Conditioning the complexity of random landscapes on marginal optima

Jaron Kent-Dobias*
Istituto Nazionale di Fisica Nucleare, Sezione di Roma I, Rome 00184, Italy

(Received 24 July 2024; accepted 2 December 2024; published 26 December 2024)

Marginal optima are minima or maxima of a function with many nearly flat directions. In settings with many competing optima, marginal ones tend to attract algorithms and physical dynamics. Often, the important family of marginal attractors is a vanishing minority compared with nonmarginal optima and other unstable stationary points. We introduce a generic technique for conditioning the statistics of stationary points in random landscapes on their marginality and apply it in three isotropic settings with qualitatively different structures: in the spherical spin-glasses, where the energy is Gaussian and its Hessian is a Gaussian orthogonal ensemble (GOE); in multispherical spin glasses, which are Gaussian but non-GOE; and in sums of squared spherical random functions, which are non-Gaussian. In these problems, we are able to fully characterize the distribution of marginal optima in the landscape, including when they are in the minority.

DOI: 10.1103/PhysRevE.110.064148

I. INTRODUCTION

Systems with rugged landscapes are important across many disciplines, from the physics of glasses and spin glasses to statistical inference problems [1]. The behavior of these systems is best understood when equilibrium or optimal solutions are studied and weighted averages can be taken statically over all possible configurations. However, such systems are also infamous for their tendency to defy equilibrium and optimal expectations in practice due to the presence of dynamic transitions or crossovers that leave physical or algorithmic dynamics stuck exploring only a subset of configurations [2,3].

In mean-field settings, it was long thought that physical and many algorithmic dynamics would get stuck at a specific energy level, called the threshold energy. The threshold energy is the energy level at which level sets of the landscape transition from containing mostly saddle points to containing mostly minima. The level set associated with this threshold energy contains mostly *marginal minima*, or minima whose Hessian matrix have a continuous spectral density over all sufficiently small positive eigenvalues. In most circumstances the spectrum is *pseudogapped*, which means that the spectral density smoothly approaches zero as zero eigenvalue is approached from above.

However, recent work found that the threshold energy is not important even for simple gradient descent dynamics [4–6]. Depending on the initial condition of the system and the nature of the dynamics, the energy reached can be above or below the threshold energy, while in some models the threshold energy is completely inaccessible to any dynamics [7]. Though it is still not known how to predict the energy level that many simple algorithms will reach, the results all share one commonality: the minima found are still marginal despite being in the minority compared to stiff minima or saddle points. This ubiquity of behavior suggests that the distribution of marginal minima can be used to bound out-of-equilibrium dynamical behavior.

Despite their importance in a wide variety of in- and out-of-equilibrium settings [8–17], it is not straightforward to condition on the marginality of minima using the traditional methods for analyzing the distribution of minima in rugged landscapes. Using the method of a Legendre transformation of the Parisi parameter corresponding to a set of real replicas, one can force the result to correspond with marginal minima by tuning the value of that parameter [18]. However, this results only in a characterization of the threshold energy and cannot characterize marginal minima at other energies where they are a minority.

The alternative approach, used to great success in the spherical spin glasses, is to start by understanding in detail the Hessian matrix at stationary points. Then, one can condition the analysis on whatever properties of the Hessian are necessary to lead to marginal minima. This strategy is successful in spherical spin glasses because it is straightforward to implement. First, the shape of the Hessian's spectrum is independent of the energy and the gradient. This is a property of models whose energy is a Gaussian random variable [19,20]. Furthermore, a natural parameter in the analysis of these models linearly shifts the spectrum of the Hessian. Therefore, tuning this parameter to a specific constant value allows one to require that the Hessian spectrum has a pseudogap and therefore that the associated minima be marginal. Unfortunately, this strategy is less straightforward to generalize to other models. Many models of interest, especially in inference problems, have Hessian statistics that are poorly understood. This is especially true for the statistics of the Hessian conditioned to lie at stationary points, which is necessary to understand in models whose energy is non-Gaussian.

Here, we introduce a generic method for conditioning the statistics of stationary points on their marginality. The technique makes use of a novel way to condition an integration

*Contact author: jaron.kent-dobias@roma1.infn.it

measure to select only configurations that result in a certain value of the smallest eigenvalue of a matrix. By requiring that the smallest eigenvalue of the Hessian at stationary points be zero and further looking for a sign that the zero eigenvalue lies at the edge of a continuous spectrum, we enforce the condition that the spectrum has a pseudogap and is therefore marginal. We demonstrate the method on the spherical spin glasses, where it is unnecessary but instructive, and on extensions of the spherical models where the technique is more useful. In related work, we compare the marginal complexity with the performance of gradient descent and approximate message-passing algorithms [21].

An outline of this paper follows. In Sec. II we introduce the technique for conditioning on the smallest eigenvalue and how to extend it to further condition on the presence of a pseudogap. We provide a simple but illustrative example using a Gaussian orthogonal ensemble (GOE) matrix with a shifted diagonal. In Sec. III we apply this technique to the problem of characterizing marginal minima in random landscapes. Section IV gives several examples of the marginal complexity applied to specific models of increasing difficulty. Finally, Sec. V summarizes this work and suggests necessary extensions.

II. CONDITIONING ON THE SMALLEST EIGENVALUE

In this section, we introduce a general method for conditioning a measure on the smallest eigenvalue of some matrix that depends on it. In Sec. II B we show how this works in perhaps the simplest example of GOE random matrices with a shifted diagonal. In the final subsection we describe how to extend this method to condition on the presence of a pseudogap at the bottom on the spectrum.

A. The general method

Consider an $N \times N$ real symmetric matrix A. An arbitrary function g of the minimum eigenvalue of A can be expressed using integrals over $\mathbf{s} \in \mathbb{R}^N$ as

$$g(\lambda_{\min}(A)) = \lim_{\beta\to\infty} \int \frac{d\mathbf{s}\,\delta(N - \|\mathbf{s}\|^2)e^{-\beta \mathbf{s}^T A\mathbf{s}}}{\int d\mathbf{s}'\,\delta(N - \|\mathbf{s}'\|^2)e^{-\beta \mathbf{s}'^T A\mathbf{s}'}} g\left(\frac{\mathbf{s}^T A\mathbf{s}}{N}\right). \tag{1}$$

In the limit of large β, each integral concentrates among vectors $\mathbf{s}$ in the eigenspace of A corresponding to the smallest eigenvalue of A. This produces

$$\begin{aligned}\lim_{\beta\to\infty} \int \frac{d\mathbf{s}\,\delta(N - \|\mathbf{s}\|^2)e^{-\beta \mathbf{s}^T A\mathbf{s}}}{\int d\mathbf{s}'\,\delta(N - \|\mathbf{s}'\|^2)e^{-\beta \mathbf{s}'^T A\mathbf{s}'}} g\left(\frac{\mathbf{s}^T A\mathbf{s}}{N}\right) &= \int \frac{d\mathbf{s}\,\delta(N - \|\mathbf{s}\|^2)\mathbb{1}_{\ker(A-\lambda_{\min}(A)I)}(\mathbf{s})}{\int d\mathbf{s}'\,\delta(N - \|\mathbf{s}'\|^2)\mathbb{1}_{\ker(A-\lambda_{\min}(A)I)}(\mathbf{s}')} g\left(\frac{\mathbf{s}^T A\mathbf{s}}{N}\right) \\ &= g(\lambda_{\min}(A)) \frac{\int d\mathbf{s}\,\delta(N - \|\mathbf{s}\|^2)\mathbb{1}_{\ker(A-\lambda_{\min}(A)I)}(\mathbf{s})}{\int d\mathbf{s}'\,\delta(N - \|\mathbf{s}'\|^2)\mathbb{1}_{\ker(A-\lambda_{\min}(A)I)}(\mathbf{s}')} \\ &= g(\lambda_{\min}(A)),\end{aligned} \tag{2}$$

as desired. The first relation extends a technique for calculating the typical minimum eigenvalue of an ensemble of matrices first introduced by Ikeda and later used by Kent-Dobias in the context of random landscapes and is similar to an earlier technique for conditioning the value of the ground state energy in random landscapes by Fyodorov and Le Doussal [21–25]. A Boltzmann distribution is introduced over a spherical model whose Hamiltonian is quadratic with interaction matrix given by A. In the limit of zero temperature, the measure will concentrate on the ground states of the model, which correspond with the eigenspace of A associated with its minimum eigenvalue $\lambda_{\min}$. The second relation uses the fact that, once restricted to the sphere $\|\mathbf{s}\|^2 = N$ and the minimum eigenspace, $\mathbf{s}^T A\mathbf{s} = \mathbf{s}^T \mathbf{s}\lambda_{\min}(A) = N\lambda_{\min}(A)$.

The relationship is formal, but we can make use of the fact that the integral expression with a Gibbs distribution can be manipulated with replica techniques, averaged over, and in general treated with a physicist's toolkit. In particular, we have specific interest in using $g(\lambda_{\min}(A)) = \delta(\lambda_{\min}(A))$, a Dirac delta function, which can be inserted into averages over ensembles of matrices A (or indeed more complicated averages) in order to create the condition that the minimum eigenvalue is zero.

B. Simple example: Shifted Gaussian orthogonal ensemble

We demonstrate the efficacy of the technique by rederiving a well-known result: the large-deviation function for pulling an eigenvalue from the bulk of the GOE spectrum. Consider an ensemble of $N \times N$ matrices $A = B + \mu I$ for B drawn from the GOE ensemble with entries whose variance is σ^2/N. We know that the bulk spectrum of A is a Wigner semicircle with radius 2σ shifted by a constant μ. Therefore, for $\mu = 2\sigma$, the minimum eigenvalue will typically be zero, while for $\mu > 2\sigma$ the minimum eigenvalue would need to be a large deviation from the typical spectrum and its likelihood will be exponentially suppressed with N. For $\mu < 2\sigma$, the bulk of the typical spectrum contains zero and therefore a larger N^2 deviation, moving an extensive number of eigenvalues, would be necessary [26]. This final case cannot be quantified by this method, but instead the nonexistence of a large deviation linear in N appears as the emergence of an imaginary part in the large deviation function.

To compute this large deviation function, we employ the method outlined in the previous subsection to calculate

$$\begin{aligned}e^{NG_{\lambda^*}(\mu)} &= P(\lambda_{\min}(B + \mu I) = \lambda^*) \\ &= \overline{\delta(N\lambda^* - N\lambda_{\min}(B + \mu I))},\end{aligned} \tag{3}$$

where the overline is the average over B, and we have defined the large-deviation function $G_{\lambda^*}(\mu)$. Using the representation of $\lambda_{\min}$ defined in (1), we have

$$e^{NG_{\lambda^*}(\mu)} = \overline{\lim_{\beta\to\infty}\int \frac{d\mathbf{s}\,\delta(N-\|\mathbf{s}\|^2)e^{-\beta\mathbf{s}^T(B+\mu I)\mathbf{s}}}{\int d\mathbf{s}'\,\delta(N-\|\mathbf{s}'\|^2)e^{-\beta\mathbf{s}'^T(B+\mu I)\mathbf{s}'}}\,\delta(N\lambda^*-\mathbf{s}^T(B+\mu I)\mathbf{s})}, \tag{4}$$

Using replicas to treat the denominator ($x^{-1}=\lim_{m\to 0}x^{m-1}$) and transforming the δ function to its Fourier representation, we have

$$e^{NG_{\lambda^*}(\mu)} = \overline{\lim_{\beta\to\infty}\lim_{m\to 0}\int d\hat{\lambda}\prod_{\alpha=1}^{m}[d\mathbf{s}^\alpha\,\delta(N-\|\mathbf{s}^\alpha\|^2)]\exp\left\{-\beta\sum_{\alpha=1}^{m}(\mathbf{s}^\alpha)^T(B+\mu I)\mathbf{s}^\alpha+\hat{\lambda}[N\lambda^*-(\mathbf{s}^1)^T(B+\mu I)\mathbf{s}^1]\right\}}, \tag{5}$$

having introduced the auxiliary parameter $\hat{\lambda}$ in the Fourier representation of the δ function. The whole expression, so transformed, is an exponential integral linear in the matrix B. Taking the average over B, we find

$$\begin{aligned} e^{NG_{\lambda^*}(\mu)} = &\lim_{\beta\to\infty}\lim_{m\to 0}\int d\hat{\lambda}\prod_{\alpha=1}^{m}[d\mathbf{s}^\alpha\,\delta(N-\|\mathbf{s}^\alpha\|^2)] \\ &\times\exp\left\{N[\hat{\lambda}(\lambda^*-\mu)-m\beta\mu]+\frac{\sigma^2}{N}\left[\beta^2\sum_{\alpha\gamma}^{m}(\mathbf{s}^\alpha\cdot\mathbf{s}^\gamma)^2+2\beta\hat{\lambda}\sum_{\alpha}^{m}(\mathbf{s}^\alpha\cdot\mathbf{s}^1)^2+\hat{\lambda}^2N^2\right]\right\}. \end{aligned} \tag{6}$$

We make the Hubbard–Stratonovich transformation to the matrix field $Q^{\alpha\beta}=\frac{1}{N}\mathbf{s}^\alpha\cdot\mathbf{s}^\beta$. This produces an integral expression of the form

$$e^{NG_{\lambda^*}(\mu)} = \lim_{\beta\to\infty}\lim_{m\to 0}\int d\hat{\lambda}\,dQ\,e^{N\mathcal{U}_{\mathrm{GOE}}(\hat{\lambda},Q|\beta,\lambda^*,\mu)}, \tag{7}$$

where the effective action $\mathcal{U}_{\mathrm{GOE}}$ is given by

$$\mathcal{U}_{\mathrm{GOE}}(\hat{\lambda},Q|\beta,\lambda^*,\mu)=\hat{\lambda}(\lambda^*-\mu)+\lim_{m\to 0}\left\{-m\beta\mu+\sigma^2\left[\beta^2\sum_{\alpha\gamma}^{m}(Q^{\alpha\gamma})^2+2\beta\hat{\lambda}\sum_{\alpha}^{m}(Q^{1\alpha})^2+\hat{\lambda}^2\right]+\frac{1}{2}\log\det Q\right\}, \tag{8}$$

and $Q^{\alpha\alpha}=1$ because of the spherical constraint. We can evaluate this integral using the saddle-point method. We make a replica symmetric ansatz for Q, because this is a 2-spin spherical model but with the first row singled out because of its unique coupling with $\hat{\lambda}$. The resulting matrix has the form

$$Q=\begin{bmatrix} 1 & \tilde{q}_0 & \tilde{q}_0 & \cdots & \tilde{q}_0 \\ \tilde{q}_0 & 1 & q_0 & \cdots & q_0 \\ \tilde{q}_0 & q_0 & 1 & \ddots & q_0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \tilde{q}_0 & q_0 & q_0 & \cdots & 1 \end{bmatrix}. \tag{9}$$

The relevant expressions in the effective action produce

$$\sum_{\alpha\beta}(Q^{\alpha\beta})^2 = m+2(m-1)\tilde{q}_0^2+(m-1)(m-2)q_0^2, \tag{10}$$

$$\sum_{\alpha}(Q^{1\alpha})^2 = 1+(m-1)\tilde{q}_0^2, \tag{11}$$

$$\begin{aligned}\log\det Q =&\,(m-2)\log(1-q_0) \\ &+\log\left[1+(m-2)q_0-(m-1)\tilde{q}_0^2\right].\end{aligned} \tag{12}$$

