IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **53** (2020) 125002 (53pp) https://doi.org/10.1088/1751-8121/ab73ac

Distribution of rare saddles in the *p*-spin energy landscape

Valentina Ros

Laboratoire de Physique de l'Ecole Normale Superieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité, Paris, France

E-mail: valentina.ros@lps.ens.fr

Received 10 December 2019, revised 30 January 2020
Accepted for publication 6 February 2020
Published 3 March 2020

Abstract

We compute the statistical distribution of index-1 saddles surrounding a given local minimum of the *p*-spin energy landscape, as a function of their distance to the minimum in configuration space and of the energy of the latter. We identify the saddles also in the region of configuration space in which they are subdominant in number (i.e. rare) with respect to local minima, by computing large deviation probabilities of the extremal eigenvalues of their Hessian. As an independent result, we determine the joint large deviation probability of the smallest eigenvalue and eigenvector of a GOE matrix perturbed with both an additive and multiplicative finite-rank perturbation.

Keywords: high-dimensional random landscapes, statistics of critical points, random matrix theory

(Some figures may appear in colour only in the online journal)

1. Introduction

High-dimensional systems are typically associated to complex, highly non-convex energy landscapes, in which the number of stationary points (local minima, maxima or saddles) increases steeply with the dimensionality. Classifying these points in terms of their energy, of their stability and of their location in the underlying configuration space is a topic that is of interest in a large variety of fields, including disordered systems [1–15, 50], ecology and biology [16–19], neural networks [20, 21], inference [22–26], game theory [27], string theory and cosmology [28, 29]. In many of these contexts, a crucial motivation for determining the distribution of stationary points is to understand how the energy functional is explored dynamically, through algorithms that proceed via local moves in configuration space, biased towards lower-energy configurations. When metastable local minima proliferate, indeed, the dynamical search of the global minimum (or optimal state) is likely hampered by the ruggedness and

glassiness of the landscape. In high-dimensional glassy systems, several features of the resulting slow dynamics (such as aging [39–41]) have been characterized in detail. However, it is still to large extent an open question [42, 43] how the system escapes dynamically from the metastable, trapping local minima via activated crossings of the surrounding energy barriers.

Addressing this question is notoriously challenging, as it requires to determine the energetic cost of the paths in the landscape connecting different local minima. It is clear that a pivotal role in fixing such cost is played by the critical points lying along the path, in particular by the saddles: characterizing how the saddles are arranged with respect to local minima and how they are connected in configuration space is therefore crucial. Key questions in this respect are: given a local minimum, what is the number and what is the energy distribution of the saddles that lie at a fixed distance from it in configuration space? Which among these saddles are *geometrically connected* to the minimum, meaning that there exist descending paths in the landscape that connect the saddle to the minimum? Do these saddles represent potential escape states for the system that is dynamically trapped in a metastable local minimum?

For random landscapes, these questions can be approached within a statistical framework. The so called spherical models are prototypical incarnations of random landscapes, that allow one for explicit analytical results [1, 30–34], including mathematically rigorous ones [35–38]. The pure p-spin is the simplest model belonging to this class: the energy functional is in this case a monomial of degree p with random coefficients and Gaussian statistics, defined on a sphere of large dimension $N \gg 1$. In this model the random fluctuations give rise to a rugged landscape, with an exponentially-large (in the dimension N) number $\mathcal{N} \sim \exp\left[N\Sigma + o(N)\right]$ of stationary points, Σ being their 'complexity'. These points are non-trivially distributed in terms of their energy and stability: local minima are typically confined below a certain energy level called the *threshold*, above which saddles of extensive index $k = O(N)$ dominate (the index being the number of unstable directions in configuration space). More precisely, at any value of energy below the threshold one typically finds an exponentially-large number $\mathcal{N}_k \sim \exp\left[N\Sigma_k + o(N)\right]$ of saddles of arbitrary non-extensive index $k = o(N)$. These saddles are distributed hierarchically, with complexities Σ_k that are strictly decreasing with k: the dominant (at the exponential scale in N) stationary points below the threshold are minima with $k = 0$, followed by index-1 saddles, index-2 saddles and so on [2, 11].

Because of the large-dimensionality of configuration space, for any given local minimum of the p-spin landscape the saddles lie in overwhelming majority at very large distance from it in configuration space, and are geometrically disconnected to it. Those saddles that are close and connected to the minimum are *atypical* in the sense that they constitute an exponentially-small (in N) fraction of the whole population: computing their complexity requires to condition explicitly to be nearby the reference minimum in configuration space. A calculation of this type was first performed in [45], where the *constrained* complexity of stationary points at fixed distance from a reference minimum was obtained through the replica formalism and within the so called annealed approximation (see also [46, 47] and the more recent [48]). More recently, the same results have been recovered within a quenched formalism exploiting the Kac–Rice formalism [44], and supplemented with the statistical analysis of the Hessian of the counted stationary points, that allowed to determine their stability. The stability analysis heavily relies on a connection with random matrix theory [11, 49, 50]. For the p-spin model, it is found that the stationary points that are closer to the minimum are typically saddles of index-1 connected geometrically to it, while those at larger distance are other local minima. As a consequence, information on the statistics of the energy barriers surrounding the minimum can be extracted from the energy distribution of the nearby index-1 saddles.

The information obtained in this way is however not fully complete, as it corresponds only to the saddles that are closest to the reference minimum. In other words, the calculation performed in [44] allows to identify only the saddles that lie in the region of configuration space where they are the typical stationary points (i.e. those having larger complexity). At larger distance from the reference minimum, it is likely that other index-1 saddles connected to the minimum are present, but are not traced as they have smaller complexity with respect to minima. The purpose of this work is to identify these saddles and determine their energy distribution and complexity.

To target the saddles in the regions of configuration space dominated by minima, we need to impose explicit constraints on the Hessian matrices of the stationary points we are counting. These matrices have the statistics of a GOE matrix deformed with finite-rank perturbations, that are generated by conditioning the stationary point to be at fixed distance from the reference minimum. Computing the complexity requires to determine the joint probability distribution of the smallest eigenvalue of such deformed GOE matrix, and of the corresponding eigenvector. Random matrix ensembles deformed with low-rank perturbations have been widely investigated in the literature: extensive effort has been devoted in particular to the characterization of the eigenvalues transitions (named *BBP transitions* after Baik, Ben Arous and Péché [51]) occurring when outliers (or isolated eigenvalues) appear in the spectrum. For deformed Wigner matrices (in particular in the case of Gaussian entries), several results have been derived on the typical value of the isolated eigenvalues [52–56], on their fluctuations [57–59] and on the typical value of the eigenvector projection along the direction of the perturbation [60, 61]. The large deviations of the isolated eigenvalue in the case of a deterministic additive perturbation have been determined in [62]. This result has been recently pushed forward in [63], by computing the *joint* large deviations of the isolated eigenvalue and of the projection of the corresponding eigenvector along the direction of the additive perturbation. This paper builds on [63] to extend the large deviation results to the case in which the GOE matrix is deformed with a combination of both an additive and multiplicative perturbations, which is relevant to characterize the statistics of the p-spin Hessian matrices at a critical point.

The paper is split into three parts: in the the first part (section 2) we present the results on the p-spin energy landscape. In the second part (section 3) we state the large deviation functions of the smallest eigenvalue and eigenvectors of a deformed GOE matrix in general form, and summarize the main steps of the derivation. The third part (section 4) is devoted to the derivation of these large deviation principles. The second and third parts of the paper are formulated in general terms, and can be read independently from the first. A more detailed summary of the structure of the paper is given at the beginning of each part. The conclusions are given in section 5.

2. Part I: rare saddles in the landscape of the spherical p-spin model

In this first part of the work, we discuss how the complexity of index-1 saddles of the spherical p-spin model is obtained, and present the results of the calculation. In section 2.1 we summarize the general formalism for the computation of the complexity and we recall the statistical properties of the Hessian of the energy landscape, evaluated at the stationary points. In section 2.2 we set up the calculation of the complexity of the *atypical* saddles, and we state the expressions of the large deviation functions for the minimal eigenvalue and eigenvector of the Hessian matrices. In section 2.3 we present the resulting complexity of the index-1 saddles at fixed given overlap from a reference minimum of the landscape, and we comment on the implications for the dynamical exploration of the landscape.

2.1. The p-spin energy landscape: total constrained complexity and Hessian statistics

2.1.1. Constrained complexity and Kac–Rice formula. We consider the energy landscape of the spherical p-spin model with $p \geqslant 3$:

$$E[\mathbf{s}] = -\sum_{i_1<i_2\cdots<i_p} J_{i_1,i_2,\ldots,i_p} s_{i_1} s_{i_2} \ldots s_{i_p}, \tag{1}$$

where $i_k \in \{1,\cdots,N\}$, the configurations $\mathbf{s} = (s_1,\cdots,s_N)$ lie on the surface of a sphere and satisfy $\sum_{i=1}^{N} s_i^2 = N$, and their closeness is measured in terms of the overlap $q(\mathbf{s},\mathbf{s}') = \mathbf{s}\cdot\mathbf{s}'/N$. The quenched random couplings $J_{i_1,i_2,\ldots,i_p}$ are independent Gaussian variables with zero mean and variance $\langle J_{\mathbf{i}}^2\rangle = p!/2N^{p-1}$. The random energy landscape (1) is therefore itself Gaussian, with zero average $\langle E[\mathbf{s}]\rangle = 0$ and covariance

$$\langle E[\mathbf{s}]\, E[\mathbf{s}']\rangle = \frac{N}{2}\left(\frac{\boldsymbol{s}\cdot\boldsymbol{s}'}{N}\right)^p \tag{2}$$

that is isotropic, meaning that it depends on $\boldsymbol{s},\boldsymbol{s}'$ only through their overlap. In the following, we denote the energy density of a configuration $\mathbf{s}$ by $\epsilon = \lim_{N\to\infty} E[\mathbf{s}]/N$. The threshold value of the energy is $\epsilon_{\rm th} = -[2(p-1)/p]^{1/2}$, while $\epsilon_{\rm gs}$ denotes the density of the ground state configurations.

At energy densities $\epsilon > \epsilon_{\rm th}$ the landscape is dominated by saddles with a huge index $k = O(N)$: this portion of the landscape is easily explored dynamically since stationary points have plenty of directions in configuration space in which the energy landscape is descending [39], and it is not of interest in the light of activated dynamics. We therefore restrict to the energy regime $\epsilon_{\rm gs} \leqslant \epsilon \leqslant \epsilon_{\rm th}$, which is dominated by stationary points that are either trapping local minima or saddles with few negative directions $k \sim o(N)$. The complexities $\Sigma_k(\epsilon)$ count the number of such stationary points of energy density ϵ and index k, at the exponential scale in N. The *total* complexity $\Sigma(\epsilon)$ is obtained as

$$\Sigma(\epsilon) = \max_k \Sigma_k(\epsilon). \tag{3}$$

For the spherical p-spin $\Sigma(\epsilon) = \Sigma_0(\epsilon)$ for all $\epsilon_{\rm gs} \leqslant \epsilon \leqslant \epsilon_{\rm th}$: at each value of energy below the *threshold* the typical (most numerous) stationary points are local minima.

In the following we aim at characterizing stationary points $\mathbf{s}$ of energy density ϵ and index k that are at overlap $q = \mathbf{s}\cdot\mathbf{s}_0/N$ with respect to some fixed local minimum $\mathbf{s}_0$ of the landscape, extracted with uniform measure among those at energy ϵ_0. We denote with $\Sigma_k(\epsilon, q|\epsilon_0)$ the corresponding complexities, and with $\Sigma(\epsilon, q|\epsilon_0)$ the total one, obtained maximizing over k. More precisely, following the notation of [44] we define rescaled spin configurations on the unit sphere, $\boldsymbol{\sigma} = \mathbf{s}/\sqrt{N}$, and introduce the rescaled energy $h[\boldsymbol{\sigma}] \equiv \sqrt{2/N}E[\sqrt{N}\boldsymbol{\sigma}]$. Given a reference local minimum $\boldsymbol{\sigma}^0$ drawn at random from the population of minima with energy ϵ_0 ($\epsilon_{\rm gs} \leqslant \epsilon_0 \leqslant \epsilon_{\rm th}$), we denote with $\mathcal{N}_{\boldsymbol{\sigma}^0}(\epsilon, q|\epsilon_0)$ the number of stationary points with energy ϵ that are at fixed overlap $\boldsymbol{\sigma}^0\cdot\boldsymbol{\sigma} = q$ with the minimum, and define the associated total quenched complexity as:

$$\Sigma(\epsilon, q|\epsilon_0) = \lim_{N\to\infty}\frac{1}{N}\Big\langle \log\mathcal{N}_{\boldsymbol{\sigma}^0}(\epsilon, q|\epsilon_0)\Big\rangle_0, \tag{4}$$

where the average $\langle\cdot\rangle_0$ is over both the local minima of energy ϵ_0 at fixed realization of the random energy field (1), and over the different realizations of the latter. Notice that for $q = 0$, which is the typical value of the overlap between an arbitrary pair of stationary points, the constraint is ineffective and (4) reproduces the well-known complexity curve of local minima $\Sigma(\epsilon) = \Sigma_0(\epsilon)$.

The total complexity (4) has been computed in [44] using the Kac–Rice formula and its generalizations [8, 11, 25]. From that calculation it followed that the quenched complexity actually coincides with its annealed counterpart computed in [46], obtained exchanging the average with the logarithm in (4). The latter be easily obtained as the large-N asymptotic of the Kac–Rice formula for the first moment of $\mathcal{N}_{\boldsymbol{\sigma}^0}$. To state the formula, we introduce the gradient $\mathbf{g}[\boldsymbol{\sigma}]$ of the energy field $h[\boldsymbol{\sigma}]$: since the functional is restricted to the sphere, its gradient lies in the $M = (N-1)-$dimensional tangent plane to the sphere at the point $\boldsymbol{\sigma}$; similarly, the Hessian matrix $\mathcal{H}[\boldsymbol{\sigma}]$ collects the components of the second derivatives of $h[\boldsymbol{\sigma}]$ along the directions corresponding to some basis $\{\mathbf{e}_i[\boldsymbol{\sigma}]\}_{i=1}^{M}$ spanning the tangent plane. In terms of these quantities, the constrained complexity reads:

$$\Sigma(\epsilon, q|\epsilon_0) = \lim_{N\to\infty} \frac{1}{N} \log \left[\int d\boldsymbol{\sigma}\, \delta\left(\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}^0 - q\right) \mathcal{E}_{\boldsymbol{\sigma}|\boldsymbol{\sigma}_0}(\epsilon, q|\epsilon_0)\; p_{\boldsymbol{\sigma}|\boldsymbol{\sigma}^0}(\mathbf{0}, \epsilon) \right], \tag{5}$$

where the integration is over the configurations $\boldsymbol{\sigma}$ at fixed overlap q with the reference minimum, $p_{\boldsymbol{\sigma}|\boldsymbol{\sigma}^0}(\mathbf{0}, \epsilon)$ denotes the joint density function of the gradient and field $(\mathbf{g}[\boldsymbol{\sigma}], h[\boldsymbol{\sigma}])$, conditioned to the values of gradient and field at $\boldsymbol{\sigma}^0$ and evaluated at $(\mathbf{0}, \sqrt{2N}\epsilon)$, and $\mathcal{E}_{\boldsymbol{\sigma}|\boldsymbol{\sigma}_0}(\epsilon, q|\epsilon_0)$ is the following expectation value

$$\mathcal{E}_{\boldsymbol{\sigma}|\boldsymbol{\sigma}_0}(\epsilon, q|\epsilon_0) = \left\langle\, |\det\mathcal{H}[\boldsymbol{\sigma}]| \;\middle|\; \begin{cases} \mathbf{g}[\boldsymbol{\sigma}^0] = 0, \mathbf{g}[\boldsymbol{\sigma}] = 0 \\ h[\boldsymbol{\sigma}^0] = \sqrt{2N}\epsilon_0, h[\boldsymbol{\sigma}] = \sqrt{2N}\epsilon \end{cases} \right\rangle. \tag{6}$$

Notice that, while in principle the quantity inside the logarithm in (5) depends on the particular local minimum $\boldsymbol{\sigma}_0$, as a consequence of the isotropy of the p-spin covariances the dependence is only on the overlap parameter q. Therefore the uniform average on the local minima at fixed value of q yields a constant factor equal to one (see the Supplemental Material of [44], in particular section G.1, for the derivation of this formula). The asymptotic of (5) is determined by computing the conditional distribution of the energy field and of its derivatives, which can be determined explicitly due to Gaussianity. In particular, the average of the Hessian determinant in (5) is done over the distribution of $\mathcal{H}[\boldsymbol{\sigma}]$ *conditioned* to the fact that $\boldsymbol{\sigma}$ is a stationary point of energy density ϵ, at fixed overlap q from another stationary point (a minimum) of energy ϵ_0, as we recall in the following section.

2.1.2. Statistics of the Hessians at overlap q from a reference minimum. In absence of conditioning (equivalently, for $q = 0$) the Hessian at a stationary point $\boldsymbol{\sigma}$ has the statistical distribution of a GOE matrix, shifted by a constant diagonal matrix that depends only on the energy density ϵ. This follows from the isotropy of the correlations (2), which translates into a matrix distribution that is itself invariant under basis rotations in the tangent plane. The energy-dependent shift follows from the spherical constraint imposed on the variables $\boldsymbol{\sigma}$, and it is such that for any $\epsilon < \epsilon_{\text{th}}$ the typical configuration of the Hessian density of states (in the large-N limit) is a semicircle which is entirely supported on the positive semi-axis, implying that typical stationary points are minima. Saddles are generated by large deviations of the smallest eigenvalues of the Hessian, that are pulled out of the bulk of the density of states and into the negative semi-axis: this happens with a large-deviation probability that is exponentially decaying in N [64], implying the exponential suppression of the complexity of saddles with respect to that of minima [2, 11].

When we enforce the point $\boldsymbol{\sigma}$ to be at finite overlap q from another local minimum $\boldsymbol{\sigma}_0$, the isotropy is broken along the direction in configuration space that connects the two stationary points. At the level of the Hessian statistics, this translates into rank-1 perturbations (both additive and multiplicative) to an otherwise GOE distributed matrix, that depend explicitly

on the parameters ϵ, ϵ_0 and q [44] (see also [12]). To express it compactly, it is convenient to choose a basis $\{\mathbf{e}_i[\boldsymbol{\sigma}]\}_{i=1}^{M}$ in the tangent plane at $\boldsymbol{\sigma}$ in such a way that the last vector $\mathbf{e}_M = (q\boldsymbol{\sigma} - \boldsymbol{\sigma}_0)/\sqrt{1-q^2}$ is the only one having a projection on $\boldsymbol{\sigma}_0$, while all the remaining ones are arbitrary vectors spanning the space orthogonal to $\boldsymbol{\sigma}, \boldsymbol{\sigma}_0$, see figure 1. With this choice of basis the conditioned Hessian is distributed as:

$$\mathcal{H}[\boldsymbol{\sigma}] \sim \mathcal{M} - \sqrt{2N}p\epsilon\, \mathbb{1}, \tag{7}$$

where $\mathbb{1}$ is the identity matrix and $\mathcal{M}$ is an M-dimensional matrix with the following properties: the $(M-1)-$dimensional block made of the entries $m_{ij(\neq M)}$ has GOE statistics with zero average and variance

$$\sigma^2(p) = p(p-1); \tag{8}$$

the elements m_{iM} for $i \neq M$ have a different variance $\Delta^2(q) < \sigma^2$ depending explicitly on the overlap parameter q, and the element m_{MM} has a non-zero average $\mu(q, \epsilon, \epsilon_0)$ and yet another variance $\tilde{\Delta}^2(q) < \Delta^2(q)$. These functions depend explicitly on p: for $p = 3$, for instance, one finds $\tilde{\Delta}^2(q) = 0$. Their explicit form is recalled in appendix A.

To further simplify the notation, we introduce an $M \times M$ deterministic matrix of the form:

$$F(q) \equiv \mathbb{1} - \left[1 - \frac{\Delta(q)}{\sigma}\right] \mathbf{e}_M \mathbf{e}_M^T, \tag{9}$$

and define a complex (purely imaginary) variable $\zeta(q)$ through the identity:

$$\frac{\Delta^4(q)}{\sigma^2} + [\zeta(q)]^2 = \tilde{\Delta}^2(q). \tag{10}$$

The matrix $\mathcal{M}$ can then be re-written as:

$$\mathcal{M} = F(q)\mathcal{X}F(q) + \left(\sqrt{N}\mu + \zeta(q)\,\xi\right) \mathbf{e}_M \mathbf{e}_M^T. \tag{11}$$

Here $\mathcal{X}$ is a GOE matrix with variance (8) and ξ is a Gaussian random variable independent of $\mathcal{X}$, having zero average and unit variance. Notice that the variance of the MM element of the perturbed matrix $F(q)\mathcal{X}F(q)$ equals to Δ^4/σ^2, which is different with respect to the variance of m_{MM}: the fluctuating variable ξ is added to compensate for this difference. As we recall in the next section, the main effect of the finite-rank perturbation in (11) is to modify the typical configuration of the density of states giving rise to an isolated eigenvalue.

2.1.3. The isolated eigenvalue of the Hessian and the saddles. When the finite-rank perturbations to the GOE matrix $\mathcal{M}$ in (7) are sufficiently strong, they generate a sub-leading correction to the density of states

$$\rho_\epsilon(\lambda) = \frac{\sqrt{4\sigma^2(p) - (\lambda + \sqrt{2}p\epsilon)^2}}{2\pi\sigma^2(p)} \tag{12}$$

of the Hessian matrix, in the form of a single eigenvalue $\lambda_0(q, \epsilon, \epsilon_0)$ that is isolated and detached from the support of (12), meaning that $\lambda_0(q, \epsilon, \epsilon_0) < -2\sigma^2 - \sqrt{2}\,p\,\epsilon$. The explicit expression of this eigenvalue has been determined in [44]. It is more conveniently given in terms of the resolvent[1] of the unperturbed GOE matrix $\mathcal{X}$ with variance σ:

[1] The resolvent is defined for $|z| > 2\sigma$ as the solution of the quadratic equation $\sigma^2 G_\sigma^2(z) - zG_\sigma(z) + 1 = 0$ satisfying $G_\sigma(z) \to 0$ as $|z| \to \infty$.

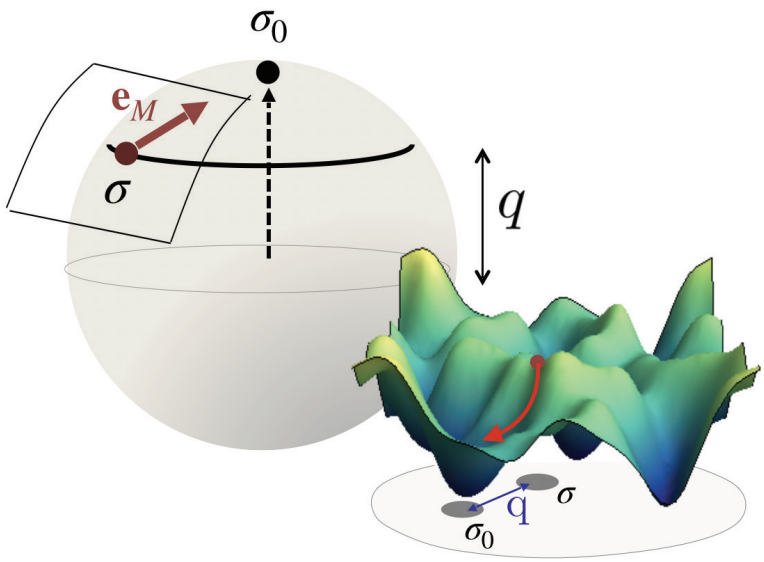

Figure 1. Schematic representation of configuration space, with the reference minimum $\boldsymbol{\sigma}_0$ and a saddle $\boldsymbol{\sigma}$ at overlap q, that is geometrically connected to the minimum. The vector $\mathbf{e}_M$ lies in the tangent plane to the sphere at $\boldsymbol{\sigma}$, along the direction connecting $\boldsymbol{\sigma}$ to the reference minimum $\boldsymbol{\sigma}_0$.

$$G_\sigma(z) = \left\langle \frac{1}{M}\mathrm{Tr}\frac{1}{z-\mathcal{X}} \right\rangle \overset{z \text{ real}}{=} \frac{1}{2\sigma^2}\left(z - \mathrm{sign}(z)\sqrt{z^2-4\sigma^2}\right) \in \left[-\frac{1}{\sigma}, \frac{1}{\sigma}\right]. \tag{13}$$

Setting

$$\lambda_0(q,\epsilon,\epsilon_0) = \lambda_{\min}^{\mathrm{typ}}(q,\epsilon,\epsilon_0) - \sqrt{2}p\epsilon, \tag{14}$$

it is found in [44] that the typical value $\lambda_{\min}^{\mathrm{typ}}(q,\epsilon,\epsilon_0)$ of the smallest eigenvalue of $\mathcal{M}$ is the solution of $\lambda - \mu(q,\epsilon,\epsilon_0) - \Delta^2(q)G_\sigma(\lambda) = 0$, and reads explicitly:

$$\lambda_{\min}^{\mathrm{typ}}(q,\epsilon,\epsilon_0) = \frac{1}{2(\sigma^2-\Delta^2)}\left(2\mu\sigma^2 - \Delta^2\mu + \Delta^2\sqrt{\mu^2 - 4(\sigma^2-\Delta^2)}\right) = \mu + \Delta^2 G_{\sigma'}(\mu), \tag{15}$$

where $G_{\sigma'}(\mu)$ has a modified variance

$$\sigma'(p,q) = \sqrt{\sigma^2(p) - \Delta^2(q)}. \tag{16}$$

Notice that this expression is independent of the Gaussian fluctuations with variance $\zeta(q)$ of the element m_{MM}. Using the equation satisfied by $\lambda_{\min}^{\mathrm{typ}}$ we get:

$$\lambda_{\min}^{\mathrm{typ}}(q,\epsilon,\epsilon_0) = G_\sigma^{-1}\left(G_{\sigma'}(\mu)\right) = \frac{1}{G_{\sigma'}(\mu)} + \sigma^2 G_{\sigma'}(\mu), \tag{17}$$

where

$$G_\sigma^{-1}(z) = \frac{1}{z} + \sigma^2 z \tag{18}$$

is the inverse of the resolvent operator, restricted to the domain $|z| < 1/\sigma$. This expression is consistent provided that the argument of G_σ^{-1} belongs to $[-1/\sigma, 1/\sigma]$. Assuming that $G_{\sigma'}(\mu) < 0$, this gives:

$$G_{\sigma'}(\mu) > -\frac{1}{\sigma} \longrightarrow \mu < -\sigma\left[1 + \frac{(\sigma')^2}{\sigma^2}\right], \tag{19}$$

which identifies the regime of parameters for which the isolated eigenvalue exists. We denote the threshold value with:

$$\mu_c(p,q) \equiv -\sigma(p)\left[1 + \left(\frac{\sigma'(p)}{\sigma(p)}\right)^2\right] = -\sqrt{p(p-1)}\left[2 - \frac{\Delta^2(q)}{p(p-1)}\right]. \tag{20}$$

In this regime, the eigenvector $\mathbf{v}_0$ associated to $\lambda_{\min}^{\text{typ}}$ has a projection $\mathbf{v}_0 \cdot \mathbf{e}_M[\boldsymbol{\sigma}]$ along the direction connecting the two stationary points which remains non-zero as $N \to \infty$. Notice that $\lim_{\sigma' \to 0} G_{\sigma'}(\mu) = 1/\mu$, implying that when $\Delta(q) \to \sigma$ the eigenvalue exists for $\mu < -\sigma$ and reduces to $\lambda_{\min}^{\text{typ}} = \mu + \sigma^2/\mu$, reproducing the well-known expression resulting from a purely additive perturbation [52–55]. In presence of a multiplicative perturbation given by the matrices $F(q)$, the same form holds with $1/\mu$ replaced with $G_{\sigma'}(\mu)$.

