\rangle} = \frac{1}{N}\sum_i \overline{\langle\sigma_i\rangle\langle\tau_i\rangle} = \frac{1}{N}\sum_i \overline{\langle\sigma_i\rangle^2} = 0 \ . \tag{29}$$

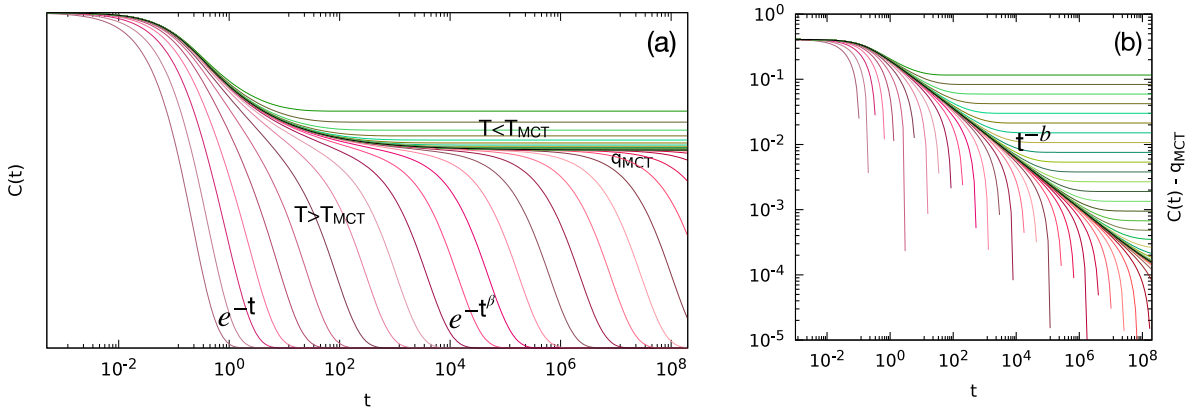

Fig. 7. A. Equilibrium overlap correlation function in the (3+4)-model at several temperatures. If $\lim_{t\to\infty} C(t) = 0$ the system is an ergodic liquid (red lines), while for $\lim_{t\to\infty} C(t) = q > 0$ it is in a dynamically arrested supercooled liquid phase (green lines). The relaxation time scale τ diverges upon approaching $T_{\rm MCT}$ from above. B. A log–log representation of $C(t) - q_{\rm MCT}$ versus time shows the power-law approach to the plateau from above and below $T_{\rm MCT}$.
Source: Plots adapted from Ref. [61].

We conclude that two typical equilibrium configurations are orthogonal on the sphere.

- **Spin glass phase** ($T < T_{\rm K}$): Few lowest-energy glassy states dominate the Gibbs measure, leading to a spin glass phase with $\langle\sigma_i\rangle \neq 0$. Following the previous reasoning, we now obtain $q = \sum_i \overline{\langle\sigma_i\rangle^2}/N > 0$.

In the language of structural glasses, the paramagnetic phase corresponds to the liquid phase, and the spin glass phase to the glass phase. The transition happening at $T_{\rm K}$ is called a *Random First Order Transition* and it has mixed character [53–57]: the transition is thermodynamically of second order (e.g. the energy is continuous at $T_{\rm K}$ so there is no latent heat), but $\langle q\rangle$ jumps from zero to a finite value at $T_{\rm K}$, hence the order parameter is discontinuous as in a first order transition. The low-temperature phase for $T < T_{\rm K}$ is also characterized by replica symmetry breaking [3,53–57,72].

3.2. Equilibrium dynamics

We now consider more carefully the equilibrium dynamics of the model. The system is initialized at time $t = 0$ in an equilibrium configuration, drawn from the Boltzmann–Gibbs measure at temperature T, and Langevin dynamics is run at the same temperature T, so the system remains in equilibrium at all times and enjoys the special properties discussed in Section 2.3. We consider the time-dependent overlap correlation function $C(t, t')$,

$$C(t,t') = \frac{1}{N}\sum_i \overline{\langle\sigma_i(t)\sigma_i(t')\rangle} = \overline{\langle\hat\sigma(t)\cdot\hat\sigma(t')\rangle}\ , \tag{30}$$

in the limit $N \to \infty$ at finite times t, t'. Because of TTI, we have $C(t + t_w, t_w) = C(t)$, using the notation of Eq. (21). This function shows several different regimes upon varying temperature, corresponding to three distinct dynamical phases.

3.2.1. $T > T_{MCT}$

At high temperature, one observes an exponential decay of $C(t) \sim e^{-(t/\tau)^\beta}$ at long times, with relaxation time τ, and possibly a non-trivial exponent β called *stretching exponent*, as illustrated in Fig. 7A. The stretching exponent decreases by approaching $T_{\rm MCT}$ from $\beta = 1$ to a smaller value $\beta \approx 0.9$; this effect is observed also in numerical simulations of real glasses [73]. Hence, the dynamics is *ergodic*. At long times $t \gg \tau$, ergodicity implies $\langle\sigma_i(t)\sigma_i(0)\rangle \sim \langle\sigma_i\rangle_{eq}\langle\sigma_i\rangle_{eq}$, see Eq. (24). Hence the overlap between the state at $t = 0$ and that at time t converges to the typical overlap of two independent equilibrium configurations, i.e. to the thermodynamic value $q = 0$ (Fig. 8B). Upon cooling, the typical relaxation time τ increases, and it diverges as a power-law upon approaching a critical temperature, i.e. $\tau = |T - T_{\rm MCT}|^{-\gamma}$ when $T \to T_{\rm MCT}^+$. At $T = T_{\rm MCT}$, the correlation function does not decay to zero anymore, but it reaches a finite limit, $\lim_{t\to\infty} C(t) = q_{\rm MCT}$. The values of $T_{\rm MCT}$ and $q_{\rm MCT}$ are obtained by finding the maximal temperature such that the equation

$$\beta^2 f'(q) = q/(1-q) \tag{31}$$

has a solution $q > 0$. This equation can be derived following two paths. Either dynamically, by closing the equilibrium DMFT equation in the long-time limit [60], or by a replica calculation that evaluates the free energy of the metastable

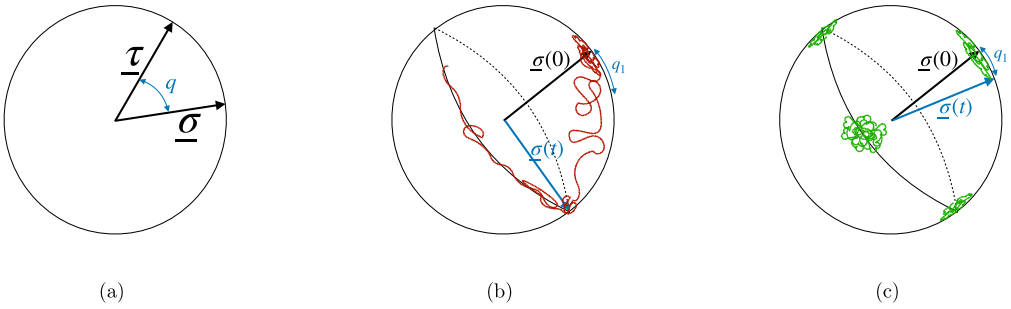

Fig. 8. A. Overlap q between two distinct configurations $\underline{\tau}$ and $\underline{\sigma}$. B. Relaxation in the liquid phase. C. States in the non-ergodic phase.

states [74,75]. See Refs. [3,61] for pedagogical reviews. The critical scaling of $C(t)$ upon approaching T_{MCT} from above is characterized by three time regimes (Fig. 7B):

$$\begin{aligned} C(t) - q_{\text{MCT}} &\propto t^{-b} , \qquad t \ll \tau , \\ C(t) - q_{\text{MCT}} &\propto -t^{a} , \qquad t \lesssim \tau , \\ C(t) &\propto \exp(-t/\tau) \qquad t \gg \tau , \end{aligned} \tag{32}$$

and the three exponents γ, a, b are related by $\lambda_{\text{MCT}} = \frac{\Gamma(1-a)^2}{\Gamma(1-2a)} = \frac{\Gamma(1-b)^2}{\Gamma(1-2b)}$ and $\gamma = \frac{1}{2a}$, with a non-universal parameter

$$\lambda_{\text{MCT}} = \frac{f'''(q_{\text{MCT}}) f'(q_{\text{MCT}})}{2 q_{\text{MCT}} f''(q_{\text{MCT}})^2} . \tag{33}$$

In the special case of pure models, $f(q) \propto q^p$, one has $q_{\text{MCT}} = (p-2)/(p-1)$, and λ_{MCT} is then identically equal to 0.5. In summary, for T slightly higher than T_{MCT}, a typical dynamical trajectory starting from an equilibrium state $\underline{\sigma}(0)$ remains close to the initial state for a long time, of the order of τ, and then relaxes away from it towards another typical equilibrium state, which is then orthogonal to $\underline{\sigma}(0)$ (Fig. 8B). At T_{MCT} and below, this relaxation process is frozen and the trajectory remains forever confined in the vicinity of the initial state, without reaching equilibrium. The critical behavior described above had been previously described within Mode-Coupling Theory, an approximate theory of the glass transition [76]. This analogy was crucial to realize that the p-spin model provides a mean-field theory of the glass transition [50,51].

3.2.2. $T_K < T < T_{MCT}$

Remarkably, it is found that $T_{\text{MCT}} > T_{\text{K}}$, so the critical divergence of the relaxation time happens at a temperature strictly higher than the thermodynamic transition [51,59,60]. For $T_{\text{K}} < T < T_{\text{MCT}}$, the dynamics is trapped around the initial state for an infinite time: the overlap correlation reaches a finite plateau, $\lim_{t\to\infty} C(t) = q_1$, with q_1 given by the largest solution of Eq. (31). In other words, a typical equilibrium trajectory remains confined in a cone of scalar product $\hat{\sigma}(t)\cdot\hat{\sigma}(0) \le q_1$ around the initial condition (Fig. 8C). But yet, we know that the system is thermodynamically a paramagnet, hence typical pairs of equilibrium configurations have overlap $q_0 = 0$. We conclude that each equilibrium configuration is surrounded by a cone of size $\approx 1 - q_1$, that is explored by the dynamics starting in that configuration, but that there are many such cones, each one corresponding to a distinct independent initial equilibrium state, so that typical pairs of cones are orthogonal on the sphere. Each of these cones defines a metastable spin-glass (SG) state. Because of these metastable states, the dynamics is *not ergodic*: the long-time limit of a typical dynamical trajectories does not reach equilibrium, because it is unable to jump out of the cone defined by the initial state. More precisely, the time to escape from the initial metastable state scales exponentially with the system size, $\tau \approx e^N$, in this regime. Hence, if the limit $N \to \infty$ is taken first, no relaxation is observed on finite times; while for finite N systems, some relaxation can be observed, but over extremely long time scales (Fig. 10B).

It is possible to show that in this regime there is an exponential number in N of equivalent SG states, each having a typical size $1 - q_1 > 0$ and being orthogonal, i.e. with $q_0 = 0$, to all other states [3,77,78]. The finite limit

$$\Sigma = \lim_{N\to\infty} \frac{1}{N} \log \#\text{SG} \tag{34}$$

is called "complexity" or entropy of the SG states, and is illustrated in Fig. 9. In the glass literature, this quantity is also known as configurational entropy [2]. A metastable SG state can be thought, in a first approximation (which is actually exact for the pure p-spin model), as a local minimum of the Hamiltonian "dressed" by thermal fluctuations [3,77,78]. Hence, the presence of an exponential number of SG states indicates an extremely rough Hamiltonian at low energies. In Ref. [61] different methods to evaluate the complexity in the p-spin model are compared.