Inserting these expressions into the effective action and taking the limit of m to zero, we arrive at

$$e^{NG_{\lambda^*}(\mu)} = \lim_{\beta\to\infty}\int d\hat{\lambda}\,dq_0\,d\tilde{q}_0\,e^{N\mathcal{U}_{\mathrm{GOE}}(\hat{\lambda},q_0,\tilde{q}_0|\beta,\lambda^*,\mu)}, \tag{13}$$

with the new effective action

$$\begin{aligned}&\mathcal{U}_{\mathrm{GOE}}(\hat{\lambda},q_0,\tilde{q}_0|\beta,\lambda^*,\mu) \\ &\quad=\hat{\lambda}(\lambda^*-\mu)+\sigma^2\left[2\beta^2\left(q_0^2-\tilde{q}_0^2\right)+2\beta\hat{\lambda}\left(1-\tilde{q}_0^2\right)+\hat{\lambda}^2\right] \\ &\quad\quad-\log(1-q_0)+\tfrac{1}{2}\log\left(1-2q_0+\tilde{q}_0^2\right).\end{aligned} \tag{14}$$

We need to evaluate the integral above using the saddle-point method, but in the limit $\beta\to\infty$. We expect the overlaps to concentrate on one as β goes to infinity. We therefore take

$$q_0 = 1-y\beta^{-1}-z\beta^{-2}+O(\beta^{-3}), \tag{15}$$

$$\tilde{q}_0 = 1-\tilde{y}\beta^{-1}-(z+\Delta z)\beta^{-2}+O(\beta^{-3}). \tag{16}$$

However, taking the limit with $y\neq\tilde{y}$ results in an expression for the action that diverges with β. To cure this, we must take $\tilde{y}=y$. The result is

$$\begin{aligned}\mathcal{U}_{\mathrm{GOE}}(\hat{\lambda},y,\Delta z|\infty,\lambda^*,\mu)=&\,\hat{\lambda}(\lambda^*-\mu)+\sigma^2[\hat{\lambda}^2+4(y+\Delta z)] \\ &+\frac{1}{2}\log\left(1-\frac{2\Delta z}{y^2}\right).\end{aligned} \tag{17}$$

Extremizing this action over the new parameters y, Δz, and $\hat{\lambda}$, we find

$$\hat{\lambda}=\frac{1}{\sigma}\sqrt{\left(\frac{\mu-\lambda^*}{2\sigma}\right)^2-1}, \tag{18}$$

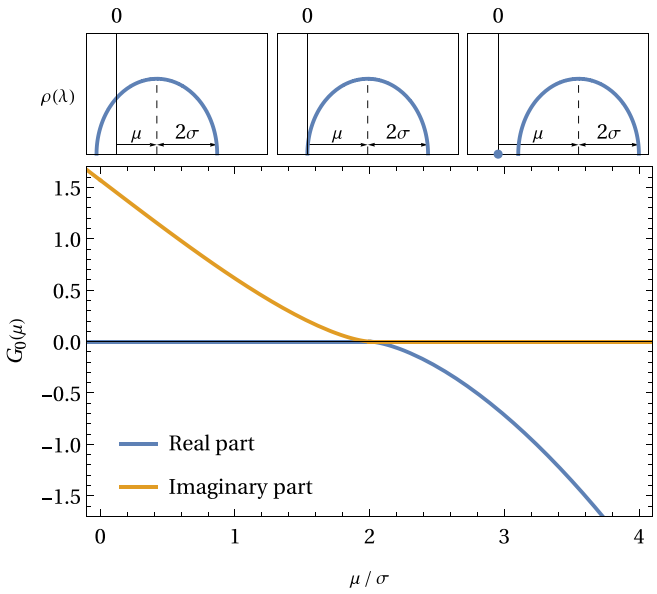

FIG. 1. The large deviation function $G_0(\mu)$ defined in (3) as a function of the shift μ to the GOE diagonal. $G_0(2\sigma) = 0$, while for $\mu > 2\sigma$ it is negative and for $\mu < 2\sigma$ it gains an imaginary part. The top panels show schematically what happens to the spectral density in each of these regimes. For $\mu < 2\sigma$, an N^2-large deviation would be required to fix the smallest eigenvalue to zero and the calculation breaks down, leading to the imaginary part. For $\mu > 2\sigma$ the spectrum can satisfy the constraint on the smallest eigenvalue by isolating a single eigenvalue at zero at the cost of an order-N-large deviation. At the transition point $\mu = 2\sigma$ the spectrum is pseudogapped.

$$y = \frac{1}{2\sigma}\left[\frac{\mu-\lambda^*}{2\sigma} + \sqrt{\left(\frac{\mu-\lambda^*}{2\sigma}\right)^2 - 1}\right]^{-1}, \tag{19}$$

$$\Delta z = \frac{1}{4\sigma^2}\left[\left(\frac{\mu-\lambda^*}{2\sigma}\right)^2 - 1 - \frac{\mu-\lambda^*}{2\sigma}\sqrt{\left(\frac{\mu-\lambda^*}{2\sigma}\right)^2 - 1}\right]. \tag{20}$$

Inserting this solution into the effective action we arrive at

$$\begin{aligned} G_{\lambda^*}(\mu) &= \underset{\hat{\lambda},y,\Delta z}{\text{extremum}}\, \mathcal{U}_{\text{GOE}}(\hat{\lambda}, y, \Delta z|\infty, \lambda^*, \mu) \\ &= -\frac{\mu-\lambda^*}{2\sigma}\sqrt{\left(\frac{\mu-\lambda^*}{2\sigma}\right)^2 - 1} \\ &\quad - \log\left[\frac{\mu-\lambda^*}{2\sigma} - \sqrt{\left(\frac{\mu-\lambda^*}{2\sigma}\right)^2 - 1}\right]. \end{aligned} \tag{21}$$

This function is plotted in Fig. 1 for $\lambda^* = 0$. For $\mu < 2\sigma$, $G_0(\mu)$ has an imaginary part. This indicates that the existence of a zero minimum eigenvalue when $\mu < 2\sigma$ corresponds to a large deviation that grows faster than N, rather like N^2, since in this regime the bulk of the typical spectrum is over zero and therefore extensively many eigenvalues must have large deviations in order for the smallest eigenvalue to be zero [26]. For $\mu \geqslant 2\sigma$ this function gives the large deviation function for the probability of seeing a zero eigenvalue given the shift μ. $\mu = 2\sigma$ is the maximum of the function with a real value and corresponds to the intersection of the typical bulk spectrum with zero, i.e., a pseudogap.

Here, we see what appears to be a general heuristic for identifying the saddle parameters for which the spectrum is pseudogapped: the equivalent of this large-deviation function will lie on the singular boundary between a purely real and complex value.

C. Conditioning on a pseudogap

We have seen that this method effectively conditions a random matrix ensemble on its lowest eigenvalue being zero. However, this does not correspond on its own to marginality. In the previous example, most values of μ where the calculation was valid correspond to matrices with a single isolated eigenvalue. However, the marginal minima we are concerned with have pseudogapped spectra, where the continuous part of the spectral density has a lower bound at zero.

Fortunately, our calculation can be modified to ensure that we consider only pseudogapped spectra. First, we insert a shift μ by hand into the "natural" spectrum of the problem at hand, conditioning the trace to have a specific value $\mu = \frac{1}{N}\operatorname{Tr} A$. Then, we choose this artificial shift so that the resulting conditioned spectra are pseudogapped. As seen the previous subsection, this can be done by starting from a sufficiently large μ and decreasing it until the calculation develops an imaginary part, signaling the breakdown of the large-deviation principle at order N.

In isotropic or zero-signal landscapes, there is another way to condition on a pseudogap. In such landscapes, the typical spectrum does not have an isolated eigenvalue. Therefore, for a given μ the bottom of the spectrum can be located by looking for the value λ^* that maximizes the (real) large deviation function. Inverting this reasoning, we can find the value $\mu = \mu_{\text{m}}$ corresponding to a marginal spectrum by requiring that the large deviation function has a maximum in λ^* at $\lambda^* = 0$, or

$$0 = \frac{\partial}{\partial\lambda^*} G_{\lambda^*}(\mu_{\text{m}})\Big|_{\lambda^*=0}. \tag{22}$$

In the example problem of Sec. II B, this corresponds precisely to $\mu_{\text{m}} = 2\sigma$, the correct marginal shift. Note that when we treat the Dirac δ function using its Fourier representation with auxiliary parameter $\hat{\lambda}$, as in the previous subsection, this condition corresponds with choosing μ such that $\hat{\lambda} = 0$.

III. MARGINAL COMPLEXITY IN RANDOM LANDSCAPES

The methods of the previous section can be used in diverse settings. However, we are interested in applying them to study stationary points in random landscapes whose Hessian spectrum has a pseudogap; that is, that are marginal. In Sec. III A we define the marginal complexity using the tools of the previous section. In Sec. III B we review several general features in a physicist's approach to computing the marginal complexity. In Sec. III C we introduce a representation of the marginal complexity in terms of an integral over a superspace, which

condenses the notation and the resulting calculation and which we will use in one of our examples in the next section.

A. Marginal complexity from Kac–Rice

The situation in the study of random landscapes is often as follows: an ensemble of smooth energy functions $H : \mathbb{R}^N \to \mathbb{R}$ defines a family of random landscapes, often with their configuration space subject to one or more constraints of the form $g(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}^N$. The typical geometry of landscapes drawn from the ensemble is studied by their complexity, or the average logarithm of the number of stationary points with certain properties, e.g., of marginal minima at a given energy.

Such problems can be studied using the method of Lagrange multipliers, with one introduced for every constraint. If the configuration space is defined by r constraints, then the problem of identifying stationary points is reduced to extremizing the Lagrangian

$$L(\mathbf{x}, \boldsymbol{\omega}) = H(\mathbf{x}) + \sum_{i=1}^{r} \omega_i g_i(\mathbf{x}) \tag{23}$$

with respect to $\mathbf{x}$ and the Lagrange multipliers $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_r\}$. To write the gradient and Hessian of the energy, which are necessary to count stationary points, care must be taken to ensure they are constrained to the tangent space of the configuration manifold. For our purposes, the Lagrangian formalism offers a solution: the gradient $\boldsymbol{\nabla} H : \mathbb{R}^N \times \mathbb{R}^r \to \mathbb{R}^N$ and Hessian $\operatorname{Hess} H : \mathbb{R}^N \times \mathbb{R}^r \to \mathbb{R}^{N\times N}$ of the energy H can be written as the simple vector derivatives of the Lagrangian L, with

$$\boldsymbol{\nabla} H(\mathbf{x}, \boldsymbol{\omega}) = \partial L(\mathbf{x}, \boldsymbol{\omega}) = \partial H(\mathbf{x}) + \sum_{i=1}^{r} \omega_i \partial g_i(\mathbf{x}), \tag{24}$$

$$\begin{aligned}\operatorname{Hess} H(\mathbf{x}, \boldsymbol{\omega}) &= \partial\partial L(\mathbf{x}, \boldsymbol{\omega}) \\ &= \partial\partial H(\mathbf{x}) + \sum_{i=1}^{r} \omega_i \partial\partial g_i(\mathbf{x}),\end{aligned} \tag{25}$$

where $\partial = \frac{\partial}{\partial \mathbf{x}}$ will always represent the derivative with respect to the vector argument $\mathbf{x}$. Note that, unlike the energy, which is a function of the configuration $\mathbf{x}$ alone, the gradient and Hessian depend also on the Lagrange multipliers $\boldsymbol{\omega}$. In situations with an extensive number of constraints, it is important to take seriously contributions of the form $\frac{\partial^2 L}{\partial \mathbf{x} \partial \boldsymbol{\omega}}$ to the Hessian [27]. However, the cases we study here have N^0 constraints and these contributions appear as finite-N corrections.

The number of stationary points in a landscape for a particular function H is found by integrating over the Kac–Rice measure

$$\begin{aligned}d\nu_H(\mathbf{x}, \boldsymbol{\omega}) = d\mathbf{x}\, d\boldsymbol{\omega}\, \delta(\mathbf{g}(\mathbf{x}))\, \delta(\boldsymbol{\nabla} H(\mathbf{x}, \boldsymbol{\omega})) \\ \times |\det \operatorname{Hess} H(\mathbf{x}, \boldsymbol{\omega})|,\end{aligned} \tag{26}$$

with a δ function of the gradient and the constraints ensuring that we count valid stationary points, and the determinant of the Hessian serving as the Jacobian of the argument to the δ function [28,29]. It is usually more interesting to condition the count on interesting properties of the stationary points, such as the energy and spectrum trace, or

$$\begin{aligned}d\nu_H(\mathbf{x}, \boldsymbol{\omega}|E, \mu) = d\nu_H(\mathbf{x}, \boldsymbol{\omega})\, \delta(NE - H(\mathbf{x})) \\ \times \delta(N\mu - \operatorname{Tr} \operatorname{Hess} H(\mathbf{x}, \boldsymbol{\omega})).\end{aligned} \tag{27}$$

We specifically want to control the value of the minimum eigenvalue of the Hessian at the stationary points. Using the method introduced in Sec. II, we can write the number of stationary points with energy E, the Hessian trace μ, and the smallest eigenvalue λ^* as

$$\begin{aligned}\mathcal{N}_H(E, \mu, \lambda^*) &= \int d\nu_H(\mathbf{x}, \boldsymbol{\omega}|E, \mu)\, \delta(N\lambda^* - \lambda_{\min}(\operatorname{Hess} H(\mathbf{x}, \boldsymbol{\omega}))) \\ &= \lim_{\beta\to\infty} \int d\nu_H(\mathbf{x}, \boldsymbol{\omega}|E, \mu) \frac{d\mathbf{s}\, \delta(N - \|\mathbf{s}\|^2) \delta(\mathbf{s}^T \partial \mathbf{g}(\mathbf{x})) e^{-\beta \mathbf{s}^T \operatorname{Hess} H(\mathbf{x},\boldsymbol{\omega})\mathbf{s}}}{\int d\mathbf{s}'\, \delta(N - \|\mathbf{s}'\|^2) \delta(\mathbf{s}'^T \partial \mathbf{g}(\mathbf{x})) e^{-\beta \mathbf{s}'^T \operatorname{Hess} H(\mathbf{x},\boldsymbol{\omega})\mathbf{s}'}} \delta(N\lambda^* - \mathbf{s}^T \operatorname{Hess} H(\mathbf{x}, \boldsymbol{\omega})\mathbf{s}),\end{aligned} \tag{28}$$

where the additional δ functions

$$\delta(\mathbf{s}^T \partial \mathbf{g}(\mathbf{x})) = \prod_{s=1}^{r} \delta(\mathbf{s}^T \partial g_i(\mathbf{x})) \tag{29}$$

ensure that the integrals involving potential eigenvectors $\mathbf{s}$ are constrained to the tangent space of the configuration manifold at the point $\mathbf{x}$.