When the parameters are such that the shifted eigenvalue $\lambda_0(q, \epsilon, \epsilon_0) < 0$, the associated stationary points are saddles of index-1. As found in [44], this happens when the overlap q with the reference minimum is large enough (for any fixed ϵ, larger than a given $q_{\text{ms}}(\epsilon|\epsilon_0)$, see figure 2): the total complexity (4) is therefore contributed by saddles for large enough q. These saddles are *geometrically connected* to the reference minimum $\boldsymbol{\sigma}_0$, meaning that their unstable direction has an $O(1)$ projection along the direction pointing towards $\boldsymbol{\sigma}_0$ in configuration space. Notice that no large deviation calculation is necessary to find these saddles, as the *typical* configurations of the Hessian have a negative mode: in other words, at these values of the overlap index-1 saddles are the typical, exponentially most numerous stationary points. At smaller values of q, the typical stationary points are instead minima with no isolated eigenvalue; in this regime the complexity of saddles has to be obtained with a large deviation calculation, by conditioning explicitly the Hessian to exhibit one negative isolated eigenvalue.

2.2. Computing the complexity of atypical saddles

2.2.1. The constrained complexity of saddles. We now give a formula for the constrained complexity of saddles at overlap q with the reference minimum, in the *annealed* approximation. Let us denote with $\mathcal{N}_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0)$ the number of stationary points $\boldsymbol{\sigma}$ having an Hessian with smallest eigenvalue taking a given value $\lambda_{\min} = \lambda$ and such that the corresponding eigenvector $\mathbf{v}_{\min}$ has a macroscopic projection $u_{\min} = |\mathbf{v}_{\min} \cdot \mathbf{e}_M[\boldsymbol{\sigma}]|^2 = u > 0$ along the direction connecting the two stationary points in configuration space. The complexity $\Sigma(\epsilon, q, \lambda, u|\epsilon_0)$ of these points in the annealed approximation is given by:

$$\Sigma(\epsilon, q, \lambda, u|\epsilon_0) = \lim_{N\to\infty} \frac{1}{N} \log \left\langle \mathcal{N}_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0) \right\rangle_0, \tag{21}$$

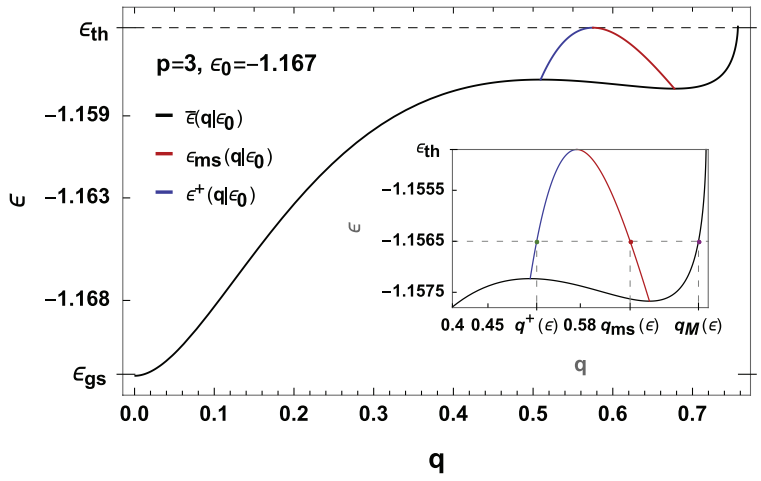

Figure 2. Plot of the energy curves $\bar{\epsilon}(q|\epsilon_0)$, $\epsilon_{\rm ms}(q|\epsilon_0)$ and $\epsilon^+(q|\epsilon_0)$. *Inset.* Zoom of the main plot. The dashed lines identify the overlaps $q^+(\epsilon|\epsilon_0)$, $q_{\rm ms}(\epsilon|\epsilon_0)$ and $q_M(\epsilon|\epsilon_0)$ for $\epsilon = -1.1565$.

where the average number can be written as:

$$\left\langle \mathcal{N}_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0) \right\rangle_0 = \int d\boldsymbol{\sigma}\, \delta\left(\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}^0 - q\right) \left\langle |\det \mathcal{H}[\boldsymbol{\sigma}]| \left| \begin{cases} \mathbf{g}[\boldsymbol{\sigma}^0] = 0, \mathbf{g}[\boldsymbol{\sigma}] = 0 \\ h[\boldsymbol{\sigma}^0] = \sqrt{2N}\epsilon_0, h[\boldsymbol{\sigma}] = \sqrt{2N}\epsilon \\ \lambda_{\rm min} = \lambda,\ u_{\rm min} = u \end{cases} \right. \right\rangle \times p_{\boldsymbol{\sigma}|\boldsymbol{\sigma}^0}(\mathbf{0}, \epsilon)\, \mathbb{G}_{\boldsymbol{\sigma}|\boldsymbol{\sigma}^0}(\lambda, u)\,. \tag{22}$$

In this modified version of the Kac–Rice formula, the expectation value of the Hessian is conditioned also to the event $\lambda_{\rm min} = \lambda$ and $u_{\rm min} = u$. The case $\lambda < 0$ corresponds to saddles with *at least* one unstable direction. The constraint on the overlap $u_{\rm min} = u$ is added to track whether the saddles are geometrically connected to $\boldsymbol{\sigma}_0$ (when $u > 0$), or whether the downhill direction is uncorrelated with the minimum $\boldsymbol{\sigma}_0$ (when $u = 0$). The function $\mathbb{G}_{\boldsymbol{\sigma}|\boldsymbol{\sigma}^0}(\lambda, u)$ is the joint distribution of $(\lambda_{\rm min}, u_{\rm min})$ induced by the statistics of the conditioned Hessian described in section 2.1.2.

In appendix B we argue that conditioning on λ and u does not modify the typical density of states of the Hessian to leading order in N, which therefore remains equal to (12). The effect of the conditioning is (at most) to generate isolated eigenvalues, that are sub-leading corrections to the density of states. As a consequence, to (exponential) order in N the expectation value of the determinant in (22) is insensitive to the conditioning on the smallest eigenvalue. Additionally, the distribution $\mathbb{G}_{\boldsymbol{\sigma}|\boldsymbol{\sigma}^0}(\lambda, u)$ depends on $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^0$ only through the parameters q, ϵ and ϵ_0, because the full distribution of the Hessian does. We re-label it as $\mathbb{G}_{\epsilon,q|\epsilon_0}(\lambda, u)$ in the following. For values of λ, u that are different with respect to the typical ones, $\mathbb{G}_{\epsilon,q|\epsilon_0}(\lambda, u)$ is a large deviation probability with a given rate function to be determined:

$$\lim_{N\to\infty} \frac{\log \mathbb{G}_{\epsilon,q|\epsilon_0}(\lambda, u)}{N} = -L_{\epsilon,q|\epsilon_0}(\lambda, u). \tag{23}$$

It follows from these considerations that we can re-write (21) as:

$$\Sigma(\epsilon, q, \lambda, u|\epsilon_0) = \Sigma(\epsilon, q|\epsilon_0) - L_{\epsilon,q|\epsilon_0}(\lambda, u), \tag{24}$$

where $\Sigma(\epsilon, q|\epsilon_0)$ is the total constrained complexity already computed in [44]. In the following, we shall consider *typical* values $u_{\rm typ}(\lambda)$ of the overlap u, defined as:

$$u_{\rm typ}(\lambda) \equiv \underset{u\in[0,1]}{\rm argmin}\ L_{\epsilon,q|\epsilon_0}(\lambda, u), \tag{25}$$

and set

$$F_{\epsilon,q|\epsilon_0}(\lambda) \equiv L_{\epsilon,q|\epsilon_0}(\lambda, u_{\rm typ}(\lambda)). \tag{26}$$

The complexity of the most numerous stationary points with $\lambda_{\rm min} = \lambda$ is then:

$$\Sigma(\epsilon, q, \lambda|\epsilon_0) = \Sigma(\epsilon, q|\epsilon_0) - F_{\epsilon,q|\epsilon_0}(\lambda), \tag{27}$$

and thus it is readily obtained from the large deviation rate $F_{\epsilon,q|\epsilon_0}(\lambda)$ of the smallest eigenvalue of an Hessian. Saddles are obtained setting $\lambda < 0$. The second and third parts of this work are devoted to the computation of the rate function $F_{\epsilon,q|\epsilon_0}(\lambda)$. In the following section, we adapt the general result to the case of the p-spin Hessians.

2.2.2. Large deviations of the smallest eigenvalue of the Hessians. In the third part of this work we derive the large deviation function of the smallest eigenvalue of matrices of the general form:

$$\mathcal{Y} = \left(\mathbb{1} - \frac{\beta}{1+\beta}\,\mathbf{e}_M\mathbf{e}_M^T\right)\mathcal{X}\left(\mathbb{1} - \frac{\beta}{1+\beta}\,\mathbf{e}_M\mathbf{e}_M^T\right) + \theta\,\mathbf{e}_M\mathbf{e}_M^T, \tag{28}$$

where $\mathcal{X}$ is a GOE matrix with variance σ^2, β is a non-negative constant and θ is a Gaussian random variable with mean $\overline{\theta} < 0$ and variance σ_θ^2. The Hessian matrices (11) follow this distribution, with

$$\sigma \to \sigma(p) \equiv \sqrt{p(p-1)}, \quad \beta \to \frac{\sqrt{p(p-1)}}{\Delta(q)} - 1, \quad \overline{\theta} \to \mu(q, \epsilon, \epsilon_0), \tag{29}$$

and

$$\sigma_\theta^2 \to \zeta^2(q) = \tilde{\Delta}^2(q) - \frac{\Delta^4(q)}{\sigma^2}, \tag{30}$$

where the explicit expressions of these functions are given in appendix A. We let $\mathcal{F}_{\epsilon,q|\epsilon_0}(\lambda)$ be the corresponding rate function for the minimal eigenvalue. Given the diagonal shift in (7), we have that the rate in (26) is obtained as:

$$F_{\epsilon,q|\epsilon_0}(\lambda) = \mathcal{F}_{\epsilon,q|\epsilon_0}(\lambda + \sqrt{2}\,p\,\epsilon). \tag{31}$$

We not adapt the general result of section 3.3 to this case. We introduce the threshold values:

$$\lambda_p^{\pm}(\epsilon, q|\epsilon_0) \equiv x^{\pm}_{\sigma(p)}\left(\mu(q, \epsilon, \epsilon_0), \frac{\sqrt{p(p-1)}}{\Delta} - 1\right) - \sqrt{2}\,p\,\epsilon, \tag{32}$$

where the functions $x_\sigma^{\pm}$ are given in (81). Given the shifted variance (16) and the critical value (20), we define the following three regimes:

- Regime A: $-2\sigma'(p, q) < \mu(q, \epsilon, \epsilon_0) < 0$
- Regime B.1: $\mu_{\rm c}(p, q) \leqslant \mu(q, \epsilon, \epsilon_0) \leqslant -2\sigma'(p, q)$
- Regime B.2: $\mu(q, \epsilon, \epsilon_0) < \mu_{\rm c}(p, q)$.

These three Regimes can be understood in terms of the typical value of the smallest eigenvalue $\lambda_{\rm min}^{\rm typ}(q,\epsilon,\epsilon_0)$ of the matrix (11): in Regime B.2 the eigenvalue isolated from the bulk of the density of states, $\lambda_{\rm min}^{\rm typ}(q,\epsilon,\epsilon_0) < -2\sigma(p)$, see (19). In regime B.1. it holds instead $\lambda_{\rm min}^{\rm typ}(q,\epsilon,\epsilon_0) = -2\sigma(p)$, and the quantities (32) are real. The Regime A corresponds to values of the parameters q,ϵ and ϵ_0 for which the quantities (32) are complex. Notice that it always holds $\mu_c(p,q) < -2\sigma'(p,q) \leqslant 0$. When $\Delta(q) \neq \sigma(p)$, we find $\sigma'(p) \to 0$: therefore, Regime A is present only when the multiplicative perturbation to the Hessian is present.

To state the form of the large deviation function, we further introduce the function:

$$\mu_p^*(x|q,\epsilon,\epsilon_0) = \theta_0^*\left(x\Big|\,\sigma(p), \frac{\sqrt{p(p-1)}}{\Delta} - 1, \mu(q,\epsilon,\epsilon_0), \zeta^2(q)\right), \tag{33}$$

where the function θ_0^* is defined in (179)[2]. Given these quantities, the large deviation function $F_{\epsilon,q|\epsilon_0}(\lambda)$ reads as follows:

- In Regime A,

$$F_{\epsilon,q|\epsilon_0}(\lambda) = \mathcal{G}_0(\lambda + \sqrt{2}\,p\,\epsilon). \tag{35}$$

- In Regime B.1,

$$F_{\epsilon,q|\epsilon_0}(\lambda) = \begin{cases} \mathcal{G}_{q,\epsilon|\epsilon_0}(\lambda + \sqrt{2}\,p\,\epsilon) & \lambda_p^-(\epsilon,q|\epsilon_0) < \lambda < \lambda_p^+(\epsilon,q|\epsilon_0) \\ \mathcal{G}_0(\lambda + \sqrt{2}\,p\,\epsilon) & \lambda < \lambda_p^- \text{ or } \lambda_p^+ < \lambda < -2\sigma(p) - \sqrt{2}p\epsilon \\ . \end{cases} \tag{36}$$

- In Regime B.2,

$$F_{\epsilon,q|\epsilon_0}(\lambda) = \begin{cases} \mathcal{G}_{q,\epsilon|\epsilon_0}(\lambda + \sqrt{2}\,p\,\epsilon) & \lambda_p^-(\epsilon,q|\epsilon_0) < \lambda < -2\sigma(p) - \sqrt{2}p\epsilon \\ \mathcal{G}_0(\lambda + \sqrt{2}\,p\,\epsilon) & \lambda < \lambda_p^-(\epsilon,q|\epsilon_0). \end{cases} \tag{37}$$

Here the large deviation function $\mathcal{G}_0(x)$ is the one of an unperturbed GOE matrix, given by [64]:

$$\mathcal{G}_0(x) = \int_x^{-2\sigma} \frac{\sqrt{z^2-4\sigma^2}}{2\sigma^2}\mathrm{d}z = \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \frac{1}{2} + \log\sigma, \tag{38}$$

where for $x < -2\sigma$:

$$\mathcal{I}(x) = \log\left(-\frac{x}{2} + \frac{1}{2}\sqrt{x^2-4\sigma^2}\right) - \frac{1}{2} + \frac{x^2}{4\sigma^2} + \frac{x}{4\sigma^2}\sqrt{x^2-4\sigma^2}. \tag{39}$$

The other rate function is obtained as:

$$\mathcal{G}_{q,\epsilon|\epsilon_0}(x) \equiv \mathcal{G}_{\theta,\beta}(x)\Big|_{\theta\to\mu_p^*,\ \beta\to\frac{\sqrt{p(p-1)}}{\Delta(q)}-1} \tag{40}$$

[2] For generic p, this function has a lengthy expression in terms of the parameters q,ϵ,ϵ_0. In the special case $p=3$, however, some simplifications occur due to the fact that $\tilde{\Delta}(q)=0$, meaning that in this case the *MM*-element of the Hessian does not fluctuate. In this particular case we find:

$$\mu_{p=3}^* = \mu - \frac{(1-q^2)x + \sqrt{4\mu^2(1+q^2)^2 - 4\mu(1+3q^4+4q^2)x + (3q^2+1)^2x^2 + 24(1-q^2)^2}}{2\,(1+q^2)}. \tag{34}$$

where the explicit form of $\mathcal{G}_{\theta,\beta}(x)$ is given in (82). In the regimes in which the large deviation function equals to $\mathcal{G}_0(x)$ it holds $u_{\rm typ}(\lambda) = 0$, while in the other regimes one finds $u_{\rm typ}(\lambda) > 0$: therefore, the latter regimes correspond to the saddles that are geometrically connected to the reference minimum. In the following, these results are used to determine statistical distribution of index-1 saddles.

2.2.3. Quenched versus annealed complexity: a comment. Before discussing the results of the complexity calculation of saddles, it is necessary to comment on the 'annealed' nature of the calculation we are performing. The complexity in (21) gives the asymptotic value of the *average* number of stationary points with the desired properties; this may in principle differ from the asymptotic value of the *typical* number of such stationary points, that is controlled by the so-called quenched complexity which is obtained exchanging the average and the logarithm in (21). The calculation of the latter is in general more involved; it requires to resort to representation of the logarithm in terms of higher moments of the number of stationary points,

$$\left\langle \log \mathcal{N}_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0) \right\rangle_0 = \lim_{n\to 0} \frac{\langle \mathcal{N}^n_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0)\rangle_0 - 1}{n}, \tag{41}$$

and to analytically continue the expression of these moments in order to take the limit $n \to 0$. As shown explicitly in [44], when computing the *total* constrained complexity $\Sigma(\epsilon, q|\epsilon_0)$ the two procedures are equivalent. The computation of the quenched complexity through the replica trick, indeed, naturally leads to the emergence of an order parameter q_1 that can be interpreted as the typical overlap between the stationary points of energy ϵ that are at overlap q from the reference minimum. The calculation shows that this overlap takes the particularly simple value $q_1 = q^2$, indicating that the stationary points have the weakest possible correlation with each others. It is this feature that implies that (i) the quenched and annealed constrained total complexities $\Sigma(\epsilon, q|\epsilon_0)$ coincide, (ii) the statistical properties of the Hessian described in section 2.1.2 can be themselves determined in an annealed setting, computing the distribution of $\mathcal{H}[\boldsymbol{\sigma}]$ over the realizations of the random energy field only, and not over all the stationary points $\boldsymbol{\sigma}$ at fixed overlap from the reference minimum. In the calculation presented here, we are assuming that the same remains true when conditioning to the value of the smallest eigenvalues of the Hessian. As we discuss in appendix C, this corresponds to assuming that the conditioning does not affect the value of the typical overlap q_1 between stationary points with those stability properties, introducing additional correlations between them. This is *a priori* not guaranteed, and it is therefore an approximation: in the same appendix, we discuss what would be the steps required to perform a quenched calculation of $\Sigma(\epsilon, q, \lambda, u|\epsilon_0)$ and comment further on the assumptions on which the annealed approximation relies.

2.3. Complexity of saddles: the results

2.3.1. Transitions in the population of saddles. For fixed energy ϵ_0 of the reference minimum $\boldsymbol{\sigma}_0$, we are interested in characterizing the properties of the *dominant* saddles (i.e. of those having higher complexity) as a function of their overlap q with $\boldsymbol{\sigma}_0$ and of their energy density ϵ. We anticipate that for the values of q, ϵ for which the complexity of saddles is non-zero, the dominant ones have always index $k = 1$. Their properties however change as a function of q, ϵ. To discuss this, it is convenient to introduce three special values of the overlap $q^+(\epsilon|\epsilon_0), q_{\rm ms}(\epsilon|\epsilon_0), q_M(\epsilon|\epsilon_0)$ and of the energy density $\epsilon^+(q|\epsilon_0), \bar{\epsilon}(q|\epsilon_0), \epsilon_{\rm ms}(q|\epsilon_0)$ defined in terms of the total constrained complexity $\Sigma(\epsilon, q|\epsilon_0)$ and of $\lambda_p^+(\epsilon, q|\epsilon_0)$ in (32) in the following way:

Table 1. Summary of the special values of the overlaps/energy densities defined in the text. Each function depends on the energy density ϵ_0 of the reference minimum.

Special overlaps and energies at fixed ϵ_0	
$\mathbf{q_M}(\epsilon\|\boldsymbol{\epsilon_0})$: overlap of stationary points at energy ϵ that are closer to the reference minimum	$\overline{\boldsymbol{\epsilon}}(q\|\epsilon_0)$: energy of deepest stationary points at overlap q with the reference minimum
$\mathbf{q}_{\rm ms}(\epsilon\|\boldsymbol{\epsilon_0})$: transition between typical ($q > q_{\rm ms}$) and atypical ($q < q_{\rm ms}$) saddles	$\boldsymbol{\epsilon}_{\rm ms}(q\|\epsilon_0)$: transition between typical ($\epsilon > \epsilon_{\rm ms}$) and atypical ($\epsilon < \epsilon_{\rm ms}$) saddles
$\mathbf{q}^+(\epsilon\|\boldsymbol{\epsilon_0})$: transition between connected ($q > q^+$) and disconnected ($q < q^+$) saddles	$\boldsymbol{\epsilon}^+(q\|\epsilon_0)$: transition between connected ($\epsilon < \epsilon^+$) and disconnected ($\epsilon > \epsilon^+$) saddles
$\mathbf{q}^*(\epsilon_0)$: overlap of the deepest saddle(s) connected to the reference minimum	$\boldsymbol{\epsilon}^*(\epsilon_0)$: energy of the deepest saddle(s) connected to the reference minimum
$\mathbf{q}^{\rm mx}_{\rm ms}(\epsilon_0)$: overlap of the farthest saddle(s) connected to the reference minimum	$\boldsymbol{\epsilon}^*_{\mathbf{1}}(\epsilon_0)$: energy of the farthest saddle(s) connected to the reference minimum

- The overlap $q_M(\epsilon|\epsilon_0)$ is the one at which the total constrained complexity becomes non-negative, i.e. for each $q > q_M(\epsilon|\epsilon_0)$ one finds $\Sigma(\epsilon, q|\epsilon_0) < 0$, implying that typically there are *no* stationary points at those values of the overlap. Similarly, the energy curve $\overline{\epsilon}(q|\epsilon_0)$ gives the energy density of the deepest stationary points found at overlap q with the reference minimum, and it is defined from $\Sigma(\overline{\epsilon}, q|\epsilon_0) = 0$: for $\epsilon < \overline{\epsilon}(q|\epsilon_0)$, typically there are *no* stationary points at overlap q with the reference minimum.
- The overlap $q_{\rm ms}(\epsilon|\epsilon_0)$ is the one at which the stationary points contributing to the total constrained complexity are marginal saddles, with an Hessian having an isolated eigenvalue that is exactly equal to zero: $\lambda_0(q_{\rm ms}, \epsilon, \epsilon_0) = 0$. In the high-overlap regime $q_{\rm ms}(\epsilon|\epsilon_0) \leqslant q \leqslant q_M(\epsilon|\epsilon_0)$ the complexity $\Sigma(\epsilon, q|\epsilon_0)$ is contributed by index-1 saddles that are geometrically connected to the reference minimum, whereas for $0 \leqslant q \leqslant q_{\rm ms}(\epsilon|\epsilon_0)$ it is contributed by local minima. The energy curve $\epsilon_{\rm ms}(q|\epsilon_0)$ gives the energy at which the typical value of the isolated eigenvalue vanishes, and it is defined by $\lambda_0(q, \epsilon_{\rm ms}, \epsilon_0) = 0$.
- The overlap $q^+(\epsilon|\epsilon_0)$ and the energy density $\epsilon^+(q|\epsilon_0)$ are defined as the points where $\lambda^+_\sigma(\epsilon, q|\epsilon_0)$ is exactly equal to zero:

$$\lambda^+_\sigma(\epsilon, q|\epsilon_0)\Big|_{\epsilon=\epsilon^+(q|\epsilon_0)} = 0 = \lambda^+_\sigma(\epsilon, q|\epsilon_0)\Big|_{q=q^+(\epsilon|\epsilon_0)}$$

A plot of the transition overlaps and energies is given in figure 2 for $\epsilon_0 = -1.167$ and $p = 3$, and the notation is summarized in table 1.

When $q \leqslant q_{\rm ms}(\epsilon|\epsilon_0)$ and local minima are the dominant stationary points, a population of saddles with finite complexity exists, with a whole range of values of $\lambda < 0$ and complexity (27). These saddles are have at least one negative mode of the Hessian, but not extensively-many of them, i.e. $k = o(N)$[3]. The complexity of the dominant ones is obtained minimizing the large deviation function in (27) over $\lambda \leqslant 0$. It can be checked that in the relevant regime of parameters the following inequality is satisfied:

$$-2\sigma'(p, q) - \mu(q, \epsilon, \epsilon_0) \geqslant 0 \tag{42}$$

and therefore that Regime B holds (see figure 3). The large deviation function to optimize is therefore (36), which is a decreasing function of λ, minimal at the boundary value $\lambda = 0$.

[3] Indeed, the bulk of the density of states is not altered by the conditioning, and it is therefore equal to a semicircle law entirely supported on the positive semi-axis for all $\epsilon < \epsilon_{\rm th}$.

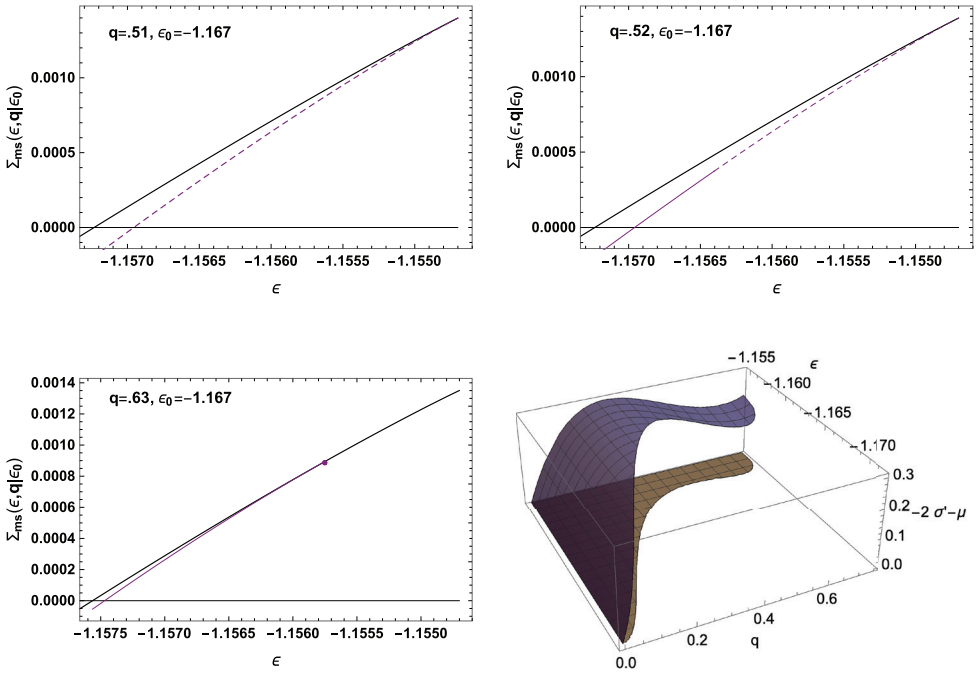

Figure 3. Total constrained entropy $\Sigma(\epsilon, q|\epsilon_0)$ (black) and entropy of the marginal saddles $\Sigma_{\rm ms}(\epsilon, q|\epsilon_0)$ (purple) for fixed $\epsilon_0 = -1.167$ and different values of q. The continuous part of the purple lines corresponds to saddles that are geometrically connected to the reference minimum (meaning that $u_{\rm typ} > 0$), while the dashed part corresponds to disconnected saddles. *Top left.* For this value of q none of the lines in figure 2 is crossed: typically the Hessian has no isolated eigenvalue, and the index-1 saddles are not connected to the minimum as $\epsilon^+(q|\epsilon_0) < \bar{\epsilon}(q|\epsilon_0)$. *Top right.* For this value of q the line $\epsilon^+(q|\epsilon_0)$ in figure 2 is crossed: the index-1 saddles at smaller energy have $u_{\rm typ} > 0$ while those at higher energy have $u_{\rm typ} = 0$. *Bottom left.* For this value of q the curve $\epsilon_{\rm ms}(q|\epsilon_0)$ is crossed: above a given energy (point in the figure) the typical stationary points are index-1 saddles with a negative isolated eigenvalue, which vanishes at the point where $\Sigma(\epsilon, q|\epsilon_0) = \Sigma_{\rm ms}(\epsilon, q|\epsilon_0)$. *Bottom right.* Plot of the function $-2\sigma'(p, q) - \mu(q, \epsilon, \epsilon_0)$ (blue surface) for $\epsilon_0 = -1.167$, $p = 3$ and $\epsilon \geqslant \bar{\epsilon}(q|\epsilon_0)$. The function is always larger than zero (gray surface), indicating that for these parameters Regime B holds.