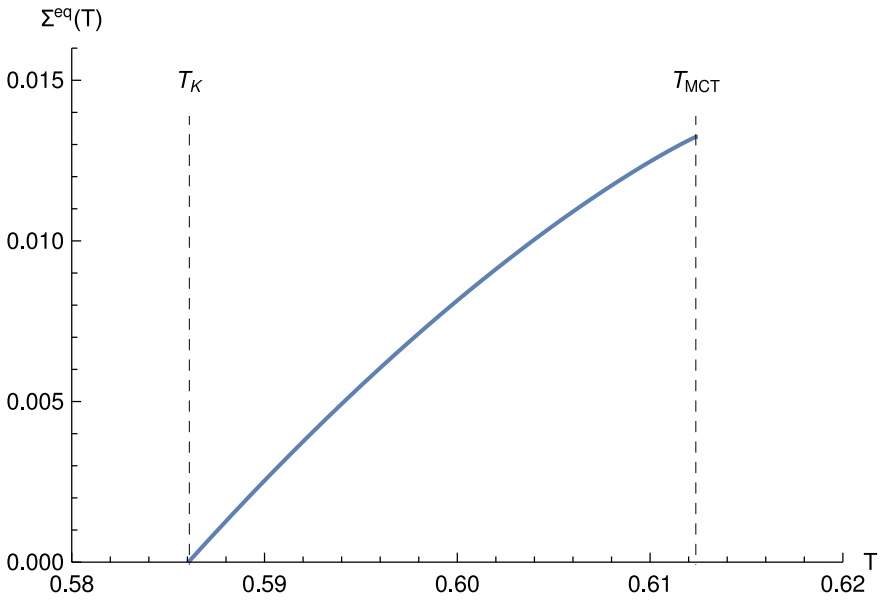

Fig. 9. Equilibrium complexity for the 3-spin model.

Finally, note that while equilibrium configurations are always mathematically well defined, in practice it is impossible to sample from the Boltzmann–Gibbs distribution for $T < T_{\mathrm{MCT}}$ by conventional means. Typically, one would take a finite N system, initialize it in a random (infinite-temperature) state, and run the dynamics at the target temperature for long enough, until equilibration is reached. But because the relaxation time is exponentially large in N, the dynamics takes an astronomically large time to equilibrate unless the system is very small, which makes the sampling practically impossible. A solution to this problem is the so-called *planting trick*, see Refs. [45,61,79] for a detailed discussion.

3.2.3. $T < T_K$

The shape of the equilibrium complexity for a p-spin model is illustrated in Fig. 9. The complexity is only defined for $T < T_{\mathrm{MCT}}$, where metastable states exist, and it is a decreasing function of temperature. This is consistent with intuition: the lower the temperature, the lower the energy, and it is reasonable to expect less local minima of the energy landscape at lower energy. The complexity is found to vanish continuously at T_{K}, indicating that at this temperature, the system in equilibrium can only be found in a sub-exponential number of SG states (which can be shown to be actually finite). Because the complexity cannot be negative (it is the logarithm of the number of states), the SG states at T_{K} are the lowest free energy states, and are thus thermodynamically stable [3]. A true phase transition to an equilibrium spin glass phase then happens at T_{K}. For any $T < T_{\mathrm{K}}$, the complexity remains identically zero and the system is found in the lowest free energy states at each T, which become, when $T \to 0$, the ground states of the Hamiltonian. Throughout this phase, the dynamics is qualitatively similar to that in the regime $T_{\mathrm{K}} < T < T_{\mathrm{MCT}}$: the correlation function relaxes to a finite plateau q_1, whose calculation now requires replica symmetry breaking [80]. From this dynamical point of view, the transition at T_{K} is akin to a first-order phase transition, in which no dynamical slowing down is observed.

4. Out-of-equilibrium dynamics

A second major breakthrough in glass physics was the exact solution by Cugliandolo and Kurchan of the out-of-equilibrium dynamics of the pure spherical p-spin model [43]. This solution established the presence of persistent *aging* in the dynamics of mean-field spin glasses, qualitatively similar to that of real structural glasses. Furthermore, it provided a geometric interpretation of aging in terms of saddles and minima of the rough energy landscape [43,77]. In this section, we review these results.

4.1. Gradient descent from an equilibrated configuration

For simplicity, we will focus on the simplest case of out-of-equilibrium dynamics, namely gradient descent (GD) dynamics from fixed initial temperature T. More precisely, the algorithm goes as follows.

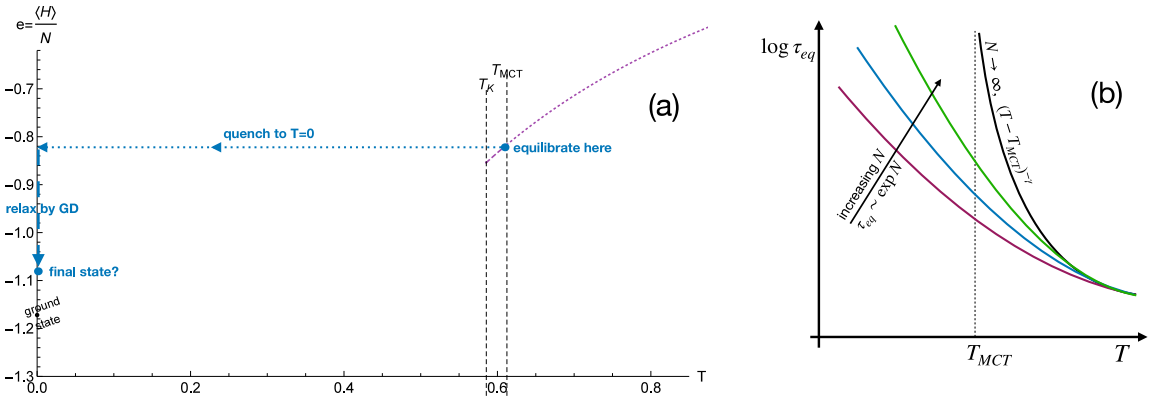

Fig. 10. A. Sketch of the gradient descent protocol. B. Sketch of the equilibrium relaxation time as a function of T and system size N.

(a) Prepare an initial configuration $\underline{\sigma}(0)$ in equilibrium at $T > T_{\rm K}$, i.e. by sampling it from the Boltzmann–Gibbs distribution

$$P_{in}(\underline{\sigma}(0)) = \frac{e^{-\beta H(\underline{\sigma}(0))}}{Z} \ . \tag{35}$$

We restrict the initial temperature to $T > T_{\rm K}$ for simplicity; with this choice, the initial state is in the phase where the system is thermodynamically a paramagnet and the Boltzmann–Gibbs distribution can be studied without the need of replica symmetry breaking [80], see Section 3. Based on the discussion of Section 3, we know that it is algorithmically hard (i.e. exponentially hard in N) to sample from Eq. (35) when $T < T_{\rm MCT}$, but in some models one can use the planting trick [45,79]. Even when this is not possible, this prescription to generate $\underline{\sigma}(0)$ is always mathematically well defined and it can be studied by analytical techniques in the thermodynamic limit [74,80,81].

(b) Run the gradient dynamics at $T = 0$, i.e. the noiseless version of the Langevin Eq. (28), keeping in mind the spherical constraint:

$$\frac{\partial \sigma_i}{\partial t} = -\frac{\partial H}{\partial \sigma_i} - \mu \sigma_i \ . \tag{36}$$

In the $T = 0$ limit, the Lagrange multiplier can be easily computed by imposing the spherical constraint:

$$|\underline{\sigma}|^2 = N \quad \Rightarrow \quad \underline{\sigma} \cdot \frac{\partial \underline{\sigma}}{\partial t} = -\underline{\sigma} \cdot \underline{\nabla} H - \mu N = 0 \quad \Rightarrow \quad \mu = -\frac{1}{N} \underline{\sigma} \cdot \underline{\nabla} H \ . \tag{37}$$

This out-of-equilibrium dynamics, illustrated in Fig. 10A, does not satisfy TTI and FDT, see Section 2.3.

The physical motivation for considering this dynamical protocol is the following. We have seen in Section 3 that the equilibrium relaxation time of the system grows quickly upon cooling. In the thermodynamic limit, it diverges as a power-law at $T_{\rm MCT}$. At finite (large enough) N, the relaxation time is independent of N for $T > T_{\rm MCT}$, it grows strongly around $T_{\rm MCT}$, and for $T < T_{\rm MCT}$ it is exponential in N, as illustrated in Fig. 10B. Consider now the simulated annealing protocol illustrated in Fig. 4, with a fixed cooling rate $\mathrm{d}T/\mathrm{d}t$. Whenever T is such that $\tau_{\rm eq}(T)\mathrm{d}T/\mathrm{d}t \ll T$, the system spends a long time at a given temperature before temperature changes, and as a result it can equilibrate easily. When instead $\tau_{\rm eq}(T)\mathrm{d}T/\mathrm{d}t \gg T$, the temperature is changing so fast compared to the equilibration time that the system is effectively being quenched athermally. Hence, one can approximate the high-T part of a constant cooling schedule by equilibrium dynamics, and the low-T part as a gradient descent dynamics; the delicate regime is when $\tau_{\rm eq}(T)\mathrm{d}T/\mathrm{d}t \approx T$ and the system is falling out of equilibrium, which happens at some cooling-rate-dependent glass transition T_g. Because $\tau_{\rm eq}(T)$ changes very rapidly with T, the crossover regime is a narrow temperature interval around T_g. As a result, the idealized description where the system is fully equilibrated down to T_g and follows zero-temperature gradient descent dynamics from there, is a good approximation of an actual simulated annealing dynamics, but it is easier to solve analytically. This is why we focus on this protocol in the rest of this section.

4.2. Hessian of the final state

The Hessian of the final state is a fundamental quantity in order to understand the long-time GD dynamics [43]. The GD dynamics ends up in a local minimum of the Hamiltonian, with the spherical constraint, hence

$$\lim_{t\to\infty} \underline{\sigma}(t) = \underline{\sigma}_\infty \ , \qquad \text{which is a solution of} \qquad \underline{\nabla} H = -\mu \underline{\sigma} \ . \tag{38}$$

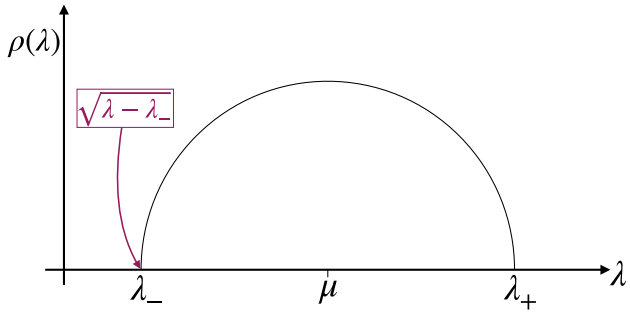

Fig. 11. Hessian spectrum in the p-spin spherical model.

In the long-time regime we can study the asymptotic relaxation. Defining $\delta\underline{\sigma}(t) = \underline{\sigma}(t) - \underline{\sigma}_\infty$, which is small at long times, the GD dynamics, from Eq. (36), can be linearized and gives

$$\frac{\partial \delta\sigma_i}{\delta t} \sim -\sum_j \frac{\partial^2 H}{\partial\sigma_i\partial\sigma_j}\delta\sigma_j - \mu\delta\sigma_i = -\sum_j M_{ij}(\underline{\sigma}_\infty)\delta\sigma_j \,, \qquad M_{ij}(\underline{\sigma}_\infty) = \left[\frac{\partial^2 H}{\partial\sigma_i\partial\sigma_j} + \mu\delta_{ij}\right]_{\underline{\sigma}=\underline{\sigma}_\infty} \,, \tag{39}$$

where $M_{ij}(\underline{\sigma}_\infty)$ is the asymptotic Hessian of the GD dynamics. Decomposing it in normal modes, the asymptotic GD dynamics becomes

$$M_{ij} = \sum_\alpha \lambda_\alpha |\underline{\sigma}_\alpha\rangle\langle\underline{\sigma}_\alpha| \qquad \Rightarrow \qquad |\delta\underline{\sigma}(t)\rangle = \sum_\alpha e^{-\lambda_\alpha t}|\underline{\sigma}_\alpha\rangle\langle\underline{\sigma}_\alpha|\delta\underline{\sigma}(0)\rangle \,, \tag{40}$$

such that the relaxation is dominated by the lowest eigenvalue λ_α of the Hessian. Note, however, that this is only an asymptotic result that holds at times possibly diverging with N, so the gradient descent dynamics is more complex than a simple relaxation along eigenmodes.