The complexity of points with a specific energy, stability, and minimum eigenvalue is defined as the average over the ensemble of functions H of the logarithm of the number $\mathcal{N}_H$ of stationary points, or

$$\Sigma_{\lambda^*}(E, \mu) = \frac{1}{N} \overline{\log \mathcal{N}_H(E, \mu, \lambda^*)}. \tag{30}$$

In practice, this can be computed by introducing replicas to treat the logarithm ($\log x = \lim_{n\to 0} \frac{\partial}{\partial n} x^n$) and introducing another set of replicas to treat each of the normalizations in the numerator ($x^{-1} = \lim_{m\to -1} x^m$). This leads to the

expression

$$\Sigma_{\lambda^*}(E,\mu) = \lim_{\beta\to\infty}\lim_{n\to 0}\frac{1}{N}\frac{\partial}{\partial n}\int\prod_{a=1}^{n}\overline{\left[d\nu_H(\mathbf{x}_a,\boldsymbol{\omega}_a|E,\mu)\,\delta\left(N\lambda^* - \left(\mathbf{s}_a^1\right)^T \operatorname{Hess} H(\mathbf{x}_a,\boldsymbol{\omega}_a)\mathbf{s}_a^1\right)\right.}$$
$$\left.\times\lim_{m_a\to 0}\left(\prod_{\alpha=1}^{m_a} d\mathbf{s}_a^\alpha\,\delta\left(N-\left\|\mathbf{s}_a^\alpha\right\|^2\right)\delta\left(\left(\mathbf{s}_a^\alpha\right)^T\partial\mathbf{g}(\mathbf{x}_a)\right)e^{-\beta(\mathbf{s}_a^\alpha)^T\operatorname{Hess}H(\mathbf{x}_a,\boldsymbol{\omega}_a)\mathbf{s}_a^\alpha}\right)\right] \tag{31}$$

for the complexity of stationary points of a given energy, trace, and smallest eigenvalue.

The marginal complexity follows from the complexity as a function of μ and λ^* in an analogous way to Sec. II C. In general, one sets $\lambda^* = 0$ and tunes μ from a sufficiently large value until the complexity develops an imaginary component, which corresponds to the bulk of the spectrum touching zero. The value $\mu = \mu_{\rm m}$ that satisfies this is the marginal stability.

In the cases studied here with zero signal to noise, a simpler approach is possible. The marginal stability $\mu = \mu_{\rm m}$ can be identified by requiring that the complexity is stationary with respect to changes in the value of the minimum eigenvalue λ^*, or

$$0 = \frac{\partial}{\partial\lambda^*}\Sigma_{\lambda^*}(E,\mu_{\rm m}(E))\big|_{\lambda^*=0}. \tag{32}$$

The marginal complexity follows by evaluating the complexity conditioned on $\lambda^* = 0$ at the marginal stability $\mu = \mu_{\rm m}(E)$,

$$\Sigma_{\rm m}(E) = \Sigma_0(E,\mu_{\rm m}(E)). \tag{33}$$

B. General features of saddle-point computation

Several elements of the computation of the marginal complexity, and indeed the ordinary dominant complexity, follow from the formulas of the above section in the same way. The physicist's approach to this problem seeks to convert all of the components of the Kac–Rice measure defined in (26) and (27) into elements of an exponential integral over configuration space. To begin with, all Dirac δ functions are expressed using their Fourier representation, with

$$\delta(\nabla H(\mathbf{x}_a,\boldsymbol{\omega}_a)) = \int\frac{d\hat{\mathbf{x}}_a}{(2\pi)^N}e^{i\hat{\mathbf{x}}_a^T\nabla H(\mathbf{x}_a,\boldsymbol{\omega}_a)}, \tag{34}$$

$$\delta(NE - H(\mathbf{x}_a)) = \int\frac{d\hat{\beta}_a}{2\pi}e^{\hat{\beta}_a[NE-H(\mathbf{x}_a)]}, \tag{35}$$

$$\delta\left(N\lambda^* - \left(\mathbf{s}_a^1\right)^T\operatorname{Hess}H(\mathbf{x}_a,\boldsymbol{\omega})\mathbf{s}_a^1\right) = \int\frac{d\hat{\lambda}_a}{2\pi}e^{\hat{\lambda}_a[N\lambda^*-(\mathbf{s}_a^1)^T\operatorname{Hess}H(\mathbf{x}_a,\boldsymbol{\omega})\mathbf{s}_a^1]}. \tag{36}$$

To do this we introduced auxiliary fields $\hat{\mathbf{x}}_a$, $\hat{\beta}_a$, and $\hat{\lambda}_a$. Because the permutation symmetry of replica vectors is preserved in replica symmetry breaking (RSB) orders, the order parameters $\hat{\beta}$ and $\hat{\lambda}$ will quickly lose their indices, since they will ubiquitously be constant over the replica index at the eventual saddle-point solution.

We would like to make a similar treatment of the determinant of the Hessian that appears in (26). The standard approach is to drop the absolute value function around the determinant. This can potentially lead to severe problems with the complexity [19]. However, it is a justified step when the parameters of the problem E, μ, and λ^* put us in a regime where the exponential majority of stationary points have the same index. This is true for maxima and minima, and for saddle points whose spectra have a strictly positive bulk with a fixed number of negative outliers. It is in particular a safe operation for the present problem of marginal minima, which lie right at the edge of disaster.

Dropping the absolute value function allows us to write

$$\det\operatorname{Hess}H(\mathbf{x}_a,\boldsymbol{\omega}_a) = \int d\bar{\boldsymbol{\eta}}_a\,d\boldsymbol{\eta}_a\,e^{-\bar{\boldsymbol{\eta}}_a^T\operatorname{Hess}H(\mathbf{x}_a,\boldsymbol{\omega}_a)\boldsymbol{\eta}_a} \tag{37}$$

using the N-dimensional Grassmann vectors $\bar{\boldsymbol{\eta}}_a$ and $\boldsymbol{\eta}_a$. For the spherical models this step is unnecessary, since there are other ways to treat the determinant keeping the absolute value signs, as in previous works [4,7]. However, other examples of ours are for models where the same techniques are impossible.

Finally, the δ function fixing the trace of the Hessian to μ in (27) must be addressed. One could treat it using a Fourier representation as in (34)–(36), but this is inconvenient because a term of the form $\operatorname{Tr}\partial\partial H(\mathbf{x})$ in the exponential integrand cannot be neatly captured in superspace representation introduced in the next section. However, in the cases we study in this paper a simplification can be made: the trace of $\partial\partial H$ can be separated into two pieces, one that is spatially independent and one that is typically small, or

$$\operatorname{Tr}\partial\partial H(\mathbf{x}) = N\mu_H^* + \Delta_H(\mathbf{x}), \tag{38}$$

where $\overline{\mu_H^*} = \mu^*$ and $\overline{\Delta_H(\mathbf{x})} = O(N^0)$. Then fixing the trace of the Hessian to μ implies that

$$\mu = \frac{1}{N}\operatorname{Tr}\operatorname{Hess}H(\mathbf{x},\boldsymbol{\omega}) = \frac{1}{N}\left(\partial\partial H(\mathbf{x}) + \sum_{i=1}^{r}\omega_i\operatorname{Tr}\partial\partial g_i(\mathbf{x})\right)$$
$$= \mu^* + \frac{1}{N}\sum_{i=1}^{r}\omega_i\operatorname{Tr}\partial\partial g_i(\mathbf{x}) + O(N^{-1}) \tag{39}$$

for typical samples H. In particular, here we study only cases with quadratic g_i, which results in a linear expression relating μ and the ω_i that is independent of $\mathbf{x}$. Since H contains the disorder of the problem, this simplification means that the effect of fixing the trace is largely independent of the disorder and mostly depends on properties of the constraint manifold.

C. Superspace representation

The ordinary Kac–Rice calculation involves many moving parts, and this method for incorporating marginality adds even more. It is therefore convenient to introduce compact and sim-

plifying notation through a superspace representation. The use of superspace in the Kac–Rice calculation is well established, as well as the deep connections with Becchi-Rouet-Stora-Tyutin (BRST) symmetry that is implied [30–32]. Appendix A introduces the notation and methods of superspace algebra. Here we describe how it can be used to simplify the complexity calculation for marginal minima.

We consider the $\mathbb{R}^{N|4}$ superspace whose Grassmann indices are $\bar\theta_1,\ \theta_1,\ \bar\theta_2,\ \theta_2$. Consider the supervector defined by

$$\boldsymbol{\phi}_a^\alpha(1,2) = \mathbf{x}_a + \bar\theta_1\boldsymbol{\eta}_a + \bar{\boldsymbol{\eta}}_a\theta_1 + i\hat{\mathbf{x}}_a\bar\theta_1\theta_1 + \mathbf{s}_a^\alpha(\bar\theta_1\theta_2 + \bar\theta_2\theta_1). \tag{40}$$

Note that this supervector does not span the whole superspace: only a couple terms from the $\bar\theta_2$, θ_2 sector are present, since the rest are unnecessary for our representation. With this supervector so defined, the replicated count of stationary points with energy E, trace μ, and smallest eigenvalue λ^* can be written as

$$\begin{aligned}\mathcal{N}_H(E,\mu,\lambda^*)^n &= \lim_{\beta\to\infty}\int d\boldsymbol{\omega}\, d\hat\beta\, d\hat\lambda \prod_{a=1}^{n}\lim_{m_a\to 0}\prod_{\alpha=1}^{m_a} d\boldsymbol{\phi}_a^\alpha \exp\Big\{\delta^{\alpha 1}N(\hat\beta E \\ &\quad + \hat\lambda\lambda^*) + \int d1\, d2\, B^\alpha(1,2)L\big(\boldsymbol{\phi}_a^\alpha(1,2),\boldsymbol{\omega}\big)\Big\}. \end{aligned}\tag{41}$$

Here we have also defined the operator

$$B^\alpha(1,2) = \delta^{\alpha 1}\bar\theta_2\theta_2(1-\hat\beta\bar\theta_1\theta_1) - \delta^{\alpha 1}\hat\lambda - \beta, \tag{42}$$

which encodes various aspects of the complexity problem. When the Lagrangian is expanded in a series with respect to the Grassmann indices and the definition of B inserted, the result of the Grassmann integrals produces exactly the content of the integrand in (31) with the substitutions (34), (35), (36), and (37) of the Dirac δ functions and the determinant made. The new measures

$$\begin{aligned} d\boldsymbol{\phi}_a^\alpha &= \left[d\mathbf{x}_a\,\delta(\mathbf{g}(\mathbf{x}_a))\,\frac{d\hat{\mathbf{x}}_a}{(2\pi)^N}\,d\boldsymbol{\eta}_a\,d\bar{\boldsymbol{\eta}}_a\,\delta^{\alpha 1} + (1-\delta^{\alpha 1})\right] \\ &\quad\times d\mathbf{s}_a^\alpha\,\delta\big(\|\mathbf{s}_a^\alpha\|^2 - N\big)\,\delta\big((\mathbf{s}_a^\alpha)^T\partial\mathbf{g}(\mathbf{x}_a)\big), \end{aligned}\tag{43}$$

$$d\boldsymbol{\omega} = \left(\prod_{i=1}^{r} d\omega_i\right)\delta\left(N\mu - \mu^* - \sum_i^r \omega_i\,\mathrm{Tr}\,\partial\partial g_i\right) \tag{44}$$

collect the individual measures of the various fields embedded in the superfield, along with their constraints. With this way of writing the replicated count, the problem of marginal complexity temporarily takes the schematic form of an equilibrium calculation with configurations $\boldsymbol{\phi}$, inverse temperature B, and energy L. This makes the intermediate pieces of the calculation dramatically simpler. Of course the intricacies of the underlying problem are not banished: near the end of the calculation, terms involving the superspace must be expanded. We will make use of this representation to simplify the analysis of the marginal complexity when analyzing random sums of squares in Sec. IV C.

IV. EXAMPLES

In this section we present analysis of marginal complexity in three random landscapes. In Sec. IV A we treat the spherical spin glasses, which reveals some general aspects of the calculation. Since the spherical spin glasses are Gaussian and have identical GOE spectra at each stationary point, the approach introduced here is overkill. In Sec. IV B we apply the methods to a multispherical spin glass, which is still Gaussian but has a non-GOE spectrum whose shape can vary between stationary points. Finally, in Sec. IV C we analyze a model of sums of squared random functions, which is non-Gaussian and whose Hessian statistics depend on the conditioning of the energy and gradient.

A. Spherical spin glasses

The spherical spin glasses are a family of models that encompass every isotropic Gaussian field on the hypersphere. Their configuration space is the sphere S^{N-1} defined by all $\mathbf{x}\in\mathbb{R}^N$ such that $0 = g(\mathbf{x}) = \frac{1}{2}(\|\mathbf{x}\|^2 - N)$. One can consider the models as defined by ensembles of centered Gaussian functions H such that the covariance between two points in the configuration space is

$$\overline{H(\mathbf{x})H(\mathbf{x}')} = Nf\left(\frac{\mathbf{x}\cdot\mathbf{x}'}{N}\right) \tag{45}$$

for some function f with positive series coefficients. Such functions can be considered to be made up of all-to-all tensorial interactions, with

$$H(\mathbf{x}) = \sum_{p=0}^{\infty}\frac{1}{p!}\sqrt{\frac{f^{(p)}(0)}{N^{p-1}}}\sum_{i_1,\ldots,i_p}^{N} J_{i_1,\ldots,i_p}x_{i_1}\cdots x_{i_p}, \tag{46}$$

and the elements of the tensors J being independently distributed with the unit normal distribution [33]. We focus on marginal minima in models with $f'(0) = 0$, which corresponds to models without a random external field. Such a random field would correspond in each individual sample H to a signal, and therefore complicate the analysis by correlating the positions of stationary points and the eigenvectors of their Hessians. Here, μ^* of (38) is zero.

The marginal optima of these models can be studied without the methods introduced in this paper, and have been in the past [4,7]. First, these models are Gaussian, so at large N the Hessian is statistically independent of the gradient and energy [19,20]. Therefore, conditioning the Hessian can be done mostly independently from the problem of counting stationary points. Second, in these models the Hessian at every point in the landscape belongs to the GOE class with the same width of the spectrum $\mu_{\mathrm{m}} = 2\sqrt{f''(1)}$. Therefore, all marginal minima in these systems have the same constant shift $\mu = \mu_{\mathrm{m}}$. Despite the fact that the complexity of marginal optima is well known by simpler methods, it is instructive to carry through the calculation for this case, since we will learn some things about its application in more nontrivial settings.

Note that in the pure version of these models with $f(q) = \frac{1}{2}q^p$, the methods of this section must be amended slightly. This is because in these models there is an exact correspondence $\mu = -pE$ between the trace of the Hessian and the energy, and therefore they cannot be fixed independently. This correspondence implies that when $\mu = \mu_{\rm m}$, the corresponding energy level $E_{\rm th} = -\frac{1}{p}\mu_{\rm m}$ contains all marginal minima. This is what gives this threshold energy such singular importance to dynamics in the pure spherical models.

The procedure to treat the complexity of the spherical models has been made in detail elsewhere [7]. Here we make only a sketch of the steps involved. First we notice that $\mu = \frac{1}{N}\omega \,\mathrm{Tr}\,\partial\partial g(\mathbf{x}) = \omega$, so that the only Lagrange multiplier ω in this problem is set directly to the shift μ. The substitutions (34), (35), and (36) are made to convert the Dirac δ functions into exponential integrals, and the substitution (37) is made to likewise convert the determinant.