It follows that for $q \leqslant q_{\rm ms}(\epsilon|\epsilon_0)$ the dominant saddles are *marginally stable*, with a single Hessian mode that is exactly equal to zero. We denote the complexity of these saddles with:

$$\Sigma_{\rm ms}(\epsilon, q|\epsilon_0) \equiv \Sigma(\epsilon, q|\epsilon_0) - \begin{cases} \mathcal{G}_{q,\epsilon|\epsilon_0}(\sqrt{2}\, p\, \epsilon) & \text{if } \lambda_p^-(\epsilon, q|\epsilon_0) < 0 < \lambda_p^+(\epsilon, q|\epsilon_0) \\ \mathcal{G}_0(\sqrt{2}\, p\, \epsilon) & \text{if } 0 < \lambda_p^-(\epsilon, q|\epsilon_0) \ \text{ or } \ \lambda_p^+(\epsilon, q|\epsilon_0) < 0 \end{cases} \tag{43}$$

where the subscript stands for 'marginal saddles'. These saddles are geometrically connected to the reference minimum only whenever $\lambda_p^-(\epsilon, q|\epsilon_0) < 0 < \lambda_p^+(\epsilon, q|\epsilon_0)$. We find that, for the values of parameters we are interested in, $\lambda_p^-(\epsilon, q|\epsilon_0) < 0$ always, and the relevant condition is $0 < \lambda_p^+(\epsilon, q|\epsilon_0)$: for $\epsilon < \epsilon^+(q|\epsilon_0)$ defined above, it holds $\lambda_p^+(\epsilon, q|\epsilon_0) > 0$ and thus the corresponding saddles satisfy $u > 0$.

As a consequence, we find that the saddles dominating the energy landscape are always index-1 saddles, with complexity:

$$\Sigma_1(\epsilon, q|\epsilon_0) = \begin{cases} 0 & \text{if } q_M(\epsilon|\epsilon_0) < q \\ \Sigma(\epsilon, q|\epsilon_0) & \text{if } q_{\rm ms}(\epsilon|\epsilon_0) \leqslant q \leqslant q_M(\epsilon|\epsilon_0) \\ \Sigma_{\rm ms}(\epsilon, q|\epsilon_0) & \text{if } q < q_{\rm ms}(\epsilon|\epsilon_0). \end{cases} \tag{44}$$

The population of dominating saddles displays three regimes, separated by two transitions: (i) at high-overlap with the reference minimum, the saddles have a single Hessian mode that is strictly negative, and are geometrically connected with the minimum; (ii) at intermediate overlaps, the saddles are marginal, and still geometrically connected to the minimum; (iii) at low overlaps, the dominant saddles are marginal, but uncorrelated to the reference minimum. Plots of the total constrained complexity $\Sigma(\epsilon, q|\epsilon_0)$ and of the complexity $\Sigma_{\rm ms}(\epsilon, q|\epsilon_0)$ of marginally stable saddles are given in figure 3, in the different regimes.

2.3.2. Iso-complexity curves and deepest saddles at fixed overlap. A convenient way to represent the saddles complexity is through iso-complexity curves $\bar{\epsilon}^1_x(q|\epsilon_0)$, see figure 4, which give the energies of the index-1 saddles having a fixed value of the complexity:

$$\Sigma_1(\bar{\epsilon}^1_x, q|\epsilon_0) = x. \tag{45}$$

The smallest of these curves $\bar{\epsilon}^1_{x=0}(q|\epsilon_0)$ corresponds to zero complexity and gives the energy of the deepest index-1 saddles found at overlap q with the reference minimum. A comparison between this energy and the energy of the deepest stationary points $\bar{\epsilon}(q|\epsilon_0)$ at the same overlap is given in figure 5. The two curves coincide for overlap larger than $q^*(\epsilon_0) \equiv q_{\rm ms}(\bar{\epsilon}^1_{x=0}|\epsilon_0)$, which is also the local minimum of the two curves, as shown explicitly[4] in [44]. Following the notation of that work, we denote the corresponding energy with $\epsilon^*(\epsilon_0)$. It follows from figure 4 that this is the energy of the deepest saddles that are geometrically connected to the reference minimum, and therefore it corresponds to the optimal (i.e. lowest) energy barrier.

For $q < q^*(\epsilon_0)$, the energy of the deepest marginal saddles $\bar{\epsilon}^1_0(q|\epsilon_0)$ is higher than the one of the deepest minima (as it follows naturally from the fact that their complexity is smaller). This curve has a local maximum at an overlap $q \equiv q^{\rm mx}_{\rm ms}(\epsilon_0)$, corresponding to an energy density $\epsilon^*_1(\epsilon_0)$. We find that this overlap coincides with the point at which $\bar{\epsilon}^1_0(q|\epsilon_0)$ intersects $\epsilon^+(q|\epsilon_0)$,

$$\epsilon^*_1(\epsilon_0) = \bar{\epsilon}^1_0(q^{\rm mx}_{\rm ms}|\epsilon_0) = \epsilon^+(q^{\rm mx}_{\rm ms}|\epsilon_0), \tag{46}$$

meaning that exactly at these overlap the deepest saddles become geometrically disconnected from the reference minimum. Notice that also the curve $\bar{\epsilon}(q|\epsilon_0)$ is maximal at the point where $\epsilon^*(q|\epsilon_0) = \epsilon^+(q|\epsilon_0)$: this overlap is smaller than $q^{\rm mx}_{\rm ms}(\epsilon_0)$, and corresponds to saddles that are not geometrically connected to the minimum.

2.3.3. Distribution of escape states and dynamical barrier. From the analysis above it follows that local minima below the threshold are surrounded by an exponential multiplicity of index-1 saddles that are geometrically connected to the minima. The energy density of these saddles is distributed over an interval $\epsilon \in [\epsilon^*(\epsilon_0), \epsilon_{\rm th}]$ whose width depends on the energy ϵ_0 of the local minimum. These connected saddles are distributed in a region of configuration space that corresponds to overlaps $q \in [q^{\rm mx}_{\rm ms}(\epsilon_0), q_M(\epsilon_0)]$: outside this interval, saddles are either absent, or the dominant ones are uncorrelated to the reference minimum, in the sense that the corresponding downhill direction in configuration space does not point towards the minimum.

[4] Actually, it is shown in [44] that for arbitrary value of x, the iso-complexity curve have local minima at overlaps q_x which coincide with the overlaps $q_{\rm ms}$ at which the typical value of the isolated eigenvalue vanishes: the transition between the marginal saddles and the saddles with a negative eigenvalue occurs exactly at the minimum of these iso-complexity curves.

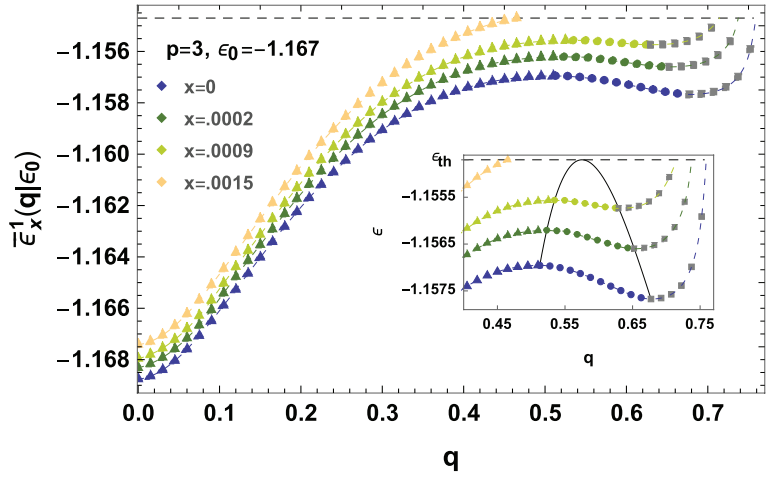

Figure 4. Iso-complexity curves of index-1 saddles, see (45). The different symbols correspond to saddles with one negative Hessian mode (squares), marginal saddles geometrically connected to the reference minimum (circles) and marginal saddles that are disconnected (triangles). The gray part of the curves correspond to the typical saddles already determined in [44]. *Inset.* Zoom of the iso-complexity curves. The black lines are the curves $\epsilon^+(q|\epsilon_0)$ and $\epsilon_{\rm ms}(q|\epsilon_0)$.

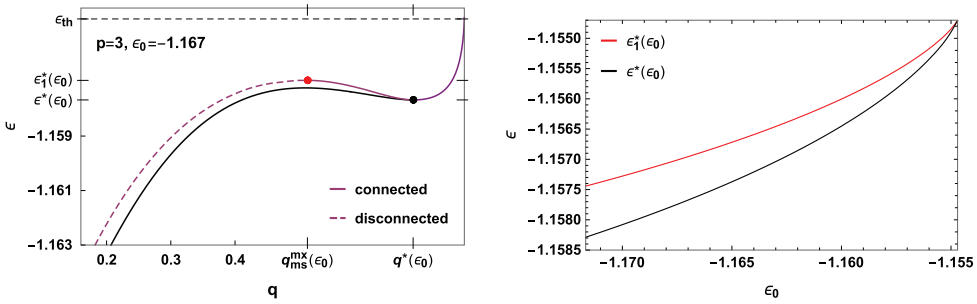

Figure 5. *Left.* Comparison between the energy of the deepest minima (black) and of the deepest saddles (purple) either correlated (solid) or uncorrelated (dashed) with the minimum. The local maximum corresponds to $\epsilon_1^*(\epsilon_0)$ (red dot), the local minimum to $\epsilon^*(\epsilon_0)$ (black dot). *Right.* Dependence of the energies $\epsilon_1^*(\epsilon_0)$ and $\epsilon^*(\epsilon_0)$ on the energy density of the reference minimum ϵ_0.

Each of the connected index-1 saddles represents a potential escape state for the system that is dynamically trapped in the reference minimum. However, it is not guaranteed that once the system escapes through a saddle, it is able to decorrelate from the initial minimum, i.e. to reach regions of configuration space that are orthogonal to it. It is indeed likely that the escape from a local minimum is a complicated dynamical process involving a sequence of jumps between minima that are sufficiently close to each others in configuration space, until decorrelation is achieved. The true 'dynamical barrier' would then correspond to the maximal energy barrier crossed in this composite process.

A lower bound to the dynamical barrier can be obtained from the zero-temperature Franz–Parisi potential [67, 68], as the energy corresponding to the local maximum of the potential curve. As shown in [44], the local maximum of the Franz–Parisi potential coincides exactly

with the local maximum of the curve $\bar{\epsilon}(q|\epsilon_0)$ (and it is thus contributed by local minima). The minimal-energy saddles at $q^*(\epsilon_0)$ correspond to a smaller energy barrier, indicating that when the system escapes through those saddles, it does not fully decorrelate from the initial local minimum. Indeed, this is consistent with the study of the dynamics [69]. On the other hand, some of the marginal saddles at smaller overlap q identified in this work satisfy the bound, see figure 5. In particular, the local maximum of the curve $\bar{\epsilon}_0^1(q|\epsilon_0)$, where the transition occurs between saddles that are geometrical connected to the minimum and saddles that are not, corresponds to an energy barrier $\epsilon_1^*(\epsilon_0) - \epsilon_0$ satisfying the bound. The dependence of $\epsilon_1^*(\epsilon_0)$ on the depth ϵ_0 of the reference minimum is shown in figure 5. These saddles represent potential candidates for the dynamical barriers: checking whether this is the case through the study of the dynamics is an interesting open problem.

3. Part II: general statements of the large deviation functions

In this second part of the paper, we give the general expressions of the large deviation functions of the smallest eigenvalue and eigenvector of GOE matrices deformed with both and additive and multiplicative perturbation along a fixed direction in configuration space. In particular, in section 3.1 we recall the general expression for the *typical* value of the isolated eigenvalue of the Hessian, and define the various large deviation functions to be determined. In sections 3.2–3.4 we report the general expressions of these large deviation functions, and discuss their interpretation in terms of a *BBP-like* transition of the second-smallest eigenvalue of the perturbed matrices. In section 3.5 we give a summary of the main steps of the calculation, which is presented in detail in the third part of the paper.

3.1. Perturbed GOE matrix: typical values and large deviations

We let $\mathcal{X}$ be a M-dimensional GOE matrix with entries x_{ij} with respect to some basis $\mathbf{e}_i$, and variances $\langle x_{ij}^2 \rangle = (\sigma^2/M)[1 + \delta_{ij}]$. This corresponds to the distribution:

$$P(\mathcal{X}) = \frac{1}{Z_M(\sigma)} \mathrm{e}^{-\frac{M}{4\sigma^2}\mathrm{Tr}\mathcal{X}^2}, \tag{47}$$

where $Z_M(\sigma)$ is the normalization. For $\beta \geqslant 0$ we define the $M \times M$ matrix:

$$F_\beta = \mathbb{1} - \frac{\beta}{1+\beta}\, \mathbf{e}_M \mathbf{e}_M^T \tag{48}$$

where $\mathbb{1}$ is the identity matrix, $\mathbf{e}_M$ is a unit vector and we set

$$\mathcal{Y} = F_\beta \mathcal{X} F_\beta + \theta\, \mathbf{e}_M \mathbf{e}_M^T, \tag{49}$$

where θ is the strength of the additive perturbation, which we take to be a fluctuating variable with distribution:

$$f_{\bar{\theta},\sigma_\theta}(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \mathrm{e}^{-\frac{M}{2\sigma_\theta^2}(\theta-\bar{\theta})^2}. \tag{50}$$

We denote with $\mu_M \leqslant \mu_{M-1} \leqslant \cdots \leqslant \mu_1$ the eigenvalues of $\mathcal{Y}$. Notice that the statistics of the rescaled, conditioned Hessian described in section 2.1.2 (up to the shift by the ϵ-dependent diagonal matrix) is the one of a matrix of the form (49) with parameters $\sigma^2 \to p(p-1)$, $\beta \to \sigma/\Delta(q) - 1$, $\bar{\theta} \to \mu(q, \epsilon, \epsilon_0)$ and $\sigma_\theta \to \zeta(q)$.

In the following, we restrict to the case $\overline{\theta} < 0$, which is of interest for the p-spin landscape problem. We denote with $\rho_M^{\text{typ}}(\mu)$ the *typical* eigenvalue density of the matrix $\mathcal{Y}$. For certain values of the parameters $\overline{\theta}, \beta$, the latter exhibits a sub-leading correction with respect to the GOE semicircle:

$$\rho_\sigma(\mu) = \frac{\sqrt{4\sigma^2 - \mu^2}}{2\pi\sigma^2}, \tag{51}$$

that corresponds to the smallest eigenvalue being isolated from the bulk of the density of states. This happens whenever:

$$\overline{\theta} \leqslant \theta_c \equiv -\sigma\left(1 + \frac{\sigma'^2}{\sigma^2}\right) = -\sigma\left(\frac{1 + 2\beta^2 + 4\beta}{[1+\beta]^2}\right) \tag{52}$$

or equivalently

$$\sigma^2 G_{\sigma'}(\overline{\theta}) \geqslant -\sigma, \tag{53}$$

where

$$\sigma' = \sigma\sqrt{\frac{\beta(\beta+2)}{(1+\beta)^2}} < \sigma \tag{54}$$

and where G_σ is the GOE resolvent:

$$G_\sigma(z) \overset{z \text{ real}}{=} \frac{1}{2\sigma^2}\left(z - \text{sign}(z)\sqrt{z^2 - 4\sigma^2}\right) \in \left[-\frac{1}{\sigma}, \frac{1}{\sigma}\right]. \tag{55}$$

In this case one has that the typical value of the smallest eigenvalue μ_M reads:

$$\mu_M^{\text{typ}} = \mu_0(\overline{\theta}, \beta) \equiv G_\sigma^{-1}(G_{\sigma'}(\overline{\theta})) = \frac{1}{G_{\sigma'}(\overline{\theta})} + \sigma^2 G_{\sigma'}(\overline{\theta}) \leqslant -2\sigma, \tag{56}$$

and thus the typical density of eigenvalues is

$$\rho_M^{\text{typ}}(\mu) = \frac{\sqrt{4\sigma^2 - \mu^2}}{2\pi\sigma^2} + \frac{1}{M}\delta\left(\mu - \mu_0(\overline{\theta}, \beta)\right) + o\left(\frac{1}{M}\right). \tag{57}$$

When (52) is not satisfied, the sub-leading contribution to (57) is absent and $\mu_M^{\text{typ}} = -2\sigma$. In absence of the multiplicative perturbation (when $\beta = 0$), we have $\sigma' \to 0$; using that

$$\lim_{\sigma' \to 0} G_{\sigma'}(x) = \frac{1}{x} \tag{58}$$

we recover the well known results for the minimal eigenvalue of a GOE matrix subject to an additive rank-1 perturbation [52, 54, 55],

$$\lim_{\beta \to 0} \mu_0(\overline{\theta}, \beta) = \overline{\theta} + \frac{\sigma^2}{\overline{\theta}}. \tag{59}$$

Notice that the typical density of states (57) does not depend on the fluctuations of θ but only on its average value. The fluctuations enter into play when looking at large deviations of μ_M. We denote with $\mathbf{v}_M$ the corresponding eigenvector, and define $u_{\text{M}} = |\mathbf{v}_M \cdot \mathbf{e}_M|^2$. We use the notation $\mathcal{P}_{\overline{\theta},\sigma_\theta,\beta}(x)$ for the distribution of the smallest eigenvalue μ_M, which is given by:

$$\mathcal{P}_{\overline{\theta},\sigma_\theta,\beta}(x) = \int_0^1 \mathrm{d}u \int_{-\infty}^{\infty} \mathrm{d}\theta f_{\overline{\theta},\sigma_\theta}(\theta)\, \tilde{\mathcal{P}}_{\theta,\beta}(x,u), \tag{60}$$

where $\tilde{\mathcal{P}}_{\theta,\beta}(x,u)$ is the joint probability density of μ_M and u_M, *conditioned* to a fixed value of the additive perturbation θ. In the following we compute the large deviation function:

$$\lim_{M\to\infty} \frac{\log \tilde{\mathcal{P}}_{\theta,\beta}(x,u)}{M} = -\mathcal{L}_{\theta,\beta}(x,u). \tag{61}$$

For each x, we determine the typical value $u_{\mathrm{typ}}(x)$ maximizing the large deviation function,

$$u_{\mathrm{typ}}(x) \equiv \underset{u\in[0,1]}{\operatorname{argmin}}\ \mathcal{L}_{\theta,\beta}(x,u) \tag{62}$$

and set

$$\overline{\mathcal{L}}_{\theta,\beta}(x) \equiv \mathcal{L}_{\theta,\beta}(x, u_{\mathrm{typ}}(x)). \tag{63}$$

The large deviation function for fluctuating θ is then obtained as:

$$\lim_{M\to\infty} \frac{\mathcal{P}_{\overline{\theta},\sigma_\theta,\beta}(x)}{M} \equiv -\mathcal{F}_{\overline{\theta},\sigma_\theta,\beta}(x) = -\min_{\theta}\left[\frac{(\theta-\overline{\theta})^2}{2\sigma_\theta^2} + \overline{\mathcal{L}}_{\theta,\beta}(x)\right]. \tag{64}$$

This large deviation function exhibits an explicit dependence on the variance σ_θ^2; nevertheless, as we shall see, its minimum is always attained at the typical value μ_M^{typ} of the smallest eigenvalue, that does not depend on σ_θ^2 and it is given by $\mu_0(\overline{\theta},\beta)$ in (56) when (52) is satisfied, and by -2σ otherwise.

3.2. Large deviation function at fixed u and θ

We begin by stating the form of the large deviation function (61). We define the constants:

$$C_2 = -2\theta\,(1+\beta)^4, \qquad C_3 = 2\beta(2+\beta), \qquad C_4(x,u) = C_2 + \frac{C_3^2}{2}xu, \tag{65}$$

and introduce:

$$\kappa_{\theta,\beta}(x,u) = \frac{\sigma^2 C_3[2+C_3(1-u)]^3}{C_4^2(x,u)(1-u)} = \frac{4\sigma^2\beta(2+\beta)[1+\beta(\beta+2)(1-u)]^3}{(1-u)[\beta^2(\beta+2)^2 ux - (\beta+1)^4\theta]^2}. \tag{66}$$

We identify the following two regimes of parameters:

$$\begin{aligned} &\text{Case A}: \quad \kappa_{\theta,\beta}(x,u) > 1 \\ &\text{Case B}: \quad \kappa_{\theta,\beta}(x,u) \leqslant 1, \end{aligned} \tag{67}$$

and define the rate functions:

$$\begin{aligned} \mathcal{L}^{(a)}_{\theta,\beta}(x,u) &= \frac{1}{4\sigma^2}\left(x^2 + C_2 xu + \frac{C_3^2}{4}x^2u^2 + C_3x^2u\right) - \mathcal{I}(x) + \frac{1}{2} - \frac{1}{2}\log\left(\frac{2\sigma^2(1-u)}{C_3(1-u)+2}\right) \\ &\quad - \frac{C_4^2(x,u)(1-u)^2}{4\sigma^2\left[2+C_3(1-u)\right]^2}, \\ \mathcal{L}^{(b)}_{\theta,\beta}(x) &= \frac{1}{4\sigma^2}\left(2x^2 + C_3x^2 + C_2x + \frac{C_3^2}{4}x^2\right) - \frac{3}{2}\mathcal{I}(x) + \frac{1}{2} + \frac{1}{2}\log\left[\frac{C_3^2x + 2C_3x + 2C_2}{4\sigma^2}\right] \end{aligned} \tag{68}$$

where:

$$\mathcal{I}(z) = \int \mathrm{d}\lambda\, \rho_\sigma(\lambda) \log|\lambda - z| = \begin{cases} \log\left(-\frac{z}{2} + \frac{1}{2}\sqrt{z^2 - 4\sigma^2}\right) - \frac{1}{2} + \frac{z^2}{4\sigma^2} + \frac{z}{4\sigma^2}\sqrt{z^2 - 4\sigma^2} & \text{if } z < -2\sigma \\ \frac{z^2}{4\sigma^2} - \frac{1}{2} + \log\sigma & \text{if } -2\sigma < z < 0 \end{cases}, \tag{69}$$

and

$$l(\theta,\beta) = 1 - \frac{1}{2}\log\left(\frac{2\sigma^4}{C_3+2}\right) - \frac{C_2^2}{4\sigma^2(C_3+2)^2} = 1 - \log\left(\frac{\sigma^2}{1+\beta}\right) - \frac{\theta^2}{2\sigma^2[1+\beta]^2}. \tag{70}$$

When Case B holds, we further define the following functions:

$$F(x,u) = -\frac{C_4(x,u)(1-u) + \sqrt{C_4^2(x,u)(1-u)^2 - \sigma^2 C_3(1-u)\left[2 + C_3(1-u)\right]^3}}{\sigma^2\left[2 + C_3(1-u)\right]^2}$$
$$\mu_1(x,u) = -\frac{2\left(C_4(1-u)\left[1 + C_3(1-u)\right] - \sqrt{(1-u)\left[C_4^2(1-u) - \sigma^2 C_3\left[2 + C_3(1-u)\right]^3\right]}\right)}{C_3(1-u)\left[2 + C_3(1-u)\right]^2}. \tag{71}$$

Notice that the functions (71) are complex in Case A, when $\kappa_{\theta,\beta}(x,u) > 1$. In terms of these quantities, the large deviation function (61) is given by the following expressions:

- When Case A holds:

$$\mathcal{L}_{\theta,\beta}(x,u) = \mathcal{L}^{(a)}_{\theta,\beta}(x,u) - l(\theta,\beta). \tag{72}$$

- When Case B holds:

$$\mathcal{L}_{\theta,\beta}(x,u) = \begin{cases} \mathcal{L}^{(a)}_{\theta,\beta}(x,u) - l(\theta,\beta) & \text{if } \sigma^2 F(x,u) \geqslant -\sigma \\ \mathcal{L}^{(b)}_{\theta,\beta}(x) - l(\theta,\beta) & \text{if } \sigma^2 F(x,u) < -\sigma \quad \text{and} \quad x \geqslant \mu_1(x,u) \\ \mathcal{L}^{(a)}_{\theta,\beta}(x,u) - l(\theta,\beta) & \text{if } \sigma^2 F(x,u) < -\sigma \quad \text{and} \quad x < \mu_1(x,u). \end{cases} \tag{73}$$

This expression is continuous at the point $x = \mu_1(x,u)$.

The constant shift equals to $l(\theta,\beta) = \mathcal{L}^{(a)}_{\theta,\beta}(x_{\mathrm{typ}}, u_{\mathrm{typ}})$, where $x_{\mathrm{typ}}, u_{\mathrm{typ}}$ are the typical values for the given parameters; it is added to ensure that $\mathcal{L}_{\theta,\beta}(x_{\mathrm{typ}}, u_{\mathrm{typ}}) = 0$. In (appendix D), we discuss how limiting cases known in the literature are recovered.

3.2.1. Interpretation in terms of the second-smallest eigenvalue. When θ is kept fixed, the typical value of the smallest eigenvalue μ_M undergoes a transition at $\theta = \theta_c$ given in (52): for $\theta \leqslant \theta_c$, it equals to $\mu_0(\theta,\beta)$, which can be equivalently written as:

$$\mu_0(\theta,\beta) = G_\sigma^{-1}(G_{\sigma'}(\bar{\theta})) = m_\sigma^+[C_2, C_3] \tag{74}$$

in terms of the constants (65), where

$$m_\sigma^\pm[a,b] = \frac{2}{b(b+2)^2}\left(-a(b+1) \pm \sqrt{a^2 - \sigma^2 b(b+2)^3}\right). \tag{75}$$

The different regimes of the large deviation function (73) in Case B can be interpreted in terms of an analogous transition of the typical value μ_{M-1}^{typ} of the *second-smallest* eigenvalue of the matrix $\mathcal{Y}$. As it will appear from the explicit calculation in section 4.1, fixing $\mu_M = x$ and $u_M = u$ leads to a modification of the joint distribution of the remaining eigenvalues $\{\mu_i\}_{i=1}^{M-1}$. In particular, the resulting joint distribution has the same form of the joint distribution of all the eigenvalues $\{\mu_\alpha\}_{\alpha=1}^{M}$ of the matrix (49), but with modified parameters $\tilde{\theta}, \tilde{\beta}$ defined by:

$$\tilde{\theta} = \frac{\theta(1-u)(1+\beta)^4 - xu(1-u)\beta^2(2+\beta)^2}{[1+\beta(1-u)(2+\beta)]^2}, \qquad (1+\tilde{\beta})^2 = 1+\beta(1-u)(2+\beta). \tag{76}$$

This is equivalent to mapping $C_3 \to C_3(1-u)$ and $C_2 \to C_4(x,u)(1-u)$. One can easily check by substitution that the function $F(x,u)$ in (71) can be written in terms of these parameters as:

$$F(x,u) = \frac{1}{\sigma^2 G_{\tilde{\sigma}}(\tilde{\theta})}, \tag{77}$$

where

$$\tilde{\sigma}^2 = \sigma^2 \left[\frac{\tilde{\beta}(\tilde{\beta}+2)}{(1+\tilde{\beta})^2}\right] \leqslant \sigma^2. \tag{78}$$

A comparison with (53) shows that the two regimes of (73) correspond to the regimes in which the typical value of the second- smallest eigenvalue sticks to the boundary of the semicircle $\rho_\sigma(\lambda)$ (when $\sigma^2 F(x,u) \geqslant -\sigma$) or is smaller that -2σ (when $\sigma^2 F(x,u) < -\sigma$). In the latter case, the typical value of the second-smallest eigenvalue takes precisely the form:

$$\begin{aligned}\mu_{M-1}^{\text{typ}}(x,u) = \mu_1(x,u) &= G_\sigma^{-1}\left(G_{\tilde{\sigma}}(\tilde{\theta})\right) = G_\sigma^{-1}\left(\frac{1}{\sigma^2 F(x,u)}\right) \\ &= m_\sigma^+[C_4(x,u)(1-u), C_3(1-u)].\end{aligned} \tag{79}$$

Notice that the argument of G_σ^{-1} is larger than $-1/\sigma$, as it should. Therefore, $\mathcal{L}_{\theta,\beta}(x,u)$ is proportional to $\mathcal{L}_{\theta,\beta}^{(a)}(x,u)$ whenever the parameters x,u are chosen in such a way that $x \leqslant \mu_{M-1}^{\text{typ}}(x,u)$, and it is proportional to $\mathcal{L}_{\theta,\beta}^{(b)}(x)$ otherwise. As it will follow from section 4.6, this last regime is relevant only whenever u is taken to be different from its typical value $u_{\text{typ}}(x)$ defined in (62): when the overlap is allowed to adjust itself to its typical value, one naturally finds that $x \leqslant \mu_{M-1}^{\text{typ}}(x, u_{\text{typ}}(x))$.