From Eq. (39), it can be shown that in the p-spin model the correlation between the Gaussian couplings and the configuration $\underline{\sigma}_\infty$ can be neglected in the thermodynamic limit, see e.g. Ref. [35]. The Hessian of typical stationary points in the energy landscape is thus a shifted GOE matrix, with $\overline{M_{ij}} = \mu\delta_{ij}$ and $\mathrm{Var}[M_{ij}] = \frac{f''(1)}{N}$. Its eigenvalue spectrum thus has a semicircular shape (Fig. 11) [43,77]:

$$\rho(\lambda) = \frac{\sqrt{(\lambda-\mu)^2 - 4f''(1)}}{\pi f''(1)} \,, \tag{41}$$

and the two edges of the spectrum are $\lambda_\pm = \mu \pm 2\sqrt{f''(1)}$. In order to evaluate the Hessian matrix, we thus need to compute the Lagrange multiplier μ from Eq. (37). In the pure model ($H = H_p$), it is proportional to the energy:

$$\mu_p = -\frac{1}{N}\sum_i \sigma_i \sum_{i_2<\cdots<i_p} J_{ii_2\ldots i_p}\sigma_{i_2}\cdots\sigma_{i_p} = -\frac{pH_p}{N} \,. \tag{42}$$

In the mixed model ($H = \sum_p \sqrt{a_p}H_p$) the above proportionality does not hold:

$$\mu = -\frac{1}{N}\sum_p \sqrt{a_p}pH_p \not\propto H \,. \tag{43}$$

Because (in the thermodynamic limit) the typical spectrum is thus only determined by the Lagrange multiplier μ, this difference has important implications in terms of the energy landscape: in the pure p-spin model, the spectrum of a typical stationary point is fully determined by its energy, while this is not the case in the mixed p-spin. We now discuss the implications of this difference in more details. As illustrated in Fig. 11, when $\lambda_- > 0$ the stationary point is a stable minimum with a gapped spectrum of strictly positive eigenvalues; when $\lambda_- < 0$, it is an unstable saddle point with negative eigenvalues; while stationary points with $\lambda_- = 0$ are local minima but have arbitrarily small eigenvalues, hence they are deemed *marginally stable*.

4.3. Pure p-spin

As we have already noticed, from Eq. (42), in pure p-spin models, with $f(q) = q^p/2$, there is a direct proportionality between the spectrum-shift μ and the energy of the minimum $e = H/N$. Recalling that $f''(1) = p(p-1)/2$, the marginality condition $\lambda_- = \mu - 2\sqrt{f''(1)} = 0$ fixes a value for μ and therefore a value for the energy e:

$$\lambda_- = 0 \quad \Longrightarrow \quad \mu_{th} = 2\sqrt{p(p-1)/2} = -pe_{th} \quad \Longrightarrow \quad e_{th} = \sqrt{2(p-1)/p} \,. \tag{44}$$

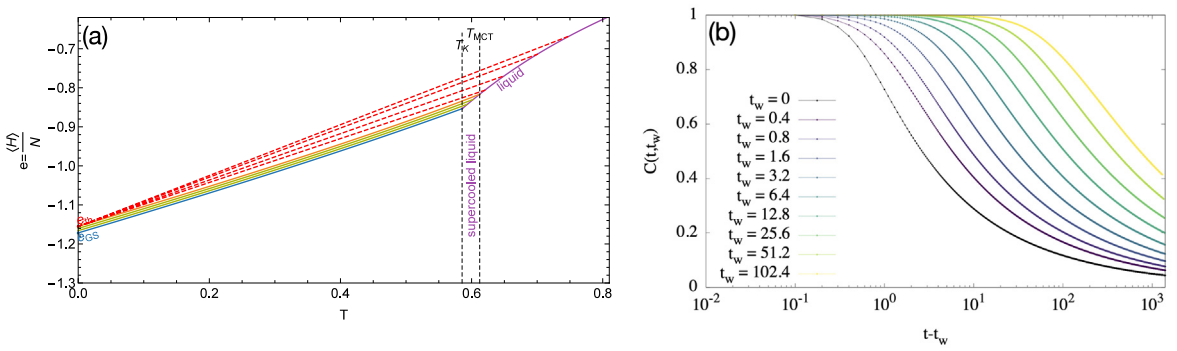

Fig. 12. 3-spin model. A. Representation of the gradient descent dynamics in the energy–temperature plane. An initial state at temperature $T > T_{\mathrm{MCT}}$ is brought to the threshold level (red dashed lines, only the initial and final point are physical). An initial state at $T < T_{\mathrm{MCT}}$ is brought below the threshold (full colored lines, that represent the evolution of the metastable state from the initial to the final temperature). B. Aging of the correlation function $C(t + t_w, t_w)$ for gradient descent dynamics from infinite temperature, as a function of t for several t_w, obtained from solving the DMFT equations corresponding to Eq. (36).
Source: Plots adapted from Ref. [61].

The energy corresponding to marginally stable minima is called the *threshold energy* [43], see also [77,78,82]. Typical stationary points of the energy landscape are stable minima for $e < e_{th}$, and unstable saddles for $e > e_{th}$.

The GD dynamics can be characterized in terms of the time-dependent energy, $e(t) = \langle H(t)\rangle/N$, and the correlation function $C(t, t')$ given by Eq. (30). Two distinct phases, depending on the initial equilibrium temperature $T > T_{\mathrm{K}}$, exist.

- $T < T_{\mathrm{MCT}}$ - If the initial temperature is below T_{MCT}, then one observes [81]:
  (a) an exponential relaxation to a nearby local minimum (also called the "inherent structure" in the structural glass literature [2]), $e(t) - e_{\mathrm{IS}}(T) \propto e^{-t/\tau}$, with a temperature-dependent final energy $e_{\mathrm{IS}}(T)$;
  (b) and a persistent memory of the initial condition, $C(t, 0) \to q_r(T)$ for $t \to \infty$, i.e. the configuration at time t remains in a cone of width $q_r(T)$ around the initial state.

  In this phase, the dynamics can be mapped onto a *restricted thermodynamics* calculation [74]. A thermodynamic calculation of $e_{\mathrm{IS}}(T)$ and $q_r(T)$ can be achieved by constraining a "slave" configuration to have fixed overlap with an equilibrium "master" configuration, and then finding the overlap that minimizes the free energy of the slave configuration, which gives $q_r(T)$. This is called a *Franz-Parisi* [74,81] or *state-following* construction [45,83]. For $T < T_{\mathrm{K}}$ the same mapping is possible but it requires a more complicated replica symmetry breaking scheme [80].
- $T > T_{\mathrm{MCT}}$: if the initial temperature is above T_{MCT}, then one observes [43]:
  (a) a power-law relaxation to the threshold energy, $e(t) - e_{th} \propto t^{-\nu}$, independently of T;
  (b) consistently with the temperature-independence of the final state, memory of the initial condition is lost, i.e. $\lim_{t\to\infty} C(t, 0) = 0$. Furthermore, $\lim_{t\to\infty} C(t + t_w, t_w) = 0$ for any t_w, i.e. memory of any finite time configuration is also lost, as illustrated in Fig. 12B. This condition is called *weak ergodicity breaking* [43,84,85].

In the pure p-spin model it is thus possible to explain the asymptotic (long-time) dynamics in terms of simple geometric properties of the energy landscape [43]. At high energies $e > e_{th}$, typical stationary points are unstable (they present negative eigenvalues of the Hessian). At the threshold energy e_{th}, they become marginally stable minima, and below e_{th} they become stable. The threshold manifold is an attractor for the asymptotic dynamics starting from high temperatures ($T > T_{\mathrm{MCT}}$), independently of the used protocol, while for low temperatures ($T < T_{\mathrm{MCT}}$) one starts in the basin of attraction of a stable minimum, which is then reached quickly by the gradient descent. This geometric transition in the energy landscape is illustrated in Fig. 13A. A more detailed calculation [82] gives the complexity associated to saddles of order one (or higher), that is found to cross that of stable minima precisely at e_{th}, see Fig. 13B. Hence, saddle points also exist below e_{th}, but they are exponentially rarer than stable minima, while the reverse is true above e_{th}.

A simple argument[2] suggests a power-law convergence $\propto t^{-2/3}$ (i.e. $\nu = 2/3$) towards the threshold energy, and goes as follows. The system, while aging towards the threshold, moves close to saddle points with rarefying number of negative eigenvalues of the Hessian. Let us assume that during this process the dynamics induces random rotations of the gradient with respect to the local Hessian eigenvectors. The time to escape from the neighborhood of a saddle is thus proportional to the probability that a random rotation brings the gradient in a negative direction. Given that the shift of the spectrum

2 Private communication from F.Ricci-Tersenghi.

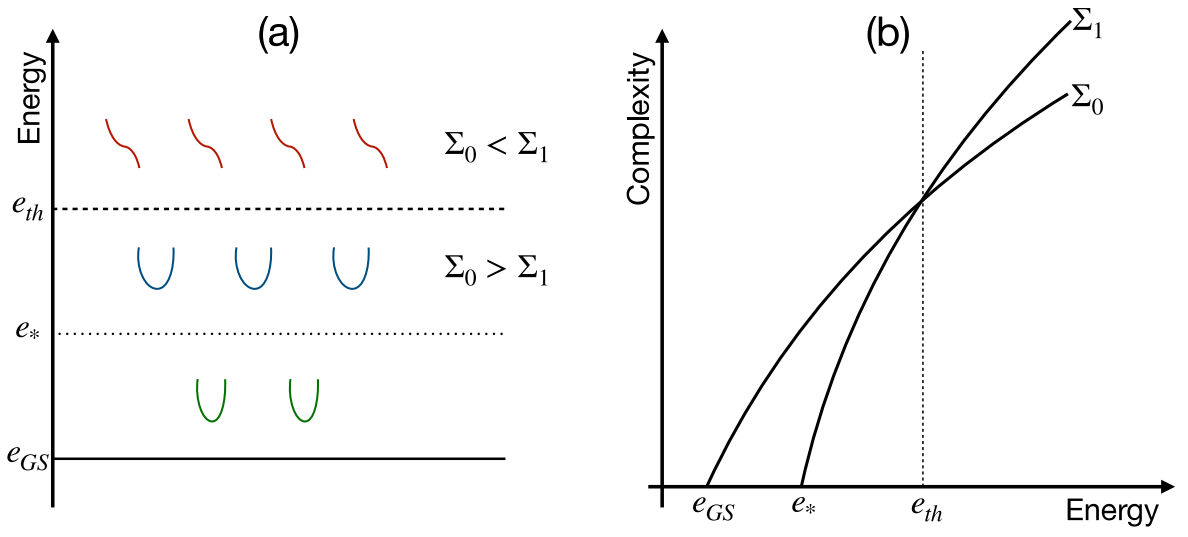

Fig. 13. A. Schematic illustration of the energy landscape in a pure p-spin model [43]. B. Complexity of minima Σ_0 and of saddles of index one Σ_1 (one negative eigenvalue of the Hessian) [82].

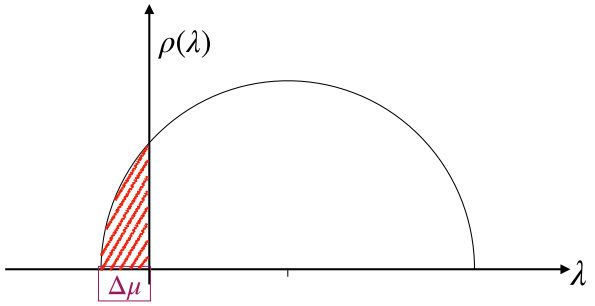

Fig. 14. Hessian spectrum during the gradient descent.

is proportional to the energy (Fig. 14),

$$\lambda_- = \Delta\mu = \mu - \mu_{th} \propto -(e - e_{th}) = -\Delta e \ , \tag{45}$$

and that $\rho(\lambda)$ vanishes as a square root at the edge (Fig. 11), the probability of extracting a negative eigenvalue for small $\Delta\mu$ is:

$$\rho(\lambda < 0) \propto \int_{\Delta\mu}^{0} \mathrm{d}\lambda \sqrt{\lambda - \Delta\mu} \propto |\Delta\mu|^{3/2} \propto \Delta e^{3/2} \ . \tag{46}$$

Therefore the time to escape is $t \propto 1/\rho(\lambda < 0) \propto \Delta e^{-3/2}$, which then gives the power-law scaling $\Delta e \propto t^{-2/3}$. Such a value of the exponent ν seems compatible with recent numerical results [47], but a previous estimate via series expansions [86] predicts a slightly higher value of $\nu \sim 0.755$. Computing the exact value of the exponent ν thus remains an open problem.