Once these substitutions have been made, the entire expression (31) is an exponential integral whose argument is a linear functional of H. This allows for the average to be taken over the disorder. If we gather all the H-dependant pieces associated with replica a into the linear functional $\mathcal{O}_a$ then the average over the ensemble of functions H gives

$$\begin{aligned}\overline{e^{\sum_a^n \mathcal{O}_a H(\mathbf{x}_a)}} &= e^{\frac{1}{2}\sum_a^n\sum_b^n \mathcal{O}_a\mathcal{O}_b \overline{H(\mathbf{x}_a)H(\mathbf{x}_b)}} \\ &= e^{N\frac{1}{2}\sum_a^n\sum_b^n \mathcal{O}_a\mathcal{O}_b f(\frac{\mathbf{x}_a\cdot\mathbf{x}_b}{N})}.\end{aligned} \tag{47}$$

The result is an integrand that depends on the many vector variables we have introduced only through their scalar products with each other. We therefore make a change of variables in the integration from those vectors to matrices that encode their possible scalar products. These matrices are

$$\begin{aligned} &C_{ab} = \frac{1}{N}\mathbf{x}_a\cdot\mathbf{x}_b, \quad R_{ab} = -i\frac{1}{N}\mathbf{x}_a\cdot\hat{\mathbf{x}}_b, \quad D_{ab} = \frac{1}{N}\hat{\mathbf{x}}_a\cdot\hat{\mathbf{x}}_b, \\ &Q_{ab}^{\alpha\gamma} = \frac{1}{N}\mathbf{s}_a^\alpha\cdot\mathbf{s}_b^\gamma, \quad \hat{X}_{ab}^\alpha = -i\frac{1}{N}\hat{\mathbf{x}}_a\cdot\mathbf{s}_b^\alpha, \quad X_{ab}^\alpha = \frac{1}{N}\mathbf{x}_a\cdot\mathbf{s}_b^\alpha, \\ &G_{ab} = \frac{1}{N}\bar{\boldsymbol{\eta}}_a\cdot\boldsymbol{\eta}_b. \end{aligned} \tag{48}$$

Order parameters that mix the normal and Grassmann variables generically vanish in these settings and we don't consider them here [34]. This transformation changes the measure of the integral, with

$$\prod_{a=1}^n d\mathbf{x}_a \frac{d\hat{\mathbf{x}}_a}{(2\pi)^N} d\bar{\boldsymbol{\eta}}_a\, d\boldsymbol{\eta}_a \prod_{\alpha=1}^{m_a} d\mathbf{s}_a^\alpha = dC\, dR\, dD\, dG\, dQ\, dX\, d\hat{X}\, (\det J)^{N/2} (\det G)^{-N}, \tag{49}$$

where J is the Jacobian of the transformation in the real-valued fields. This Jacobian takes a block form

$$J = \begin{bmatrix} C & iR & X_1 & \cdots & X_n \\ iR & D & i\hat{X}_1 & \cdots & i\hat{X}_n \\ X_1^T & i\hat{X}_1^T & Q_{11} & \cdots & Q_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_n^T & i\hat{X}_n^T & Q_{n1} & \cdots & Q_{nn} \end{bmatrix}. \tag{50}$$

The Grassmann integrals produces their own inverted Jacobian. The matrix that make up the blocks of the matrix J are such that C, R, and D are $n \times n$ matrices indexed by their lower indices, Q_{ab} is an $m_a \times m_b$ matrix indexed by its upper indices, while X_a is an $n \times m_a$ matrix with one lower and one upper index.

These steps follow identically to those more carefully outlined in the cited papers [4,7]. Following them in the present case, we arrive at a form for the complexity of stationary points with fixed energy E, stability μ, and lowest eigenvalue λ^* with

$$\begin{aligned}\Sigma_{\lambda^*}(E,\mu) = \lim_{\beta\to\infty}\lim_{n\to 0}\lim_{m_1\cdots m_n\to 0} \frac{1}{N}\frac{\partial}{\partial n}\int dC\, dR\, dD\, dG\, dQ\, dX\, d\hat{X}\, d\hat{\beta}\, d\hat{\lambda} \\ \times \exp\Bigg\{ nN\mathcal{S}_{\rm SSG}(\hat{\beta}, C, R, D, G|E,\mu) + nN\mathcal{U}_{\rm SSG}(\hat{\lambda}, Q, X, \hat{X}|\beta,\lambda^*,\mu,C) \\ + \frac{N}{2}\log\det\left[I - \begin{bmatrix} Q_{11} & \cdots & Q_{1n} \\ \vdots & \ddots & \vdots \\ Q_{n1} & \cdots & Q_{nn} \end{bmatrix}^{-1} \begin{bmatrix} X_1^T & i\hat{X}_1^T \\ \vdots & \vdots \\ X_n^T & i\hat{X}_n^T \end{bmatrix} \begin{bmatrix} C & iR \\ iR & D \end{bmatrix}^{-1} \begin{bmatrix} X_1 \cdots X_n \\ i\hat{X}_1 \cdots i\hat{X}_n \end{bmatrix} \right] \Bigg\}. \end{aligned} \tag{51}$$

The exponential integrand is split into two effective actions coupled only by a residual determinant. The first of these actions is the usual effective action for the complexity of the spherical spin glasses, or

$$\mathcal{S}_{\rm SSG}(\hat{\beta}, C, R, D, G|E, \mu) = \hat{\beta}E + \lim_{n\to 0}\frac{1}{n}\Big\{-\mu\,\mathrm{Tr}(R+G) + \frac{1}{2}\sum_{ab}\big[\hat{\beta}^2 f(C_{ab}) + (2\hat{\beta}R_{ab} - D_{ab})f'(C_{ab}) + \big(R_{ab}^2 - G_{ab}^2\big)f''(C_{ab})\big] + \frac{1}{2}\log\det\begin{bmatrix} C & iR \\ iR^T & D\end{bmatrix} - \log\det G\Big\}. \tag{52}$$

The second of these actions is analogous to the effective action (8) from the GOE example of Sec. II B and contains the contributions from the marginal pieces of the calculation, and is given by

$$\begin{aligned}&\mathcal{U}_{\rm SSG}(\hat{\lambda}, Q, X, \hat{X}|\beta, \lambda^*, \mu, C)\\ &= \hat{\lambda}\lambda^* + \lim_{n\to 0}\lim_{m_1\cdots m_n\to 0}\frac{1}{n}\Bigg\{\frac{1}{2}\log\det Q - \sum_{a=1}^{n}\Bigg(\sum_{\alpha=1}^{m_a}\beta\mu Q_{aa}^{\alpha\alpha} + \hat{\lambda}\mu Q_{aa}^{11}\Bigg) + 2\sum_{ab}^{n} f''(C_{ab})\\ &\quad\times\Bigg[\beta\sum_{\alpha}^{m_a}\Bigg(\beta\sum_{\gamma}^{m_b}\big(Q_{ab}^{\alpha\gamma}\big)^2 - \hat{\beta}\big(X_{ab}^{\alpha}\big)^2 - 2X_{ab}^{\alpha}\hat{X}_{ab}^{\alpha}\Bigg) + \hat{\lambda}\big(\hat{\lambda}\big(Q_{ab}^{11}\big)^2 - \hat{\beta}\big(X_{ab}^{1}\big)^2 - 2X_{ab}^{1}\hat{X}_{ab}^{1}\big) + \beta\hat{\lambda}\Bigg(\sum_{\alpha}^{m_a}Q_{ab}^{\alpha 1} + \sum_{\alpha}^{m_b}Q_{ab}^{1\alpha}\Bigg)\Bigg]\Bigg\}.\end{aligned} \tag{53}$$

The fact that the complexity can be split into two relatively independent pieces in this way is a characteristic of the isotropic and Gaussian nature of the spherical spin glass. In Sec. IV C we study a model whose energy is isotropic but not Gaussian and where such a decomposition is impossible.

There are some dramatic simplifications that emerge from the structure of this particular problem. First, notice that the dependence on the parameters X and $\hat{X}$ are purely quadratic. Therefore, there will always be a saddle-point condition where they are both zero. In this case without a fixed or random field, we except this solution to be correct. We can reason about why this is so: X, for instance, quantifies the correlation between the typical position of stationary points and the direction of their typical eigenvectors. In a landscape without a signal, where no direction is any more important than any other, we expect such correlations to be zero: where a state is located does not give any information as to the orientation of its soft directions. On the other hand, in the spiked case, or with an external field, the preferred direction can polarize both the direction of typical stationary points *and* their soft eigenvectors. Therefore, in these instances one must account for solutions with nonzero X and $\hat{X}$.

We similarly expect that $Q_{ab} = 0$ for $a \neq b$. For the contrary to be true, eigenvectors at independently sampled stationary points would need to have their directions correlated. This is expected in situations with a signal, where such correlations would be driven by a shared directional bias towards the signal. In the present situation, where there is no signal, such correlations do not exist.

When we take $X = \hat{X} = 0$ and $Q_{ab}^{\alpha\beta} = \delta_{ab}Q^{\alpha\beta}$, we find that

$$\mathcal{U}_{\rm SSG}(\hat{\lambda}, Q, 0, 0|\beta, \lambda^*, \mu, C) = \mathcal{U}_{\rm GOE}(\hat{\lambda}, Q|\beta, \lambda^*, \mu), \tag{54}$$

with $\sigma^2 = f''(1)$. That is, the effective action for the terms related to fixing the eigenvalue in the spherical Kac–Rice problem is exactly the same as that for the GOE problem. This is perhaps not so surprising, since we established from the beginning that the Hessian of the spherical spin glasses belongs to the GOE class.

The remaining analysis of the eigenvalue-dependent part $\mathcal{U}_{\rm SSG}$ follows precisely the same steps as were made in Sec. II B for the GOE example. The result of the calculation is also the same: the exponential factor containing $\mathcal{U}_{\rm SSG}$ produces precisely the large deviation function $G_{\lambda^*}(\mu)$ of (21) [again with $\sigma^2 = f''(1)$]. The remainder of the integrand depending on $\mathcal{S}_{\rm SSG}$ produces the ordinary complexity of the spherical spin glasses without conditions on the Hessian eigenvalue. We therefore find that

$$\Sigma_{\lambda^*}(E, \mu) = \Sigma(E, \mu) + G_{\lambda^*}(\mu). \tag{55}$$

We find the marginal complexity by solving

$$0 = \frac{\partial}{\partial\lambda^*}\Sigma_{\lambda^*}(E, \mu_{\rm m}(E))\Big|_{\lambda^*=0} = \frac{\partial}{\partial\lambda^*}G_{\lambda^*}(\mu_{\rm m}(E))\Big|_{\lambda^*=0}, \tag{56}$$

which gives $\mu_{\rm m}(E) = 2\sigma = 2\sqrt{f''(1)}$ independent of E, as we presaged above. Since $G_0(\mu_{\rm m}) = 0$, this gives finally

$$\Sigma_{\rm m}(E) = \Sigma_0(E, \mu_{\rm m}(E)) = \Sigma(E, \mu_{\rm m}). \tag{57}$$

The marginal complexity in these models is thus simply the ordinary complexity evaluated at a fixed trace $\mu_{\rm m}$ of the Hessian.

B. Multispherical spin glasses

The multispherical spin glasses are a simple extension of the spherical ones, where the configuration space is taken to be the union of more than one hypersphere. Here we consider the specific case where the configuration space is the union of two $(N-1)$-spheres, with $\Omega = S^{N-1}\times S^{N-1}$. The two spheres give rise to two constraints: for $\mathbf{x} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}]$ with components $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathbb{R}^N$, the constraints are $0 = g_1(\mathbf{x}) = \frac{1}{2}(\|\mathbf{x}^{(1)}\|^2 - N)$ and $0 = g_2(\mathbf{x}) = \frac{1}{2}(\|\mathbf{x}^{(2)}\|^2 - N)$. These two constraints are fixed by two Lagrange multipliers ω_1 and ω_2.

The energy in our multispherical spin glass is given by

$$H(\mathbf{x}) = H_1(\mathbf{x}^{(1)}) + H_2(\mathbf{x}^{(2)}) - \epsilon\mathbf{x}^{(1)}\cdot\mathbf{x}^{(2)}. \tag{58}$$

The energy H_i of each individual sphere is taken to be a centered Gaussian random function with a covariance given in the usual spherical spin glass way for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$ by

$$\overline{H_i(\mathbf{x})H_j(\mathbf{x}')} = N\delta_{ij} f_i\left(\frac{\mathbf{x}\cdot\mathbf{x}'}{N}\right), \tag{59}$$

with the functions f_1 and f_2 not necessarily the same. As for the spherical spin glasses, μ^* of (38) is zero.

In this problem, there is an energetic competition between the independent spin glass energies on each sphere and their tendency to align or anti-align through the interaction term. These models have more often been studied with random fully connected couplings between the spheres, for which it is possible to also use configuration spaces involving spheres of different sizes [35–41]. The deterministically coupled model was previously studied as a thought experiment [7].

We again make use of the method of Lagrange multipliers to find stationary points on the constrained configuration space. The Lagrangian and its gradient and Hessian are

$$L(\mathbf{x}, \boldsymbol{\omega}) = H(\mathbf{x}) + \tfrac{1}{2}\omega_1(\|\mathbf{x}^{(1)}\|^2 - N) + \tfrac{1}{2}\omega_2(\|\mathbf{x}^{(2)}\|^2 - N), \tag{60}$$

$$\nabla H(\mathbf{x}, \boldsymbol{\omega}) = \begin{bmatrix} \partial_1 H_1(\mathbf{x}^{(1)}) - \epsilon\mathbf{x}^{(2)} + \omega_1\mathbf{x}^{(1)} \\ \partial_2 H_2(\mathbf{x}^{(2)}) - \epsilon\mathbf{x}^{(1)} + \omega_2\mathbf{x}^{(2)} \end{bmatrix}, \tag{61}$$

$$\operatorname{Hess} H(\mathbf{x}, \boldsymbol{\omega}) = \begin{bmatrix} \partial_1\partial_1 H_1(\mathbf{x}^{(1)}) + \omega_1 I & -\epsilon I \\ -\epsilon I & \partial_2\partial_2 H_2(\mathbf{x}^{(2)}) + \omega_2 I \end{bmatrix}, \tag{62}$$

where $\partial_1 = \frac{\partial}{\partial \mathbf{x}^{(1)}}$ and $\partial_2 = \frac{\partial}{\partial \mathbf{x}^{(2)}}$. Like in the spherical spin glasses, fixing the trace of the Hessian to μ is equivalent to a constraint on the Lagrange multipliers. However, in this case it corresponds to $\mu = \omega_1 + \omega_2$, and therefore they are not uniquely fixed by fixing μ.

Since the energy in the multispherical models is Gaussian, the properties of the matrix $\partial\partial H$ are again independent of the energy and gradient. This means that the form of the Hessian is parameterized solely by the values of the Lagrange multipliers ω_1 and ω_2, just as $\mu = \omega$ alone parameterized the Hessian in the spherical spin glasses. Unlike that case, however, the Hessian takes different shapes with different spectral widths depending on their precise combination. In Appendix C we derive a variational form for the spectral density of the Hessian in these models using standard methods.

Because of the independence of the Hessian, the method introduced in this article is not necessary to characterize the marginal minima of this system. Rather, we could take the spectral density derived in Appendix C and find the Lagrange multipliers ω_1 and ω_2 corresponding with marginality by tuning the edge of the spectrum to zero. In some ways the current method is more convenient than this, since it is a purely variational method and therefore can be reduced to a single root-finding exercise.

Unlike the constraints on the configurations $\mathbf{x}$, the constraint on the tangent vectors $\mathbf{s} = [\mathbf{s}^{(1)}, \mathbf{s}^{(2)}] \in \mathbb{R}^{2N}$ remains the same spherical constraint as before, which implies $N = \|\mathbf{s}\|^2 = \|\mathbf{s}^{(1)}\|^2 + \|\mathbf{s}^{(2)}\|^2$. Defining intra- and inter-sphere overlap matrices

$$Q_{ab}^{ij,\alpha\gamma} = \frac{1}{N}\mathbf{s}_a^{(i),\alpha}\cdot\mathbf{s}_b^{(j),\gamma}, \tag{63}$$

this problem no longer has the property that the diagonal of the Qs is one, but instead that $1 = Q_{aa}^{11,\alpha\alpha} + Q_{aa}^{22,\alpha\alpha}$. This is the manifestation of the fact that a normalized vector in the tangent space of the multispherical model need not be equally spread over the two subspaces but can be concentrated in one or the other.