Case A can be analogously interpreted in terms of the effective parameters (76). Indeed, we find that (66) can be re-written as:

$$\kappa_{\theta,\beta}(x,u) = \frac{4\tilde{\sigma}^2}{\tilde{\theta}^2}, \tag{80}$$

and therefore Case A corresponds to the regime in which $-2\tilde{\sigma} < \tilde{\theta} < 0$. In this regime, the functions $m_\sigma^+[C_4(x,u)(1-u), C_3(1-u)]$ are complex (and exactly equal at the boundary value $\tilde{\theta} = -2\tilde{\sigma}$).

When interpreted in terms of the second-smallest eigenvalue, the large deviation function $\mathcal{L}_{\theta,\beta}(x,u)$ displays the same three regimes that will appear in section 3.3, with the substitutions $\theta \to \tilde{\theta}$, $\beta \to \tilde{\beta}$ and $\sigma' \to \tilde{\sigma}$.

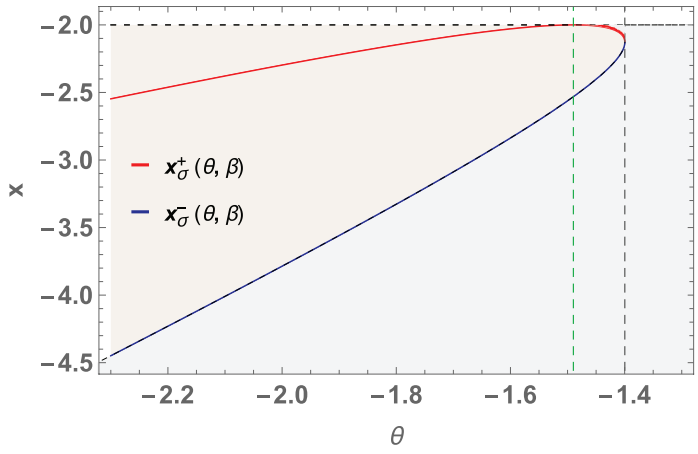

Figure 6. Plot of the functions $x_\sigma^\pm(\theta,\beta)$ for $\beta = .4$ and $\sigma = 1$. The different colors correspond to the regions of parameters where the large deviation function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ equals either to $\mathcal{G}_{\theta,\beta}(x)$ (red) or to $\mathcal{G}_0(x)$ (blue). The dashed vertical lines denote $\theta = \theta_c$ (green) and $\theta = -2\sigma'$ (black). The plot shows three regimes: (i) Regimes B.2, for $\theta < \theta_c$: the function $x_\sigma^+(\theta,\beta)$ equals to the typical value for the smallest eigenvalue of the matrix, i.e. $x_\sigma^+(\theta,\beta) = \mu_0(\theta,\beta)$. At $\theta = \theta_c$ it becomes equal to -2σ, signaling that the smallest eigenvalue is reabsorbed into the bulk; (ii) Regime B.1: the smallest eigenvalue is typically not out of the bulk, $x_\sigma^+(\theta,\beta)$ gives the analytic continuation of the isolated eigenvalue into the second Riemann sheet in the complex plane. This ends at $\theta = -2\sigma'$, when $x_\sigma^+(\theta,\beta) = x_\sigma^-(\theta,\beta)$; (iii) Regime A: for $\theta > -2\sigma'$, both functions are complex.

3.3. Large deviation function optimized over u at fixed θ

We now state the form of the large deviation function (63), obtained by optimizing (61) over the overlap u, at fixed x. The behavior of the resulting function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ depends on whether the parameters θ, β are such that μ_M^{typ} is typically out of the bulk of the semicircle of not, and whether the parameter x is taken to be larger or smaller than the following two thresholds:

$$x_\sigma^\pm(\theta,\beta) = \frac{(1+\beta)[1+2\beta(\beta+2)]\theta \pm \sqrt{(1+\beta)^2\theta^2 - 4\beta(\beta+2)\sigma^2}}{2\beta(\beta+1)(\beta+2)} = m_\sigma^\pm[C_2, C_3]. \tag{81}$$

For fixed β and as a function of θ, these thresholds have three regimes, see caption in figure: 6, that correspond to three different regimes for the large deviation function $\overline{\mathcal{L}}_{\theta,\beta}(x)$:

- Regime A: $-2\sigma' \leqslant \theta$: in this regime the functions $x_\sigma^\pm(\theta,\beta)$ in (81) are complex;
- Regime B.1: $\theta_c \leqslant \theta \leqslant -2\sigma'$ or equivalently $\sigma^2 G_{\sigma'}(\theta) < -\sigma$ and $\theta \leqslant -2\sigma'$: in this regime $\mu_M^{\text{typ}} = -2\sigma$;
- Regime B.2: $\theta \leqslant \theta_c = -\sigma[1 + (\sigma')^2/\sigma^2]$ or equivalently $\sigma^2 G_{\sigma'}(\theta) \geqslant -\sigma$. In this regime the typical value μ_M^{typ} of the smallest eigenvalue is out of bulk and equals to $\mu_0(\theta,\beta)$, see (52) and (74).

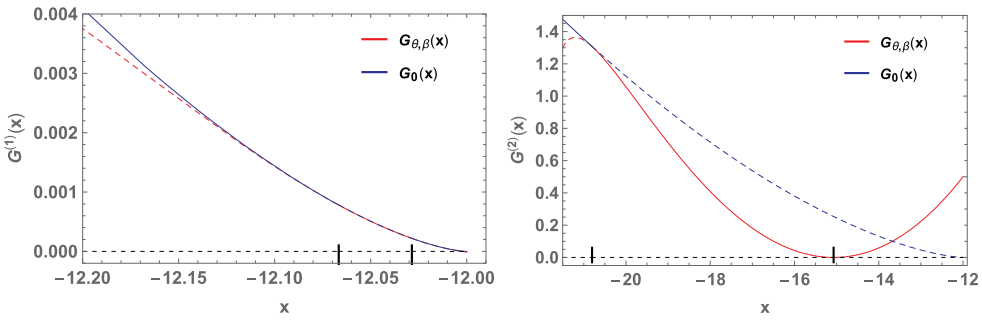

Figure 7. Plots of the large deviations function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ for the smallest eigenvalue (solid lines), for values of parameters for which its typical value is at the boundary of the semicircle (*left*), or it is out of the bulk (*right*). The black ticks mark the values $x_\sigma^\pm(\theta,\beta)$. When the large deviation function coincides with $\mathcal{G}_{\theta,\beta}(x)$ (red curves), the eigenvector corresponding to the minimal eigenvalue has a typical projection along the special direction that is $u_{\rm typ}(x) > 0$.

Given the functions:

$$\mathcal{G}_0(x) = \int_x^{-2\sigma} \frac{\sqrt{z^2-4\sigma^2}}{2\sigma^2}\mathrm{d}z = \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \frac{1}{2} + \log\sigma,$$
$$\mathcal{G}_{\theta,\beta}(x) = \frac{[1+\beta(\beta+2)]^2}{4\sigma^2}x^2 - \frac{\mathcal{I}(x)}{2} + \frac{(1+\beta)^4\theta^2 - 2\theta x}{4\sigma^2} + \frac{1}{2}\log[\beta(\beta+2)x - (1+\beta)^2\theta], \tag{82}$$

it holds:

$$\overline{\mathcal{L}}_{\theta,\beta}(x) = \begin{cases} \mathcal{G}_0(x) & \text{in Regime A} \\ \mathcal{G}^{(1)}_{\theta,\beta}(x) & \text{in Regime B.1} \\ \mathcal{G}^{(2)}_{\theta,\beta}(x) & \text{in Regime B.2,} \end{cases} \tag{83}$$

where in the Regime B.1 one has:

$$\mathcal{G}^{(1)}_{\theta,\beta}(x) = \begin{cases} \mathcal{G}_0(x) & \text{if } x < x_\sigma^-(\theta,\beta) \text{ or } x_\sigma^+(\theta,\beta) < x < -2\sigma \\ \mathcal{G}_{\theta,\beta}(x) & \text{if } x_\sigma^-(\theta,\beta) < x < x_\sigma^+(\theta,\beta), \end{cases} \tag{84}$$

while in Regime B.2 one has:

$$\mathcal{G}^{(2)}_{\theta,\beta}(x) = \begin{cases} \mathcal{G}_{\theta,\beta}(x) & \text{if } x_\sigma^-(\theta,\beta) < x < -2\sigma \\ \mathcal{G}_0(x) & \text{if } x < x_\sigma^-(\theta,\beta). \end{cases} \tag{85}$$

The function $\mathcal{G}^{(1)}_{\theta,\beta}(x)$ has a minimum at $x_{\rm typ} = -2\sigma$, while $\mathcal{G}^{(2)}_{\theta,\beta}(x)$ vanishes at:

$$x_{\rm typ} = G_\sigma^{-1}(G_{\sigma'}(\theta)) = \mu_0(\theta,\beta), \tag{86}$$

that is indeed the typical value of μ_M in this regime of parameters. Both large deviation functions are continuous at $x_\sigma^\pm(\theta,\beta)$. We notice that the explicit expressions of $\mu_0(\theta,\beta)$ and of $x_\sigma^+(\theta,\beta)$ coincide, even though $\mu_0(\theta,\beta)$ is well defined only in the regime $\theta \leqslant \theta_{\rm c}$ (otherwise the resolvent in (86) would not be invertible), while $x_\sigma^+(\theta,\beta)$ is defined in the opposite regime of parameters. The coincidence of the two expressions follows from a symmetry of the GOE resolvent evaluated on the real axis, as we discuss more precisely in section 4.3 . Plots of the large deviation function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ are given in figure 7.

For what concerns the typical overlaps, $\mathcal{G}_{\theta,\beta}(x)$ corresponds to a non-trivial typical overlap $u_{\rm typ}(x) > 0$ with the special direction, while $\mathcal{G}_0(x)$ corresponds to zero overlap, i.e. in Regime B.1 we have

$$u_{\rm typ}(x) = \begin{cases} 0 & \text{if } x_\sigma^+(\theta,\beta) < x < -2\sigma \\ u_{\theta,\beta}^+(x) & \text{if } x_\sigma^-(\theta,\beta) < x < x_\sigma^+(\theta,\beta) \\ 0 & \text{if } x < x_\sigma^-(\theta,\beta), \end{cases} \tag{87}$$

while in Regime B.2 it holds:

$$u_{\rm typ}(x) = \begin{cases} u_{\theta,\beta}^+(x) & \text{if } x_\sigma^-(\theta,\beta) < x < -2\sigma \\ 0 & \text{if } x < x_\sigma^-(\theta,\beta). \end{cases} \tag{88}$$

The expression for $u_{\theta,\beta}^+(x)$ is given explicitly in (162). Notice that when the eigenvector associated to the smallest eigenvalue is uncorrelated with the special direction ($u_{\rm typ}(x) = 0$), the large deviation function $\mathcal{G}_0(x)$ coincides with the one in absence of perturbations [64]. In the limit of a purely additive perturbation $\beta \to 0$, the Regime A disappears as $\sigma' \to 0$. Moreover, one finds $x_\sigma^+ \to \theta + \sigma^2/\theta$ and $x_\sigma^- \to -\infty$. The typical value of the overlap, when positive, reduces to $u_{\rm typ}(x) \to 1 - [x\theta - \sqrt{\theta^2(x^2 - 4\sigma^2)}]/(2\theta^2)$. The known results are recovered [63].

3.4. Large deviations for fluctuating θ

We finally state the expression for the large deviation function $\mathcal{F}_{\overline{\theta},\sigma_\theta,\beta}(x)$ in (64), obtained optimizing over the Gaussian fluctuations of θ. In Regime A the optimization is trivial. In Regime B we shall show that all the inequalities in the previous section survive with the substitution $\theta \to \overline{\theta}$, meaning that we can identify once more the three regimes:

- Regime A: $-2\sigma' \leqslant \overline{\theta}$ (the functions $x_\sigma^\pm(\overline{\theta},\beta)$ are complex);
- Regime B.1: $\sigma^2 G_{\sigma'}(\overline{\theta}) < -\sigma$, meaning $-\theta_{\rm c} \leqslant \overline{\theta} \leqslant -2\sigma'$. In this regime $\mu_M^{\rm typ}(\overline{\theta},\beta) = -2\sigma$;
- Regime B.2: $\sigma^2 G_{\sigma'}(\overline{\theta}) \geqslant -\sigma$, meaning $\overline{\theta} < \theta_{\rm c}$. In this regime $\mu_M^{\rm typ}(\overline{\theta},\beta) = \mu_0(\theta,\beta)$, see (52), (74).

The large deviation function for fluctuating θ reads:

$$\mathcal{F}_{\overline{\theta},\sigma_\theta,\beta}(x) = \begin{cases} \mathcal{G}_0(x) & \text{in Regime A} \\ \mathcal{F}^{(1)}_{\overline{\theta},\sigma_\theta,\beta}(x) & \text{in Regime B.1} \\ \mathcal{F}^{(2)}_{\theta,\beta}(x) & \text{in Regime B.2,} \end{cases} \tag{89}$$

where

$$\mathcal{F}^{(1)}_{\overline{\theta},\sigma_\theta,\beta}(x) = \begin{cases} \mathcal{G}_0(x) & \text{if } x < x_\sigma^-(\overline{\theta},\beta) \text{ or } x_\sigma^+(\overline{\theta},\beta) < x < -2\sigma \\ \mathcal{G}_{\theta^*,\beta}(x) & \text{if } x_\sigma^-(\overline{\theta},\beta) < x < x_\sigma^+(\overline{\theta},\beta). \end{cases} \tag{90}$$

and

$$\mathcal{F}^{(2)}_{\overline{\theta},\sigma_\theta,\beta}(x) = \begin{cases} \mathcal{G}_{\theta^*,\beta}(x) & \text{if } x_\sigma^-(\overline{\theta},\beta) < x < -2\sigma \\ \mathcal{G}_0(x) & \text{if } x < x_\sigma^-(\overline{\theta},\beta). \end{cases} \tag{91}$$

Therefore the large deviation function has the same form as in the previous section with $\theta \to \overline{\theta}$, except for $\mathcal{G}_{\theta,\beta}$ which has to be computed at the shifted point $\theta \to \theta^* = \theta_0^*(x)$ whose explicit expression is given in (179). Notice that, as it should,

$$\theta_0^*(x) \overset{\sigma_\theta \to 0}{\longrightarrow} \overline{\theta}. \tag{92}$$

3.5. Derivation of the large deviations: the idea of the calculation

In this section we summarize the skeleton of the derivation of the large deviation functions, whose details are presented in the following. The starting point is the derivation of the joint density of the eigenvalues μ_α of the matrix $\mathcal{Y}$ given in (49), and of the corresponding eigenvector components along $\boldsymbol{e}_M$. We set $\mu_M \leqslant \mu_{M-1} \leqslant \cdots \mu_1$ and let $\mathbf{v}_\alpha$ be the matrix eigenvectors, and $u_\alpha = |\mathbf{v}_\alpha \cdot \boldsymbol{e}_M|^2 \in [0,1]$. As we derive in the following, the joint probability density of μ_α, u_α reads:

$$P_{\theta,\beta}(\mu_\alpha, u_\alpha) = \frac{\mathrm{e}^{-MV(\mu_\alpha,u_\alpha)}}{\mathcal{Z}_M[\theta,\beta]} \prod_{\gamma<\alpha} (\mu_\gamma - \mu_\alpha) \prod_{\alpha=1}^{M} \theta\left(\mu_\alpha - \mu_{\alpha+1}\right)\ \delta\left(\sum_{\alpha=1}^{M} u_\alpha - 1\right) \prod_\alpha \frac{1}{u_\alpha^{1/2}}, \quad (93)$$

with $\mathcal{Z}_M[\theta,\beta]$ a normalization and

$$V(\mu_\alpha, u_\alpha) = \frac{1}{4\sigma^2}\left[\sum_\alpha \mu_\alpha^2 + \frac{C_3^2}{4}\left(\sum_\alpha \mu_\alpha u_\alpha\right)^2 + C_2 \sum_\alpha \mu_\alpha u_\alpha + C_3 \sum_\alpha \mu_\alpha^2 u_\alpha\right], \quad (94)$$

with the constants given in (65). Therefore, the effect of the additive and multiplicative perturbations is to introduce a coupling between the μ_α and u_α through the confinement potential $V(\mu_\alpha, u_\alpha)$. The joint probability density of $\mu_M = x$ and $u_M = u$ has to be obtained integrating over all the other eigenvalues and eigenvector projections, as:

$$\frac{\mathcal{P}_{\theta,\beta}(x,u)}{p(u)} = \frac{\mathrm{e}^{-\frac{M}{4\sigma^2}\left[x^2 + C_2 x u + \frac{C_3^2}{4} x^2 u^2 + C_3 x^2 u\right]}}{\mathcal{Z}_M^*[\theta,\beta]} \int \prod_{\alpha=1}^{M-1} \mathrm{d}\mu_\alpha [(\mu_\alpha - x)\theta(\mu_\alpha - x)] F(\vec{\mu})\, I_{x,u}(\vec{\mu}). \quad (95)$$

In this formula $p(u)$ is the distribution of a single eigenvector component for a GOE matrix, $F(\vec{\mu})$ is the measure on the remaining $M-1$ eigenvalues:

$$F(\vec{\mu}) = \prod_{\alpha>\gamma=1}^{M-1} (\mu_\gamma - \mu_\alpha)\theta(\mu_\gamma - \mu_\alpha)\mathrm{e}^{-\frac{M}{4\sigma^2}\left[\sum_{\alpha=1}^{M-1}\mu_\alpha^2\right]}, \quad (96)$$

$I_{x,u}(\vec{\mu})$ is the integral over the remaining eigenvectors components, and the normalization is rescaled as $\mathcal{Z}_M^*[\theta,\beta] = \mathcal{Z}_M \Gamma(M/2)/\pi^{M/2}$. From the explicit expression:

$$I_{x,u}(\vec{\mu}) = \int_{-\infty}^{\infty} \prod_{\alpha=1}^{M-1} de_\alpha \frac{\Gamma\left(\frac{M-1}{2}\right)}{\pi^{\frac{M-1}{2}}} \delta\left(\sum_{\alpha=1}^{M-1} e_\alpha^2 - 1\right) \times$$
$$\times\, \mathrm{e}^{-\frac{M}{4\sigma^2}\left[C_4(x,u)(1-u)\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2 + \frac{[C_3(1-u)]^2}{4}\left(\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2\right)^2 + C_3(1-u)\sum_{\alpha=1}^{M-1}\mu_\alpha^2 e_\alpha^2\right]} \quad (97)$$

one sees that (97), up to normalization constants, has the same structure as (93) with $C_3 \to C_3(1-u)$ and $C_2 \to C_4(x,u)(1-u)$. Therefore, at fixed x and u the distribution of the remaining eigenvalues is the one of a GOE matrix perturbed exactly as the original one, with modified parameters given in (76). We made use of this observation in the interpretation of the large deviation function.

Given (95), the core of the calculation is the computation of the integrals over the matrix eigenvalues and eigenvectors. This is done in three steps: (i) introducing two auxiliary fields y, λ the integration over the e_α becomes Gaussian and can be performed; (ii) the integration over the μ_α is performed solving a variational problem for the eigenvalue density, both for its continuous part and for the isolated eigenvalue generated by the perturbations; (iii) the auxiliary parameters y, λ are fixed with a saddle point calculation.

More precisely, the integration over the eigenvector components and over the continuous part of the eigenvalue density leads to the following expression for the joint probability:

$$\mathcal{P}_{\theta,\beta}(x,u) \sim \mathcal{A}_M \, \mathrm{e}^{-M\Psi_0(x,u)} \int_{\xi \geqslant x} \mathrm{d}\xi \mathrm{e}^{-M\left[\frac{\xi^2}{4\sigma^2} - \mathcal{I}(\xi)\right]} \int_{\mathcal{D}(\xi)} \mathrm{d}y \mathrm{d}\lambda \, \mathrm{e}^{M\phi(y,\lambda)}, \quad (98)$$

where the remaining integrals are over the auxiliary parameters (with $\phi(y,\lambda)$ their action) and over the variable ξ, which represents the value of the second-smallest eigenvalue μ_{M-1} of the matrix. The integration over this eigenvalue has to be done separately, since for certain values of parameters the effective perturbations (76) give rise to an outlier in the spectrum, that corresponds to its smaller eigenvalue μ_{M-1}. The two integrals are coupled by the fact that (λ, y) belong to a domain $\mathcal{D}(\xi)$ that depends explicitly on the value of ξ. All the integration can be performed with a saddle point approximation. Depending on the values of ξ, the solutions (λ^*, y^*) of the minimization problem for the action $\phi(y,\lambda)$ are either within the domain, or outside the domain; in that case, the ξ-dependent boundary values have to be taken. Once the optimization over the auxiliary parameters is performed, performing the integral over ξ with a saddle point approximation we are left with:

$$\mathcal{P}_{\beta,\theta}(x,u) = \mathcal{A}_M \mathrm{e}^{-M\left[\Psi_0(x,u) + \inf_{-2\sigma \geqslant \xi \geqslant x} \Psi_1(x,u,\xi)\right]}, \quad (99)$$

where $\Psi_1(x,u,\xi)$ is (up to additive terms that are constant in ξ) the large deviation function for the smallest eigenvalue of a matrix perturbed according to (76). The optimization over ξ depends on wether x is larger or smaller than the typical value μ^{typ}_{M-1} of this eigenvalue: the different cases correspond to the different regimes of (73). In particular, when x, u are such that $\mu^{\mathrm{typ}}_{M-1} = -2\sigma$ (meaning that $\sigma^2 F(x,u) \geqslant -\sigma$), the optimum of (99) is attained at $\xi^* = -2\sigma$. When instead $\mu^{\mathrm{typ}}_{M-1} \equiv \mu_1(x,u) < -2\sigma$ (meaning that $\sigma^2 F(x,u) < -\sigma$), the optimum is at $\xi^* = \mu_1(x,u)$ if $x < \mu_1(x,u)$, or at the boundary value $\xi^* = x$ otherwise.

The other large deviation functions follow straightforwardly from an optimization over the overlap u and over the additive perturbation θ.

4. Part III: detailed derivation of large deviation functions

In this part of the paper, we present the derivation of the results summarized above. In particular, in section 4.1 we show how the joint distribution of eigenvalues μ_α and eigenvectors components u_α is modified by adding a combination of additive and multiplicative perturbations. In section 4.2 we re-write the joint distribution $\mathcal{P}_{\theta,\beta}(x,u)$ as the integral of an action depending on the configuration of the second-smallest eigenvalue ξ, and over two additional auxiliary parameters λ, y. In sections 4.3 and 4.4 we solve the saddle point equations for the auxiliary parameters λ, y, and in section 4.5 we optimize over the value of the second-smallest eigenvalue. Finally, in section 4.6 we determine the optimal value of the overlap $u_{\mathrm{typ}}(x)$, and in section 4.7 we optimize over the fluctuations of the additive perturbation θ. Additional details on the calculation are given in the Appendices.