4.4. Mixed p-spin

We next consider a mixed p-spin model that presents a RFOT [83], e.g. the (3+4)-spin model, with $f(q) = \frac{1}{2}(q^3 + q^4)$, that was studied in Ref. [47]. The energy and the Lagrange multiplier are:

$$e = e_3 + e_4 \ , \qquad \mu = -3e_3 - 4e_4 \ , \tag{47}$$

where $e_p = \langle H_p \rangle / N$ with $p = 3, 4$. In this case, an exponential number of marginally stable states is found in a finite range of energies. To see this, one can compute (an approximation of) the complexity $\Sigma(e, \mu)$, which gives the number

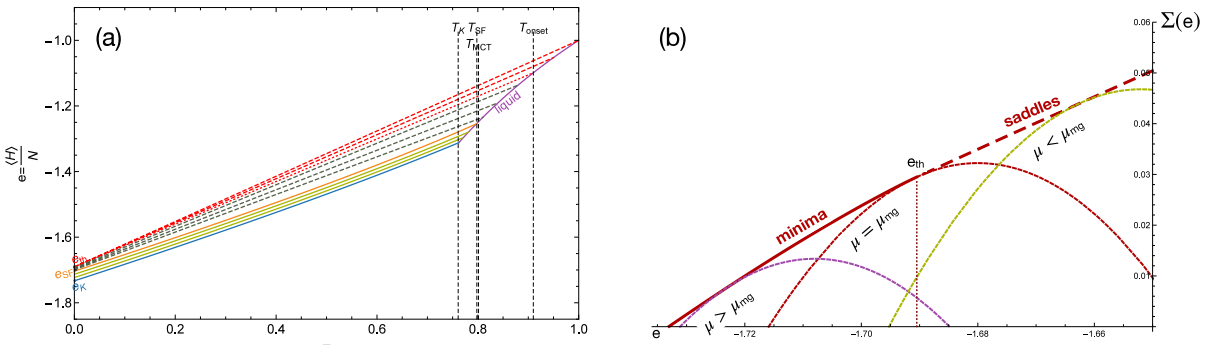

Fig. 15. (3+4)-spin model. A. Representation of the gradient descent protocol in the energy–temperature plane, as in Fig. 12A. B. Complexity as a function of the energy for several values of μ. Dashed lines represent the complexity $\Sigma(e,\mu)$ for fixed μ. The envelope is $\Sigma(e) = \max_\mu \Sigma(e,\mu)$, which is represented as a full line when $\mu^*(e) = \text{argmax}_\mu \Sigma(e,\mu) > \mu_{mg}$ (stable minima dominate) and as a dashed line when $\mu^*(e) < \mu_{mg}$ (unstable saddles dominate).
Source: Plots adapted from Ref. [47,61].

of stationary points with energy e and Lagrange parameter μ, which controls the shape of the spectrum [47,79]. For fixed μ, $\Sigma(e,\mu)$ is a parabola and there exists a family of possible energies e, as illustrated in Fig. 15B. This is true in particular for marginal states with $\mu = \mu_{mg} = 2\sqrt{f''(1)}$, which are therefore present in a finite range of energies. At fixed energy e, the dominant states are those with $\mu^*(e) = \text{argmax}_\mu \Sigma(e,\mu)$ (Fig. 15B). When $\mu^*(e) > \mu_{mg}$, the most numerous states (exponentially in N) are stable minima, while when $\mu^*(e) < \mu_{mg}$ the dominant states are unstable saddles. The value e_{th} such that $\mu^*(e_{th}) = \mu_{mg}$ thus corresponds to the geometrical transition that separates the minima-dominated and saddle-dominated regions of the landscape. Note that in this case there is a single marginal value of μ_{mg} that corresponds, however, to a range of energies, which is why we only use the term "threshold" when talking about the energy e_{th}.

Because of this different structure of the energy landscape, in the mixed p-spin with a RFOT transition a new phase emerges, which displays both memory and aging [47]. One thus finds three distinct temperature regimes.

- $T < T_{SF}$: Memorious exponential dynamics with $C(t,0) \to q_r(T)$ for $t \to \infty$ and exponential relaxation, $e(t) - e_{IS}(T) \propto e^{-t/\tau}$ to the inherent structure. Also in this case, $e_{IS}(T)$ and $q_r(T)$ can be computed via the state following or Franz-Parisi approach. Because relaxation is exponential, transient (aging) effects become quickly unobservable.
- $T_{SF} < T < T_{onset}$: Memorious power-law dynamics with $\lim_{t\to\infty} C(t,0) > 0$ and $e(t) - e_{IS}(T) \propto t^{-\nu}$ relaxation to a below-threshold energy $e_{IS}(T) < e_{th}$, which depends on the initial temperature, consistently with the observation that memory is preserved. Final states are marginal, i.e. $\mu(t) \to \mu_{mg}$, and persistent aging is observable in $C(t+t_w, t_w)$.
- $T > T_{onset}$: Memoryless power-law relaxation to the threshold energy, with $\lim_{t\to\infty} C(t,0) = 0$, $e(t) - e_{th} \propto t^{-\nu}$, and persistent aging with weak ergodicity breaking.

The high-temperature and low-temperature phases are identical to those of the pure p-spin, but in this case there is an intermediate phase. Furthermore, in the pure p-spin one has $T_{onset} = T_{SF} = T_{MCT}$, hence the equilibrium and out-of-equilibrium dynamics both have a phase transition at the same temperature, while in the mixed p-spin the equilibrium transition at T_{MCT} is unrelated to the out-of-equilibrium ones at T_{SF} and T_{onset}.

The RFOT class of mixed p-spin models presents a behavior strongly reminiscent to that of structural glasses [2,73,87,88], as illustrated in Fig. 16. For initial temperatures above T_{onset}, gradient descent converges to the T-independent energy e_{th} at which marginal states dominate the landscape. For $T_{onset} > T > T_{SF}$, T-dependent atypical marginal states, close to the initial condition, are found instead. Finally, for $T < T_{SF}$, a stable inherent structure close to the initial state is found. The geometric interpretation of this relaxation is not fully understood. In particular, the problem of computing the final energy $e_{IS}(T)$ in the intermediate phase without having to solve explicitly the dynamical equations remains open.

4.5. Summary

The study of the out-of-equilibrium dynamics of the spherical p-spin model, both pure [43] and mixed [47], is extremely useful to get a first insight on the behavior of models with complex (or rough) energy landscapes. In particular:

- A generic scenario for the geometric, thermodynamic and dynamics properties of glassy systems at the mean-field level is obtained.

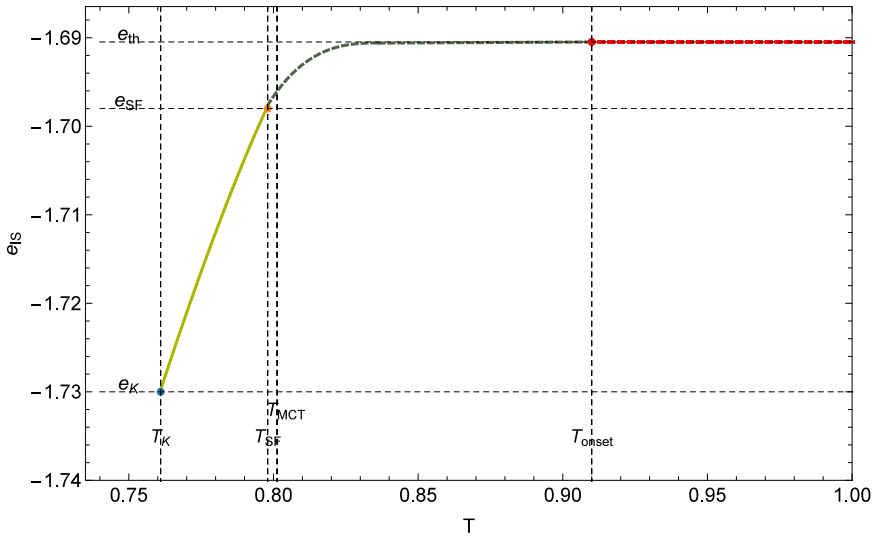

Fig. 16. Asymptotic energy vs initial temperature in the (3+4)-spin model.
Source: Plot adapted from Ref. [47].

- A simple analytic solution of these models can be obtained using the replica method and dynamical mean field theory (DMFT), which gives dynamical equations with fully explicit kernels.
- In the high-temperature regime ($T > T_{\text{onset}}$), equilibration is easy, and gradient descent converges slowly ($t^{-\nu}$ with $\nu < 1$) to the geometric threshold.
- In the low-temperature regime ($T < T_{\text{SF}}$), equilibration is hard (exp(N)), but once it is achieved, gradient descent converges fast ($e^{-t/\tau}$) to a stable close minimum.
- An intermediate-temperature regime ($T_{\text{SF}} < T < T_{\text{onset}}$) exists, where equilibration can be easy ($T > T_{\text{MCT}}$) or hard ($T < T_{\text{MCT}}$), but in any case gradient descent converges slowly ($t^{-\nu}$) to a marginally stable minimum close to the initial condition.
- Because of this rich structure, it is exponentially hard in N to find the ground state and low enough energy states by simulated annealing.

This kind of study can be generalized to other models, different search algorithms, and other cooling protocols, leading to an even richer phenomenology. The impossibility of equilibration below T_{MCT} is a general property of mean-field systems [89].

5. Constraint satisfaction problems and the jamming transition

As we discussed in Sections 1.2 and 1.4, the *jamming* transition is a characteristic phase transition of particle systems with finite-range interactions, such as granulars and emulsions [5–8]. It occurs when the particle density increases so much that a rigid network of contact interactions is formed, and it is closely related to the SAT/UNSAT transition of constraint satisfaction problems [13].

As recognized in Ref. [14], the simplest toy model of the jamming transition is the *perceptron* [90,91]. In the formulation of Ref. [14], its degrees of freedom, like for the p-spin model, are a set of N continuous spins $\underline{\sigma}$ with a spherical constraint $|\underline{\sigma}| = \sqrt{N}$. This vectorial spin degree of freedom interacts with $M = \alpha N$ obstacles $\underline{\xi}_m$ that act as a quenched disorder. The constraint satisfaction problem is defined by the following Hamiltonian or cost function

$$H(\underline{\sigma}) = \sum_{m=1}^{M} \frac{1}{2} h_\mu^2 \Theta(-h_m) \,, \qquad h_m = \frac{\underline{\xi}_m \cdot \underline{\sigma}}{\sqrt{N}} - \kappa \,, \tag{48}$$

being $h_m \geq 0$ the constraint imposed by the mth obstacle, and κ a real parameter representing the diameter of the obstacles. Indeed, one can imagine that an obstacle is located in position $\underline{\Xi}_m = -\sqrt{N}\underline{\xi}_m/|\underline{\xi}_m|$ on the sphere, and the constraint is then equivalent to $\hat{\Xi}_m \cdot \hat{\sigma} < -\kappa/|\underline{\xi}_m|$, which imposes (for negative κ) that the vector $\underline{\sigma}$ should stay outside of a cone around the obstacle $\underline{\Xi}_m$ [14].