The calculation of the marginal complexity in this problem follows very closely to that of the spherical spin glasses in the previous subsection. We immediately make the simplifying assumptions that the soft directions of different stationary points are typically uncorrelated and therefore $X = \hat{X} = 0$ and the overlaps Q between eigenvectors are only nonzero when in the same replica. The result for the complexity has the schematic form of (51), but with different effective actions depending now on overlaps inside each of the two spheres and between the two spheres. The effective action for the traditional complexity of the multispherical spin glass is

$$\begin{aligned}
&\mathcal{S}_{\mathrm{MSG}}(\hat{\beta}, C^{11}, R^{11}, D^{11}, G^{11}, C^{22}, R^{22}, D^{22}, G^{22}, C^{12}, R^{12}, R^{21}, D^{12}, G^{12}, G^{21}|E, \omega_1, \omega_2) \\
&\quad = \hat{\beta}\left(E - E_1 - E_2 - \epsilon c_d^{12}\right) + \mathcal{S}_{\mathrm{SSG}}(\hat{\beta}, C^{11}, R^{11}, D^{11}, G^{11}|E_1, \omega_1) \\
&\qquad + \mathcal{S}_{\mathrm{SSG}}(\hat{\beta}, C^{22}, R^{22}, D^{22}, G^{22}|E_2, \omega_2) + \lim_{n\to 0}\frac{1}{n}\Bigg\{\epsilon\,\mathrm{Tr}(R^{12} + R^{21} + G^{12} + G^{21} - \hat{\beta}C^{12}) \\
&\qquad + \frac{1}{2}\log\det\left(I - \begin{bmatrix} C^{11} & iR^{11} \\ iR^{11} & D^{11} \end{bmatrix}^{-1}\begin{bmatrix} C^{12} & iR^{12} \\ iR^{21} & D^{12} \end{bmatrix}\begin{bmatrix} C^{22} & iR^{22} \\ iR^{22} & D^{22} \end{bmatrix}^{-1}\begin{bmatrix} C^{12} & iR^{21} \\ iR^{21} & D^{12} \end{bmatrix}\right) - \log\det(I - (G^{11}G^{22})^{-1}G^{12}G^{21})\Bigg\},
\end{aligned} \tag{64}$$

which is the sum of two effective actions (52) for the spherical spin glass associated with each individual sphere, and some coupling terms. The order parameters are defined the same as in the spherical spin glasses, but now with raised indices to indicate whether the vectors come from one or the other spherical subspace. The effective action for the eigenvalue-dependent

part of the complexity is likewise given by

$$\mathcal{U}_{\mathrm{MSG}}(\hat{q},\hat{\lambda},Q^{11},Q^{22},Q^{12}|\beta,\lambda^*,\omega_1,\omega_2)$$
$$=\lim_{m\to 0}\left\{\sum_{\alpha=1}^{m}[\hat{q}^{\alpha}(Q^{11,\alpha\alpha}+Q^{22,\alpha\alpha}-1)-\beta(\omega_1 Q^{11,\alpha\alpha}+\omega_2 Q^{22,\alpha\alpha}-2\epsilon Q^{12,\alpha\alpha})]-\hat{\lambda}(\omega_1 Q^{11,11}+\omega_2 Q^{22,11}-2\epsilon Q^{12,11})\right.$$
$$\left.+\sum_{i=1,2}f_i''(1)\left[\beta^2\sum_{\alpha\gamma}^{m}(Q^{ii,\alpha\gamma})^2+2\beta\hat{\lambda}\sum_{\alpha}^{m}(Q^{ii,1\alpha})^2+\hat{\lambda}^2(Q^{ii,11})^2\right]+\frac{1}{2}\log\det\begin{bmatrix}Q^{11} & Q^{12}\\ Q^{12} & Q^{22}\end{bmatrix}\right\}. \tag{65}$$

The new variables $\hat{q}^{\alpha}$ are Lagrange multipliers introduced to enforce the constraint that $Q^{11,\alpha\alpha}+Q^{22,\alpha\alpha}=1$. Because of this constraint, the diagonal of the Q matrices cannot be taken to be one as in Sec. II B. Instead we take each of the matrices Q^{11}, Q^{22}, and Q^{12} to have the planted replica symmetric form of (9), but with the diagonal not necessarily equal to one, so

$$Q^{ij}=\begin{bmatrix}\tilde{q}_d^{ij} & \tilde{q}_0^{ij} & \tilde{q}_0^{ij} & \cdots & \tilde{q}_0^{ij}\\ \tilde{q}_0^{ij} & q_d^{ij} & q_0^{ij} & \cdots & q_0^{ij}\\ \tilde{q}_0^{ij} & q_0^{ij} & q_d^{ij} & \ddots & q_0^{ij}\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ \tilde{q}_0^{ij} & q_0^{ij} & q_0^{ij} & \cdots & q_d^{ij}\end{bmatrix}. \tag{66}$$

This requires us to introduce two new order parameters $\tilde{q}_d^{ij}$ and q_d^{ij} per pair (i,j), in addition to the off-diagonal order parameters $\tilde{q}_0^{ij}$ and q_0^{ij} already present in (9). We also need two separate Lagrange multipliers $\hat{q}$ and $\hat{\tilde{q}}$ to enforce the tangent space normalization $q_d^{11}+q_d^{22}=1$ and $\tilde{q}_d^{11}+\tilde{q}_d^{22}=1$ for the tilde and untilde replicas, respectively, which will in general take different values at the saddle point. When this ansatz is inserted into the expression (65) for the effective action and the limit of $m\to 0$ is taken, we find

$$\mathcal{U}_{\mathrm{MSG}}\big(\hat{q},\hat{\tilde{q}},\hat{\lambda},\tilde{q}_d^{11},\tilde{q}_0^{11},q_d^{11},q_0^{11},\tilde{q}_d^{22},\tilde{q}_0^{22},q_d^{22},q_0^{22},\tilde{q}_d^{12},\tilde{q}_0^{12},q_d^{12},q_0^{12}\big|\beta,\lambda^*,\omega_1,\omega_2\big)$$
$$=\sum_{i=1,2}\left\{f_i''(1)\big[\beta^2\big((\tilde{q}_d^{ii})^2-(q_d^{ii})^2+2(q_0^{ii})^2-2(\tilde{q}_0^{ii})^2\big)+2\beta\hat{\lambda}\big((\tilde{q}_d^{ii})^2-(\tilde{q}_0^{ii})^2\big)+\hat{\lambda}^2(\tilde{q}_d^{ii})^2\big]-\hat{\lambda}\tilde{q}_d^{ii}\omega_i-\beta(\tilde{q}_d^{ii}-q_d^{ii})\omega_i\right\}$$
$$+\frac{1}{2}\log\big[(2q_0^{12}\tilde{q}_0^{12}-\tilde{q}_0^{12}(\tilde{q}_d^{12}+q_d^{12})-2\tilde{q}_0^{11}q_0^{22}+\tilde{q}_d^{11}\tilde{q}_0^{22}+\tilde{q}_0^{11}q_d^{22})(2q_0^{12}\tilde{q}_0^{12}-\tilde{q}_0^{12}(\tilde{q}_d^{12}+q_d^{12})-2q_0^{11}\tilde{q}_0^{22}+q_d^{11}\tilde{q}_0^{22}$$
$$+\tilde{q}_0^{11}\tilde{q}_d^{22})+2\big(3(q_0^{12})^2-(\tilde{q}_0^{12})^2-2q_0^{12}q_d^{12}-3q_0^{11}q_0^{22}+q_d^{11}q_0^{22}+\tilde{q}_0^{11}\tilde{q}_0^{22}+q_0^{11}q_d^{22}\big)\big((\tilde{q}_0^{12})^2-(\tilde{q}_d^{12})^2-\tilde{q}_0^{11}\tilde{q}_0^{22}+\tilde{q}_d^{11}\tilde{q}_d^{22}\big)$$
$$-\big(2(q_0^{12})^2-(\tilde{q}_0^{12})^2-(q_d^{12})^2-2q_0^{11}q_0^{22}+\tilde{q}_0^{11}\tilde{q}_0^{22}+q_d^{11}q_d^{22}\big)\big((\tilde{q}_0^{12})^2-(\tilde{q}_d^{12})^2-\tilde{q}_0^{11}\tilde{q}_0^{22}+\tilde{q}_d^{11}\tilde{q}_d^{22}\big)\big]$$
$$-\log\big[(q_d^{11}-q_0^{11})(q_d^{22}-q_0^{22})-(q_d^{12}-q_0^{12})^2\big]+2\epsilon\big[\hat{\lambda}\tilde{q}_d^{12}+\beta(\tilde{q}_d^{12}-q_d^{12})\big]-\hat{q}(q_d^{11}+q_d^{22}-1)+\hat{\tilde{q}}(\tilde{q}_d^{11}+\tilde{q}_d^{22}-1). \tag{67}$$

To make the limit to zero temperature, we once again need an ansatz for the asymptotic behavior of the overlaps. These take the form $q_0^{ij}=q_d^{ij}-y_0^{ij}\beta^{-1}-z_0^{ij}\beta^{-2}$. Notice that in this case, the asymptotic behavior of the off-diagonal elements is to approach the value of the diagonal rather than to approach one. We also require $\tilde{q}_d^{ij}=q_d^{ij}-\tilde{y}_d^{ij}\beta^{-1}-\tilde{z}_d^{ij}\beta^{-2}$, i.e., that the tilde diagonal terms also approach the same diagonal value as the untilde terms, but with potentially different rates.

As before, in order for the logarithmic term to stay finite, there are necessary constraints on the values y. These are

$$\tfrac{1}{2}\big(y_d^{11}-\tilde{y}_d^{11}\big)=y_0^{11}-\tilde{y}_0^{11}, \tag{68}$$
$$\tfrac{1}{2}\big(y_d^{22}-\tilde{y}_d^{22}\big)=y_0^{22}-\tilde{y}_0^{22}, \tag{69}$$
$$\tfrac{1}{2}\big(y_d^{12}-\tilde{y}_d^{12}\big)=y_0^{12}-\tilde{y}_0^{12}. \tag{70}$$

One can see that when the diagonal elements are all equal, this requires the ys for the off-diagonal elements to be equal, as in the GOE case. Here, since the diagonal elements are not necessarily equal, we have a more general relationship.

When the β dependence of the q variables is inserted into the effective action (67) and the limit $\beta\to\infty$ taken, we find an expression that is too large to report here. However, it can be extremized over all of the variables in the problem just as in the previous examples to find the values of the Lagrange multipliers ω_1 and ω_2 corresponding to marginal minima. Figure 2(a) shows examples of the ω_1 and ω_2 corresponding to marginal spectra for a variety of couplings ϵ when the covariances of the energy on the two spherical subspaces are such that $1=f_1''(1)=f_2''(1)$. Figure 2(b) shows the Hessian spectra associated with some specific pairs (ω_1,ω_2). When $\epsilon=0$ and the two spheres are uncoupled, we find the result for two independent spherical spin glasses: if either $\omega_1=2\sqrt{f''(1)}=$

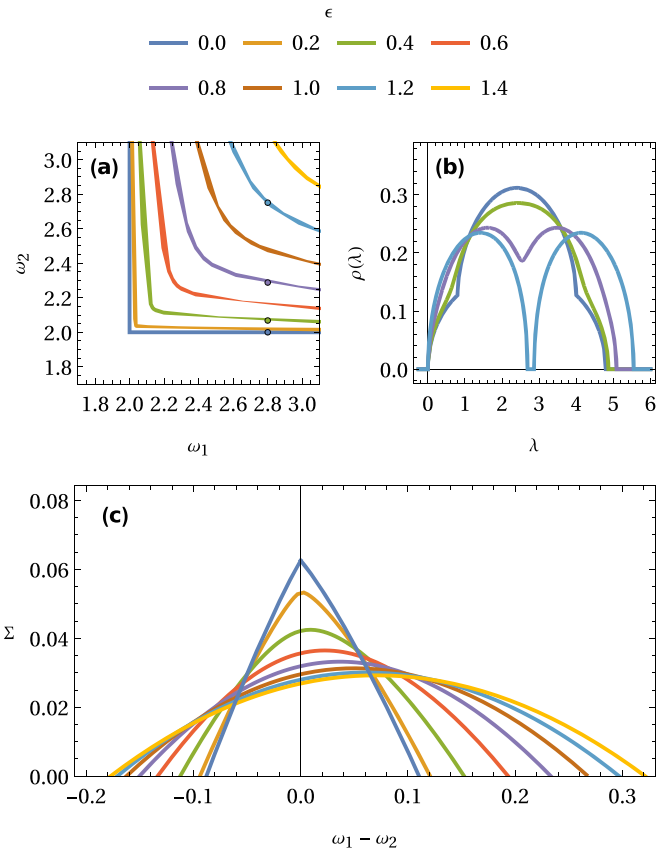

FIG. 2. Properties of marginal minima in the multispherical model. (a) Values of the Lagrange multipliers ω_1 and ω_2 corresponding to a marginal spectrum for multispherical spin glasses with $\sigma_1^2 = f_1''(1) = 1$, $\sigma_2^2 = f_2''(1) = 1$, and various ϵ. (b) Spectra corresponding to the parameters ω_1 and ω_2 marked by the circles in panel (a). (c) The complexity of marginal minima in a multispherical model with $f_1(q) = \frac{1}{6}q^3$ and $f_2(q) = \frac{1}{12}q^4$ for a variety of ϵ. Since $f_1''(1) = f_2''(1) = 1$, the marginal values correspond precisely to those in panels (a) and (b).

2 or $\omega_2 = 2\sqrt{f''(1)} = 2$ and the other Lagrange multiplier is larger than two, then we have a marginal minimum made up of the Cartesian product of a marginal minimum on one subspace and a stable minimum on the other.

Fig. 2(c) shows the complexity of marginal minima in an example where both H_1 and H_2 correspond to pure p-spin models, with $f_1(q) = \frac{1}{6}q^3$ and $f_2(q) = \frac{1}{12}q^4$. Despite having different covariance functions, these both satisfy $1 = f_1''(1) = f_2''(1)$ and therefore have marginal minima for Lagrange multipliers that satisfy the relationships in Fig. 2(a). In the uncoupled system with $\epsilon = 0$, the most common type of marginal stationary point consists of independently marginal stationary points in the two subsystems, with $\omega_1 = \omega_2 = 2$. As ϵ is increased, the most common type of marginal minimum drifts toward points with $\omega_1 > \omega_2$.

Multispherical spin glasses may be an interesting platform for testing ideas about which among the possible marginal minima can attract dynamics and which cannot. In the limit where $\epsilon = 0$ and the configurations of the two spheres are independent, the minima found dynamically should be marginal on both subspaces. Just because technically on the expanded configuration space the Cartesian product of a deep stable minimum on one sphere and a marginal minimum on the other is a marginal minimum on the whole space doesn't mean the deep and stable minimum is any easier to find. This intuitive idea that is precise in the zero-coupling limit should continue to hold at small nonzero coupling, and perhaps reveal something about the inherent properties of marginal minima that do not tend to be found by algorithms.