4.1. The joint density of the smallest eigenvalue and eigenvector projection

Let μ_α be the eigenvalues of the matrix $\mathcal{Y}$ given in (49), with $\mu_M \leqslant \mu_{M-1} \leqslant \cdots \mu_1$. Let $\mathbf{v}_\alpha$ be the corresponding eigenvalues and $u_\alpha = |\mathbf{v}_\alpha \cdot \boldsymbol{e}_M|^2 \in [0,1]$. We consider θ to be fixed. We first argue that the joint probability density of μ_α, u_α reads:

$$P_{\theta,\beta}(\mu_\alpha,u_\alpha)=\frac{\mathrm{e}^{-MV(\mu_\alpha,u_\alpha)}}{\mathcal{Z}_M[\theta,\beta]}\prod_{\gamma<\alpha}(\mu_\gamma-\mu_\alpha)\prod_{\alpha=1}^{M}\theta\left(\mu_\alpha-\mu_{\alpha+1}\right)\,\delta\left(\sum_{\alpha=1}^{M}u_\alpha-1\right)\prod_\alpha\frac{1}{u_\alpha^{1/2}},\tag{100}$$

with $\mathcal{Z}_M[\theta,\beta]$ a normalization and

$$V(\mu_\alpha,u_\alpha)=\frac{1}{4\sigma^2}\left[\sum_\alpha\mu_\alpha^2+\frac{C_3^2}{4}\left(\sum_\alpha\mu_\alpha u_\alpha\right)^2+C_2\sum_\alpha\mu_\alpha u_\alpha+C_3\sum_\alpha\mu_\alpha^2u_\alpha\right],\tag{101}$$

with the constants given in (65). As a matter of fact, for the GOE matrix $\mathcal{X}$ the joint density of the ordered eigenvalues λ_α and of the eigenvectors squared components $z_\alpha=|\mathbf{e}\cdot\mathbf{w}_\alpha|^2$ along an arbitrary direction $\mathbf{e}$ is factorized, and reads:

$$\begin{aligned}p_{\mathrm{GOE}}(\lambda_\alpha,z_\alpha)=&\frac{M!}{Z_M(\sigma)}\mathrm{e}^{-M\sum_{\alpha=1}^{M}\frac{\lambda_\alpha^2}{4\sigma^2}}\prod_{\alpha<\gamma}|\lambda_\gamma-\lambda_\alpha|\prod_\alpha\theta(\lambda_\alpha-\lambda_{\alpha+1})\\&\times\frac{\Gamma(M/2)}{(\Gamma(1/2))^M}\,\delta\left(\sum_{\alpha=1}^{M}z_\alpha-1\right)\prod_\alpha\frac{1}{z_\alpha^{1/2}}\end{aligned}\tag{102}$$

with $Z_M(\sigma)$ a normalization. The distribution (100) is obtained through the change of variable:

$$\mathcal{X}=F_\beta^{-1}\left(\mathcal{Y}-\theta\mathbf{e}_M\mathbf{e}_M^T\right)F_\beta^{-1}=F_\beta^{-1}\mathcal{Y}F_\beta^{-1}-\theta(1+\beta)^2\mathbf{e}_M\mathbf{e}_M^T,\tag{103}$$

where

$$F_\beta^{-1}=\mathbb{1}+\beta\,\mathbf{e}_M\mathbf{e}_M^T.\tag{104}$$

The confinement potential is modified, since

$$\mathrm{Tr}\mathcal{X}^2=\mathrm{Tr}\left(F_\beta^{-1}\mathcal{Y}F_\beta^{-1}\right)^2+\theta^2(1+\beta)^4-2\theta(1+\beta)^2\,\mathrm{Tr}\left(F_\beta^{-1}\mathcal{Y}F_\beta^{-1}\mathbf{e}_M\mathbf{e}_M^T\right).\tag{105}$$

Using that:

$$\begin{aligned}&\mathrm{Tr}(F_\beta^{-1}\mathcal{Y}F_\beta^{-1}\mathbf{e}_M\mathbf{e}_M^T)=(1+\beta)^2\,\mathbf{e}_M\cdot\mathcal{Y}\cdot\mathbf{e}_M\\&\mathrm{Tr}\left(F_\beta^{-1}\mathcal{Y}F_\beta^{-1}\right)^2=\mathrm{Tr}\mathcal{Y}^2+(4\beta^2+4\beta^3+\beta^4)\,(\mathbf{e}_M\cdot\mathcal{Y}\cdot\mathbf{e}_M)^2+(4\beta+2\beta^2)\mathbf{e}_M\cdot\mathcal{Y}^2\cdot\mathbf{e}_M,\end{aligned}\tag{106}$$

one finds

$$\mathrm{Tr}\mathcal{X}^2=\mathrm{Tr}\mathcal{Y}^2+\frac{C_3^2}{2}\,(\mathbf{e}_M\cdot\mathcal{Y}\cdot\mathbf{e}_M)^2+C_2\mathbf{e}_M\cdot\mathcal{Y}\cdot\mathbf{e}_M+C_3\mathbf{e}_M\cdot\mathcal{Y}^2\cdot\mathbf{e}_M+\theta^2(1+\beta)^4\tag{107}$$

with the constants given in (65). The confinement potential for the eigenvalues of $\mathcal{Y}$ is therefore given by (101) and depends explicitly on their eigenvector components u_α (the constant term $\theta^2(1+\beta)^4$ is absorbed in the normalization). On the other hand, it can be easily argued that the joint measure of the eigenvector components and of the eigenvalues is left invariant by the change of variables (for an additive rank-1 perturbation, this was shown in [65] following [66]). As a consequence, the only effect of the additive and multiplicative perturbations is to introduce a coupling between the μ_α and u_α through the confinement term. From (100) we can then get that the joint density of $\mu_M=x,u_M=u$ reads:

$$\frac{\mathcal{P}_{\theta,\beta}(x,u)}{p(u)} = \frac{\mathrm{e}^{-\frac{M}{4\sigma^2}\left[x^2+C_2xu+\frac{C_3^2}{4}x^2u^2+C_3x^2u\right]}}{\mathcal{Z}_M^*[\theta,\beta]} \int \prod_{\alpha=1}^{M-1} \mathrm{d}\mu_\alpha[(\mu_\alpha - x)\theta(\mu_\alpha - x)]F(\vec{\mu})\, I_{x,u}(\vec{\mu}). \tag{108}$$

In this formula $p(u)$ the distribution of a single eigenvector component:

$$p(u) = \frac{\Gamma(M/2)}{\sqrt{\pi}\Gamma\left(\frac{M-1}{2}\right)} \frac{(1-u)^{\frac{M-3}{2}}}{\sqrt{u}} \sim (1-u)^{\frac{M}{2}}, \tag{109}$$

$\mathcal{Z}_M^*[\theta,\beta] = \mathcal{Z}_M\Gamma(M/2)/\pi^{M/2}$ is a rescaled normalization, $F(\vec{\mu})$ is the measure on the remaining $M-1$ eigenvalues:

$$F(\vec{\mu}) = \prod_{\alpha>\gamma=1}^{M-1} (\mu_\gamma - \mu_\alpha)\theta(\mu_\gamma - \mu_\alpha)\mathrm{e}^{-\frac{M}{4\sigma^2}\left[\sum_{\alpha=1}^{M-1}\mu_\alpha^2\right]}, \tag{110}$$

while $I_{x,u}(\vec{\mu})$ is an integral over the remaining u_α:

$$I_{x,u}(\vec{\mu}) = \int_0^\infty \prod_{\alpha=1}^{M-1} \mathrm{d}u_\alpha\, p(\vec{u}|u)\mathrm{e}^{-\frac{M}{4\sigma^2}\left[C_4(x,u)\sum_{\alpha=1}^{M-1}\mu_\alpha u_\alpha + \frac{C_3^2}{4}\left(\sum_{\alpha=1}^{M-1}\mu_\alpha u_\alpha\right)^2 + C_3\sum_{\alpha=1}^{M-1}\mu_\alpha^2 u_\alpha\right]}. \tag{111}$$

Here $C_4(x,u) = C_2 + (C_3^2/2)\,xu$, and $p(\vec{u}|u)$ is the uniform distribution on a sphere of radius $1-u$ in dimension $M-1$:

$$p(\vec{u}|u) = \frac{\Gamma\left(\frac{M-1}{2}\right)}{\pi^{\frac{M-1}{2}}(1-u)^{\frac{M-1}{2}-1}} \prod_{\alpha=1}^{M-1} \frac{1}{u_\alpha^{1/2}}\, \delta\left(\sum_{\alpha=1}^{M-1} u_\alpha - (1-u)\right). \tag{112}$$

Notice that $0 \leqslant u_\alpha \leqslant 1$, but the distribution (100) can be integrated on the whole positive semi-axis because the delta enforces this constraint automatically. Explicitly, we can write:

$$I_{x,u}(\vec{\mu}) = \int_{-\infty}^{\infty} \prod_{\alpha=1}^{M-1} de_\alpha \frac{\Gamma\left(\frac{M-1}{2}\right)}{\pi^{\frac{M-1}{2}}} \delta\left(\sum_{\alpha=1}^{M-1} e_\alpha^2 - 1\right) \times \mathrm{e}^{-\frac{M}{4\sigma^2}\left[C_4(x,u)(1-u)\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2 + \frac{[C_3(1-u)]^2}{4}\left(\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2\right)^2 + C_3(1-u)\sum_{\alpha=1}^{M-1}\mu_\alpha^2 e_\alpha^2\right]}. \tag{113}$$

As anticipated, the distribution (113), up to normalization constants, has the same structure as (100) but with modified constants $C_3 \to C_3(1-u)$ and $C_2 \to C_4(x,u)(1-u)$. This implies that, fixing x and u, the distribution of the remaining eigenvalues is the one of a GOE matrix perturbed exactly as the original one, with modified parameters given in (76).

4.2. Integration over the remaining eigenvectors and eigenvalues

As we show in appendix E, (113) can be re-written in the following more convenient form:

$$I_{x,u}(\vec{\mu}) = -\frac{\Gamma\left(\frac{M-1}{2}\right)}{\pi^{\frac{M-1}{2}}} \sqrt{\frac{M^3\,4\sigma^2}{\pi C_3^2(1-u)^2}} \left[\frac{2\sigma^2}{C_3(1-u)}\right]^{\frac{M-3}{2}} \int\int_{-i\infty}^{i\infty} \mathrm{d}y\,\mathrm{d}\lambda \mathrm{e}^{M\left(\frac{y^2}{\sigma^2}-\lambda\right)} I_2 \tag{114}$$

with

$$I_2(y,\lambda,\vec{\mu}) = \int_{-\infty}^{\infty} \prod_{\alpha=1}^{M-1} de_\alpha \mathrm{e}^{-\frac{M}{2}\sum_{\alpha=1}^{M-1} e_\alpha^2\left[\mu_\alpha^2 + \left(\frac{C_4(x,u)}{C_3} - 2y\right)\mu_\alpha - 2\lambda\frac{2\sigma^2}{C_3(1-u)}\right]}. \tag{115}$$

The parameters y and λ in (114) are auxiliary fields that will be fixed through a saddle point calculation, while the integrals (115) are decoupled Gaussian integrals whose convergence imposes some constraints on the domain of y, λ. In particular, given the functions

$$\mu^{\pm}_{x,u}(y,\lambda) = -\frac{1}{2}\left(\frac{C_4(x,u)}{C_3} - 2y\right) \pm \frac{1}{2}\sqrt{8\lambda\frac{2\sigma^2}{C_3(1-u)} + \left(\frac{C_4(x,u)}{C_3} - 2y\right)^2}, \tag{116}$$

the condition for the convergence of the integrals in (115) reads (assuming that λ, y are real):

$$\mu_\alpha^2 + \left(\frac{C_4(x,u)}{C_3} - 2y\right)\mu_\alpha - 2\lambda\frac{2\sigma^2}{C_3(1-u)} = [\mu_\alpha - \mu^+_{x,u}(y,\lambda)][\mu_\alpha - \mu^-_{x,u}(y,\lambda)] > 0 \quad \forall\alpha. \tag{117}$$

For a given configuration of eigenvalues μ_α, we denote with $\mathcal{D}[\mu_\alpha]$ the domain of λ, y for which (117) is satisfied. Performing the Gaussian integration, (108) becomes equal to:

$$\frac{\mathcal{P}_{\theta,\beta}(x,u)}{p(u)} = \frac{\alpha_M(u)}{\mathcal{Z}_M\,\Gamma\left(\frac{M}{2}\right)}\left[\frac{2\pi\sigma^2}{e\,C_3(1-u)}\right]^{\frac{M}{2}} \mathrm{e}^{-\frac{M}{4\sigma^2}\left[x^2 + C_2 xu + \frac{C_3^2}{4}x^2u^2 + C_3x^2u\right]}\mathcal{J}_{\theta,\beta}(x,u) \tag{118}$$

where

$$\alpha_M(u) = -\frac{M^2}{2\sigma^2}\,\Gamma\left(\frac{M-1}{2}\right)\left(\frac{2e}{M}\right)^{\frac{M}{2}}\sqrt{\frac{C_3(1-u)}{\pi}} \tag{119}$$

scales less than exponentially with M,

$$\mathcal{J}_{\theta,\beta}(x,u) = \int\prod_{\alpha=1}^{M-1}\mathrm{d}\mu_\alpha\,\mathbf{1}_{x<\mu_{M-1}<\cdots\leqslant\mu_1}\int\mathrm{d}y\,\mathrm{d}\lambda\,\mathbf{1}_{\lambda,y\in\mathcal{D}[\mu_\alpha]}\,\mathrm{e}^{-M^2\tilde{S}_1[\vec{\mu}]-M\tilde{S}_0[y,\lambda,\vec{\mu}]} \tag{120}$$

where $\mathbf{1}$ is the indicator function, and the actions have the following expression:

$$\begin{aligned}
\tilde{S}_1[\vec{\mu}] &= \frac{1}{4\sigma^2}\frac{1}{M}\sum_{\alpha=1}^{M-1}\mu_\alpha^2 - \frac{1}{M^2}\sum_{\alpha>\gamma=1}^{M-1}\log(\mu_\gamma - \mu_\alpha),\\
\tilde{S}_0[y,\lambda,\vec{\mu}] &= \lambda - \frac{y^2}{\sigma^2} + \frac{1}{2M}\sum_{\alpha=1}^{M-1}\log[(\mu_\alpha - \mu^+_{x,u})(\mu_\alpha - \mu^-_{x,u})] - \frac{1}{M}\sum_{\alpha=1}^{M-1}\log(\mu_\alpha - x).
\end{aligned} \tag{121}$$

Notice that the action $\tilde{S}_1[\vec{\mu}]$ is the one corresponding to the joint distribution of the eigenvalues of an unperturbed GOE matrix, and is given by one-point functions of the eigenvalues. These actions can be expressed in terms of the eigenvalue density $\nu(\mu) = \sum_{\alpha=1}^{M}\delta(\mu - \mu_\alpha)$, performing the change of variable $\vec{\mu} \to \nu(\mu)$. Naturally, the density $\nu(\mu)$ can have both a continuous part and some poles, corresponding to the isolated eigenvalues. The dominating term of $\tilde{S}_1$ depends only on the continuous part of $\nu(\mu)$, and reproduces exactly then term that one would get from an unperturbed GOE; therefore, the corresponding action is zero at the typical density $\nu^{\mathrm{typ}}_{\mathrm{cont}}(\mu) = \rho_\sigma(\mu)$ corresponding to the semicircle law (51). Any contribution to $\nu(\mu)$ coming from isolated poles is of $O(1/M)$, and gives rise to sub-leading contributions to $\tilde{S}_1$ that have to be added to the linear term in M of the exponent in (118).

To proceed with the calculation, we assume that *only one* of these poles can be present, corresponding to the second-smallest eigenvalue μ_{M-1} of the matrix. We show that, under this assumption, the saddle-point equations obtained by minimizing the linear term in M of the resulting action fix this eigenvalue to its typical value $\mu^{\mathrm{typ}}_{M-1}(x,u)$ at fixed x and u, which

is either -2σ when $\sigma^2 F(x,u) \geqslant -\sigma$, or (79) otherwise, consistently with the results in section 3.2.1. We subsequently need to check that the hypothesis is consistent, meaning that whenever the second eigenvalue is fixed to its typical value, the third-smallest eigenvalue typically sticks to the boundary of the semicircle $\mu_{M-2}^{\text{typ}} = -2\sigma$. We discuss this check in appendix I.

We therefore assume that the only eigenvalue that can take values smaller that -2σ is μ_{M-1} and integrate over the remaining ones, getting:

$$\mathcal{J}_{\theta,\beta}(x,u) = A_M \int_{\xi \geqslant x} \mathrm{d}\xi\, h(\xi,x) \int \mathrm{d}y\, \mathrm{d}\lambda\, \mathbf{1}_{\lambda, y \in \mathcal{D}[\xi]} \times \mathrm{e}^{-M\left[\frac{\xi^2}{4\sigma^2} - \int \mathrm{d}\mu \log[(\mu-\xi)(\mu-x)]\rho_\sigma(\mu) + \lambda - \frac{y^2}{\sigma^2} + \frac{1}{2}\int \mathrm{d}\mu \log[(\mu-\mu^+_{x,u})(\mu-\mu^-_{x,u})]\rho_\sigma(\mu)\right]}, \tag{122}$$

where $h(\xi,x) = (\xi - x)/[(\xi - \mu^+_{x,u})(\xi - \mu^-_{x,u})]^{1/2}$, and A_M contains constant terms coming from the change of variables $\vec{\mu} \to \nu(\mu)$. Combining everything, asymptotically at the exponential scale in M we find:

$$\mathcal{P}_{\theta,\beta}(x,u) \sim \mathcal{A}_M\, \mathrm{e}^{-M\Psi_0(x,u)} \int_{\xi \geqslant x} \mathrm{d}\xi \mathrm{e}^{-M\left[\frac{\xi^2}{4\sigma^2} - \mathcal{I}(\xi)\right]} \int_{\mathcal{D}(\xi)} \mathrm{d}y \mathrm{d}\lambda\, \mathrm{e}^{M\phi(y,\lambda)}, \tag{123}$$

with

$$\Psi_0(x,u) = \frac{1}{4\sigma^2}\left(x^2 + C_2 xu + \frac{C_3^2}{4}x^2u^2 + C_3 x^2 u\right) - \frac{1}{2}\log\left(\frac{2\sigma^2}{C_3}\right) - \mathcal{I}(x) + \frac{1}{2},$$
$$\phi(y,\lambda) = \frac{y^2}{\sigma^2} - \lambda - \frac{1}{2}\int \mathrm{d}\mu \rho_\sigma(\mu) \log\left[(\mu - \mu^-_{x,u}(y,\lambda))(\mu - \mu^+_{x,u}(y,\lambda))\right]. \tag{124}$$

The expression for $\mathcal{I}(z)$ is given in (69) (and we are using that $x \leqslant -2\sigma$), and we made use of the identity:

$$\frac{\sqrt{z^2 - 4\sigma^2}}{2} = \sigma^2 G(z) - \frac{z}{2} \qquad \text{for} \quad z < -2\sigma. \tag{125}$$

The constant $\mathcal{A}_M$ in (123) contains exponential contributions that have to be determined from the condition:

$$\mathcal{P}_{\theta,\beta}(x_{\text{typ}}, u_{\text{typ}}) \sim O(1), \tag{126}$$

where $x_{\text{typ}}, u_{\text{typ}}$ are the typical values of μ_M and u_M at fixed θ, β.

Finally, we comment on the domain $\mathcal{D}$. The latter changes depending on whether the roots $\mu^\pm_{x,u}$ are real or complex. We can distinguish the following two cases :

- Case A: the roots $\mu^\pm_{x,u}(y,\lambda)$ are complex: this happens whenever the discriminant is negative, corresponding to

$$\lambda < -\frac{1}{8}\left(\frac{C_4(x,u)}{C_3} - 2y\right)^2 \frac{C_3(1-u)}{2\sigma^2} \leqslant 0. \tag{127}$$

  In this case the condition (117) is always met, and one can set

$$\mathcal{D} = \{(y,\lambda) :\ \lambda \leqslant 0 \quad \text{and} \quad y \in \mathbb{R}\}. \tag{128}$$

- Case B: The roots $\mu^-_{x,u}(y,\lambda) \leqslant \mu^+_{x,u}(y,\lambda)$ are real; for $\theta < 0$, one can self-consistently check that $\mu^+_{x,u}(y,\lambda) < 0$. A necessary condition for (117) to hold true is that $\mu^+_{x,u}(y,\lambda) \leqslant -2\sigma$, meaning that the support of the continuous part of the eigenvalue distribution lies to the right of $\mu^+_{x,u}$. Additionally, we have to impose that the eigenvalues that do not belong to the continuous part of the eigenvalue density satisfy the condition. This implies that either $\xi < \mu^-_{x,u}(y,\lambda)$ or $\mu^+_{x,u}(y,\lambda) < \xi$, meaning:

$$\mathcal{D}(\xi) = \left\{(y,\lambda) : \xi \leqslant \mu^-_{x,u}(y,\lambda) \text{ or } \mu^+_{x,u}(y,\lambda) \leqslant \xi \text{ and } \mu^+_{x,u}(y,\lambda) < -2\sigma\right\}. \tag{129}$$

4.3. Saddle point equations for the auxiliary fields I: inside the domain

In this section we discuss the saddle point equations for $\phi(y,\lambda)$ in (124), at fixed values of ξ. To simplify the notation, we denote $\mu^\pm_{x,u}$ simply with $\mu^\pm$.

The minimization of $\phi(y,\lambda)$ gives the following two equations:

$$\begin{aligned} \frac{C_3(1-u)}{2\sigma^2}\left(\mu^+ - \mu^-\right) &= G_\sigma(\mu^-) - G_\sigma(\mu^+) \\ \frac{4}{\sigma^2}y + \frac{C_3(1-u)}{2\sigma^2}\left(\mu^+ + \mu^-\right) &= G_\sigma(\mu^-) + G_\sigma(\mu^+). \end{aligned} \tag{130}$$

Summing and subtracting these equations we get the relations:

$$\begin{aligned} \frac{C_3(1-u)}{2\sigma^2}\mu^+ + \frac{2}{\sigma^2}y &= G_\sigma(\mu^-) \\ \frac{C_3(1-u)}{2\sigma^2}\mu^- + \frac{2}{\sigma^2}y &= G_\sigma(\mu^+). \end{aligned} \tag{131}$$

Assuming that G_σ can be inverted, these can be re-written as:

$$\begin{aligned} \mu^+(\lambda,y) &= G_\sigma^{-1}\left(\frac{C_3(1-u)}{2\sigma^2}\mu^-(\lambda,y) + \frac{2}{\sigma^2}y\right) \\ \mu^-(\lambda,y) &= G_\sigma^{-1}\left(\frac{C_3(1-u)}{2\sigma^2}\mu^+(\lambda,y) + \frac{2}{\sigma^2}y\right). \end{aligned} \tag{132}$$

As we show in appendix F, regardless of whether $\mu^\pm_{x,u}$ are complex or real, the solutions of these equations is given by:

$$y^* = \frac{C_4(x,u)C_3(1-u)^2}{2\left[2 + C_3(1-u)\right]^2}, \qquad \lambda^* = -\frac{\left[\sigma^2(C_3(1-u)+2)^2 + C_4^2(1-u)^2\right]}{\sigma^2(C_3(1-u)+2)^3}. \tag{133}$$

When $\mu^\pm$ are computed at the saddle point solutions y^*, λ^*, the correspondent action is given by:

$$\phi(\lambda^*,y^*) = \frac{1}{2} - \frac{1}{2}\log\left[\sigma^2\left(1 + \frac{2}{C_3(1-u)}\right)\right] + \frac{C_4^2(x,u)(1-u)^2}{4\sigma^2\left[2 + C_3(1-u)\right]^2}. \tag{134}$$

We now discuss the conditions under which the GOE resolvent can be inverted, and the saddle point solutions lie in the right domain $\mathcal{D}[\xi]$. If Case A holds, the equation are always

invertible given that the resolvent is never singular. When $\mu^{\pm}_{x,u}$ are real, i.e. when Case B holds, to write (132) it must hold:

$$\left|\frac{C_3(1-u)}{2\sigma^2}\mu^{\pm}+\frac{2}{\sigma^2}y\right|\leqslant\frac{1}{\sigma}. \tag{135}$$

Since for $\mu^{\pm}<0$ and thus $G(\mu^{\pm})<0$, these conditions become

$$F^{\pm}\equiv\frac{C_3(1-u)}{2\sigma^2}\mu^{\pm}(\lambda^*,y^*)+\frac{2}{\sigma^2}y^*\geqslant-\frac{1}{\sigma}, \tag{136}$$

and given that $F^+>F^-$ one has to impose that

$$F(x,u)\equiv F^-=-\frac{C_4(1-u)+\sqrt{C_4^2(1-u)^2-\sigma^2C_3(1-u)\left[2+C_3(1-u)\right]^3}}{\sigma^2\left[2+C_3(1-u)\right]^2}\geqslant-\frac{1}{\sigma}. \tag{137}$$

As we anticipated in (77), this condition is equivalent to $[G_{\tilde{\sigma}}(\tilde{\theta})]^{-1}\geqslant-\sigma$, which is the condition under which the typical value of the second-smallest eigenvalue $\mu^{\rm typ}_{M-1}$ *is not* out of the bulk. In this case we find:

$$\xi^{\pm}_{\sigma}(x,u)\equiv\mu^{\pm}_{x,u}(\lambda^*,y^*)=m^{\pm}_{\sigma}[C_4(x,u)(1-u),C_3(1-u)], \tag{138}$$

where $m^{\pm}_{\sigma}$ are given in (75). In this regime of parameters, the saddle point solution (y^*,λ^*) lies within $\mathcal{D}[\xi]$ iff

$$\xi\geqslant\mu^+(y^*,\lambda^*)=G^{-1}_{\sigma}\left(F(x,u)\right)=G^{-1}_{\sigma}\left(\frac{1}{\sigma^2G_{\tilde{\sigma}}(\tilde{\theta})}\right)\quad\text{or}\quad\xi\leqslant\mu^-(y^*,\lambda^*), \tag{139}$$

and has to be discarded otherwise.

When (135) is not met and $F^+\geqslant-1/\sigma>F^-$, the equation for $\mu^+(y,\lambda)$ still admits a solution $\mu^+(y^*,\lambda^*)$, which nevertheless belongs to the second Riemann sheet in the complex plane. This is due to the fact that the quadratic equation for the resolvent of a GOE matrix $\sigma^2G^2_{\sigma}(z)-zG_{\sigma}(z)+1=0$ admits another solution

$$G^{(II)}_{\sigma}(z)=\frac{1}{2\sigma^2}\left(z+\text{sign}(z)\sqrt{z^2-4\sigma^2}\right) \tag{140}$$

for z real, which is obtained from $G_{\sigma}(z)$ changing the sign in front of the square root. This function is defined on the second Riemann sheet, and it takes values in $|z|>1/\sigma$. Its inverse is again given by $G^{-1}_{\sigma}(z)=z^{-1}+\sigma^2z$, but now evaluated in this domain $|z|>1/\sigma$.

When $F^+\geqslant-1/\sigma>F^-$, $\mu^+(y^*,\lambda^*)$ solves the second of equation (131) with $G_{\sigma}\to G^{II}_{\sigma}$. In this case, the saddle point solution (y^*,λ^*) can still be considered, and it lies within the integration domain $\mathcal{D}[\xi]$ iff $\xi<\mu^-(y^*,\lambda^*)$. Notice that this is the regime of parameters in which the typical value of the second-smallest eigenvalue *is* out of the bulk. Using the results of section 3.2.1 we know that the latter can be written as:

$$\mu_1(x,u)\equiv\mu^{\rm typ}_{M-1}=G^{-1}_{\sigma}\left(G_{\tilde{\sigma}}(\tilde{\theta})\right)=G^{-1}_{\sigma}\left(\frac{1}{\sigma^2F(x,u)}\right), \tag{141}$$

where here the argument of G^{-1}_{σ} is larger than $-1/\sigma$.

We notice that the explicit expression of the typical value (141) is exactly the same as the one of $\xi^+_{\sigma}(x,u)=\mu^+_{x,u}(\lambda^*,y^*)=G^{-1}_{\sigma}\left(F(x,u)\right)$. The two expressions coincide due to the following symmetry of the function G^{-1}_{σ} on the real axis:

$$G_\sigma^{-1}(x) = G_\sigma^{-1}\left(\frac{1}{\sigma^2 x}\right). \tag{142}$$

Therefore, the threshold value $\xi_\sigma^+(x,u)$ can be though of as the analytic continuation of the expression for $\mu_1(x,u)$, extended to a regime of parameters for which typically the eigenvalue *is not* out of the bulk of the semicircle[5]. The difference between the two quantities is that in this regime, while $\mu_1(x,u)$ lies in the first Riemann sheet, $\xi_\sigma^+(x,u)$ lies in the second.

Finally, the case $F^+ < -1/\sigma$ needs not to be considered, since F^+ becomes complex before reaching the threshold value $F^+ = -1/\sigma$[6]. In summary, the saddle point solutions is acceptable whenever the resulting $\mu^\pm(y^*,\lambda^*)$ are either complex (case A), or when they satisfy any of the two conditions (139).

4.4. Saddle point equations for the auxiliary fields II: boundary of domain

When Case B holds but (139) is not met, y^*, λ^* do not belong to the domain $\mathcal{D}[\xi]$, and the rate function ϕ has to be computed at boundary manifold, where one of the two following equalities hold:

$$\xi = -\frac{1}{2}\left(\frac{C_4}{C_3} - 2y\right) \pm \frac{1}{2}\sqrt{8\lambda\frac{2\sigma^2}{C_3(1-u)} + \left(\frac{C_4}{C_3} - 2y\right)^2} = \mu_{x,u}^\pm(y,\lambda). \tag{144}$$

This is an equation relating y, λ. Assuming that (144) holds for some $\lambda = \lambda_{\text{ext}}(\xi)$ and $y = y_{\text{ext}}(\xi)$, taking its square we get the relations:

$$\begin{aligned}
\lambda_{\text{ext}}(y;\xi) &= \frac{C_3(1-u)}{4\sigma^2}\left[\xi^2 + \left(\frac{C_4(x,u)}{C_3} - 2y\right)\xi\right], \\
y_{\text{ext}}(\lambda;\xi) &= \frac{1}{2}\left[\frac{C_4(x,u)}{C_3} - \frac{4\sigma^2\lambda}{C_3(1-u)\xi} + \xi\right].
\end{aligned} \tag{145}$$

Substituting the first of these equations into $\phi(\lambda,y)$ and minimizing over y we get:

$$\frac{2y}{\sigma^2} + \frac{C_3(1-u)}{2\sigma^2}\xi - G\left(-\xi - \frac{C_4}{C_3} + 2y\right) = 0. \tag{146}$$

The two equations are solved by:

$$\begin{aligned}
y_{\text{ext}}(\xi) &= -\frac{C_3\sigma^2}{C_3\xi(C_3(1-u)+2)+2C_4} - \frac{1}{4}C_3(1-u)\xi, \\
\lambda_{\text{ext}}(\xi) &= \frac{C_3\xi(1-u)}{4\sigma^2}G^{-1}\left(\frac{2C_3}{2C_4 + C_3\xi\,[2 + C_3(1-u)]}\right).
\end{aligned} \tag{147}$$

If the second equation in (145) is used we get an equivalent result, see appendix F for the details. The rate function $\phi(y,\lambda)$ computed at (147) reads

[5] Notice that when $\beta \to 0$, we correctly recover $F \to \mu(1-u)/\sigma^2$ and thus (137) reduces to $\mu(1-u) > -\sigma$. The other threshold value $\xi_\sigma^-(x,u)$ diverges to $-\infty$ in this limit.