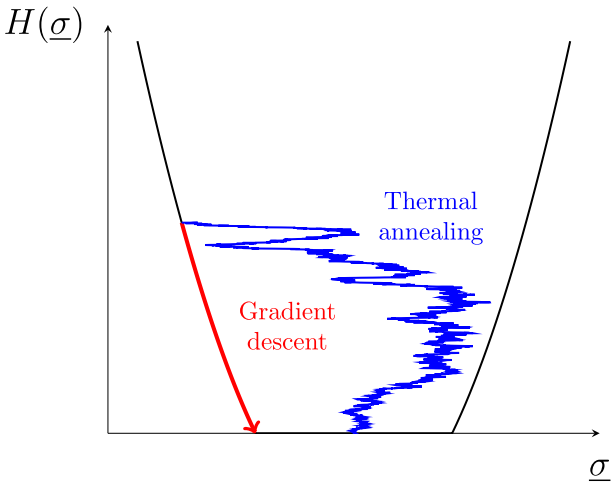

Fig. 17. An energy lake in the SAT phase of a CCSP. GD dynamics converges to the shore, while thermal annealing can explore the lake's interior.

The perceptron with $\kappa \geq 0$ is instead one of the simplest *classifiers*, i.e. a machine that allows one to separate points in a phase space by assigning them a binary label [25,92]. Imagine that one is given a set of M images of cats and dogs represented by a sequence of N bits encoded in the vectors $\underline{x}_m$, $m = 1, \dots, M$. Knowing the labels $y_m \in \{\pm 1\}$ corresponding to a cat/dog in the mth image, the goal of supervised learning is to find a vector $\underline{\sigma}$ such that $y_m = \text{sgn}\left(\underline{\sigma} \cdot \underline{x}_m\right)$, i.e. to correctly classify the training labeled images $\{\underline{x}_m, y_m\}$. Calling $\underline{\xi}_m = y_m \underline{x}_m$, the constraint becomes $\underline{\sigma} \cdot \underline{\xi}_m > 0$ for all $m = 1, \dots, M$. One can show that a stricter constraint, i.e.

$$\underline{\sigma} \cdot \underline{\xi}_m > \kappa \sqrt{N} \,, \qquad \forall\, m = 1, \dots, M \,, \qquad \kappa > 0 \,, \tag{49}$$

allows one to preserve the correct classification if the images are slightly corrupted by noise [25,92].

In the simplest setting, the *random perceptron*, the input vectors components $x_{i,m}$ are drawn from a white Gaussian distribution with zero mean and unit variance, and a random label $y_m \in \{\pm 1\}$ with probability $1/2$ is assigned to each of them. The obstacles (or patterns) $\underline{\xi}_m$ then have the same statistics of the $\underline{x}_m$, i.e. they are Gaussian vectors with i.i.d. components of zero mean and unit variance. In this case, there is nothing to be learned because the labels are random, but one can ask – as a benchmark question – if the machine can learn random noise. This problem is related to the bias/variance tradeoff and gives a bound on the generalization error [93].

We conclude that the random perceptron, whose behavior is fully determined by the parameters κ and α, corresponds to a simple random noise classification problem for $\kappa \geq 0$ [25,92], and to a particle that has to avoid random obstacles for $\kappa < 0$ [14]. In both cases, it is an instance of a continuous-variable constraint satisfaction problem (CCSP) [94], like the sphere packing problem, which consists of finding a configuration of N particle positions $\underline{x}$ such that no two particles are overlapping, namely $h_{ij} = |x_i - x_j| - \ell > 0$, $\forall i, j$. In Section 4 we analyzed the connection between the out-of-equilibrium dynamics and the energy landscape of the p-spin model. An important additional ingredient brought in by the perceptron (and more generally by CCSPs) is that its energy landscape can have "lakes", i.e. regions where the energy is identically vanishing. These regions where $H(\underline{\sigma}) = 0$ represent the set of solutions of the CCSP, as shown in Fig. 17.

The jamming transition can then be related to a SAT/UNSAT transition: increasing the fraction of obstacles α at fixed κ, the lakes of the energy landscape shrink until, at the jamming point, their volume vanishes and we are only left with UNSAT minima with $H(\underline{\sigma}) > 0$. At every value of α, κ, one can determine the *equilibrium* probability of having a SAT instance, $P_{\text{SAT}}^{\text{eq}}$, as the probability, over the choice of the random obstacles $\underline{\xi}_m$, that the absolute minimum of $H(\underline{\sigma})$ (i.e. the thermodynamic zero-temperature ground state) is at zero energy. The equilibrium SAT probability is almost unity at low obstacle density α (unjammed, SAT phase) and almost vanishing at high α (jammed, UNSAT phase), and the transition becomes sharp in the thermodynamic limit around a critical value α_c [13,14,25,92], see Fig. 18A. Alternatively, one can estimate a *non-equilibrium* $P_{\text{SAT}}^{\text{GD}}$ as the probability that GD dynamics reaches zero energy starting from a random initial condition, the probability being calculated over the choice of initial condition and the realization of the random obstacles. It is found that $P_{\text{SAT}}^{\text{GD}}$ has the same behavior of $P_{\text{SAT}}^{\text{eq}}$ illustrated in Fig. 18A, but with an *a priori* lower critical value $\alpha_c^{\text{GD}} \leq \alpha_c$. The two values only coincide for $\kappa > 0$. In the particles literature, the jamming transition is usually defined via the gradient descent protocol, hence it coincides with $\alpha_c^{\text{GD}} = \alpha_J$ [5,6].

Furthermore, the convergence time of GD dynamics diverges at α_c^{GD}, see Fig. 18B, indicating that finding energy minima in the very dilute of very dense case is much easier than in the near-critical case [95]. This scenario is quite general in search algorithms and applies to several problems (such as the coloring), and it also provides a practical way to construct difficult instances of constraint satisfaction problems [18–22].

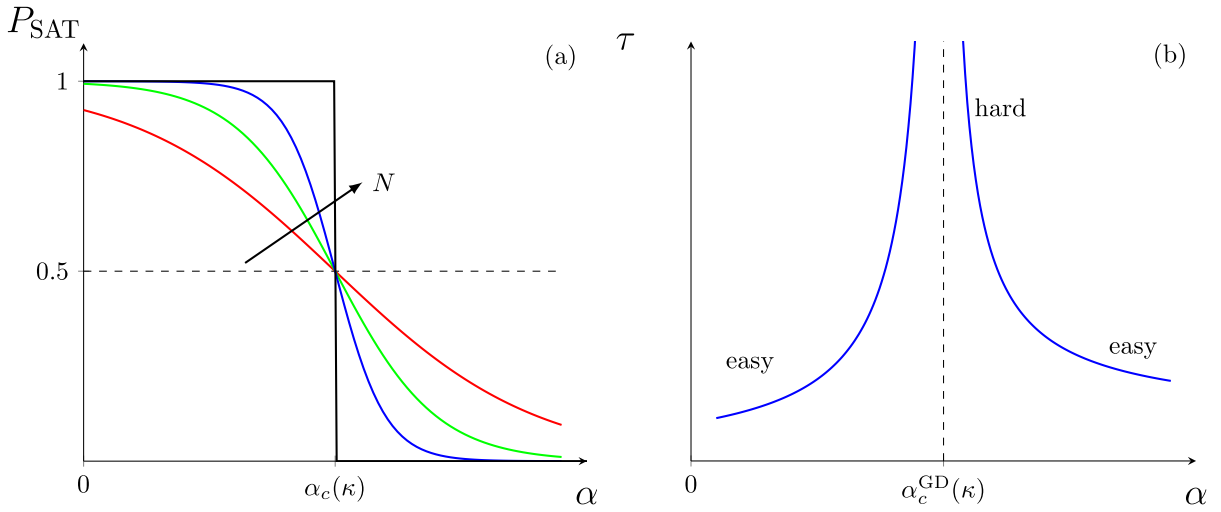

Fig. 18. A. Probability of finding a SAT configuration, $\min_{\underline{\sigma}} H(\underline{\sigma}) = 0$, as a function of the constraint fraction α. B. Characteristic time needed for gradient descent to decide is the system is SAT or UNSAT, and its divergence around the jamming transition.

Unfortunately, the solution of the random perceptron problem is technically more difficult than the p-spin. Thermodynamic calculations involving the replica method [14,25,92,94], the cavity method [96] or perturbative expansions [97,98] are still possible but a bit more involved. Complexity calculations have not been performed yet, and DMFT equations are much more complicated because the memory kernels are given by functionals over a space of trajectories [99,100]: their solution is an open problem at present, for which new numerical algorithms are required. Interestingly, the DMFT equations for the perceptron are very similar to those for infinite-dimensional particle systems [100,101]. Therefore, a consistent solution may be easily generalized to both cases. From now on, we will focus for simplicity on the GD dynamics starting from equilibrium configurations at infinite temperature, $T = \infty$.

5.1. Classification of the minima

The GD dynamics leads the system towards a local minimum of the energy landscape $H(\underline{\sigma})$: we can then repeat the study of Section 4.2 in order to characterize the vibrational spectrum of the Hessian matrix in typical stationary points [102]. As in Section 4.2, taking into account the spherical constraint via a Lagrange multiplier μ, stationary points are solutions of

$$\frac{\partial H}{\partial \sigma_i} + \mu \sigma_i = 0 \quad \text{with} \quad \mu = -\frac{1}{N}\underline{\sigma} \cdot \underline{\nabla} H = \alpha \left([h^2] + \kappa\, [h] \right) \,, \tag{50}$$

being $[f(h)] \equiv \frac{1}{M}\sum_m f(h_m)\Theta(-h_m)$ the average over the contacts (i.e. violated constraints) for a fixed configuration $\underline{\sigma}$ and obstacle realization. The Hessian matrix is then given by

$$M_{ij} = \frac{\partial^2 H}{\partial \sigma_i \partial \sigma_j} + \mu \delta_{ij} = \frac{1}{N}\sum_{m=1}^{M} \xi_i^m \xi_j^m \Theta(-h_m) + \mu \delta_{ij} \,. \tag{51}$$

While, in a local minimum, the obstacle positions ξ_i^m and gaps h_m are obviously correlated because $\underline{\sigma}$ is a solution of Eq. (50), one can argue that the correlations can be neglected at the leading order in N [102], similarly to the p-spin case. The matrix M is then a $N \times N$ matrix given by a sum of a number $\sum_m \Theta(-h_m)$ of independent random Gaussian projectors, i.e. it is a random Wishart matrix with a fraction

$$c = \alpha\, [1] = \frac{1}{N}\sum_m \Theta(-h_m) \tag{52}$$

of uncorrelated patterns, being c the *isostaticity index* indicating the ratio between the number of binding constraints and the degrees of freedom [102]. The eigenvalues of the Hessian matrix therefore follow the Marchenko–Pastur distribution

$$\begin{aligned} \rho(\lambda) &= (1-c)\Theta(1-c)\delta(\lambda+\mu) + \frac{1}{2\pi}\frac{\sqrt{(\lambda-\lambda_-)(\lambda_+-\lambda)}}{\lambda+\mu}\mathbb{1}_{[\lambda_-,\lambda_+]}(\lambda) \,, \\ \lambda_\pm &= \left(\sqrt{c} \pm 1\right)^2 - \mu \,. \end{aligned} \tag{53}$$