C. Sums of squared random functions

In this subsection we consider perhaps the simplest example of a non-Gaussian landscape: the problem of sums of squared random functions. This problem has a close resemblance to nonlinear least squares optimization. Though, for reasons we will see it is easier to make predictions for nonlinear *most* squares, i.e., the problem of maximizing the sum of squared terms. We again take a spherical configuration space with $\mathbf{x} \in S^{N-1}$ and $0 = g(\mathbf{x}) = \frac{1}{2}(\|\mathbf{x}\|^2 - N)$ as in the spherical spin glasses. The energy is built from a set of $M = \alpha N$ random functions $V_k : \mathbf{S}^{N-1} \to \mathbb{R}$ that are centered Gaussians with covariance

$$\overline{V_i(\mathbf{x})V_j(\mathbf{x}')} = \delta_{ij} f\left(\frac{\mathbf{x}\cdot\mathbf{x}'}{N}\right). \tag{71}$$

Each of the V_k is an independent spherical spin glass. The total energy is minus the sum of squares of the V_k, or

$$H(\mathbf{x}) = -\frac{1}{2}\sum_{k=1}^{M} V_k(\mathbf{x})^2. \tag{72}$$

The landscape complexity and large deviations of the ground state for the least-squares version of this problem were recently studied in a linear context, with $f(q) = \sigma^2 + aq$ [42–45]. Some results on the ground state of the general nonlinear problem can also be found in Ref. [46], and a solution to the equilibrium problem can be found in Ref. [47]. Those works indicate that the low-lying minima of the least squares problem tend to be either replica symmetric or full replica symmetry breaking. To avoid either a trivial analysis or a very complex one, we instead focus on maximizing the sum of squares, or minimizing (72).

The minima of (72) have a more amenable structure for study than the maxima, as they are typically described by a 1RSB-like structure. There is a heuristic intuition for this: in the limit of $M \to 1$, this problem is just minus the square of a spherical spin glass landscape. The distribution and properties of stationary points low and high in the spherical spin glass are not changed, except that their energies are stretched and maxima are transformed into minima. Therefore, the bottom of the landscape doesn't qualitatively change. The top, however, consists of the zero-energy level set in the spherical spin glass. This level set is well connected, and so the highest states should also be well connected and flat.

Focusing on the bottom of the landscape and therefore dealing with a 1RSB-like problem makes our analysis easier. Algorithms will tend to be stuck in the ways they are in hard optimization problems, and we will be able to predict where.

Therefore, we will study the most squares problem rather than the least squares one. We calculate the complexity of minima of (72) in Appendix D, which corresponds to maximizing the sum of squares, under a replica symmetric ansatz (which covers 1RSB-like problems) for arbitrary covariance f, and we calculate the complexity of marginal minima in this section.

As in the previous sections, we used the method of Lagrange multipliers to analyze stationary points on the constrained configuration space. The Lagrangian and its associated gradient and Hessian are

$$L(\mathbf{x},\omega) = -\frac{1}{2}\left(\sum_k^M V_k(\mathbf{x})^2 - \omega(\|\mathbf{x}\|^2 - N)\right), \tag{73}$$

$$\nabla H(\mathbf{x},\omega) = -\sum_k^M V_k(\mathbf{x})\partial V_k(\mathbf{x}) + \omega\mathbf{x}, \tag{74}$$

$$\operatorname{Hess} H(\mathbf{x},\omega) = -\sum_k^M [\partial V_k(\mathbf{x})\partial V_k(\mathbf{x}) - V_k(\mathbf{x})\partial\partial V_k(\mathbf{x})] + \omega I. \tag{75}$$

Unlike in the spherical and multispherical spin glasses, the value μ^* defined in (38) giving the typical value of $\frac{1}{N}\operatorname{Tr}\partial\partial H$ is not always zero. Instead $\mu^* = -f'(0)$, nonzero where there is a linear term in V. Fixing the trace of the Hessian is therefore equivalent to constraining the value of the Lagrange multiplier $\omega = \mu + f'(0)$.

The derivation of the marginal complexity for this model is complicated, but can be made schematically like that of the derivation of the equilibrium free energy by use of superspace coordinates. Following the framework outlined in Sec. III C, the replicated number of stationary points conditioned on energy E, trace μ, and minimum eigenvalue λ^* is given by

$$\mathcal{N}(E,\mu,\lambda^*)^n = \int d\hat\beta\, d\hat\lambda \prod_{a=1}^n \lim_{m_a\to 0}\prod_{\alpha=1}^{m_a} d\boldsymbol{\phi}_a^\alpha \times \exp\left\{\delta^{\alpha 1}N(\hat\beta E + \hat\lambda\lambda^*) - \frac{1}{2}\int d1\, d2\left[B^\alpha(1,2)\sum_{k=1}^M V_k\big(\boldsymbol{\phi}_a^\alpha(1,2)\big)^2 - (\mu + f'(0))\big\|\boldsymbol{\phi}_a^\alpha(1,2)\big\|^2\right]\right\}, \tag{76}$$

The first step to evaluate this expression is to linearize the dependence on the random functions V. This is accomplished by inserting into the integral a Dirac δ function fixing the value of the energy for each replica, or

$$\delta\big(V_k\big(\boldsymbol{\phi}_a^\alpha(1,2)\big) - v_{ka}^\alpha(1,2)\big) = \int d\hat v_{ka}^\alpha \exp\left[i\int d1\, d2\, \hat v_{ka}^\alpha(1,2)\big(V_k\big(\boldsymbol{\phi}_a^\alpha(1,2)\big) - v_{ka}^\alpha(1,2)\big)\right], \tag{77}$$

where we have introduced auxiliary superfields $\hat v$. With this inserted into the integral, all other instances of V are replaced by v, and the only remaining dependence on the disorder is from the term $\hat v V$ arising from the Fourier representation of the Dirac δ function. This term is linear in V, and therefore the random functions can be averaged over to produce

$$\overline{\exp\left[i\sum_k^M\sum_a^n\sum_\alpha^{m_a}\int d1\, d2\, \hat v_{ka}^\alpha(1,2)V_k\big(\boldsymbol{\phi}_a^\alpha(1,2)\big)\right]} = -\frac{1}{2}\sum_{ab}^n\sum_{\alpha\gamma}^{m_a}\sum_k^M\int d1\, d2\, d3\, d4\, \hat v_{ka}^\alpha(1,2) f\big(\boldsymbol{\phi}_a^\alpha(1,2)\cdot\boldsymbol{\phi}_b^\gamma(3,4)\big)\hat v_{kb}^\gamma(3,4). \tag{78}$$

The entire integrand is now factorized in the indices k and quadratic in the superfields v and $\hat v$ with the kernel

$$\begin{bmatrix} B^\alpha(1,2)\delta(1,3)\delta(2,4)\delta_{ab}\delta^{\alpha\gamma} & i\delta(1,3)\,\delta(2,4)\delta_{ab}\delta^{\alpha\gamma} \\ i\delta(1,3)\,\delta(2,4)\delta_{ab}\delta^{\alpha\gamma} & f\big(\boldsymbol{\phi}_a^\alpha(1,2)\cdot\boldsymbol{\phi}_b^\gamma(3,4)\big)\end{bmatrix}. \tag{79}$$

The integration over v and $\hat v$ results in a term in the effective action of the form

$$-\frac{M}{2}\log\operatorname{sdet}\left[\delta(1,3)\,\delta(2,4)\delta_{ab}\delta^{\alpha\gamma} + B^\alpha(1,2)f\big(\boldsymbol{\phi}_a^\alpha(1,2)\cdot\boldsymbol{\phi}_b^\gamma(3,4)\big)\right]. \tag{80}$$

When expanded, the supermatrix $\boldsymbol{\phi}_a^\alpha(1,2)\cdot\boldsymbol{\phi}_b^\gamma(3,4)$ is constructed of the scalar products of the real and Grassmann vectors that make up $\boldsymbol{\phi}$. The change of variables to these order parameters again results in the Jacobian of (50), contributing

$$\frac{N}{2}\log\det J - \frac{N}{2}\log\det G^2 \tag{81}$$

to the effective action.

Up to this point, the expressions are general and independent of a given ansatz. However, we expect that the order parameters X and $\hat X$ are zero, since again we are in a setting with no signal or external field. Applying this ansatz here avoids a dramatically more complicated expression for the effective action. We also will apply the ansatz that $Q_{ab}^{\alpha\gamma}$ is zero for $a \neq b$, which is equivalent to assuming that the soft directions of typical pairs of stationary points are uncorrelated, and further that $Q^{\alpha\gamma} = Q_{aa}^{\alpha\gamma}$ independently of the index a, implying that correlations in the tangent space of typical stationary points are the same.

Given this ansatz, taking the superdeterminant in (80) yields

$$-\frac{M}{2}\log\det\left\{\left[f'(C)\odot D-\hat{\beta}I+\left(R^{\circ 2}-G^{\circ 2}+I\sum_{\alpha\gamma}2(\delta^{\alpha 1}\hat{\lambda}+\beta)(\delta^{\gamma 1}\hat{\lambda}+\beta)(Q^{\alpha\gamma})^2\right)\odot f''(C)\right]f(C)+(I-R\odot f'(C))^2\right\}$$
$$-n\frac{M}{2}\log\det[\delta_{\alpha\gamma}-2(\delta_{\alpha 1}\hat{\lambda}+\beta)Q^{\alpha\gamma}]+M\log\det[I+G\odot f'(C)]. \tag{82}$$

where once again $\odot$ is the Hadamard product and $A^{\circ n}$ gives the Hadamard power of A. We can already see one substantive difference between the structure of this problem and that of the spherical models: the effective action in this case mixes the order parameters G due to the Grassmann variables with the ones C, R, and D due to the other variables. Notice further that the dependence on Q due to the marginal constraint is likewise no longer separable into its own term. This is the realization of the fact that the Hessian is no longer independent of the energy and gradient.

Now we have reduced the problem to an extremal one over the order parameters $\hat{\beta}$, $\hat{\lambda}$, C, R, D, G, and Q, it is time to make an ansatz for the form of order we expect to find. We will focus on a regime where the structure of stationary points is replica symmetric, and further where typical pairs of stationary points have no overlap. This requires that $f(0)=0$, or that there is no constant term in the random functions. This gives the ansatz

$$C=I,\quad R=rI,\quad D=dI,\quad G=gI. \tag{83}$$

We further take a planted replica symmetric structure for the matrix Q, identical to that in (9). This results in

$$\Sigma_{\lambda^*}(E,\mu)=\frac{1}{N}\lim_{n\to 0}\frac{\partial}{\partial n}\int d\hat{\beta}\,d\hat{\lambda}\,dr\,dd\,dg\,dq_0\,d\tilde{q}_0 e^{nN\mathcal{S}_{\rm RSS}(\hat{\beta},\hat{\lambda},r,d,g,q_0,\tilde{q}_0|\lambda^*,E,\mu,\beta)}, \tag{84}$$

with an effective action

$$\begin{aligned}&\mathcal{S}_{\rm RSS}(\hat{\beta},\hat{\lambda},r,d,g,q_0,\tilde{q}_0|\lambda^*,E,\mu,\beta)\\&=\hat{\beta}E-(\mu+f'(0))(r+g+\hat{\lambda})+\hat{\lambda}\lambda^*+\frac{1}{2}\log\left(\frac{d+r^2}{g^2}\times\frac{1-2q_0+\tilde{q}_0^2}{(1-q_0)^2}\right)\\&\quad-\frac{\alpha}{2}\log\left(\frac{1-4f'(1)\left[\beta(1-q_0)+\frac{1}{2}\hat{\lambda}-\beta(\beta+\hat{\lambda})\left(1-2q_0+\tilde{q}_0^2\right)f'(1)\right]}{[1-2(1-q_0)\beta f'(1)]^2}\right.\\&\quad\left.\times\frac{f(1)[f'(1)d-\hat{\beta}-f''(1)(r^2-g^2+4q_0^2\beta^2-4\tilde{q}_0^2\beta(\beta+\hat{\lambda})+4\beta\hat{\lambda}+2\hat{\lambda}^2)]+(1-rf'(1))^2}{[1+gf'(1)]^2}\right).\end{aligned} \tag{85}$$

We expect as before the limits of q_0 and $\tilde{q}_0$ as β goes to infinity to approach one, defining their asymptotic expansion like in (15) and (16). Upon making this substitution and taking the zero-temperature limit, we find

$$\begin{aligned}\mathcal{S}_{\rm RSS}(\hat{\beta},\hat{\lambda},r,d,g,y,\Delta z|\lambda^*,E,\mu,\infty)&=\hat{\beta}E-(\mu+f'(0))(r+g+\hat{\lambda})+\hat{\lambda}\lambda^*+\frac{1}{2}\log\left(\frac{d+r^2}{g^2}\times\frac{y^2-2\Delta z}{y^2}\right)\\&\quad-\frac{\alpha}{2}\log\left(\frac{1-2(2y+\hat{\lambda})f'(1)+4(y^2-2\Delta z)f'(1)^2}{[1-2yf'(1)]^2}\right.\\&\quad\left.\times\frac{f(1)[f'(1)d-\hat{\beta}-f''(1)(r^2-g^2+8(y\hat{\lambda}+\Delta z)+2\hat{\lambda}^2)]+[1-rf'(1)]^2}{[1+gf'(1)]^2}\right).\end{aligned} \tag{86}$$

We can finally write the complexity with fixed energy E, stability μ, and minimum eigenvalue λ^* as

$$\Sigma_{\lambda^*}(E,\mu)=\operatorname*{extremum}_{\hat{\beta},\hat{\lambda},r,d,g,y,\Delta z}\mathcal{S}_{\rm RSS}(\hat{\beta},\hat{\lambda},r,d,g,y,\Delta z|\lambda^*,E,\mu,\infty). \tag{87}$$

Note that, unlike the previous two examples, the effective action in this case does not split into two largely independent pieces, one relating to the eigenvalue problem and one relating to the ordinary complexity. Instead, the order parameters related to the eigenvalue problem are mixed throughout the effective action with those of the ordinary complexity. This is a signal of the fact that the sum of squares problem is not Gaussian, while the previous two examples are. In all non-Gaussian problems, conditioning on properties of the Hessian cannot be done independently from the complexity, and the method introduced in this paper becomes necessary.

The marginal complexity can be derived from (87) using the condition (32) to fix μ to the marginal stability $\mu_{\rm m}(E)$ and then evaluating the complexity at that stability as in (33). Figure 3 shows the marginal complexity in a sum-of-squares model with $\alpha=\frac{3}{2}$ and $f(q)=q^2+q^3$. Also shown is the dominant complexity computed in Appendix D. As the figure demonstrates, the range of energies at which marginal minima are found can differ significantly from those implied by the dominant complexity, with the lowest energy significantly

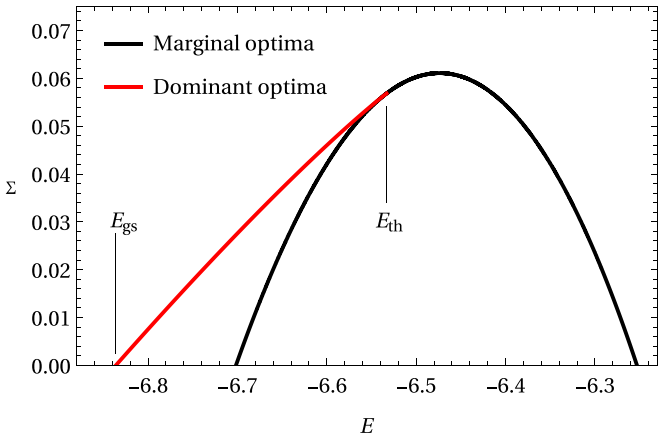

FIG. 3. Dominant and marginal complexity in the nonlinear sum of squares problem for $\alpha = \frac{3}{2}$ and $f(q) = q^2 + q^3$. The ground-state energy E_{gs} and the threshold energy E_{th} are marked on the plot.

higher than the ground state and the highest energy significantly higher than the threshold.

Figure 4 shows the associated marginal stability $\mu_m(E)$ for the same model. Recall that the definition of the marginal stability in (32) is that which eliminates the variation of $\Sigma_{\lambda^*}(E, \mu)$ with respect to λ^* at the point $\lambda^* = 0$. Unlike in the Gaussian spherical spin glass, in this model $\mu_m(E)$ varies with energy in a nontrivial way. The figure also shows the dominant stability, which is the stability associated with the dominant complexity and coincides with the marginal stability only at the threshold energy.