[6] Indeed, exactly at the point when $F^\pm$ develop a complex part and we transition to Case A, the functions take the value:

$$F^\pm(x,u) = -\frac{C_4(1-u)}{\sigma^2[2+C_3(1-u)]^3} = -\frac{1}{\sigma}\sqrt{\frac{C_3(1-u)}{2+C_3(1-u)}} \geqslant -\frac{1}{\sigma}. \tag{143}$$

$$\phi(y_{\text{ext}},\lambda_{\text{ext}}) = -\frac{(1-u)\xi\left[4C_4 + C_3\xi(4+C_3(1-u))\right]}{16\sigma^2} - \frac{\mathcal{I}(\xi)}{2} + \frac{1}{2}\log\left[\frac{2C_3}{C_3\xi(C_3(1-u)+2)+2C_4}\right], \tag{148}$$

as we derive in the same appendix.

4.5. The variational problem for the second-smallest eigenvalue

Combining (123) with the results of the last two section, we find that:

$$\mathcal{P}_{\beta,\theta}(x,u) = \mathcal{A}_M \mathrm{e}^{-M\left[\Psi_0(x,u)+\inf_{-2\sigma\geqslant\xi\geqslant x}\Psi_1(x,u,\xi)\right]} = \mathrm{e}^{-M\left[\Psi_0(x,u)+\inf_{-2\sigma\geqslant\xi\geqslant x}\Psi_1(x,u,\xi)-l(\theta,\beta)\right]} \tag{149}$$

with $l(\theta,\beta)$ defined by $\mathcal{A}_M = \exp\left(Ml(\theta,\beta)+o(M)\right)$, and

$$\Psi_1(x,u,\xi) = \frac{1}{4\sigma^2}\xi^2 - \int \mathrm{d}\mu\rho_\sigma(\mu)\log(\mu-\xi) - \Phi(x,u;\xi). \tag{150}$$

The function $\Phi(x,u;\xi)$ is given by:

$$\Phi(x,u;\xi) = \begin{cases} \phi_1(x,u) & \text{if Case A, Cond 1 or Cond 4} \\ \phi_2(x,u;\xi) & \text{if Cond 2 or Cond 3} \end{cases} \tag{151}$$

where

$$\begin{aligned} \phi_1 &\equiv \frac{1}{2} - \frac{1}{2}\log\left[\sigma^2\left(1+\frac{2}{C_3(1-u)}\right)\right] + \frac{C_4^2(x,u)(1-u)^2}{4\sigma^2\left[2+C_3(1-u)\right]^2} \\ \phi_2 &\equiv \frac{1}{2}\log\left[\frac{2C_3}{C_3\xi(C_3(1-u)+2)+2C_4(x,u)}\right] - \frac{(1-u)\xi\left[4C_4+C_3\xi(4+C_3(1-u))\right]}{16\sigma^2} - \frac{\mathcal{I}(\xi)}{2} \end{aligned} \tag{152}$$

where the conditions are:

$$\begin{aligned} &\text{Cond 1}: \quad \sigma^2 F(x,u) \geqslant -\sigma \text{ and } \xi \geqslant \xi_\sigma^+(x,u) \text{ or } \xi \leqslant \xi_\sigma^-(x,u) \\ &\text{Cond 2}: \quad \sigma^2 F(x,u) \geqslant -\sigma \text{ and } \xi_\sigma^-(x,u) < \xi < \xi_\sigma^+(x,u) \\ &\text{Cond 3}: \quad \sigma^2 F(x,u) < -\sigma \text{ and } \xi_\sigma^-(x,u) < \xi \\ &\text{Cond 4}: \quad \sigma^2 F(x,u) < -\sigma \text{ and } \xi_\sigma^-(x,u) \geqslant \xi. \end{aligned} \tag{153}$$

The function Ψ_1 in (150) is (up to constants) the large deviation function for the second eigenvalue, that we need to optimize in the domain $[x,-2\sigma]$. We can distinguish the following two cases:

- If $\sigma^2 F(x,u) \geqslant -\sigma$, typically the second-smallest eigenvalue *is not* out of the bulk. In this case the large deviation function has *three* regimes:

$$\Psi_1(x,u,\xi) = \begin{cases} \frac{1}{4\sigma^2}\xi^2 - \mathcal{I}(\xi) - \phi_1(x,u) & \text{if } \xi_\sigma^+ \leqslant \xi \leqslant -2\sigma \\ \frac{1}{4\sigma^2}\xi^2 - \mathcal{I}(\xi) - \phi_2(x,u,\xi) & \text{if } \xi_\sigma^- < \xi < \xi_\sigma^+ \\ \frac{1}{4\sigma^2}\xi^2 - \mathcal{I}(\xi) - \phi_1(x,u) & \text{if } \xi \leqslant \xi_\sigma^- \end{cases} \tag{154}$$

  and it is always minimal at $\xi = -2\sigma$, meaning that:

$$\inf_{-2\sigma\geqslant\xi\geqslant x}\Psi_1(x,u,\xi) = 1 - \mathcal{I}(-2\sigma) - \phi_1(x,u) = \frac{1}{2} - \log\sigma - \phi_1(x,u). \tag{155}$$

- If $\sigma^2 F(x,u) < -\sigma$, typically the second-smallest eigenvalue *is* out of the bulk, and takes value $\mu_1(x,u)$. In this case for any $\xi \in [\xi_\sigma^-, -2\sigma]$ it holds:

$$\Psi_1 = \frac{\xi^2}{4\sigma^2} - \frac{\mathcal{I}(\xi)}{2} + \frac{(1-u)\xi\left[4C_4 + C_3\xi(4 + C_3(1-u))\right]}{16\sigma^2} - \frac{1}{2}\log\left[\frac{2C_3}{C_3\xi(C_3(1-u)+2)+2C_4}\right], \tag{156}$$

which has a minimum at $\mu_1(x,u)$; indeed the derivative of Ψ_1 is proportional to:

$$\frac{2\sigma^2 C_3(C_3(1-u)+2)}{C_3\xi(C_3(1-u)+2)+2C_4} + \frac{\xi C_3(1-u)(C_3(1-u)+4)}{2} + C_4(1-u) - \xi + \sqrt{\xi^2 - 4\sigma^2} = 0. \tag{157}$$

Among the solutions to this equation, the only one that does not diverge in the limit $\beta \to 0$ is precisely given by $\mu_1(x,u)$. Depending on the position of x with respect to $\mu_1(x,u)$, the infimum is either attained at the minimum or at the boundary, meaning:

$$\inf_{-2\sigma \geqslant \xi \geqslant x} \Psi_1(x,u,\xi) = \begin{cases} \Psi_1(x,u,\xi = x) & \text{if } \mu_1(x,u) \leqslant x \\ \Psi_1(x,u,\mu_1(x,u)) & \text{if } x < \mu_1(x,u). \end{cases} \tag{158}$$

In appendix G, we comment on the consistence between the large deviation function for the second-smallest eigenvalue $\Psi_1(x,u;\xi)$ and known results in the literature [62] valid in the limit $\beta \to 0$. To conclude this section, we simplify the resulting expressions by noticing that Ψ_1 in (156) one has the identity:

$$\begin{aligned}\Psi_1(x,u;y \to x) =& \frac{x^2}{4\sigma^2} - \frac{1}{2}\mathcal{I}(x) - \frac{1}{2}\log\left[\frac{2C_3}{C_3^2 x + 2C_3 x + 2C_2}\right] \\ &+ \frac{(1-u)x}{4\sigma^2}\left(C_4(x,u) + C_3 x + \frac{C_3^2}{4}x(1-u)\right),\end{aligned} \tag{159}$$

implying that in the relevant regime, $\Psi_0(x,u) + \Psi_1(x,u;y \to x) = \mathcal{L}^{(b)}_{\theta,\beta}(x)$ given in (68). Similarly, as we show in the same appendix it holds:

$$\Psi_1(x,u;y \to \xi_\sigma^+) = -\frac{1}{2}\log\left(\frac{C_3(1-u)}{C_3(1-u)+2}\right) - \frac{C_4^2(x,u)(1-u)^2}{4\sigma^2\left[2 + C_3(1-u)\right]^2} = \frac{1}{2} - \log\sigma - \phi_1(x,u), \tag{160}$$

and thus in all other regimes the sum $\Psi_0(x,u) + \Psi_1$ equals to $\mathcal{L}^{(a)}_{\theta,\beta}(x,u)$ given again in (68). Combining all this we recover the results stated in section 3.2, up to the constant $l(\theta,\beta)$. In the following subsection, we determine the typical value of the overlap parameter u at fixed x, and compute the constant $l(\theta,\beta)$.

4.6. Optimization over the overlap u

We now discuss the optimization of the large deviation function $\mathcal{L}_{\theta,\beta}(x,u)$ over the overlap $u \in [0,1]$. The functions to be optimized change across the different regimes. Since $\mathcal{L}^{(b)}_{\theta,\beta}(x)$ does not depend on u, the integral over u in this case does not give any exponential contribution. We therefore focus on $\mathcal{L}^{(a)}_{\theta,\beta}(x,u)$ and identify the solutions of

$$\frac{\partial \mathcal{L}^{(a)}_{\theta,\beta}(x,u)}{\partial u} = 0 \tag{161}$$

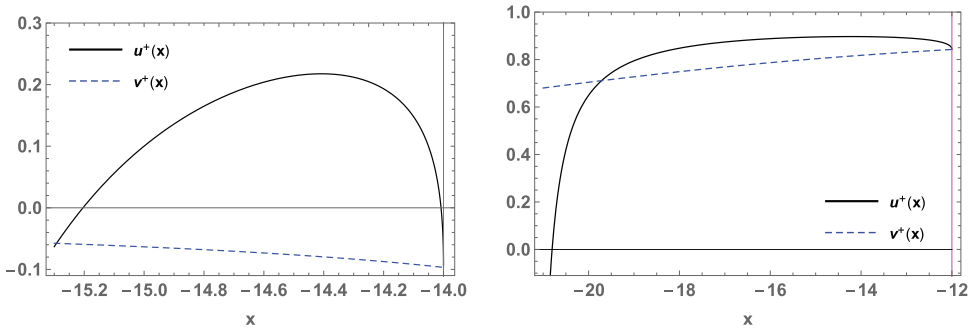

Figure 8. Plots of $u^{+}_{\theta,\beta}(x)$ for values of parameters in which the smallest eigenvalue is typically at the boundary of the semicircle (*left*) or out of the bulk (*right*). The dashed blue curve denotes $v^{+}_{\theta,\beta}(x)$, see the discussion in appendix H. The region where $u^{+}_{\theta,\beta}(x) \geqslant v^{+}_{\theta,\beta}(x)$ corresponds to the regime of parameters in which $\sigma^2 F(x,u) + \sigma > 0$.

that lie within the unit interval. This variational equation is quadratic in u, with two solutions

$$u^{\pm}_{\theta,\beta}(x) = \frac{4C_2^2 + 4C_2[C_3(C_3+3)+1]x + C_3(C_3+2)[(C_3(C_3+4)+2)x^2 + 4\sigma^2]}{4C_2^2 + 4C_2C_3(C_3+3)x + C_3^2\{(C_3+2)(C_3+4)x^2 + 4\sigma^2\}} \pm \frac{2\sqrt{(x^2-4\sigma^2)\,(2C_2 + C_3(C_3+2)x)^2}}{4C_2^2 + 4C_3C_3(C_3+3)x + C_3^2\{(C_3+2)(C_3+4)x^2 + 4\sigma^2\}}, \tag{162}$$

of which the relevant one satisfying $0 \leqslant u \leqslant 1$ for at least some values of x is $u^{+}_{\theta,\beta}(x)$, which corresponds to a minimum of $\mathcal{L}^{(a)}_{\theta,\beta}(x,u)$. Notice that in the limit $C_3 \to 0$ (equivalently, $\beta \to 0$) corresponding to a purely additive perturbation, this reduces to (using $\theta < 0$):

$$u^{+}_{\theta,\beta}(x) \to 1 - \frac{x + \sqrt{x^2 - 4\sigma^2}}{2\theta}, \tag{163}$$

which agrees with the known results [63]. When $u^{+}_{\theta,\beta}(x)$ is non-negative, we always find $u^{+}_{\theta,\beta}(x) < 1$. Therefore we can set:

$$u^{(a)}_{\text{typ}}(x) \equiv \max\left\{0, u^{+}_{\theta,\beta}(x)\right\}. \tag{164}$$

Here the superscript denotes that $u^{(a)}_{\text{typ}}(x)$ is obtained assuming $\mathcal{L}_{\theta,\beta}(x,u) \propto \mathcal{L}^{(a)}_{\theta,\beta}(x,u)$. In order to discuss the form of $u^{(a)}_{\text{typ}}(x)$, we find it convenient to separate the three following regimes of the parameters θ, β:

- Regime A: When $-2\sigma' < \theta < 0$, we find that $u^{+}_{\theta,\beta}(x) < 0$ and thus $u^{(a)}_{\text{typ}}(x) = 0$.
- Regime B1: When $\theta_c < \theta < -2\sigma'$ with $\theta_c = -\sigma[1 + 4\beta + 2\beta^2]/(1+\beta)^2$, given in (52), we find that the function $u^{+}_{\theta,\beta}(x)$ behaves as in figure 8 (*left*): it is non-monotonic in x[7], and at $x = -2\sigma$ it takes the value

$$u^{+}_{\theta,\beta}(-2\sigma) = \frac{(1+\beta)^4\theta + \sigma(\beta+1)^2[1+2\beta(2+\beta)]}{(1+\beta)^4\theta + \sigma\beta(\beta+2)[3+2\beta(2+\beta)]}, \tag{165}$$

[7] This is due to the fact that the coefficient of the quadratic term in the equation for u depends on x, and vanishes at a value of x which corresponds to the poles of (162). So at this value of x one has a divergence of the solution for u to $-\infty$ [the pole diverges to $-\infty$ when $C_3 \to 0$]. The divergence gives the non-monotonicity.

which is always negative. Indeed, $\theta \leqslant -2\sigma'$ implies that the denominator is always negative, while the numerator changes sign exactly at $\theta = \theta_c$. The function $u^+_{\theta,\beta}(x)$ vanishes exactly at the points $x^\pm_\sigma(\theta,\beta)$ given in (81), and it positive in the regime $x^-_\sigma(\theta,\beta) < x < x^+_\sigma(\theta,\beta)$. Therefore in this regime the optimization of $\mathcal{L}^{(a)}_{\theta,\beta}(x,u)$ subject to the constraint $u \in [0,1]$ gives:

$$u^{(a)}_{\rm typ}(x) = \begin{cases} 0 & \text{if } x^+_\sigma(\theta,\beta) < x < -2\sigma \\ u^+_{\theta,\beta}(x) & \text{if } x^-_\sigma(\theta,\beta) < x < x^+_\sigma(\theta,\beta) \\ 0 & \text{if } x < x^-_\sigma(\theta,\beta). \end{cases} \tag{166}$$

Notice that these values of θ coincide with the regime in which typically the smallest eigenvalue of the perturbed matrix is at the boundary of the semicircle.

- Regime B2: When $\theta \leqslant \theta_c$ the function $u^+_{\theta,\beta}(x)$ behaves as in figure 8 (*right*): it is again non-monotonic in x, but it is positive at $x = -2\sigma$, with only one zero at $x = x^-_\sigma(\theta,\beta)$. Therefore in this case:

$$u^{(a)}_{\rm typ}(x) = \begin{cases} u^+_{\theta,\beta}(x) & \text{if } x^-_\sigma(\theta,\beta) < x < -2\sigma \\ 0 & \text{if } x < x^-_\sigma(\theta,\beta). \end{cases} \tag{167}$$

Notice that these values of θ coincide with the regime in which typically the smallest eigenvalue of the perturbed matrix is smaller than -2σ, and equals to $\mu_0(\theta,\beta)$.

In order for $u^{(a)}_{\rm typ}(x)$ to be the correct solution for the optimal overlap, we have to check self-consistently that the conditions that imply $\mathcal{L}_{\theta,\beta}(x,u) \propto \mathcal{L}^{(a)}_{\theta,\beta}(x,u)$ are satisfied when $u \to u^{(a)}_{\rm typ}$. In appendix H we perform this self-consistent check, showing that when the optimization over u is performed, the relevant rate function is always $\mathcal{L}^{(a)}_{\theta,\beta}$: when the overlap u is allowed to take its typical value, one always finds that the typical value of the second-smallest eigenvalue is out of the bulk and *larger* than x, which is the large-deviation value of the smallest one, as it is natural to expect. Using the above expressions, we find:

$$\begin{aligned} \mathcal{L}^{(a)}_{\theta,\beta}(x,u^+_{\theta,\beta}(x)) &= 1 - \log\sigma^2 + \frac{1}{2}\log\frac{C_3}{2} + \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \tilde\phi_2(x), \\ \mathcal{L}^{(a)}_{\theta,\beta}(x,0) &= 1 - \log\sigma^2 + \frac{1}{2}\log\frac{C_3}{2} + \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \tilde\phi_1(x) \end{aligned} \tag{168}$$

with:

$$\begin{aligned} \tilde\phi_1 &= \frac{1}{2} - \frac{1}{2}\log\left(\frac{\sigma^2(C_3+2)}{C_3}\right) + \frac{C_2^2}{4\sigma^2(2+C_3)^2} \\ \tilde\phi_2(x) &= -\frac{x}{16\sigma^2}\left[4C_2 + C_3 x(4+C_3)\right] + \frac{1}{2}\log\left(\frac{2C_3}{2C_2 + C_3 x(2+C_3)}\right) - \frac{\mathcal{I}(x)}{2}, \end{aligned} \tag{169}$$

implying that for $\theta_c < \theta < -2\sigma'$ we have:

$$\mathcal{L}^{(a)}_{\theta,\beta}(x,u_{\rm typ}(x)) - \left(1 - \log\sigma^2 + \frac{1}{2}\log\frac{C_3}{2}\right)$$
$$\begin{cases} \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \tilde\phi_1 & \text{if } x^+_\sigma(\mu,\beta) < x < -2\sigma \\ \frac{x^2}{4\sigma^2} - \frac{\mathcal{I}(x)}{2} - \left(\tilde\phi_2(x) + \frac{\mathcal{I}(x)}{2}\right) & \text{if } x^-_\sigma(\mu,\beta) < x < x^+_\sigma(\mu,\beta) \\ \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \tilde\phi_1 & \text{if } x < x^-_\sigma(\mu,\beta), \end{cases} \tag{170}$$

while for $\theta < \theta_c$:

$$\mathcal{L}^{(a)}_{\theta,\beta}(x, u^{(a)}_{\rm typ}(x)) - \left(1 - \log\sigma^2 + \frac{1}{2}\log\frac{C_3}{2}\right) = \begin{cases} \frac{x^2}{4\sigma^2} - \frac{\mathcal{I}(x)}{2} - \left(\tilde{\phi}_2 + \frac{\mathcal{I}(x)}{2}\right) & \text{if } x^-_\sigma(\mu,\beta) < x < -2\sigma \\ \frac{x^2}{4\sigma^2} - \mathcal{I}(x) - \tilde{\phi}_1 & \text{if } x < x^-_\sigma(\mu,\beta). \end{cases} \tag{171}$$

The expression (170) agrees -up to a constant- with the one for the large deviation function of the second-smallest eigenvalue that appears in the calculation (see equation (154)), provided one keeps in mind the substitution $C_4(1-u) \to C_2$ and $C_3(1-u) \to C_2$, see section 3.2.1. Similarly, (171) is consistent with equation (156).

To conclude this section, we determine the constant $l(\theta,\beta) = \mathcal{L}^{(a)}_{\theta,\beta}(x_{\rm typ}, u_{\rm typ})$. When $\theta_c < \theta$, the typical value of the smallest eigenvalue is $x_{\rm typ} = -2\sigma$ and $u_{\rm typ} = 0$ leading to:

$$\mathcal{L}^{(a)}_{\theta,\beta}(-2\sigma, 0) = 1 - \frac{1}{2}\log\left(\frac{2\sigma^4}{C_3+2}\right) - \frac{C_2^2}{4\sigma^2(C_3+2)^2} = 1 - \log\left(\frac{\sigma^2}{1+\beta}\right) - \frac{\theta^2}{2\sigma^2[1+\beta]^2}. \tag{172}$$

When $\theta < \theta_c$ instead we have

$$x_{\rm typ} = \mu_0(\theta,\beta) = G^{-1}_\sigma\left(G_{\sigma'}(\theta)\right), \quad G_{\sigma'}(\theta) = \frac{\sqrt{C_2^2 - C_3(C_3+2)^3\sigma^2} - C_2}{C_3(C_3+2)\sigma^2} \tag{173}$$

and $u_{\rm typ} = u^{(a)}_{\rm typ}(x_{\rm typ})$. Using that $G_{\sigma'}(\theta)[2C_2 + C_3 x^+_\sigma(2+C_3)] = 2(C_3+2)$, we obtain that also in this regime:

$$\mathcal{L}^{(a)}_{\theta,\beta}(x_{\rm typ}, u_{\rm typ}) = 1 - \frac{1}{2}\log\left(\frac{2\sigma^4}{C_3+2}\right) - \frac{C_2^2}{4\sigma^2(C_3+2)^2} = 1 - \log\left(\frac{\sigma^2}{1+\beta}\right) - \frac{\theta^2}{2\sigma^2[1+\beta]^2}. \tag{174}$$

Thus, we recover (70). The final expression for the function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ in (83) is obtained as $\overline{\mathcal{L}}_{\theta,\beta}(x) = \mathcal{L}^{(a)}_{\theta,\beta}(x, u_{\rm typ}(x)) - l(\theta,\beta)$, substituting the expressions above.

4.7. Optimization over the Gaussian fluctuations of θ

The above calculations are performed for fixed $\theta < 0$. In this section, we allow for fluctuations of θ and determine the rate function in (64):

$$\mathcal{F}_{\overline{\theta},\sigma_\theta,\beta}(x) = \min_\theta\left[\frac{(\theta-\overline{\theta})^2}{2\sigma_\theta^2} + \overline{\mathcal{L}}_{\theta,\beta}(x)\right], \tag{175}$$

focusing on Regime B. Viewed as a function of θ and at fixed x, the rate function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ in (83) takes different forms depending on whether θ is such that $x^\pm_\sigma(\theta,\beta)$ are smaller or larger than x. More precisely, we find that

$$\begin{aligned} x \leqslant x^-_\sigma(\theta,\beta) &\longrightarrow \theta \geqslant \theta^*_+(x) = \frac{x + 2\beta(\beta+2)x + \sqrt{x^2 - 4\sigma^2}}{2(1+\beta)^2} \\ x \geqslant x^+_\sigma(\theta,\beta) &\longrightarrow \theta \leqslant \theta^*_-(x) = \frac{x + 2\beta(\beta+2)x - \sqrt{x^2 - 4\sigma^2}}{2(1+\beta)^2}. \end{aligned} \tag{176}$$

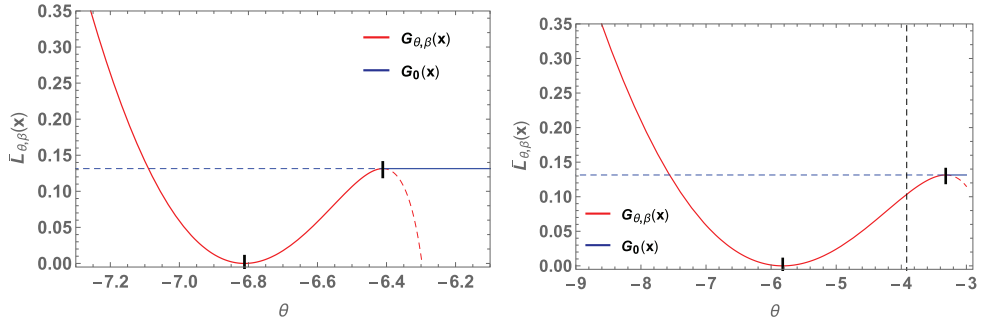

Figure 9. *Left.* Large deviation function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ as a function of θ for $\beta = 2, \sigma = 3$ and $x = -7 < x^*_\sigma(\beta)$. The ticks correspond to the local minimum and maximum attained at θ^*_- and θ^*_+, respectively. In this case the local maximum $\theta^*_+ < \theta_c$. *Right.* Large deviation function $\overline{\mathcal{L}}_{\theta,\beta}(x)$ for $\beta = 0.2, \sigma = 3$ and $x = -7 > x^*_\sigma(\beta)$. The dashed vertical lines marks θ_c, which is smaller than θ^*_+ in this case.

Notice that $\theta^*_\pm(x)$ are also the stationary points satisfying

$$\frac{\partial}{\partial\theta}\left[\mathcal{G}_{\theta,\beta}(x)\right] = 0. \tag{177}$$

In particular, $\theta^*_-(x)$ is a local minimum of $\mathcal{G}_{\theta,\beta}$: when the additive perturbation equals to $\theta^*_-(x)$, then x is precisely the typical value of the smallest eigenvalue, i.e. $x = G_\sigma^{-1}\left(G_{\sigma'}(\theta^*_-(x))\right)$, see (17). The point $\theta^*_+(x)$ is a local maximum of $\mathcal{G}_{\theta,\beta}$. For $x < -2\sigma$ it holds $\theta^*_-(x) < \theta^*_+(x)$ and $\theta^*_-(x) < \theta_c$. The position of the local maximum $\theta^*_+(x)$ with respect to θ_c depends instead on x: $\theta^*_+(x) < \theta_c$ for $x < x^*_\sigma(\beta)$ while $\theta^*_+(x) > -\theta_c$ for $x^*_\sigma(\beta) < x < -2\sigma$, with $x^*_\sigma(\beta) = -2\sigma - \sigma[\beta(1+\beta)^2(2+\beta)]^{-1}$. Therefore, viewed as a function of θ the rate $\overline{\mathcal{L}}_{\theta,\beta}(x)$ in Regime B reads, see figure 9:

$$\overline{\mathcal{L}}_{\theta,\beta}(x) = \begin{cases} \mathcal{G}_{\theta,\beta}(x) & \text{if } \theta \leqslant \theta^*_+(x) \\ \mathcal{G}_0(x) & \text{if } \theta > \theta^*_+(x). \end{cases} \tag{178}$$

The Gaussian weight in (175) shifts the local minimum from $\theta^*_-(x)$ to

$$\theta^*_0(x|\sigma,\beta,\overline{\theta},\sigma^2_\theta) \equiv \frac{2\overline{\theta}\sigma^2 + 2\beta(\beta+2)\sigma^2(\overline{\theta}+x) + [2\beta(2+\beta)+1](\beta+1)^4 x\sigma^2_\theta - \sqrt{T}}{4(1+\beta)^2\sigma^2 + 2(1+\beta)^6\sigma^2_\theta} \tag{179}$$

with

$$\begin{aligned} T =& 4\sigma^4(1+\beta)^4[\overline{\theta}^2 - 2\sigma^2_\theta] + 4\sigma^2(1+\beta)^2\overline{\theta}x[(1+\beta)^4\sigma^2_\theta - 2\beta(2+\beta)\sigma^2] \\ &+[(1+\beta)^4 x\sigma^2_\theta - 2\beta(\beta+2)\sigma^2 x]^2 - 4(1+\beta)^8\sigma^2\sigma^4_\theta. \end{aligned} \tag{180}$$

Henceforth we denote $\theta^*_0(x|\sigma,\beta,\overline{\theta},\sigma^2_\theta)$ simply with $\theta^*_0(x)$. This point lies in the correct domain provided that:

$$\theta^*_0(x) \leqslant \theta^*_+(x). \tag{181}$$

We find that, irrespectively of the value of the variance σ_θ, the two curves in (181) meet at *at most* two values of x, see figure 10, that are given precisely by:

$$x = x^\pm_\sigma(\overline{\theta},\beta), \tag{182}$$

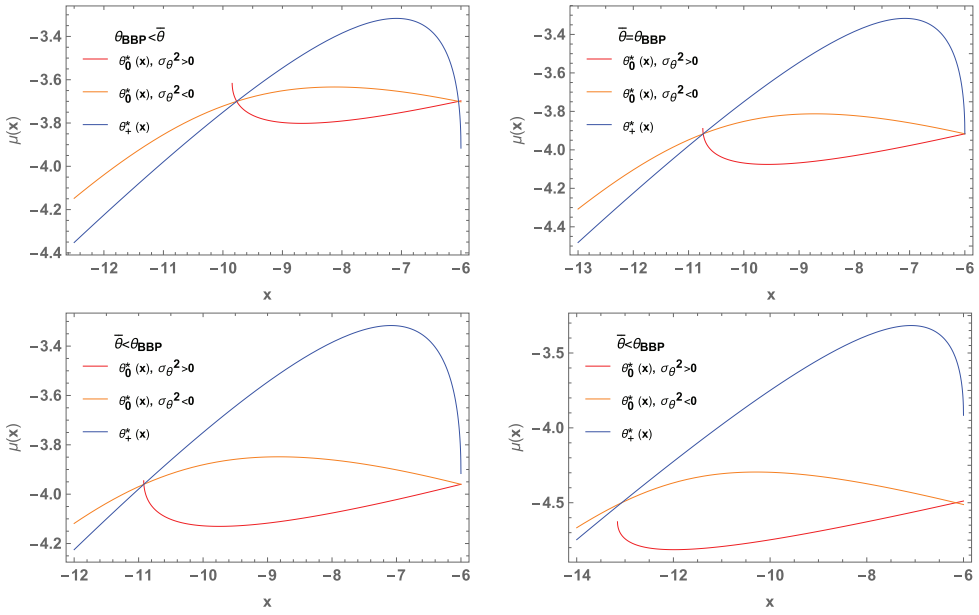

Figure 10. Comparison between the function $\theta_+^*(x)$ and $\theta_0^*(x)$ for either positive and negative values of $\sigma_\theta^2 = \pm 0.8$ and $\beta = .2, \sigma = 3$ (giving $\theta_c = -3.92$) and $\overline{\theta} = -3.7$ (*Top Left*), $\overline{\theta} = \theta_c$ (*Top Right*), $\overline{\theta} = -3.96$ (*Bottom Left*) and $\overline{\theta} = -4.5$ (*Bottom Right*).