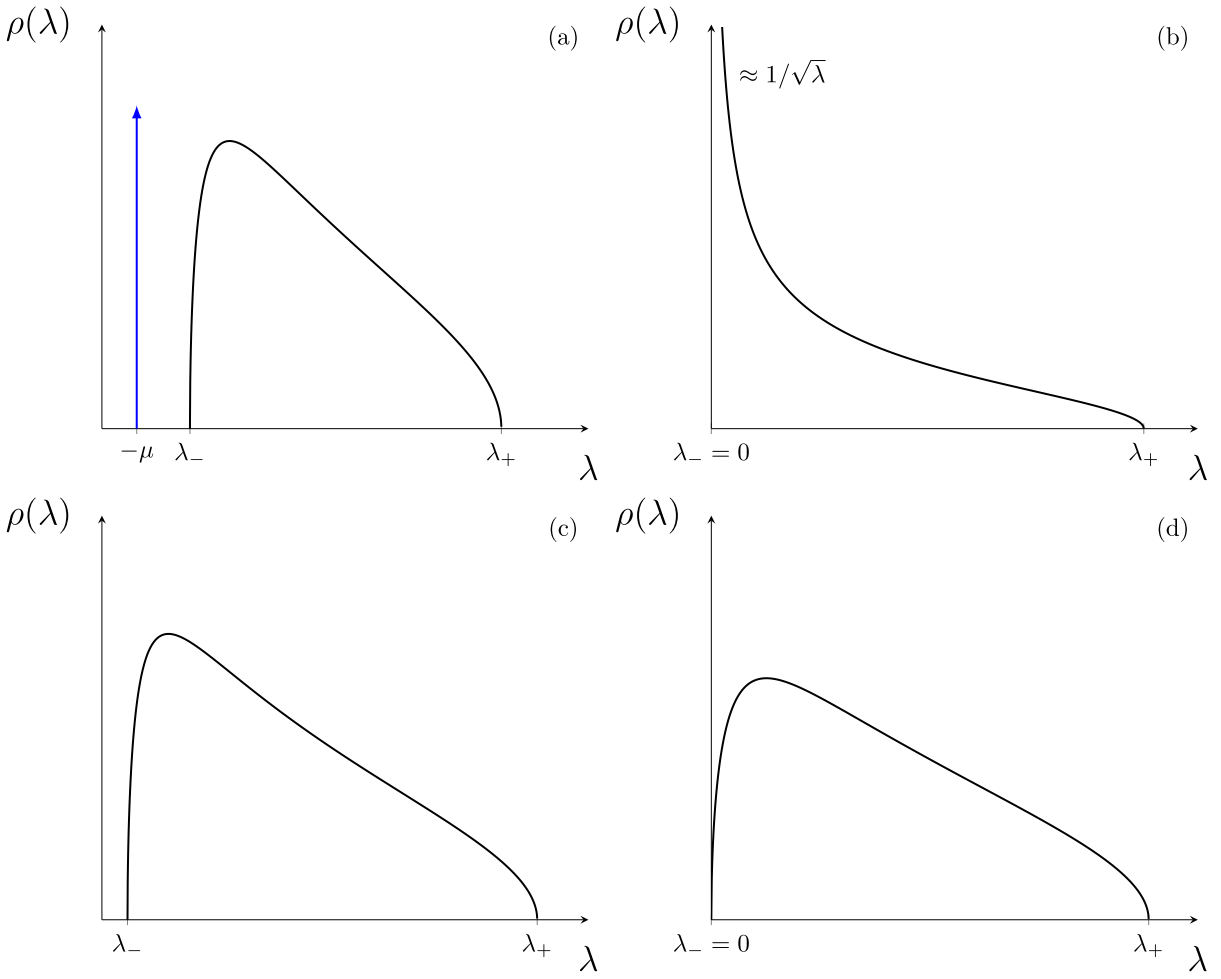

Fig. 19. Classification of the minima of the random perceptron. A. Hypostatic case. B. Isostatic case. C. Hyperstatic case, stable. D. Hyperstatic case, marginal.

Like in the p-spin case, the spectrum only depends on global parameters; here, in addition to μ that provides a global shift, it also depends on c. The isostaticity index then shows its relevance; it is worth noting that the value $c = 1$ implies that the number of contacts is equivalent to the degrees of freedom of the system, so for $c \geq 1$ the matrix $\frac{\partial^2 H}{\partial \sigma_i \partial \sigma_j}$ has full rank, while for $c < 1$ it has a rank equal to Nc and it thus has a number $N(1-c)$ of zero modes. We can now classify the possible scenarios, all shown in Fig. 19.

A. *Hypostatic case* $c < 1$. Here, there are $N(1-c)$ modes with eigenvalue $-\mu$, hence we need $\mu \leq 0$ for stability. For $\mu < 0$, all the eigenvalues are positive, and the distribution $\rho(\lambda)$ is stable and gapped (Fig. 19A). Note that this kind of minima can only exist for $\kappa > 0$, because $\mu = \alpha\left([h^2] + \kappa\,[h]\right)$ and $[h] \leq 0$ by construction. When $\mu = 0$, the isolated eigenvalue vanishes, generating a finite density of zero modes.

B. *Isostatic case* $c = 1$. Here, when in addition $\mu = 0$, the system is marginally stable. The Marchenko–Pastur distribution diverges at $\lambda \to 0^+$ as $\rho(\lambda) \sim \lambda^{-1/2}$, which implies that the density of states with vibrational frequency $\omega = \sqrt{\lambda} \sim 0$ (soft modes) is constant at low frequencies, $D(\omega \sim 0) \sim$ const (Fig. 19B). The case $c = 1$ and $\mu > 0$, which is in principle possible, is not observed in practice [102], because isostaticity is only realized at the jamming transition where $[h] = [h^2] = \mu = 0$.

C, D. *Hyperstatic case* $c > 1$. The isolated eigenvalue is absent and the stability condition $\lambda_- \geq 0$ requires $\left(\sqrt{c} - 1\right)^2 \geq \mu$. Close to isostaticity, we have $\delta c = c - 1$ and $[h^2] \ll |\,[h]\,| \sim p$, where p is identified with the mechanical pressure in particle systems [94], and the stability criterion thus implies $\delta c^2 \geq \text{const} \times p$, as derived in [103]. The system is therefore stable if the above criterion is satisfied ($\lambda_- > 0$, Fig. 19C), and becomes marginally stable when $\lambda_- = 0$ (Fig. 19D). In the latter case, the distribution goes as $\rho(\lambda) \sim \sqrt{\lambda}$ at small λ, and the density of soft modes goes as $D(\omega \sim 0) \sim \omega^2$ [104].

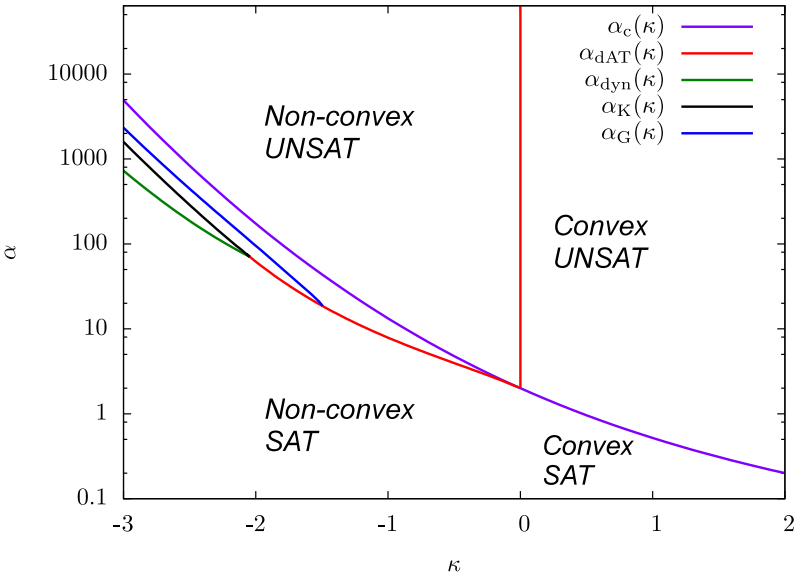

Fig. 20. Phase diagram of the random perceptron.
Source: Adapted from Ref. [94]

5.2. Phase diagram

The zero-temperature equilibrium phase diagram (i.e. the ground state structure) of the random perceptron has been deeply investigated for both positive [25,92] and negative [14,94] κ, and it is illustrated in Fig. 20. The control parameters are κ and α, and the first one governs the convexity of the solution space, which is convex at $\kappa > 0$ and non-convex at $\kappa < 0$ [14]. The phase diagram can thus be divided into four main regions: the convex region for $\kappa > 0$ and the non-convex region for $\kappa < 0$ are both separated into a SAT phase at low $\alpha < \alpha_c(\kappa)$ and an UNSAT phase at $\alpha > \alpha_c(\kappa)$. In the convex region, only the jamming transition at $\alpha_c(\kappa)$ is present; conversely, in the non-convex region the phase diagram is much richer and several phase transitions occur within the SAT phase below the jamming line [14,94].

5.2.1. Convex UNSAT phase

The phase diagram can be obtained from replica computations [14,25,92,94]; in the convex region, the replica-symmetric (RS) solution is stable both in the SAT and UNSAT case, and the transition line can be identified as

$$\alpha_c(\kappa) = \left[\int_{-\infty}^{0} \frac{\mathrm{d}h}{\sqrt{2\pi}} e^{-(h+\kappa)^2/2} h^2\right]^{-1} . \tag{54}$$

The easiest phase to analyze is the convex UNSAT one, $\alpha > \alpha_c(\kappa)$ and $\kappa > 0$ [25,92]. The energy minimum is unique because of convexity, and the free energy converges to the ground state energy when $T \to 0$, which reads

$$e_{RS} = \frac{1}{2}\left(\sqrt{\frac{\alpha}{\alpha_c(\kappa)}} - 1\right)^2 . \tag{55}$$

The GD dynamics therefore converges exponentially to the unique minimum, i.e. $e(t) - e_{RS} \propto e^{-t/\tau}$ at long times, being τ the relaxation time [95,100,105]. The RS computation also provides the contact number

$$c = \alpha\,[1] = \alpha \int_{-\infty}^{0} \frac{\mathrm{d}h}{\sqrt{2\pi}} e^{-(h+\kappa)^2/2} . \tag{56}$$

The perceptron is therefore hypostatic on the convex jamming line, and becomes isostatic only at $\kappa = 0$, i.e. at the boundary of the non-convex phase [14]. The spectrum is gapped in the UNSAT phase at $\alpha > \alpha_c(\kappa)$, and the long-time limit values of c, μ, e and other observables coincide with those of the unique ground state and can thus be computed exactly both from replica calculations [14,25,92,94] and from DMFT [100,105]. When $\alpha \to \alpha_c(\kappa)^+$, one has $\mu \to 0$ with $c < 1$ and a finite fraction of zero modes is thus present at jamming (Fig. 19A).

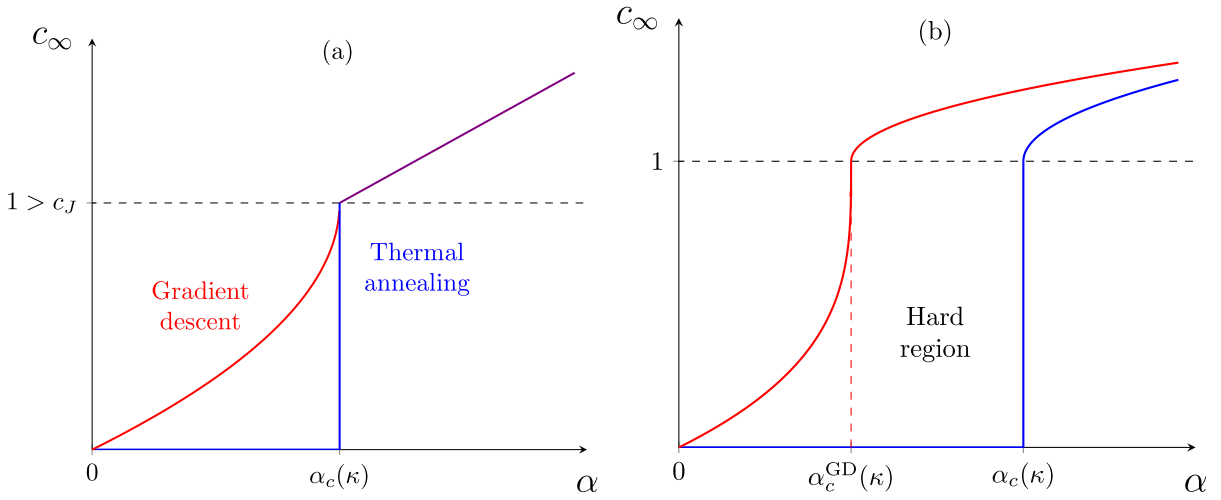

Fig. 21. Isostatic index c as computed from GD dynamics (red) and thermodynamics (blue). A. Convex region: the two methods coincide in the UNSAT phase, while in the SAT phase thermodynamics gives $c = 0$ while dynamics gives $0 < c < 1$. B. Non-convex region: the two methods do not agree and there is a hard region where solutions of the CSP exist, but the gradient descent is unable to find them.