Because this version of the model has no signal, we were able to use the heuristic (32) to fix the marginal stability. However, we could also have used the more general method for finding a pseudogapped Hessian spectrum by locating the value of μ at which the complexity develops an imaginary part, as described in Sec. II C and pictured in Fig. 1. The real and imaginary parts of the complexity $\Sigma_0(E, \mu)$ are plotted in Fig. 5 as a function of μ at fixed energy. The figure also shows the marginal stability μ_m predicted by the variational

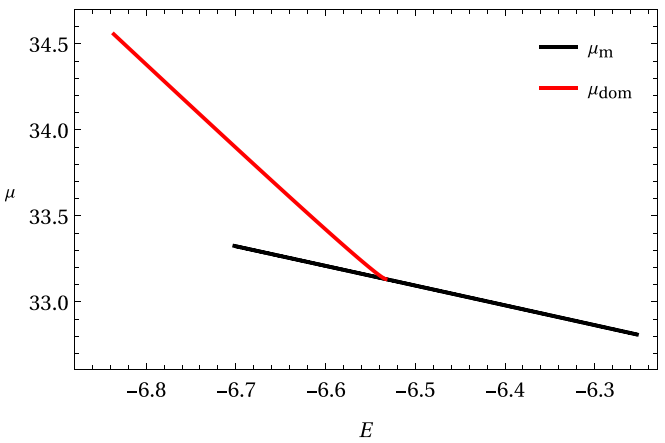

FIG. 4. The stability, or shift of the trace, for dominant and marginal optima in the nonlinear sum of squares problem for $\alpha = \frac{3}{2}$ and $f(q) = q^2 + q^3$.

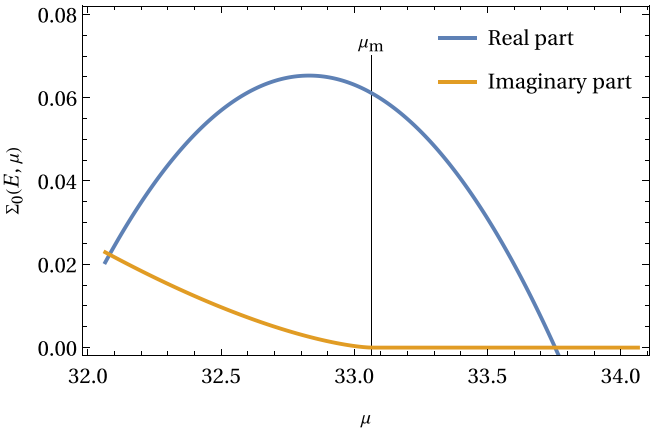

FIG. 5. Real and imaginary parts of the complexity $\Sigma_0(E, \mu)$ with fixed minimum eigenvalue $\lambda^* = 0$ as a function of μ in the nonlinear sum of squares problem with $\alpha = \frac{3}{2}$, $f(q) = q^2 + q^3$, and $E \simeq -6.47$. The vertical line depicts the value of the marginal stability μ_m.

approach (32). The marginal stability corresponds to precisely the point at which an imaginary part develops in the complexity. This demonstrates that the principles we used to determine the marginal stability continue to hold even in non-Gaussian cases where the complexity and the condition to fix the minimum eigenvalue are tangled together.

In a related paper, we use a sum of squared random functions model to explore the relationship between the marginal complexity and the performance of two generic algorithms: gradient descent and approximate message passing [21]. We show that the range of energies where the marginal complexity is positive does effectively bound the performance of these algorithms. At the moment the comparison is restricted to models with small polynomial powers appearing in $f(q)$ and with small α for computational reasons. However, using the dynamical mean-field theory results already found for these models it should be possible to make comparisons in a wider family of models [48,49].

The results for the marginal complexity are complimentary to rigorous results on the performance of algorithms in the least squares case, which focus on bounds for α and the parameters of f necessary for zero-energy solutions to exist and be found by algorithms [50,51]. After more work to evaluate the marginal complexity in the full RSB case, it will be interesting to compare the bounds implied by the distribution of marginal minima with those made by other means.

V. CONCLUSIONS

We have introduced a method for conditioning complexity on the marginality of stationary points. This method is general, and permits conditioning without first needing to understand the statistics of the Hessian at stationary points. We used our approach to study marginal complexity in three different models of random landscapes, showing that the method works and can be applied to models whose marginal complexity was not previously known. In related work, we further show that marginal complexity in the third model of

sums of squared random functions can be used to effectively bound algorithmic performance [21].

There are some limitations to the approach we relied on in this paper. The main limitation is our restriction to signalless landscapes, where there is no symmetry-breaking favored direction. This allowed us to treat stationary points with isolated eigenvalues as atypical and therefore find the marginal stability $\mu_{\rm m}$ using a variational principle. However, most models of interest in inference have a nonzero signal strength and therefore often have typical stationary points with an isolated eigenvalue. As we described, marginal complexity can still be analyzed in these systems by tuning the shift μ until the large-deviation principle breaks down and an imaginary part of the complexity appears. However, this is an inconvenient approach. It is possible that a variational approach can be preserved by treating the direction toward and the directions orthogonal to the signal differently. This problem merits further research.

Finally, the problem of predicting which marginal minima are able to attract some dynamics and which cannot attract any dynamics looms large over this work. As we discussed briefly at the end of Sec. IV B, in some simple contexts it is easy to see why certain marginal minima are not viable, but at the moment we do not know how to generalize this. Ideas related to the self-similarity and stochastic stability of minima have recently been suggested as a route to understanding this problem, but this approach is still in its infancy [52].

The title of our paper and that of Müller *et al.* suggest they address the same topic, but this is not the case [53]. That work differs in three important and fundamental ways. First, it describes minima of the Thouless, Anderson, and Palmer (TAP) free energy and involves peculiarities specific to the TAP. Second, it describes dominant minima which happen to be marginal, not a condition for finding subdominant marginal minima. Finally, it focuses on minima with a single soft direction (which are the typical minima of the low temperature Sherrington–Kirkpatrick TAP free energy), while we aim to avoid such minima in favor of ones that have a pseudogap (which we argue are relevant to out-of-equilibrium dynamics). The fact that the typical minima studied by Müller *et al.* are not marginal in this latter sense may provide an intuitive explanation for the seeming discrepancy between the proof that the low-energy Sherrington–Kirkpatrick model cannot be sampled [54] and the proof that a message-passing algorithm can find near-ground states [55]: the algorithm finds the atypical low-lying states that are marginal in the sense considered here but cannot find the typical ones considered by Müller *et al.*

ACKNOWLEDGMENTS

J.K.-D. was supported by a DynSysMath Specific Initiative of the INFN.

APPENDIX A: A PRIMER ON SUPERSPACE

In this appendix we review the algebra of superspace [56]. The superspace $\mathbb{R}^{N|2D}$ is a vector space with N real indices and $2D$ Grassmann indices $\bar\theta_1, \theta_1, \ldots, \bar\theta_D, \theta_D$. The Grassmann indices anticommute like fermions. Their integration is defined by

$$\int d\theta\, \theta = 1, \quad \int d\theta\, 1 = 0. \tag{A1}$$

Because the Grassmann indices anticommute, their square is always zero. Therefore, any series expansion of a function with respect to a given Grassmann index will terminate exactly at linear order, while a series expansion with respect to n Grassmann variables will terminate exactly at nth order. If f is an arbitrary superspace function, then the integral of f with respect to a Grassmann index can be evaluated using this property of the series expansion by

$$\int d\theta\, f(a+b\theta) = \int d\theta\, [f(a) + f'(a) b\theta] = f'(a) b. \tag{A2}$$

This kind of behavior of integrals over the Grassmann indices makes them useful for compactly expressing the Kac–Rice measure. To see why, consider the specific superspace $\mathbb{R}^{N|2}$, where an arbitrary vector can be expressed as

$$\boldsymbol{\phi}(1) = \mathbf{x} + \bar\theta_1 \boldsymbol{\eta} + \bar{\boldsymbol{\eta}}\theta_1 + \bar\theta_1\theta_1 i\hat{\mathbf{x}}, \tag{A3}$$

where $\mathbf{x}, \hat{\mathbf{x}} \in \mathbb{R}^N$ and $\bar{\boldsymbol{\eta}}, \boldsymbol{\eta}$ are N-dimensional Grassmann vectors. The dependence of $\boldsymbol{\phi}$ on 1 indicates the index of Grassmann variables $\bar\theta_1, \theta_1$ inside, since we will sometimes want to use, e.g., $\boldsymbol{\phi}(2)$ defined identically save for substitution by $\bar\theta_2, \theta_2$. Consider the series expansion of an arbitrary function f of this supervector:

$$\begin{aligned} f(\boldsymbol{\phi}(1)) &= f(\mathbf{x}) + (\bar\theta_1\boldsymbol{\eta} + \bar{\boldsymbol{\eta}}\theta_1 + \bar\theta_1\theta_1 i\hat{\mathbf{x}})^T \partial f(\mathbf{x}) \\ &\quad + \tfrac{1}{2}(\bar\theta_1\boldsymbol{\eta} + \bar{\boldsymbol{\eta}}\theta_1)^T \partial\partial f(\mathbf{x})(\bar\theta_1\boldsymbol{\eta} + \bar{\boldsymbol{\eta}}\theta_1) \\ &= f(\mathbf{x}) + (\bar\theta_1\boldsymbol{\eta} + \bar{\boldsymbol{\eta}}\theta_1 + \bar\theta_1\theta_1 i\hat{\mathbf{x}})^T \partial f(\mathbf{x}) \\ &\quad - \bar\theta_1\theta_1 \bar{\boldsymbol{\eta}}^T \partial\partial f(\mathbf{x})\boldsymbol{\eta}, \end{aligned} \tag{A4}$$

where the last step we used the fact that the Hessian matrix is symmetric and that squares of Grassmann indicies vanish. Using the integration rules defined above, we find

$$\int d\theta_1\, d\bar\theta_1\, f(\boldsymbol{\phi}(1)) = i\hat{\mathbf{x}}^T \partial f(\mathbf{x}) - \bar{\boldsymbol{\eta}}^T \partial\partial f(\mathbf{x})\boldsymbol{\eta}. \tag{A5}$$

These two terms are precisely the exponential representation of the Dirac δ function of the gradient and determinant of the Hessian (without absolute value sign) that make up the basic Kac–Rice measure, so that we can write

$$\begin{aligned} &\int d\mathbf{x}\, \delta(\nabla H(\mathbf{x}))\, \det \mathrm{Hess}\, H(\mathbf{x}) \\ &\quad = \int d\mathbf{x}\, d\bar{\boldsymbol{\eta}}\, d\boldsymbol{\eta}\, \frac{d\hat{\mathbf{x}}}{(2\pi)^N}\, e^{i\hat{\mathbf{x}}^T \nabla H(\mathbf{x}) - \bar{\boldsymbol{\eta}}^T \mathrm{Hess}\, H(\mathbf{x})\boldsymbol{\eta}} \\ &\quad = \int d\boldsymbol{\phi}\, e^{\int d1\, H(\boldsymbol{\phi}(1))}, \end{aligned} \tag{A6}$$

where we have written the measures $d1 = d\theta_1\, d\bar\theta_1$ and $d\boldsymbol{\phi} = d\mathbf{x}\, d\bar{\boldsymbol{\eta}}\, d\boldsymbol{\eta}\, \frac{d\hat{\mathbf{x}}}{(2\pi)^N}$. Besides some deep connections to the physics of BRST, this compact notation dramatically simplifies the analytical treatment of the problem. The energy of stationary points can also be fixed using this notation by writing

$$\int d\boldsymbol{\phi}\, d\hat\beta\, e^{\hat\beta E + \int d1\, (1 - \hat\beta\bar\theta_1\theta_1) H(\boldsymbol{\phi}(1))}, \tag{A7}$$

which a small calculation confirms results in the same expression as (35).

The reason why this transformation is a simplification is because there are a large variety of superspace algebraic and integral operations with direct corollaries to their ordinary real counterparts. For instance, consider a super linear operator $M(1,2)$, which like the super vector $\boldsymbol{\phi}$ is made up of a linear combination of $N \times N$ regular or Grassmann matrices indexed by every nonvanishing combination of the Grassmann indices $\bar{\theta}_1$, θ_1, $\bar{\theta}_2$, θ_2. Such a supermatrix acts on supervectors by ordinary matrix multiplication and convolution in the Grassmann indices, i.e.,

$$(M\boldsymbol{\phi})(1) = \int d2\, M(1,2)\boldsymbol{\phi}(2). \tag{A8}$$

The identity supermatrix is given by

$$\delta(1,2) = \big(\bar{\theta}_1 - \bar{\theta}_2\big)(\theta_1 - \theta_2)I. \tag{A9}$$

Integrals involving superfields contracted into such operators result in schematically familiar expressions, like that of the standard Gaussian:

$$\int d\boldsymbol{\phi} e^{-\frac{1}{2}\int d1d2\boldsymbol{\phi}(1)^T M(1,2)\boldsymbol{\phi}(2)} = (\mathrm{sdet}\, M)^{-1/2}, \tag{A10}$$

where the usual role of the determinant is replaced by the superdeterminant. The superdeterminant can be defined using the ordinary determinant by writing a block version of the matrix M. If $\mathbf{e}(1) = \{1, \bar{\theta}_1\theta_1\}$ is the basis vector of the even subspace of the superspace and $\mathbf{f}(1) = \{\bar{\theta}_1, \theta_1\}$ is that of the odd subspace, dual bases $\mathbf{e}^\dagger(1) = \{\bar{\theta}_1\theta_1, 1\}$ and $\mathbf{f}^\dagger(1) = \{-\theta_1, \bar{\theta}_1\}$ can be defined by the requirement that

$$\int d1\, e_i^\dagger(1)e_j(1) = \delta_{ij}, \quad \int d1\, f_i^\dagger(1)f_j(1) = \delta_{ij}, \tag{A11}$$

$$\int d1 e_i^\dagger(1)f_j(1) = 0, \quad \int d1 f_i^\dagger(1)e_j(1) = 0. \tag{A12}$$

With such bases and dual bases defined, we can form a block representation of M in analogy with the matrix form of an operator in quantum mechanics by

$$\int d1d2 \begin{bmatrix} \mathbf{e}^\dagger(1)M(1,2)\mathbf{e}(2) & \mathbf{e}^\dagger(1)M(1,2)\mathbf{f}(2) \\ \mathbf{f}^\dagger(1)M(1,2)\mathbf{e}(2) & \mathbf{f}^\dagger(1)M(1,2)\mathbf{f}(2) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{A13}$$

where each of the blocks is a $2N \times 2N$ real matrix. Then the superdeterminant of M is given by

$$\mathrm{sdet}\, M = \det(A - BD^{-1}C)\det(D)^{-1}, \tag{A14}$$

which is the same as the normal expression for the determinant of a block matrix save for the inverse of $\det D$. Likewise, the supertrace of M is is given by

$$\mathrm{sTr}\, M = \mathrm{Tr}\, A - \mathrm{Tr}\, D. \tag{A15}$$

The same method can be used to calculate the superdeterminant and supertrace in arbitrary superspaces, where for $\mathbb{R}^{N|2D}$ each basis has 2^{2D-1} elements. For instance, for $\mathbb{R}^{N|4}$ we have

$$\begin{aligned} \mathbf{e}(1,2) &= \{1, \bar{\theta}_1\theta_1, \bar{\theta}_2\theta_2, \bar{\theta}_1\theta_2, \bar{\theta}_2\theta_1, \bar{\theta}_1\bar{\theta}_2, \theta_1\theta_2, \bar{\theta}_1\theta_1\bar{\theta}_2\theta_2\}, \\ \mathbf{f}(1,2) &= \{\bar{\theta}_1, \theta_1, \bar{\theta}_2, \theta_2, \bar{\theta}_1\theta_1\bar{\theta}_2, \bar{\theta}_2\theta_2\theta_1, \bar{\theta}_1\theta_1\theta_2, \bar{\theta}_2\theta_2\theta_1\}, \end{aligned} \tag{A16}$$

with the dual bases defined analogously to those above.