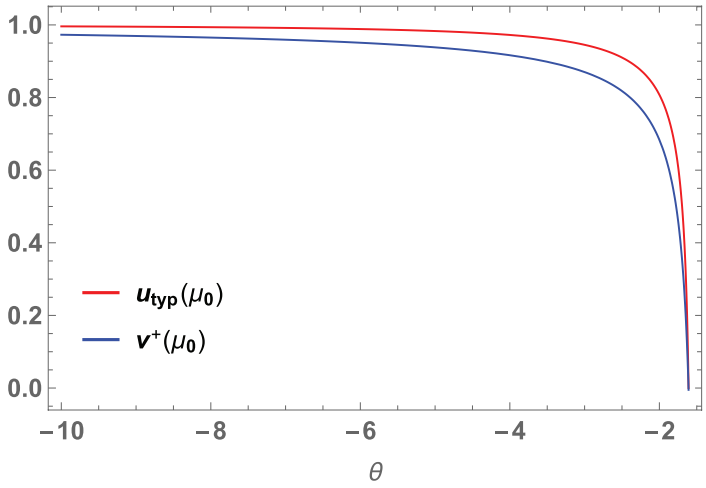

Figure I1. Values of $u^+_{\theta,\beta}(\mu_0)$ and $v^+_{\theta,\beta}(\mu_0)$ for $\sigma = 1$, $\beta = 0.6$ and $\theta < \theta_c = -1.61$. The plot shows that the inequality (I.2) is always satisfied, implying (I.1).

where $x_\sigma^\pm$ are as in (81) and we are assuming that $\theta_0^*(x)$ is real. When the curve meet, they equal to:

$$\theta_0^*(x_\sigma^\pm(\overline{\theta}, \beta)) = \overline{\theta}. \tag{183}$$

More precisely, as it appears from figure 10, we find that:

- When $\theta_c \leqslant \overline{\theta}$, the two functions in (181) cross at both $x_\sigma^\pm$, and the solution $\theta_0^*(x)$ is to be retained for $x \in [x_\sigma^-, x_\sigma^+]$; at the boundary of the interval one has $\theta_0^* = \overline{\theta}$, that is the solution to be kept for all x outside the interval;
- When $\overline{\theta} < \theta_c$ the solutions cross only at x_σ^- for $\sigma_\theta^2 < 0$ (orange curves), and the solution $\theta_0^*(x)$ is to be retained for $x > x_\sigma^-$; for $\sigma_\theta^2 < 0$ a transition occurs: if σ_θ becomes large enough the solution $\theta_0^*(x)$ becomes complex before crossing at x_σ^- (as it follows from section 2.1.2, this regime of positive variance is not of direct interest for applications to the p-spin landscape).

Evaluating the rate functions at the correct value of θ, we recover (91) and (90). Notice that the fact that the conditions are unaltered provided one performs the substitution $\theta \to \overline{\theta}$ is consistent with the observation that x_σ^+ is related to the typical value of the smallest eigenvalue, that should not be shifted by fluctuations of order $1/\sqrt{M}$ of the MM element of the matrix. For the purely additive case, this is proved in [60], see the Remark 2.16 in that paper.

5. Summary and conclusions

Characterizing the geometry of high-dimensional landscapes in terms of the distribution of their stationary points is a fundamental step to understand quantitatively the dynamical exploration of the landscape. This is particularly true when the landscape is rugged with plenty of minima separated by energy barriers, and the dynamics is expected to be dominated by activated processes. In this work we have considered a prototypical landscape, that of the spherical p-spin model, and we have determined the statistics of the index-1 saddles surrounding an arbitrary local minimum, as a function of its energy. In particular, we have identified the range of energy densities and overlaps in which an exponentially large population of saddles is found, and computed their complexity. This completes the analysis initiated in [44], where only the saddles at shorter distance from the reference minimum were obtained. We have found that the dominant saddles at larger distance are marginally stable, with a single eigenvalue of the Hessian exactly equal to zero. Moreover, we have characterized a transition occurring in the population of dominant saddles, separating a regime in which they are geometrically connected to the local minimum and a regime in which they are not, meaning that the corresponding downhill direction in the landscape points in a random direction in configuration space that is not correlated to the direction connecting the saddle to the local minimum.

A relevant question to address once the saddles are identified concerns the properties (typical energy and overlap) of the minima that are connected to the reference one through a given index-1 saddle. For the saddles that are closer to the reference minimum, these properties are determined in [69], where it is shown that the closest saddles connect the reference minimum to other minima that are quite close to it in configuration space. Therefore, escaping through these saddles the system is likely unable to decorrelate from the first trapping minimum. It is an interesting open question whether the same holds true also for the saddles at larger distance from the minimum, whose statistics is determined in this work. An alternative possibility (which is not ruled out by known results, see the discussion in section 2.3.3) is that the marginal saddles found in this work allow the system to decorrelate, i.e. to reach regions of configuration space that are orthogonal to the reference minimum. This would open interesting scenarios for the activated dynamics in this model, allowing the system to decorrelate from the trapping minimum while staying at energies that lie *below* the threshold value. How to validate or rule out this scenario through numerical simulations [70, 71] and how to embed this type of processes within simple phenomenological models [72, 73] are open directions to explore.

On the technical side, the landscape analysis performed in this work required to extend the large deviation principles derived in [63] to the case of a GOE matrix deformed with both an additive and a multiplicative finite-rank perturbation. The resulting large deviation functions display features similar to the ones obtained in case of a purely additive perturbation: in particular, we find that the different regimes of these functions have an interpretation in terms of a *BBP-like* transition of the second-smallest eigenvalue of the perturbed matrix, as it happens in the purely additive case [63]. Some new feature emerge nonetheless as a consequence of the multiplicative part of the perturbation: for instance, when the smallest eigenvalue is fixed to values of x for which the second-smallest eigenvalue is not an outlier but lies within the bulk of the eigenvalue density (see figure 7 *left*), the large deviations are affected by the finite rank perturbation only in an intermediate regime $x \in [x_\sigma^-, x_\sigma^+]$, while they coincide with the unperturbed GOE large deviation for both small-enough and large-enough x. Correspondingly, the correlation of the smallest eigenvector with the direction of the perturbation (measured by $u_{\mathrm{typ}}(x)$) displays a non-monotonic behavior in x. The scale x_σ^- appears only in presence of a multiplicative perturbation, and diverges to $x_\sigma^- \to -\infty$ in the limit of a purely additive perturbation.

Another interesting question is how to recover these results within the replica approach; the formalism developed in [74], which allows one to target stationary points with exactly one zero mode in the Hessian, may be suited in this respect. Obtaining a rigorous proof of these results, and more generally of the fact that the annealed *constrained* complexities of saddles are exact for the p-spin model, are also interesting open problems.

Acknowledgments

I thank G Biroli for the many insightful discussions on this problem and on related topics, and for the useful feedback on the manuscript. This work is supported by the Simons Foundation collaboration Cracking the Glass Problem (No. 454935 to G Biroli).

Appendix A. The statistics of conditioned Hessian

In this appendix, we recall the explicit expressions of the functions $\Delta, \tilde{\Delta}$ and μ defining the statistics of the Hessian matrices discussed in section 2.1.2. We recall that $\sigma^2 = p(p-1)$. The variances of the elements m_{iM} $(i \neq M)$ of the matrix $\mathcal{M}$ are given by:

$$\Delta^2(q) = p(p-1)\left[1 - \frac{(p-1)(1-q^2)q^{2p-4}}{1-q^{2p-2}}\right] \leqslant \sigma^2. \tag{A.1}$$

The element m_{MM} has a different variance given by:

$$\tilde{\Delta}^2(q) = p(p-1)\,\frac{b_1(q)}{b_2(q)}, \tag{A.2}$$

with

$$\begin{aligned} b_1(q) &= p(p-1)q^{4p} - (p-1)(p-2)^2 q^{2p+2} + (p-1)^2(p-2)q^{2p+8} + 2q^8 - p\left(3p^2 - 13p + 14\right)q^{2p+6} \\ &\quad + (p-2)(p-3)q^{4p+4} + \left(3p^3 - 14p^2 + 17p - 6\right)q^{2p+4} - 2(p-1)(p-2)q^{4p+2} \\ b_2(q) &= q^4\left[q^4 - (p-1)^2 q^{2p} + q^{4p} + 2p(p-2)q^{2p+2} - (p-1)^2 q^{2p+4}\right] \end{aligned} \tag{A.3}$$

and we find that in general $\tilde{\Delta}^2(q) < \Delta^2(q)$. For $p = 3$, in particular, one finds $\tilde{\Delta}^2(q) = 0$. Finally, the element m_{MM} has a non-zero average given by:

$$\mu(q,\epsilon,\epsilon_0) \equiv \frac{\sqrt{2}(p-1)\,p\left(1-q^2\right)\left(a_0(q)\epsilon_0 - a_1(q)\epsilon\right)}{q^{6-p} + q^{3p+2} - q^{p+2}\left((p-1)^2(q^4+1) - 2(p-2)\,pq^2\right)} \tag{A.4}$$

with

$$\begin{aligned} a_1 &= q^{3p} + q^{p+2}\left(p-2-(p-1)q^2\right) \\ a_0 &= q^4 + q^{2p}\left(1-p+(p-2)q^2\right). \end{aligned} \tag{A.5}$$

Appendix B. Computing the expectation value of the Hessian determinant

In section 2.2.1 we use the fact that the expectation value of the Hessian determinant in the Kac–Rice formula, conditioned to the values of the smallest eigenvalue $\lambda_{\min} = \lambda$ and of $u_{\min} = u$, to leading order in N is independent of this conditioning. To show this, we first notice that the diagonal shift in (7) is independent of the conditioning, which only affects the matrix $\mathcal{M}$. We let μ_α be the eigenvalues of $\mathcal{M}$, ordered as $\mu_M \leqslant \mu_{M-1} \leqslant \cdots \leqslant \mu_1$. Setting $x = \lambda + \sqrt{2}\,p\,\epsilon$, we condition $\mathcal{M}$ to the event $\mu_M = x$ and to the overlap u, and denote with $P_{\epsilon,q,\epsilon_0}\left(\{\mu_\alpha\}_{\alpha=1}^{M-1}\,|x,u\right)$ the joint distribution of the remaining eigenvalues. We can therefore write:

$$\begin{aligned} &\left\langle |\det\mathcal{H}[\boldsymbol{\sigma}]| \,\middle|\, \begin{cases} \mathbf{g}[\boldsymbol{\sigma}^0] = 0, \mathbf{g}[\boldsymbol{\sigma}] = 0 \\ h[\boldsymbol{\sigma}^0] = \sqrt{2N}\epsilon_0, h[\boldsymbol{\sigma}] = \sqrt{2N}\epsilon \\ \lambda_{\min} = \lambda,\ u_{\min} = u \end{cases} \right\rangle = |\lambda| \int \prod_{\alpha=1}^{M-1} \mathrm{d}\lambda_\alpha\,|\lambda_\alpha| \\ &\times P_{\epsilon,q,\epsilon_0}\left(\left\{\lambda_\alpha + \sqrt{2}\,p\,\epsilon\right\}\middle|\,\lambda + \sqrt{2}\,p\,\epsilon, u\right). \end{aligned} \tag{B.1}$$

As we derive in more generality in section 4.1, the joint distribution $P_{\epsilon,q,\epsilon_0}\left(\{\mu_\alpha\}_{\alpha=1}^{M-1}\,|x,u\right)$ has the same structure as the one of the eigenvalues of the *unconditioned* matrix $\mathcal{M}$, i.e. it equals to the joint distribution of eigenvalues of a matrix perturbed with both an additive and a multiplicative rank-1 perturbation along the same direction in configuration space. The values of the additive and multiplicative perturbations depend explicitly on the parameters x and u, see section 3.2.1. These perturbations do not modify the typical eigenvalue density of the matrix $\mathcal{M}$ to leading order in N, which remains a GOE semicircle of the form $\rho_\sigma(\mu) = \sqrt{4\sigma^2 - \mu^2}/2\pi\sigma^2$: their only effect is to generate (for certain values of parameters) isolated eigenvalues, that correspond to sub-leading corrections of order $1/N$ to the eigenvalue density. Nevertheless, these perturbation do not matter when computing (B.1) to leading exponential order in N, as only the bulk of the density of states does. In particular, using the fact that the determinant is a 1-point function of the eigenvalues, and computing (B.1) with a saddle point in the space of eigenvalue densities we get:

$$\left\langle |\det\mathcal{H}[\boldsymbol{\sigma}]| \,\middle|\, \begin{cases} \mathbf{g}[\boldsymbol{\sigma}^0] = 0, \mathbf{g}[\boldsymbol{\sigma}] = 0 \\ h[\boldsymbol{\sigma}^0] = \sqrt{2N}\epsilon_0, h[\boldsymbol{\sigma}] = \sqrt{2N}\epsilon \\ \lambda_{\min} = \lambda,\ u_{\min} = u \end{cases} \right\rangle = \mathrm{e}^{\left[\frac{M}{2}\log M + \int \mathrm{d}\lambda\,\rho_\sigma(\lambda+\sqrt{2}\,p\,\epsilon)\log|\lambda| + o(N)\right]}, \tag{B.2}$$

which is exactly the same contribution that we would obtain from the unconstrained Hessian. Notice that this contribution does not depend neither on the geometrical conditioning on q, nor on the conditioning to the value of the smallest eigenvalue.

Appendix C. Generalized Kac–Rice formula for the quenched complexity

The general expression of the higher moments appearing in (41) is given by:

$$\left\langle \mathcal{N}^n_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0) \right\rangle_0 = \int \prod_{a=1}^{n} d\boldsymbol{\sigma}^{(a)}\, \delta\left(\boldsymbol{\sigma}^{(a)} \cdot \boldsymbol{\sigma}^0 - q\right) \times p_{\vec{\boldsymbol{\sigma}}|\boldsymbol{\sigma}^0}(\mathbf{0}, \epsilon) \mathbb{G}^{(n)}_{\vec{\boldsymbol{\sigma}}|\boldsymbol{\sigma}^0}\left(\vec{\lambda}, \vec{u}\right) \times \left\langle \prod_{a=1}^{n} \left|\det \mathcal{H}[\boldsymbol{\sigma}^{(a)}]\right| \Big| \begin{cases} \mathbf{g}[\boldsymbol{\sigma}^0] = 0, \mathbf{g}[\boldsymbol{\sigma}^{(a)}] = 0 \\ h[\boldsymbol{\sigma}^0] = \sqrt{2N}\epsilon_0, h[\boldsymbol{\sigma}^{(a)}] = \sqrt{2N}\epsilon \\ \lambda^{(a)}_{\min} = \lambda,\ u^{(a)}_{\min} = u \end{cases} \right\rangle \tag{C.1}$$

where $p_{\vec{\boldsymbol{\sigma}}|\boldsymbol{\sigma}^0}$ now denotes the *joint* distribution of all gradients $\mathbf{g}[\boldsymbol{\sigma}^{(a)}]$ and all energy fields $h[\boldsymbol{\sigma}^{(a)}]$, each Hessian $\mathcal{H}[\boldsymbol{\sigma}^{(a)}]$ in the expectation value is conditioned to gradients, energy fields and smallest Hessian eigenvalues at all the other points $\boldsymbol{\sigma}^{(b)}$, and $\mathbb{G}^{(n)}_{\vec{\boldsymbol{\sigma}}|\boldsymbol{\sigma}^0}\left(\vec{\lambda}, \vec{u}\right)$ is the *joint* probability distribution of the smallest eigenvalues and of the correspondent eigenvector components of these conditioned Hessians. Following the reasoning elucidated in [25, 44] one can show that, as a consequence of the isotropy of the correlations of the random energy field, all these statistical distributions depend on the points $\boldsymbol{\sigma}^{(a)}$ only through their mutual overlaps $q_{ab} \equiv N(\boldsymbol{\sigma}^{(a)} \cdot \boldsymbol{\sigma}^{(b)})$. Introducing an $n \times n$ symmetric overlap matrix $\hat{Q}$ with components $Q_{ab} = \delta_{ab} + (1 - \delta_{ab})q_{ab}$ we can parametrize the above integral as:

$$\left\langle \mathcal{N}^n_{\boldsymbol{\sigma}^0}(\epsilon, q, \lambda, u|\epsilon_0) \right\rangle_0 = \int \prod_{a<b=1}^{n} dq_{ab} \exp\left[NS_n(\epsilon, q, \hat{Q}|\epsilon_0) + o(Nn)\right]\, \mathbb{G}^{(n)}_{\epsilon,q,\hat{Q}|\epsilon_0}\left(\vec{\lambda}, \vec{u}\right). \tag{C.2}$$

This integral can now be computed with a saddle-point approximation, optimizing over the matrix $\hat{Q}$. The total constrained complexity is contributed by stationary points for which $\lambda_{\min}$ and $u_{\min}$ take their typical values, implying that the joint distribution $\mathbb{G}^{(n)}_{\epsilon,q,\hat{Q}|\epsilon_0}\left(\vec{\lambda}, \vec{u}\right)$ does not scale exponentially with N but it is of $O(1)$. In that case the saddle point of the remaining action is attained at $q_{ab} \equiv q_1 = q^2$ [44]. In presence of the conditioning, to compute (C.2) one has to determine the large deviations of the smallest eigenvalues and eigenvectors of all the n Hessian matrices. This will in general depend on the parameters q_{ab}: to prove that the annealed calculation is correct, one has to show that this dependence is such that the saddle point value $q_{ab} \equiv q_1 = q^2$ is not shifted by additional contributions coming from this large deviation function, that are exponentially large in N. Notice that for all values of $q_{ab} \neq 0$ the Hessian matrices are coupled with each others: therefore, determining the joint distribution $\mathbb{G}^{(n)}_{\epsilon,q,\hat{Q}|\epsilon_0}\left(\vec{\lambda}, \vec{u}\right)$ to linear exponential order in N, and its generic dependence on the parameters q_{ab}, is a highly non-trivial task.

Appendix D. Large deviations at fixed θ, u: limiting cases

From the above expressions, we can easily recover the limiting cases of the large deviations for an unperturbed GOE [64] and for a purely additive perturbation [63]. In the case in which all the perturbations vanish, $\mu_1(x, u)$ diverges and $F(x, u) \to 0$. The function $\mathcal{L}^{(a)}_{\theta,\beta}(x, u)$ tends to:

$$\mathcal{L}^{(a)}_{\theta,\beta}(x,u) - l(\theta,\beta) \overset{\theta,\beta\to 0}{\longrightarrow} -\frac{x}{4\sigma^2}\sqrt{x^2-4\sigma^2} - \log\left(-\frac{x}{2}+\frac{1}{2}\sqrt{x^2-4\sigma^2}\right) + \log\sigma, \tag{D.1}$$

which for $u = 0$ coincides exactly with the large deviations for the smallest eigenvalue of an orthogonal matrix with variance σ^2, given by:

$$\mathcal{G}_0(x) = \int_x^{-2\sigma} \frac{\sqrt{z^2-4\sigma^2}}{2\sigma^2}\mathrm{d}z = \left(\frac{x^4}{4\sigma^2} - \mathcal{I}(x) + \frac{1}{2}\right) + \log\sigma - 1, \tag{D.2}$$

where $\mathcal{I}(z)$ is defined in (69). This function vanishes at $x = -2\sigma$, which is indeed the typical value of the smallest eigenvalue.

In the case of a purely additive perturbation $\beta = 0$, the relevant case is Case B. For a negative perturbation $\theta < 0$, it holds $\sigma^2 F(x,u) \to \theta(1-u)$, and the typical value of the second-smallest eigenvalue, when smaller than -2σ, becomes:

$$\mu_1(x,u) \overset{\beta\to 0}{\longrightarrow} \theta(1-u) + \frac{\sigma^2}{\theta(1-u)} \equiv \mu_1(u), \tag{D.3}$$

consistently with the fact that in this case the effective perturbation induced by fixing x is an additive perturbation with strength $\tilde{\theta} = \theta(1-u)$, see equation (76). Therefore, for $\theta(1-u) \geqslant -\sigma$

$$\mathcal{L}^{(a)}_{\theta,\beta}(x,u) \overset{\beta\to 0}{\longrightarrow} \frac{x^2}{4\sigma^2} - \frac{\theta x u}{2\sigma^2} - \mathcal{I}(x) - \frac{1}{2}\log(1-u) - \left[-\frac{1}{2} + \log\sigma + \frac{\theta^2(1-u)^2}{4\sigma^2}\right], \tag{D.4}$$

which coincides[8] with what is found in [63]. For $\theta(1-u) < -\sigma$ we have instead:

$$\mathcal{L}_{\theta,\beta}(x,u) + l(\theta,\beta) \overset{\beta\to 0}{\longrightarrow}$$
$$\begin{cases} a(x,u) - \left[-\frac{x^2}{4\sigma^2} + \frac{\mathcal{I}(x)}{2} + \frac{\theta(1-u)x}{2\sigma^2} - \frac{1}{2}\log\left(\frac{\theta(1-u)}{\sigma^2}\right) - \frac{1}{2}\right] & \text{if } x \geqslant \mu_1(u) \\ a(x,u) - \left[-\frac{y^2}{4\sigma^2} + \frac{\mathcal{I}(y)}{2} + \frac{\theta(1-u)y}{2\sigma^2} - \frac{1}{2}\log\left(\frac{\theta(1-u)}{\sigma^2}\right) - \frac{1}{2}\right]\Big|_{y=\xi^+_\sigma(u)} & \text{if } x < \mu_1(u) \end{cases} \tag{D.5}$$

with

$$a(x,u) = \frac{x^2}{4\sigma^2} - \theta\frac{xu}{2\sigma^2} - \mathcal{I}(x) - \frac{1}{2}\log(1-u), \tag{D.6}$$

which again coincides with the result in [63].

Appendix E. Introduction of the auxiliary fields y, λ

In this appendix we show how the representation (114) is obtained. First, using the Hubbard–Stratonovich transformation we set:

$$\mathrm{e}^{-\frac{M}{2}\frac{C_3^2(1-u)^2}{8\sigma^2}\left(\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2\right)^2} = \sqrt{\frac{4\sigma^2}{\pi C_3^2(1-u)^2}}\int_{-\infty}^{\infty}\mathrm{d}y\,\mathrm{e}^{-\frac{4\sigma^2 y^2}{C_3^2(1-u)^2} + i\sqrt{M}y\left(\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2\right)} \tag{E.1}$$

[8] See the combination of equation (1) in [63] and the beginning of section 7; in particular $C' = -1/2 + \log\sigma$ and $\sigma = 1$ in that work.

so that the integral (113) can be re-written as:

$$I_{x,u}(\vec{\mu}) = \frac{\Gamma\left(\frac{M-1}{2}\right)}{\pi^{\frac{M-1}{2}}}\sqrt{\frac{4M\sigma^2}{\pi C_3^2(1-u)^2}}\int_{-\infty}^{\infty} \mathrm{d}y\, \mathrm{e}^{-\frac{4M\sigma^2 y^2}{C_3^2(1-u)^2}}$$
$$\times\left[\int\prod_{\alpha=1}^{M-1} de_\alpha \delta\left(\sum_{\alpha=1}^{M-1} e_\alpha^2 - 1\right)\mathrm{e}^{-\frac{M}{2}\left[\frac{C_4(x,u)(1-u)}{2\sigma^2}-2iy\right]\sum_{\alpha=1}^{M-1}\mu_\alpha e_\alpha^2 + \frac{C_3(1-u)}{2\sigma^2}\sum_{\alpha=1}^{M-1}\mu_\alpha^2 e_\alpha^2}\right]. \tag{E.2}$$

Exponentiating the constraint, we can re-write the quantity in square brackets as:

$$re\cdot = -iM\left(\frac{2\sigma^2}{C_3(1-u)}\right)^{\frac{M-1}{2}}\int_{-i\infty}^{i\infty}\mathrm{d}\lambda \mathrm{e}^{-M\lambda}$$
$$\times\int\prod_{\alpha=1}^{M-1} de_\alpha \mathrm{e}^{-\frac{M}{2}\sum_{\alpha=1}^{M-1} e_\alpha^2\left[\mu_\alpha^2 + \left(\frac{C_4(x,u)}{C_3}-2iy\frac{2\sigma^2}{C_3(1-u)}\right)\mu_\alpha - 2\lambda\frac{2\sigma^2}{C_3(1-u)}\right]}. \tag{E.3}$$

The representation (114) is obtained with the change of variable:

$$y' = iy\frac{2\sigma^2}{C_3(1-u)}. \tag{E.4}$$

Appendix F. Derivation of the solutions for the auxiliary fields y, λ

In this appendix we report the derivation of the solutions (133) of the saddle point equations for y, λ, as well as of (147). We begin with the derivation of (133), starting from equation (132). We introduce the notation:

$$a = \frac{C_3(1-u)}{2\sigma^2}, \qquad b = \frac{2}{\sigma^2}. \tag{F.1}$$

If $\mu^{\pm}$ are complex, they can be written as:

$$\mu^{\pm} = -\frac{1}{2}\left(\frac{C_4}{C_3}-2y\right) \pm \frac{i}{2}\sqrt{-8\lambda\frac{2\sigma^2}{C_3(1-u)} - \left(\frac{C_4}{C_3}-2y\right)^2} \equiv X \pm iY, \tag{F.2}$$

where now the quantity under the square root is positive. The two equation (132) are one the adjoint of the other, and read explicitly:

$$X + iY = \frac{aX + by + iaY}{(aX+by)^2 + a^2Y^2} + \sigma^2\left(aX + by - iaY\right), \tag{F.3}$$

and equating real and imaginary parts (assuming $Y \neq 0$) gives:

$$\frac{1}{(aX+by)^2 + a^2Y^2} = \frac{1+a\sigma^2}{a} \longrightarrow (aX+by)\left(a^{-1}+2\sigma^2\right) = X, \tag{F.4}$$

that gives the solution for y^*. The first equation (F.4) allows then to solve for λ as:

$$8\frac{\lambda^* 2\sigma^2}{C_3(1-u)} = -\frac{16\left[\sigma^2(C_3(1-u)+2)^2 + C_4^2(1-u)^2\right]}{C_3(1-u)(C_3(1-u)+2)^3}. \tag{F.5}$$