5.2.2. Convex SAT phase

The convex SAT case is, somehow surprisingly, more complex than the UNSAT case, because in the SAT phase the solutions found by thermal annealing and GD dynamics do not coincide. Replica computations [14,25,92,94] indeed predict the existence of a single convex lake of solutions with $H(\underline{\sigma}) = 0$, surrounded by a convex energy landscape. In this scenario, during thermal annealing the system starts from a high-energy configuration and explores the lake uniformly when temperature is very low. When temperature is finally switched off, the algorithm thus stops in a randomly chosen configuration inside the lake, where $c = e = \mu = 0$, and all the directions are flat - i.e. $\rho(\lambda) = \delta(\lambda)$, see Fig. 17. Conversely, GD dynamics starts from the same configuration but follows the steepest descent path that ends up on the lake "shore", implying a final value of $\mu = e = 0$, as for thermal annealing, but a finite fraction of contacts $0 < c < 1$ that have precisely $h = 0$ [100], as illustrated in Fig. 21A. The convergence to the final zero-energy state is exponential as in the UNSAT case. Therefore, the spectrum of the final state of the GD dynamics shows a Marchenko–Pastur distribution at finite λ together with a finite fraction of zero modes. Numerical results for the GD dynamics are given in [95]. The difference between the final states of thermal and athermal dynamics may have important consequences in machine learning applications, see e.g. [27,29].

The relaxational dynamics of the convex problem also shows features of dynamic criticality: indeed the relaxation time τ of the GD dynamics diverges when approaching the jamming transition from both directions [95], see Fig. 18B. It has been numerically shown that $\tau \sim 1/\lambda_1$, being λ_1 the first non-zero eigenvalue of the Hessian matrix [95]. When approaching the transition from the UNSAT phase, λ_1 is given by the isolated eigenvalue $\lambda_1 = -\mu \sim |[h]| \sim \delta\alpha$, therefore one has $\tau \sim \delta\alpha^{-1}$. On the other hand, from the SAT phase one would naively obtain $\lambda_1 = \lambda_-$, which remains finite when $\alpha \to \alpha_c(\kappa)$. However, the relaxation has been observed to be dominated by an isolated, low-frequency eigenmode, which is not captured by the density $\rho(\lambda)$ [95,106–109]. Computing this isolated mode analytically remains an open problem.

5.2.3. Non-convex phase

When $\kappa < 0$, the energy landscape becomes non-convex and rough, and correspondingly the phase diagram becomes more complex; we refer to Ref. [94] for details. Concerning the jamming transition, there are two main differences with respect to the convex case:

- the jamming transition is *isostatic*, i.e. $c = 1$ identically all along the jamming line, contrarily to the hypostaticity observed in the convex phase;
- thermodynamics and GD dynamics *do not agree* on the location of the transition, with $\alpha_c^{\mathrm{GD}}(\kappa) < \alpha_c(\kappa)$ as computed from replica calculations, because the energy landscape is developing growing complexity with many local energy minima in which GD dynamics gets trapped before the jamming transition occurs at the thermodynamic level — see Fig. 21B. This gap defines a *hard region* $\alpha_c^{\mathrm{GD}}(\kappa) < \alpha < \alpha_c(\kappa)$, where zero-energy solutions exist but the GD dynamics cannot find them [33].

Furthermore, in the UNSAT phase one expects that the energy landscape of the perceptron is very similar to that of the mixed p-spin, hence GD should display the same power-law relaxation described for the p-spin in Section 4, namely $e(t) - e_{th} \sim t^{-\nu}$ with $\nu < 1$, associated to memoryless persistent aging with weak ergodicity breaking. This has not been

carefully checked in the perceptron yet, but it has been investigated in numerical simulations of spherical particles, with $\nu \simeq 0.84$ in $d = 2$ and $\nu \simeq 0.70$ in $d = 3$ [110,111]. Also, the final state is hyperstatic with $c > 1$, and similarly to the mixed p-spin, we expect it to be marginal, hence $\lambda_- = 0$ and the spectrum goes as $\rho(\lambda) \sim \sqrt{\lambda}$, see Fig. 19D. Note that because the relaxation of the energy is a power-law, the relaxation time τ is formally infinite throughout this phase; this is a consequence of the marginal stability of the final state.

Furthermore, precisely at the non-convex jamming transition, the system exhibits other non-trivial critical relations. The distribution of gaps h and forces f exhibit a universal power-law behavior [112]

$$P(h \to 0^+) \sim h^{-\gamma} \ , \quad P(f \to 0^+) \sim f^{\theta} \ , \tag{57}$$

with $\gamma \simeq 0.41269\cdots$ and $\theta \simeq 0.42311\cdots$. These critical exponents have been obtained analytically by replica calculations [113], hence at the thermodynamical jamming point. However, they have also been computed by numerical simulations of GD dynamics in sphere packings and in the perceptron (hence at the dynamical jamming transition), obtaining perfect agreement within numerical precision, despite the different nature of the transition [114–119]. Furthermore, scaling relations based on marginal stability [112,120,121] are also satisfied by the exponents predicted by the replica method. These results hints at a strong universality of the jamming transition, despite the fact that no direct mapping of the perceptron on the sphere packing problem exists [122].

The non-convex SAT phase, on the other hand, shows more similarities with the convex case. As in the latter, the energy decay follows an exponential law; the relaxation time has been derived from a scaling argument [106,108], with the result

$$\tau \sim \delta c^{-\beta} \ , \qquad \beta = \frac{4 + 2\theta}{1 + \theta} \simeq 3.41\cdots \ , \tag{58}$$

which is in agreement with some numerical results, although with considerable uncertainty [109]. Moreover, a derivation of this exponent within dynamical mean-field theory is still missing.

Finally, we mention that a modified perceptron [123] can be used to describe a different universality class of non-convex hypostatic jamming, found for example in ellipsoids [124].

5.3. Summary

In this section, we discussed the complex behavior of constraint satisfaction problems, based on the paradigmatic perceptron model. The latter is closely related to the sphere packing problem, and their DMFT equations almost coincide in the infinite-dimensional limit. Their energy landscape is even richer than the p-spin one.

We have seen that the jamming transition is associated with a SAT/UNSAT transition, and that, in the non-convex case, a hard region exists where GD dynamics and thermodynamics do not give the same result. All the results obtained for the *non-convex* perceptron also apply to the sphere packing problem of particles in infinite spatial dimensions, $d \to \infty$.

However, DMFT equations are particularly difficult to solve: the analytical understanding of mean-field dynamics is still poor, and the currently known results are limited to the long-time limit of the GD dynamics in the UNSAT phase [100,105]. The numerical solution of the DMFT equations has been possible only at very short times for the sphere packing problem [100], while more encouraging results have been found in classification problems [28,29]. To conclude, we recall some of the main open problems:

- the derivation of the β exponent of the relaxation time $\tau \sim \delta c^{-\beta}$ in the SAT phase;
- the derivation of the ν exponent in the energy relaxation $e - e_{th} \sim t^{-\nu}$ in the non-convex UNSAT phase;
- the understanding of why jamming criticality is so strongly universal that mean-field thermodynamic predictions are also consistent with dynamical results in all dimensions [121,125];
- the effect of finite dimensionality, especially focused on the role of localized modes [109].

While the toy models discussed here provide a good starting point to understand many phenomena that are characteristic of disordered systems with complex landscapes, many problems remain open and a lot of exciting results are surely yet to come.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

We warmly thank the organizers of the summer school *Fundamental Problems in Statistical Physics XV*, held in Brunico, Italy, in July 2021, and of the summer school *Glassy Systems and Inter-Disciplinary Applications*, held in Cargese, France, in June 2021, for the invitation to lecture there (F.Z.) and for the invitation to participate in Cargese (A.M., G.F.). We also thank all the participants to those schools for many discussions and for providing a welcoming and stimulating environment.

Our work has received support by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n. 723955 - GlassUniversality) and by a grant from the Simons Foundation (#454955, Francesco Zamponi).

References

[1] A. Cavagna, Phys. Rep. 476 (2009) 51.
[2] F. Sciortino, J. Stat. Mech. Theory Exp. 2005 (2005) P05015.
[3] T. Castellani, A. Cavagna, J. Stat. Mech. Theory Exp. 2005 (2005) P05012.
[4] T.R. Kirkpatrick, D. Thirumalai, J. Phys. A: Math. Gen. 22 (1989) L149.
[5] C.S. O'Hern, S.A. Langer, A.J. Liu, S.R. Nagel, Phys. Rev. Lett. 88 (2002) 075507.
[6] C.S. O'Hern, L.E. Silbert, A.J. Liu, S.R. Nagel, Phys. Rev. E 68 (2003) 011306.
[7] A.J. Liu, S.R. Nagel, Annu. Rev. Condens. Matter Phys. 1 (2010) 347.
[8] A. Liu, S. Nagel, W. Van Saarloos, M. Wyart, in: L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, W. van Saarloos (Eds.), Dynamical Heterogeneities and Glasses, Oxford University Press, 2011, arXiv:1006.2365.
[9] J.H. Conway, N.J.A. Sloane, Sphere Packings, Lattices and Groups, Spriger-Verlag, 1993.
[10] S. Torquato, F.H. Stillinger, Rev. Modern Phys. 82 (2010) 2633.
[11] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error Correcting Codes, Vol. 16, Elsevier, 1977.
[12] W.C. Huffman, V. Pless, Fundamentals of Error-Correcting Codes, Cambridge University Press, 2010.
[13] F. Altarelli, R. Monasson, G. Semerjian, F. Zamponi, in: A. Biere, M. Heule, H. van Maaren, T. Walsh (Eds.), Handbook of Satisfiability, Frontiers in Artificial Intelligence and Applications, IOS Press, 2009, arXiv:0802.1829.
[14] S. Franz, G. Parisi, J. Phys. A 49 (2016) 145001.
[15] D.L. Applegate, R.E. Bixby, V. Chvá, W.J. Cook, The Traveling Salesman Problem: A Computational Study, Princeton University Press, 2011.
[16] Several practical applications of the TSP and an interactive visualization for learning/teaching the problem can be found at http://www.math.uwaterloo.ca/tsp/index.html.
[17] T.R. Jensen, B. Toft, Graph Coloring Problems, John Wiley & Sons, 2011.
[18] P.C. Cheeseman, B. Kanefsky, W.M. Taylor, IJCAI, Vol. 91, 1991, pp. 331–337.
[19] D. Mitchell, B. Selman, H. Levesque, AAAI, Vol. 92, Citeseer, 1992, pp. 459–465.
[20] S. Kirkpatrick, B. Selman, Science 264 (1994) 1297.
[21] B. Selman, S. Kirkpatrick, Artificial Intelligence 81 (1996) 273.
[22] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, Nature 400 (1999) 133.
[23] A. Engel, C. Van den Broeck, Statistical Mechanics of Learning, Cambridge University Press, 2001.
[24] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Rev. Modern Phys. 91 (2019) 045002.
[25] E. Gardner, Europhys. Lett. 4 (1987) 481.
[26] M. Geiger, L. Petrini, M. Wyart, arXiv:2012.15110, 2020.
[27] Stéphane D'Ascoli, Maria Refinetti, Giulio Biroli, Florent Krzakala, Double trouble in double descent: bias and variance(s) in the lazy regime, in: Hal Daumé III, Aarti Singh (Eds.), Proceedings of the 37th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, 119, PMLR, 2020, pp. 2280–2290.
[28] F. Mignacco, F. Krzakala, P. Urbani, L. Zdeborová, J. Stat. Mech. Theory Exp. 2021 (2021) 124008.
[29] F. Mignacco, P. Urbani, arXiv:2112.10852, 2021.
[30] E. Schneidman, M.J. Berry, R. Segev, W. Bialek, Nature 440 (2006) 1007.
[31] S. Cocco, S. Leibler, R. Monasson, Proc. Natl. Acad. Sci. 106 (2009) 14058.
[32] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D.S. Marks, C. Sander, R. Zecchina, J.N. Onuchic, T. Hwa, M. Weigt, Proc. Natl. Acad. Sci. 108 (2011) E1293.
[33] L. Zdeborová, F. Krzakala, Adv. Phys. 65 (2016) 453.
[34] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, M. Weigt, Rep. Progr. Phys. 81 (2018) 032601.
[35] V. Ros, G.B. Arous, G. Biroli, C. Cammarota, Phys. Rev. X 9 (2019) 011003.
[36] Stefano Sarao Mannelli, Florent Krzakala, Pierfrancesco Urbani, Lenka Zdeborova, Passed & spurious: descent algorithms and local minima in spiked matrix-tensor models, in: Kamalika Chaudhuri, Ruslan Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, 97, PMLR, 2019, pp. 4333–4342.
[37] S.S. Mannelli, G. Biroli, C. Cammarota, F. Krzakala, P. Urbani, L. Zdeborová, Phys. Rev. X 10 (2020) 011057.
[38] F. Mignacco, P. Urbani, L. Zdeborová, Mach. Learn.: Sci. Technol. 2 (2021) 035029.
[39] F. Antenucci, S. Franz, P. Urbani, L. Zdeborová, Phys. Rev. X 9 (2019) 011020.
[40] J. Trinquier, G. Uguzzoni, A. Pagnani, F. Zamponi, M. Weigt, Nature Commun. 12 (2021) 1.
[41] S. Cocco, R. Monasson, L. Posani, S. Rosay, J. Tubiana, Physica A 504 (2018) 45.
[42] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Science 220 (1983) 671.
[43] L.F. Cugliandolo, J. Kurchan, Phys. Rev. Lett. 71 (1993) 173.
[44] A. Montanari, F. Ricci-Tersenghi, Phys. Rev. B 70 (2004) 134406.
[45] L. Zdeborová, F. Krzakala, Phys. Rev. B 81 (2010) 224205.
[46] F. Krzakala, L. Zdeborová, J. Phys. Conf. Ser. 473 (2013) 12022.
[47] G. Folena, S. Franz, F. Ricci-Tersenghi, Phys. Rev. X 10 (2020) 031045.
[48] L. Cugliandolo, in: J. Barrat, M. Feigelman, J. Kurchan, J. Dalibard (Eds.), Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter, Springer-Verlag, 2003, arXiv:cond-mat/0210312.
[49] J. Kurchan, arXiv:0901.1271, 2009.
[50] T.R. Kirkpatrick, P.G. Wolynes, Phys. Rev. A 35 (1987) 3072.
[51] T.R. Kirkpatrick, D. Thirumalai, Phys. Rev. Lett. 58 (1987) 2091.
[52] T.R. Kirkpatrick, D. Thirumalai, Phys. Rev. B 36 (1987) 5388.