If on the other hand $\mu^{\pm}$ are real, they can be written as

$$\mu^{\pm} = -\frac{1}{2}\left(\frac{C_4}{C_3} - 2y\right) \pm \frac{1}{2}\sqrt{8\lambda\frac{2\sigma^2}{C_3(1-u)} + \left(\frac{C_4}{C_3} - 2y\right)^2} \equiv X \pm \sqrt{Y}, \tag{F.6}$$

where the quantity under the square root is again positive. Equation (132) read in this case:

$$\left(X \pm \sqrt{Y}\right)\left(aX + by \mp a\sqrt{Y}\right) = 1 + \sigma^2\left(aX + by \mp a\sqrt{Y}\right)^2, \tag{F.7}$$

which are equivalent to:

$$aX^2 + byX - aY - 1 - \sigma^2(aX + by)^2 - \sigma^2 a^2 Y = \mp\sqrt{Y}\left(by + 2a\sigma^2(aX + by)\right). \tag{F.8}$$

Summing and subtracting these two equations, we get two linear equations for λ^*, y^*:

$$\begin{aligned} by + 2a\sigma^2(aX + by) &= 0 \\ aX^2 + byX - aY - 1 - \sigma^2(aX + by)^2 - \sigma^2 a^2 Y &= 0, \end{aligned} \tag{F.9}$$

which are again solved by (133). Notice that $\lambda^* < 0$, which implies that μ^+ is the largest of the two real solutions, and it is negative. The action $\phi(\lambda^*, y^*)$ is obtained noticing that the saddle point equations imply:

$$\mathcal{I}(\mu^+) + \mathcal{I}(\mu^-) = \log\left[\sigma^2\left(1 + \frac{2}{C_3(1-u)}\right)\right] - 1 + a\mu^+\mu^- + \frac{by^*}{2}\left(\mu^+ + \mu^-\right). \tag{F.10}$$

We now come to the derivation of (147). First, taking y as a free parameter we find that the expression for $\lambda_{\text{ext}}(y;\xi)$ in (145) implies that:

$$\begin{aligned} y < \frac{C_4(x,u)}{2C_3} + \xi &\longrightarrow \xi = \mu^+_{x,u}(y, \lambda_{\text{ext}}(y;\xi)) \\ y > \frac{C_4(x,u)}{2C_3} + \xi &\longrightarrow \xi = \mu^-_{x,u}(y, \lambda_{\text{ext}}(y;\xi)) \end{aligned} \tag{F.11}$$

and at the threshold value:

$$y = \frac{C_4(x,u)}{2C_3} + \xi \longrightarrow \xi = \mu^-_{x,u}(y, \lambda_{\text{ext}}(y;\xi)) = \mu^+_{x,u}(y, \lambda_{\text{ext}}(y;\xi)). \tag{F.12}$$

Using that in both cases:

$$\begin{aligned} \mu^-_{\text{ext}} &= \mu^+ - \sqrt{8\lambda\frac{2\sigma^2}{C_3(1-u)} + \left(\frac{C_4}{C_3} - 2y\right)^2} = -\xi - \frac{C_4}{C_3} + 2y = -\frac{4\sigma^2\lambda}{C_3(1-u)\xi}, \\ \mu^+_{\text{ext}} &= \mu^- + \sqrt{8\lambda\frac{2\sigma^2}{C_3(1-u)} + \left(\frac{C_4}{C_3} - 2y\right)^2} = -\xi - \frac{C_4}{C_3} + 2y = -\frac{4\sigma^2\lambda}{C_3(1-u)\xi}, \end{aligned} \tag{F.13}$$

for any value of y we get that the action evaluated at $\lambda_{\text{ext}}(y;\xi)$ reduces to:

$$\tilde{\phi}(y) = \frac{y^2}{\sigma^2} - \frac{C_3(1-u)}{4\sigma^2}\left[\xi^2 + \left(\frac{C_4}{C_3} - 2y\right)\xi\right] - \frac{1}{2M}\sum_{\alpha=1}^{M-1}\log(\mu_\alpha - \xi) - \frac{1}{2M}\sum_{\alpha=1}^{M-1}\log(\mu_\alpha - \mu^{\pm}_{\text{ext}}(y)). \tag{F.14}$$

Equivalently, if $y_{\rm ext}(\lambda;\xi)$ is used one finds

$$\tilde{\phi}(\lambda) = \frac{1}{4\sigma^2}\left(\frac{C_4}{C_3} - \frac{4\sigma^2\lambda}{C_3(1-u)\xi} + \xi\right)^2 - \lambda - \frac{1}{2M}\sum_{\alpha=1}^{M-1}\log(\mu_\alpha - \xi) - \frac{1}{2M}\sum_{\alpha=1}^{M-1}\log(\mu_\alpha - \mu^{\pm}_{\rm ext}(\lambda)). \tag{F.15}$$

These functions can be further optimized in y or λ, by solving the equations:

$$\begin{aligned}
&\frac{2y}{\sigma^2} + \frac{C_3(1-u)}{2\sigma^2}\xi - G\left(-\xi - \frac{C_4}{C_3} + 2y\right) = 0\\
&\frac{2C_4}{C_3} + 2\xi + C_3(1-u)\xi - \lambda\frac{8\sigma^2}{C_3(1-u)\xi} - 2\sigma^2 G\left(-\frac{4\sigma^2\lambda}{C_3(1-u)\xi}\right) = 0.
\end{aligned} \tag{F.16}$$

Note that in the first equation the argument of the resolvent is positive, in the second equation it is negative because $\lambda < 0$. Both these equations are linear for a GOE (the coefficients of the quadratic terms simplify, with solutions given in (147). The threshold condition (F.12) becomes equivalent to:

$$\xi(C_3(1-u)+4) + \frac{4C_3\sigma^2}{C_3\xi(C_3(1-u)+2)+2C_4} + \frac{2C_4}{C_3} = 0. \tag{F.17}$$

To determine (148) we use that $\mu^{\pm}_{\rm ext} = \xi$ (with $\pm$ chosen depending on the value of $y_{\rm ext}$), as well as (F.13) and the saddle point condition for $y_{\rm ext}$, we get

$$\mathcal{I}(\mu^{\mp}_{\rm ext}) = \log\left[\left(\frac{C_3(1-u)}{2}+1\right)\xi + \frac{C_4(x,u)}{C_3}\right] - \frac{1}{2} - \frac{\xi + \frac{C_4}{C_3} - 2y_{\rm ext}}{2}\left[\frac{2y_{\rm ext}}{\sigma^2} + \frac{C_3(1-u)}{2\sigma^2}\xi\right], \tag{F.18}$$

and calling

$$H(\xi) = \frac{2C_3}{C_3\xi(C_3(1-u)+2)+2C_4} \tag{F.19}$$

this is:

$$\mathcal{I}(\mu^{\mp}_{\rm ext}) = \log\left(\frac{1}{H(\xi)}\right) - \frac{1}{2} - \frac{H(\xi)}{2}G^{-1}\left[-H(\xi)\right] = \log\left(\frac{1}{H(\xi)}\right) + \frac{\sigma^2}{2}H^2(\xi). \tag{F.20}$$

The expression (148) is obtained using that:

$$\frac{y_{\rm ext}^2}{\sigma^2} - \lambda_{\rm ext} = -\frac{(1-u)\xi}{16\sigma^2}\left[4C_4 + C_3\xi(4 + C_3(1-u))\right] + \frac{\sigma^2}{4}H^2. \tag{F.21}$$

To conclude the appendix, we remark that equation (160) follows from the general identity:

$$\mathcal{I}(x) = \log\left(-\frac{1}{G_\sigma(x)}\right) - \frac{1}{2} + \frac{x}{2}G_\sigma(x), \tag{F.22}$$

using that:

$$\xi^+_\sigma = G^{-1}_\sigma\left(\frac{1}{\sigma^2 F(x,u)}\right), \quad 2C_4(x,u) + \xi^+_\sigma C_3(2 + C_3(1-u)) = -\frac{2\sigma^2 F(x,u)[2 + C_3(1-u)]}{(1-u)} \tag{F.23}$$

as well as:

$$\frac{(\xi_\sigma^+)^2}{4\sigma^2}\left(1+\frac{(1-u)C_3}{4}[4+C_3(1-u)]\right)+\frac{\xi_\sigma^+}{4\sigma^2}\left(C_4(1-u)-\frac{1}{F(x,u)}\right)$$
$$=-\frac{1}{4}-\frac{(1-u)^2(2C_2+C_3^2\,u\,x)^2}{16\sigma^2[2+C_3(1-u)]^2}. \tag{F.24}$$

Appendix G. Large deviations for the second-smallest eigenvalue: the case of purely additive perturbation

In this appendix we compare the large deviation function for the second-smallest eigenvalue computed in section 4.5 with the results given in [62] for the large deviations in the case of a purely additive perturbation. Indeed, for $\beta = 0$ and x, u fixed, the eigenvalue μ_{M-1} is the smallest eigenvalue of a matrix subject to an additive rank-1 perturbation of strength $\tilde{\theta} = \theta(1-u)$, see equation (76). In this limit, given that $\sigma^2 F(x,u) \to \theta(1-u)$ and that $\xi_\sigma^- \to -\infty$, the two cases discussed in section 4.5 reduce to the following:

- If $\theta(1-u) < -\sigma$, typically the second-smallest eigenvalue *is* out of the bulk and

$$\Psi_1(x,u,\xi) \to \frac{1}{4\sigma^2}\xi^2 - \frac{1}{2}\mathcal{I}(\xi) - \frac{\theta(1-u)\xi}{2\sigma^2} + \frac{1}{2}\log(-2\theta) - \frac{1}{2}\log C_3, \tag{G.1}$$

  and the logarithmic divergence due to C_3 gets canceled by another term in Ψ_0. This function (up to constants that do not depend on ξ) matches with $L_\theta^\beta(\xi)$ in Th. 1.1 of [62]. It has a minimum in $\xi^* = \theta(1-u) + \sigma^2/(\theta(1-u))$, that is indeed the typical value of the smallest eigenvalue of a GOE matrix subject to the additive perturbation of strength $\tilde{\theta} = \theta(1-u)$.
- If $\theta(1-u) > -\sigma$, typically the second-smallest eigenvalue *is not* out of the bulk. In this case the large deviation function has only two regimes:

$$\Psi_1(x,u,\xi) \to \begin{cases} \frac{1}{4\sigma^2}\xi^2 - \mathcal{I}(\xi) - \phi_1(x,u) & \text{if } \xi \geqslant \xi^* \\ \frac{1}{4\sigma^2}\xi^2 - \frac{1}{2}\mathcal{I}(\xi) - \theta\frac{(1-u)\xi}{2\sigma^2} + \frac{1}{2}\log(-2\theta) - \frac{1}{2}\log C_3 & \text{if } \xi < \xi^*, \end{cases} \tag{G.2}$$

  which matches with the function $M_\theta^\beta(x)$ of [62] (up to constants that do not depend on ξ).

The difference in the constants comes from the fact that the large deviation function $\Psi_1(x,u,\xi)$ in section 4.5 is not normalized to zero at the typical value.

Appendix H. Self-consistent checks on $u_{\text{typ}}(x)$

In this appendix we check under which conditions the rate functions to be optimized is $\mathcal{L}^{(a)}(\theta,\beta)$. If Case A holds (see (67)), this is always the case. In Case B, in order to perform the check we need to determine the sign of the function:

$$\tilde{F}(x,v) = \sigma^2 F(x,v) + \sigma, \tag{H.1}$$

evaluated at $v = u^{(a)}_{\rm typ}(x)$. This function is quadratic in v, with two roots given by:

$$v^{\pm}_{\theta,\beta}(x) = \frac{-2C_2 + C_3^2(4\sigma + x) + 6C_3\sigma \pm \sqrt{\left(2C_2 + C_3^2 x\right)^2 - 4C_3\sigma(6C_2 + C_3(3C_3+4)x) + 4C_3^2\sigma^2}}{2C_3^2(2\sigma + x)}. \tag{H.2}$$

Again, it is convenient to consider the above regimes of θ, β:

- Regime A: In this case $u^{(a)}_{\rm typ} = 0$. Plugging $u = 0$ into (67), it can be checked that the condition to be in Case A becomes:

$$\frac{4\sigma^2\beta(\beta+2)}{\theta^2(1+\beta)^2} > 1, \tag{H.3}$$

  which is always satisfied for $-2\sigma' < \theta < 0$. Therefore, in this regime Case A holds and $\mathcal{L}^{(a)}_{\theta,\beta}(x,u)$ is the right large deviation function to be optimized.
- Regime B1: When $\theta_{\rm c} < \theta < -2\sigma'$, we find that $v^-_{\theta,\beta}(x) \geqslant v^+_{\theta,\beta}(x)$; moreover, when real, $\tilde{F}(x,v) \geqslant 0$ for $v^+_{\theta,\beta}(x) \leqslant v \leqslant v^-_{\theta,\beta}(x)$. The function $v^-_{\theta,\beta}(x)$ is a monotonically increasing function of x which satisfies $v^-_{\theta,\beta}(x) \overset{x\to-\infty}{\to} 1$ and $v^-_{\theta,\beta}(x) \overset{x\to-2\sigma}{\to} \infty$. Similarly, $v^+_{\theta,\beta}(x)$ is monotonic and satisfies $v^+_{\theta,\beta}(x) \overset{x\to-\infty}{\to} 0$, while $v^+_{\theta,\beta}(-2\sigma) = u^+_{\theta,\beta}(-2\sigma) < 0$, implying that $v^+_{\theta,\beta}(x) < 0$ for any x. Therefore, we always have $v^+_{\theta,\beta}(x) < u^{(a)}_{\rm typ}(x) < v^-_{\theta,\beta}(x)$, which implies that $\tilde{F}(x, u^{(a)}_{\rm typ}(x)) \geqslant 0$. Therefore, also in this regime the solution is self-consistent, meaning that the correct large-deviation function to optimize is $\mathcal{L}^{(a)}_{\theta,\beta}(x,u)$.
- Regime B2: When $\theta < \theta_{\rm c}$, we find $v^+_{\theta,\beta}(-2\sigma) = u^+_{\theta,\beta}(-2\sigma) > 0$ and $v^+_{\theta,\beta}(x) > 0$ for any x. The functions $v^+_{\theta,\beta}(x)$ and $u^+_{\theta,\beta}(x)$ cross at a point $x^-_\sigma(\mu,\beta) < x^{**} < -2\sigma$, where $\tilde{F}(x, u^+_{\theta,\beta}(x))$ becomes negative. It can be checked that $\mu_1(x, u_{typ}(x)) - x \geqslant 0$ for $x < x^{**}$, meaning that also in this regime the function to be optimized is again $\mathcal{L}^{(a)}_{\theta,\beta}$.

Appendix I. Self-consistent check: at most one isolated eigenvalue is generated

In the derivation of the large deviation function, we made the assumption that for any value of the parameters θ, β and for any choice of x and u, the $M-1$ eigenvalues $\mu_{M-1}, \cdots, \mu_1$ typically arrange themselves in such a way that *at most one* of them, namely μ_{M-1}, is found to be smaller than -2σ and isolated from the continuous part of the density of states. In order to validate this hypothesis self-consistently, we have to check that when μ_{M-1} takes its typical value, the third-smallest eigenvalue satisfies $\mu^{\rm typ}_{M-2} = -2\sigma$. As we pointed out several times already, once the values of μ_M and u_M are fixed to x, u the distribution of the remaining $M-1$ eigenvalues is the one of a GOE matrix perturbed with both an additive and multiplicative perturbation along a given direction ,with parameters $\tilde{\theta}$ and $\tilde{\beta}$ (that depend explicitly on x and u, see equation (76)). Similarly, when the values of μ_{M-1} and u_{M-1} are kept fixed, the distribution of the remaining $M-2$ eigenvalues is again the one of a perturbed GOE matrix with modified parameters. Our goal is to argue that when fixing $\mu_{M-1} = \mu^{\rm typ}_{M-1}$ and $u^{\rm typ}_{M-1}$, then $\mu^{\rm typ}_{M-2} = -2\sigma$. This is totally equivalent to stating that, when μ_M and u_M are fixed to their typical value, then $\mu^{\rm typ}_{M-1} = -2\sigma$. This is trivially true when $\mu^{\rm typ}_M = -2\sigma$, i.e. when $\theta \geqslant \theta_{\rm c}$. In the regime $\theta < \theta_{\rm c}$, then $\mu^{\rm typ}_M = \mu_0(\theta,\beta)$ and $u_{\rm typ} = u^+_{\theta,\beta}(\mu_0)$. In order for the second eigenvalue to stick to the boundary of the semicircle, it must hold:

$$\sigma^2 F(\mu_0, u^+_{\theta,\beta}(\mu_0)) + \sigma \geqslant 0, \tag{I.1}$$

see section 3.2.1. From the discussion in appendix H it follows that this is guaranteed if, for arbitrary values of σ, β and for $\theta < \theta_c$, we find:

$$v^+_{\theta,\beta}(\mu_0) \leqslant u^+_{\theta,\beta}(\mu_0). \tag{I.2}$$

This inequality can be checked graphically: in figure I1 we give an example for a fixed value of β, σ. Very similar results are obtained for different values of β, σ.

ORCID iDs

Valentina Ros https://orcid.org/0000-0001-8189-8809

References

[1] Crisanti A and Sommers H-J 1995 Thouless–Anderson–Palmer approach to the spherical p-spin spin glass model *J. Physique* I **5** 805
[2] Cavagna A, Giardina I and Parisi G 1998 Stationary points of the Thouless–Anderson–Palmer free energy *Phys. Rev.* B **57** 11251
[3] Monasson R 1995 Structural glass transition and the entropy of the metastable states *Phys. Rev. Lett.* **75** 2847
[4] Bray A J and Dean D S 2007 Statistics of critical points of Gaussian fields on large-dimensional spaces *Phys. Rev. Lett.* **98** 150201
[5] Crisanti A, Leuzzi L, Parisi G and Rizzo T 2004 Spin-glass complexity *Phys. Rev. Lett.* **92** 127203
[6] Rizzo T 2005 TAP complexity, the cavity method and supersymmetry *J. Phys. A: Math. Gen.* **38** 3287
[7] Aspelmeier T, Bray A J and Moore M A 2005 Complexity of ising spin glasses *Phys. Rev. Lett.* **92** 087203
[8] Fyodorov Y V 2004 Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices *Phys. Rev. Lett.* **92** 240601
[9] Fyodorov Y V and Williams I 2007 Replica symmetry breaking condition exposed by random matrix calculation of landscape complexity *J. Stat. Phys.* **129** 1081
[10] Fyodorov Y V and Nadal C 2012 Critical behavior of the number of minima of a random landscape at the glass transition point and the Tracy–Widom distribution *Phys. Rev. Lett.* **109** 167203
[11] Auffinger A, Ben Arous G and Cerný J 2013 Random matrices and complexity of spin glasses *Commun. Pure Appl. Math.* **66** 165
[12] Subag E 2017 The complexity of spherical p-spin models-a second moment approach *Ann. Probab.* **45** 3385
[13] Majumdar S N and Martin O C 2006 Statistics of the number of minima in a random energy landscape *Phys. Rev.* E **74** 061112
[14] Fyodorov Y V, Le Doussal P, Rosso A and Texier C 2018 Exponential number of equilibria and depinning threshold for a directed polymer in a random potential *Ann. Phys., NY* **397** 1
[15] Fyodorov Y V and Le Doussal P 2019 Manifolds in high dimensional random landscape: complexity of stationary points and depinning (arXiv:1908.09217)
[16] May R M 1972 Will a large complex system be stable? *Nature* **238** 413
[17] Fyodorov Y V and Khoruzhenko B A 2016 Nonlinear analogue of the May–Wigner instability transition *Proc. Natl Acad. Sci.* **113** 6827
[18] Onuchic J N, Luthey-Schulten Z and Wolynes P G 1997 Theory of protein folding: the energy landscape perspective *Ann. Rev. Phys. Chem.* **48** 545
[19] Stadler P F 2002 Fitness landscapes *Biological Evolution and Statistical Physics* (Berlin: Springer) p 183
[20] Kadmon J and Sompolinsky H 2015 Transition to chaos in random neuronal networks *Phys. Rev.* X **5** 041030

[21] Wainrib G and Touboul J 2013 Topological and dynamical complexity of random neural networks *Phys. Rev. Lett.* **110** 118101
[22] Rong G and Ma T 2017 On the optimization landscape of tensor decompositions *Advances in Neural Information Processing Systems*
[23] Ben Arous G, Mei S, Montanari A and Nica M 2017 The landscape of the spiked tensor model (arXiv:1711.05424)
[24] Fyodorov Y V 2019 A spin glass model for reconstructing nonlinearly encrypted signals corrupted by noise *J. Stat. Phys.* **175** 789–818
[25] Ros V, Ben Arous G, Biroli G and Cammarota C 2019 Complex energy landscapes in spiked-tensor and simple glassy models: ruggedness, arrangement of local minima and phase transitions *Phys. Rev.* X **9** 011003
[26] Mannelli S S, Krzakala F, Urbani P and Zdeborova L 2019 Passed and spurious: descent algorithms and local minima in spiked matrix-tensor models *Int. Conf. on Machine Learning* pp 4333–42
[27] Galla T and Farmer J D 2013 Complex dynamics in learning complicated games *Proc. Natl Acad. Sci.* **110** 1232–6
[28] Douglas M R, Shiffman B and Zelditch S 2004 Critical points and supersymmetric vacua I *Commun. Math. Phys.* **252** 325
[29] Aazami A and Easther R 2006 Cosmology from random multifield potentials *J. Cosmol. Astropart. Phys.* **2006** 013
[30] Gross D J and Mézard M 1984 The simplest spin glass *Nucl. Phys.* B **240** 431
[31] Derrida B 1980 Random-energy model: limit of a family of disordered models *Phys. Rev. Lett.* **45** 79
[32] Crisanti A and Sommers H J 1992 The spherical p-spin interaction spin glass model: the statics *Z. Phys.* B **87** 341
[33] Barrat A, Franz S and Parisi G 1997 Temperature evolution and bifurcations of metastable states in mean-field spin glasses, with connections with structural glasses *J. Phys. A: Math. Gen.* **30** 5593
[34] Crisanti A and Leuzzi L 2006 Spherical 2 + p spin-glass model: an analytically solvable model with a glass-to-glass transition *Phys. Rev.* B **73** 014412
[35] Subag E and Zeitouni O 2017 The extremal process of critical points of the pure p-spin spherical spin glass model *Probab. Theory Relat. Fields* **168** 773
[36] Subag E 2017 The geometry of the Gibbs measure of pure spherical spin glasses *Inventiones Math.* **210** 135
[37] Auffinger A and Chen W-K 2018 On the energy landscape of spherical spin glasses *Adv. Math.* **330** 553
[38] Subag E 2018 Free energy landscapes in spherical spin glasses (arXiv:1804.10576)
[39] Cugliandolo L F and Kurchan J 1993 Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model *Phys. Rev. Lett.* **71** 173
[40] Bouchaud J-P, Cugliandolo L F, Kurchan J and Mezard M 1998 Out of equilibrium dynamics in spin-glasses and other glassy systems *Spin Glasses and Random Fields* (Singapore: World Scientific) p 161
[41] Folena G, Franz S and Ricci-Tersenghi F 2019 Memories from the ergodic phase: the awkward dynamics of spherical mixed p-spin models (arXiv:1903.01421)
[42] Berthier L and Biroli G 2011 Theoretical perspective on the glass transition and amorphous materials *Rev. Mod. Phys.* **83** 587
[43] Franz S 2005 First steps of a nucleation theory in disordered systems *J. Stat. Mech.* P04001
[44] Ros V, Biroli G and Cammarota C 2019 Complexity of energy barriers in mean-field glassy systems *Europhys. Lett.* **126** 20003
[45] Cavagna A, Giardina I and Parisi G 1997 An investigation of the hidden structure of states in a mean-field spin-glass model *J. Phys. A: Math. Gen.* **30** 7021
[46] Cavagna A, Garrahan J P and Giardina I 1999 Quenched complexity of the mean-field p-spin spherical model with external magnetic field *J. Phys. A: Math. Gen.* **32** 711
[47] Cavagna A, Giardina I and Parisi G 1997 Structure of metastable states in spin glasses by means of a three replica potential *J. Phys. A: Math. Gen.* **30** 4449
[48] Barbier D and Cugliandolo L 2019 A constrained TAP approach for disordered spin models: application to the mixed spherical case (arXiv:1911.12052)
[49] Fyodorov Y V 2015 High-dimensional random fields and random matrix theory *Markov Process. Relat. Fields* **21** 483

[50] Fyodorov Y V and Le Doussal P 2018 Hessian spectrum at the global minimum of high-dimensional random landscapes *J. Phys. A: Math. Theor.* **51** 474002
[51] Baik J, Ben Arous G and Péché S 2005 Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices *Ann. Probab.* **33** 1643–97
[52] Edwards S F and Jones R C 1976 The eigenvalue spectrum of a large symmetric random matrix *J. Phys. A: Math. Gen.* **9** 1595
[53] Kosterlitz J M, Thouless D J and Jones R C 1976 Spherical model of a spin-glass *Phys. Rev. Lett.* **36** 1217
[54] Péché S 2006 The largest eigenvalue of small rank perturbations of Hermitian random matrices *Probab. Theory Relat. Fields* **134** 127
[55] Féral D and Péché S 2007 The largest eigenvalue of rank one deformation of large Wigner matrices *Commun. Math. Phys.* **272** 185
[56] Bassler K E, Forrester P J and Frankel N E 2009 Eigenvalue separation in some random matrix models *J. Math. Phys.* **50** 033302
[57] Capitaine M, Donati-Martin C and Féral D 2009 The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations *Ann. Probab.* **37** 1–47
[58] Pizzo A, Renfrew D and Soshnikov A 2013 On finite rank deformations of Wigner matrices *Ann. Probab. Stat.* **49** 64–94
[59] Benaych-Georges F, Guionnet A and Maida M 2011 Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices *Electron. J. Probab.* **16** 1621–62
[60] Benaych-Georges F and Nadakuditi R R 2009 The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices *Adv. Math.* **227** 494
[61] Noiry N 2019 Measures of spiked random matrices (arXiv:1903.11731)
[62] Maida M 2007 Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles *Electron. J. Probab.* **12** 1131
[63] Biroli G and Guionnet A 2019 Large deviations for the largest eigenvalues and eigenvectors of spiked random matrices (arXiv:1904.01820)
[64] Ben Arous G, Dembo A and Guionnet A 2001 Aging of spherical spin glasses *Probab. Theory Relat. Fields* **120** 1
[65] Bogomolny E 2017 Modification of the Porter–Thomas distribution by rank-one interaction *Phys. Rev. Lett.* **118** 022501
[66] Aleiner I L and Matveev K A 1998 Shifts of random energy levels by a local perturbation *Phys. Rev. Lett.* **80** 814
[67] Franz S and Parisi G 1995 Recipes for metastable states in spin glasses *J. Physique* I **5** 1401
[68] Franz S and Parisi G 1998 Effective potential in glassy systems: theory and simulations *Physica* A **261** 317
[69] Ros V, Biroli G and Cammarota C 2020 in preparation
[70] Stariolo D A and Cugliandolo L F 2019 Activated dynamics of the Ising p-spin disordered model with finite number of variables *Europhys. Lett.* **127** 16002
[71] Baity-Jesi M, Achard-de Lustrac A and Biroli G 2018 Activated dynamics: an intermediate model between the random energy model and the p-spin model *Phys. Rev.* **E 98** 012133
[72] Bouchaud J P 1992 Weak ergodicity breaking and aging in disordered systems *J. Physique* I **2** 1705
[73] Dyre J C 1987 Master-equation appoach to the glass transition *Phys. Rev. Lett.* **58** 792
[74] Müller M, Leuzzi L and Crisanti A 2006 Marginal states in mean-field glasses *Phys. Rev.* B **74** 134431