[53] T.R. Kirkpatrick, P.G. Wolynes, Phys. Rev. B 36 (1987) 8552.
[54] B. Derrida, Phys. Rev. B 24 (1981) 2613.
[55] D.J. Gross, M. Mézard, Nuclear Phys. B 240 (1984) 431.
[56] D. Gross, I. Kanter, H. Sompolinsky, Phys. Rev. Lett. 55 (1985) 304.
[57] E. Gardner, Nuclear Phys. B 257 (1985) 747.
[58] T.R. Kirkpatrick, D. Thirumalai, P.G. Wolynes, Phys. Rev. A 40 (1989) 1045.
[59] A. Crisanti, H.-J. Sommers, Z. Phys. B Condensed Matter 87 (1992) 341.
[60] A. Crisanti, H. Horner, H.-J. Sommers, Z. Phys. B Condensed Matter 92 (1993) 257.
[61] G. Folena, The mixed p-spin model : selecting, following and losing states, 2020, https://tel.archives-ouvertes.fr/tel-02883385.
[62] H. Sompolinsky, A. Zippelius, Phys. Rev. Lett. 47 (1981) 359.
[63] H. Sompolinsky, A. Zippelius, Phys. Rev. B 25 (1982) 6860.
[64] S. Franz, F. Tria, J. Stat. Phys. 122 (2006) 313.
[65] A. Auffinger, G.B. Arous, Ann. Probab. 41 (2013) 4214.
[66] A. Auffinger, G.B. Arous, J. Č, Comm. Pure Appl. Math. 66 (2013) 165.
[67] E. Subag, Invent. Math. 210 (2017) 135.
[68] A. Auffinger, W.-K. Chen, Adv. Math. 330 (2018) 553.
[69] G.B. Arous, E. Subag, O. Zeitouni, Comm. Pure Appl. Math. 73 (2020) 1732.
[70] J.M. Kosterlitz, D.J. Thouless, R.C. Jones, Phys. Rev. Lett. 36 (1976) 1217.
[71] D. Barbier, L.F. Cugliandolo, G.S. Lozano, N. Nessi, Europhys. Lett. 132 (2020) 50002.
[72] M. Mézard, G. Parisi, M.A. Virasoro, Spin Glass Theory and beyond, World Scientific, 1987.
[73] S. Sastry, P.G. Debenedetti, F.H. Stillinger, Nature 393 (1998) 554.
[74] S. Franz, G. Parisi, J. Physique I 5 (1995) 1401.
[75] R. Monasson, Phys. Rev. Lett. 75 (1995) 2847.
[76] W. Götze, Complex Dynamics of Glass-Forming Liquids: A Mode-Coupling Theory, Vol. 143, OUP Oxford, 2008.
[77] J. Kurchan, G. Parisi, M.A. Virasoro, J. Physique I 3 (1993) 1819.
[78] A. Crisanti, H.-J. Sommers, J. Physique I 5 (1995) 805.
[79] G. Folena, S. Franz, F. Ricci-Tersenghi, J. Stat. Mech. Theory Exp. 2021 (2021) 033302.
[80] A. Barrat, S. Franz, G. Parisi, J. Phys. A: Math. Gen. 30 (1997) 5593.
[81] A. Barrat, R. Burioni, M. Mézard, J. Phys. A: Math. Gen. 29 (1996) L81.
[82] A. Cavagna, I. Giardina, G. Parisi, Phys. Rev. B 57 (1998) 11251.
[83] Y. Sun, A. Crisanti, F. Krzakala, L. Leuzzi, L. Zdeborová, J. Stat. Mech. Theory Exp. 2012 (2012) P07002.
[84] J.-P. Bouchaud, J. Physique I 2 (1992) 1705.
[85] J. Bouchaud, L. Cugliandolo, J. Kurchan, M. Mezard, in: A. Young (Ed.), Spin Glasses and Random Fields, World Scientific Pub Co Inc, 1998, arXiv.org:cond-mat/9702070.
[86] S. Franz, E. Marinari, G. Parisi, J. Phys. A: Math. Gen. 28 (1995) 5437.
[87] M. Ozawa, T. Kuroiwa, A. Ikeda, K. Miyazaki, Phys. Rev. Lett. 109 (2012) 205701.
[88] P. Charbonneau, P.K. Morse, Phys. Rev. Lett. 126 (2021) 088001.
[89] A. Montanari, G. Semerjian, J. Stat. Phys. 125 (2006) 23.
[90] F. Rosenblatt, Psychol. Rev. 65 (1958) 386.
[91] H.-D. Block, Rev. Modern Phys. 34 (1962) 123.
[92] E. Gardner, B. Derrida, J. Phys. A: Math. Gen. 21 (1988) 271.
[93] Alia Abbaras, Benjamin Aubin, Florent Krzakala, Lenka Zdeborová, Rademacher complexity and spin glasses: a link between the replica and statistical theories of learning, in: Jianfeng Lu, Rachel Ward (Eds.), Proceedings of The First Mathematical and Scientific Machine Learning Conference, in: Proceedings of Machine Learning Research, 107, PMLR, 2020, pp. 27–54.
[94] S. Franz, G. Parisi, M. Sevelev, P. Urbani, F. Zamponi, M. Sevelev, SciPost Phys. 2 (2017) 019.
[95] S. Hwang, H. Ikeda, Phys. Rev. E 101 (2020) 052308.
[96] M. Mézard, J. Phys. A: Math. Gen. 22 (1989) 2181.
[97] A. Altieri, S. Franz, G. Parisi, J. Stat. Mech. Theory Exp. 2016 (2016) 093301.
[98] A. Altieri, Phys. Rev. E 97 (2018) 012103.
[99] E. Agoritsas, G. Biroli, P. Urbani, F. Zamponi, J. Phys. A 51 (2018) 085002.
[100] A. Manacorda, F. Zamponi, J. Phys. A 55 (2022) 334001.
[101] T. Maimbourg, J. Kurchan, F. Zamponi, Phys. Rev. Lett. 116 (2016) 015902.
[102] S. Franz, G. Parisi, P. Urbani, F. Zamponi, Proc. Natl. Acad. Sci. 112 (2015) 14539.
[103] M. Wyart, S. Nagel, T. Witten, Europhys. Lett. 72 (2005) 486.
[104] E. DeGiuli, A. Laversanne-Finot, G. Düring, E. Lerner, M. Wyart, Soft Matter 10 (2014) 5628.
[105] A. Sclocchi, P. Urbani, Phys. Rev. E 105 (2022) 024134.
[106] E. Lerner, G. Düring, M. Wyart, Europhys. Lett. 99 (2012) 58003.
[107] A. Ikeda, T. Kawasaki, L. Berthier, K. Saitoh, T. Hatano, Phys. Rev. Lett. 124 (2020) 058001.
[108] H. Ikeda, J. Chem. Phys. 153 (2020) 126102.
[109] Y. Nishikawa, A. Ikeda, L. Berthier, J. Stat. Phys. 182 (2021) 1.
[110] R.N. Chacko, P. Sollich, S.M. Fielding, Phys. Rev. Lett. 123 (2019) 108001.
[111] Y. Nishikawa, M. Ozawa, A. Ikeda, P. Chaudhuri, L. Berthier, Phys. Rev. X 12 (2022) 021001.
[112] M. Wyart, Phys. Rev. Lett. 109 (2012) 125502.
[113] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, F. Zamponi, Nature Commun. 5 (2014) 3725.
[114] P. Charbonneau, E.I. Corwin, G. Parisi, F. Zamponi, Phys. Rev. Lett. 114 (2015) 125504.
[115] Y. Kallus, Phys. Rev. E 93 (2016) 012902.
[116] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, F. Zamponi, Annu. Rev. Condens. Matter Phys. 8 (2017) 265.
[117] S. Franz, A. Sclocchi, P. Urbani, Phys. Rev. Lett. 123 (2019) 115702.
[118] S. Franz, A. Sclocchi, P. Urbani, SciPost Phys. 9 (2020) 012.
[119] P. Charbonneau, E.I. Corwin, R.C. Dennis, R.D.H. Rojas, H. Ikeda, G. Parisi, F. Ricci-Tersenghi, Phys. Rev. E 104 (2021) 014102.
[120] E. DeGiuli, E. Lerner, C. Brito, M. Wyart, Proc. Natl. Acad. Sci. 111 (2014) 17054.
[121] M. Müller, M. Wyart, Ann. Rev. Cond. Mat. 6 (2015) 177.
[122] G. Biroli, P. Charbonneau, Y. Hu, H. Ikeda, G. Szamel, F. Zamponi, J. Phys. Chem. B 125 (2021) 6244.
[123] H. Ikeda, P. Urbani, F. Zamponi, J. Phys. A 52 (2019) 344001.
[124] C. Brito, H. Ikeda, P. Urbani, M. Wyart, F. Zamponi, Proc. Natl. Acad. Sci. 115 (2018) 11736.
[125] S. Franz, A. Sclocchi, P. Urbani, J. Stat. Mech. Theory Exp. 2021 (2021) 023208.