APPENDIX B: BECCHI-ROUET-STORA-TYUTIN SYMMETRY

When the trace μ is not fixed, there is an unusual symmetry in the dominant complexity of minima [30–32]. This arises from considering the Kac–Rice formula as a kind of gauge fixing procedure [57]. Around each stationary point consider making the coordinate transformation $\mathbf{u} = \nabla H(\mathbf{x})$. Then, in the absence of fixing the trace of the Hessian to μ, the Kac–Rice measure becomes

$$\int d\nu(\mathbf{x}, \boldsymbol{\omega}|E) = \int \sum_\sigma d\mathbf{u}\delta(\mathbf{u})\delta(NE - H(\mathbf{x}_\sigma)), \tag{B1}$$

where the sum is over stationary points σ. This integral has a symmetry of its measure of the form $\mathbf{u} \mapsto \mathbf{u} + \delta\mathbf{u}$. Under the nonlinear transformation that connects $\mathbf{u}$ and $\mathbf{x}$, this implies a symmetry of the measure in the Kac–Rice integral of $\mathbf{x} \mapsto \mathbf{x} + (\mathrm{Hess}\, H)^{-1}\delta\mathbf{u}$. This symmetry, while exact, is nonlinear and difficult to work with.

When the absolute value function has been dropped and Grassmann vectors introduced to represent the determinant of the Hessian, this symmetry can be simplified considerably. Due to the expansion properties of Grassmann integrals, any appearance of $-\bar{\boldsymbol{\eta}}\boldsymbol{\eta}^T$ in the integrand resolves to $(\mathrm{Hess}\, H)^{-1}$. The symmetry of the measure can then be written

$$\mathbf{x} \mapsto \mathbf{x} - \bar{\boldsymbol{\eta}}\boldsymbol{\eta}^T\delta\mathbf{u} = \mathbf{x} + \bar{\boldsymbol{\eta}}\delta\boldsymbol{\epsilon}, \tag{B2}$$

where $\delta\boldsymbol{\epsilon} = -\boldsymbol{\eta}^T\delta\mathbf{u}$ is a Grassmann number. This establishes that $\delta\mathbf{x} = \bar{\boldsymbol{\eta}}\delta\boldsymbol{\epsilon}$, now linear. The rest of the transformation can be built by requiring that the action is invariant after expansion in $\delta\boldsymbol{\epsilon}$. This gives

$$\delta\mathbf{x} = \bar{\boldsymbol{\eta}}\delta\boldsymbol{\epsilon}, \quad \delta\hat{\mathbf{x}} = -i\hat{\beta}\bar{\boldsymbol{\eta}}\delta\boldsymbol{\epsilon}, \quad \delta\boldsymbol{\eta} = -i\hat{\mathbf{x}}\delta\boldsymbol{\epsilon}, \quad \delta\bar{\boldsymbol{\eta}} = 0, \tag{B3}$$

so that the differential form of the symmetry is

$$\mathcal{D} = \bar{\boldsymbol{\eta}} \cdot \frac{\partial}{\partial\mathbf{x}} - i\hat{\beta}\bar{\boldsymbol{\eta}} \cdot \frac{\partial}{\partial\hat{\mathbf{x}}} - i\hat{\mathbf{x}} \cdot \frac{\partial}{\partial\boldsymbol{\eta}}, \tag{B4}$$

The Ward identities associated with this symmetry give rise to relationships among the order parameters. These identities come from applying the differential symmetry to Grassmann-valued order parameters and are

$$\begin{aligned} 0 &= \frac{1}{N}\mathcal{D}\langle\mathbf{x}_a \cdot \boldsymbol{\eta}_b\rangle = \frac{1}{N}[\langle\bar{\boldsymbol{\eta}}_a \cdot \boldsymbol{\eta}_b\rangle - i\langle\mathbf{x}_a \cdot \hat{\mathbf{x}}_b\rangle] \\ &= G_{ab} + R_{ab}, \end{aligned} \tag{B5}$$

$$\begin{aligned} 0 &= \frac{i}{N}\mathcal{D}\langle\hat{\mathbf{x}}_a \cdot \boldsymbol{\eta}_b\rangle = \frac{1}{N}[\hat{\beta}\langle\bar{\boldsymbol{\eta}}_a \cdot \boldsymbol{\eta}_b\rangle + \langle\hat{\mathbf{x}}_a \cdot \hat{\mathbf{x}}_b\rangle] \\ &= \hat{\beta}G_{ab} + D_{ab}. \end{aligned} \tag{B6}$$

These identities establish $G_{ab} = -R_{ab}$ and $D_{ab} = \hat{\beta}R_{ab}$, allowing elimination of the matrices G and D in favor of R. Fixing the trace to μ explicitly breaks this symmetry, and the simplification is lost.

APPENDIX C: SPECTRAL DENSITY IN THE MULTISPHERICAL SPIN GLASS

In this Appendix we derive an expression for the asymptotic spectral density of the Hessian in the two-sphere multispherical spin glass that we describe in Sec. IV B. We use a typical approach of employing replicas to compute the resolvent [58]. The resolvent for the Hessian of the multispherical model is given by an integral over $\mathbf{y} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}] \in \mathbb{R}^{2N}$ as

$$\begin{aligned} G(\lambda) &= \lim_{n\to 0} \int \|\mathbf{y}_1\|^2 \prod_{a=1}^{n} d\mathbf{y}_a \, \exp\left\{-\frac{1}{2}\mathbf{y}_a^T(\operatorname{Hess} H(\mathbf{x},\boldsymbol{\omega}) - \lambda I)\mathbf{y}_a\right\} \\ &= \lim_{n\to 0} \int \left(\left\|\mathbf{y}_1^{(1)}\right\|^2 + \left\|\mathbf{y}_1^{(2)}\right\|^2\right) \prod_{a=1}^{n} d\mathbf{y}_a \, \exp\left\{-\frac{1}{2}\begin{bmatrix}\mathbf{y}_a^{(1)}\\ \mathbf{y}_a^{(2)}\end{bmatrix}^T \left(\begin{bmatrix}\partial_1\partial_1 H_1(\mathbf{x}^{(1)}) + \omega_1 I & -\epsilon I \\ -\epsilon I & \partial_2\partial_2 H_2(\mathbf{x}^{(2)}) + \omega_2 I\end{bmatrix} - \lambda I\right)\begin{bmatrix}\mathbf{y}_a^{(1)}\\ \mathbf{y}_a^{(2)}\end{bmatrix}\right\}. \end{aligned} \tag{C1}$$

If $Y_{ab}^{(ij)} = \frac{1}{N}\mathbf{y}_a^{(i)} \cdot \mathbf{y}_b^{(j)}$ is the matrix of overlaps of the vectors $\mathbf{y}$, then a short and standard calculation involving the average over H and the change of variables from $\mathbf{y}$ to Y yields

$$\overline{G(\lambda)} = N \lim_{n\to 0} \int dY \left(Y_{11}^{(11)} + Y_{11}^{(22)}\right) e^{nN\mathcal{S}(Y)}, \tag{C2}$$

where the effective action $\mathcal{S}$ is given by

$$\begin{aligned} \mathcal{S}(Y) = \lim_{n\to 0} \frac{1}{n}\Bigg\{ &\frac{1}{4}\sum_{ab}^{n}\left[f_1''(1)\left(Y_{ab}^{(11)}\right)^2 + f_2''(1)\left(Y_{ab}^{(22)}\right)^2\right] + \frac{1}{2}\sum_{a}^{n}\left[2\epsilon Y_{aa}^{(12)} + (\lambda-\omega_1)Y_{aa}^{(11)} + (\lambda-\omega_2)Y_{aa}^{(22)}\right] \\ &+ \frac{1}{2}\log\det\begin{bmatrix} Y^{(11)} Y^{(12)} \\ Y^{(12)} Y^{(22)}\end{bmatrix}\Bigg\}, \end{aligned} \tag{C3}$$

Making the replica symmetric ansatz $Y_{ab}^{(ij)} = y^{(ij)}\delta_{ab}$ for each of the matrices $Y^{(ij)}$ yields

$$\mathcal{S}(y) = \tfrac{1}{4}[f_1''(1)(y^{(11)})^2 + f_2''(1)(y^{(22)})^2] + \epsilon y^{(12)} + \tfrac{1}{2}[(\lambda-\omega_1)y^{(11)} + (\lambda-\omega_2)y^{(22)}] + \tfrac{1}{2}\log(y^{(11)}y^{(22)} - y^{(12)}y^{(12)}), \tag{C4}$$

while the average resolvent becomes

$$\overline{G(\lambda)} = N(y^{(11)} + y^{(22)}) \tag{C5}$$

for $y^{(11)}$ and $y^{(22)}$ evaluated at a saddle point of $\mathcal{S}$. The spectral density at large N is then given by the discontinuity in its imaginary point on the real axis, or

$$\rho(\lambda) = \frac{1}{2\pi i N}(\overline{G(\lambda + i0^+)} - \overline{G(\lambda + i0^-)}). \tag{C6}$$

APPENDIX D: COMPLEXITY OF DOMINANT OPTIMA FOR SUMS OF SQUARED RANDOM FUNCTIONS

Here we share an outline of the derivation of formulas for the complexity of dominant optima in sums of squared random functions of Sec. IV C. While in this paper we only treat problems with a replica symmetric structure, formulas for the effective action are generic to any RSB structure and provide a starting point for analyzing the challenging full RSB setting.

Using the $\mathbb{R}^{N|2}$ superfields

$$\boldsymbol{\phi}_a(1) = \mathbf{x}_a + \bar{\theta}_1\boldsymbol{\eta}_a + \bar{\boldsymbol{\eta}}_a\theta_1 + i\hat{\mathbf{x}}_a\bar{\theta}_1\theta_1, \tag{D1}$$

the replicated count of stationary points can be written

$$\mathcal{N}(E,\mu)^n = \int d\hat{\beta}\prod_{a=1}^{n} d\boldsymbol{\phi}_a \exp\left[N\hat{\beta}E - \frac{1}{2}\int d1\left(B(1)\sum_{k=1}^{M} V_k(\boldsymbol{\phi}_a(1))^2 - (\mu + f'(0))\|\boldsymbol{\phi}_a(1)\|^2\right)\right] \tag{D2}$$

for $B(1) = 1 - \hat{\beta}\bar{\theta}_1\theta_1$. The derivation of the complexity follows from here nearly identically to that in Appendix A.2 of Fyodorov and Tublin with superoperations replacing standard ones [44]. First we insert Dirac δ functions to fix each of the M energies $V_k(\boldsymbol{\phi}_a(1))$ as

$$\delta(V_k(\boldsymbol{\phi}_a(1)) - v_{ka}(1)) = \int d\hat{v}_{ka} \, \exp\left[i\int d1\, \hat{v}_{ka}(1)(V_k(\boldsymbol{\phi}_a(1)) - v_{ka}(1))\right]. \tag{D3}$$

The squared V_k appearing in the energy can now be replaced by the variables v_k, leaving the only remaining dependence on the disordered V in the contribution of (D3), which is linear. The average over the disorder can then be computed, which yields

$$\overline{\exp\left[i\sum_{k=1}^{M}\sum_{a=1}^{n}\int d1\,\hat{v}_{ka}(1)V_k(\boldsymbol{\phi}_a(1))\right]} = \exp\left[-\frac{1}{2}\sum_{k=1}^{M}\sum_{ab}^{n}\int d1\,d2\,\hat{v}_{ka}(1)f\left(\frac{\boldsymbol{\phi}_a(1)\cdot\boldsymbol{\phi}_b(2)}{N}\right)\hat{v}_{kb}(2)\right]. \tag{D4}$$

The result is factorized in the indices k and Gaussian in the superfields v and $\hat{v}$ with kernel

$$\begin{bmatrix} B(1)\delta_{ab}\delta(1,2) & i\delta_{ab}\delta(1,2) \\ i\delta_{ab}\delta(1,2) & f\left(\frac{\boldsymbol{\phi}_a(1)\cdot\boldsymbol{\phi}_b(2)}{N}\right) \end{bmatrix}. \tag{D5}$$

Making the M independent Gaussian integrals, we find

$$\mathcal{N}(E,\mu)^n = \int d\hat{\beta}\left(\prod_{a=1}^{n} d\boldsymbol{\phi}_a\right)\exp\left\{nN\hat{\beta}E + \frac{\mu+f'(0)}{2}\sum_{a}^{n}\int d1\,\|\boldsymbol{\phi}_a\|^2 - \frac{M}{2}\log\operatorname{sdet}\left[\delta_{ab}\delta(1,2) + B(1)f\left(\frac{\boldsymbol{\phi}_a(1)\cdot\boldsymbol{\phi}_b(2)}{N}\right)\right]\right\}. \tag{D6}$$

We make a change of variables from the fields $\boldsymbol{\phi}$ to matrices $\mathbb{Q}_{ab}(1,2) = \frac{1}{N}\boldsymbol{\phi}_a(1)\cdot\boldsymbol{\phi}_b(2)$. This transformation results in a change of measure of the form

$$\prod_{a=1}^{n} d\boldsymbol{\phi}_a = d\mathbb{Q}\,(\operatorname{sdet}\mathbb{Q})^{\frac{N}{2}} = d\mathbb{Q}\,\exp\left[\frac{N}{2}\log\operatorname{sdet}\mathbb{Q}\right]. \tag{D7}$$

We therefore have

$$\mathcal{N}(E,\mu)^n = \int d\hat{\beta}\,d\mathbb{Q}\,\exp\left\{nN\hat{\beta}E + N\frac{\mu+f'(0)}{2}\operatorname{sTr}\mathbb{Q} + \frac{N}{2}\log\operatorname{sdet}\mathbb{Q} - \frac{M}{2}\log\operatorname{sdet}\left[\delta_{ab}\delta(1,2) + B(1)f(\mathbb{Q}_{ab}(1,2))\right]\right\}. \tag{D8}$$

We now need to blow up our supermatrices into our physical order parameters. We have from the definition of ϕ and $\mathbb{Q}$ that

$$\mathbb{Q}_{ab}(1,2) = C_{ab} - G_{ab}(\bar{\theta}_1\theta_2 + \bar{\theta}_2\theta_1) - R_{ab}(\bar{\theta}_1\theta_1 + \bar{\theta}_2\theta_2) - D_{ab}\bar{\theta}_1\theta_2\bar{\theta}_2\theta_2, \tag{D9}$$

where C, R, D, and G are the matrices defined in (48). Other possible combinations involving scalar products between fermionic and bosonic variables do not contribute at physical saddle points [34]. Inserting this expansion into the expression above and evaluating the superdeterminants and supertrace, we find

$$\mathcal{N}(E,\mu)^n = \int d\hat{\beta}\,dC\,dR\,dD\,dG\,e^{nN\mathcal{S}_{\mathrm{KR}}(\hat{\beta},C,R,D,G)}, \tag{D10}$$

where the effective action is given by

$$\mathcal{S}_{\mathrm{KR}}(\hat{\beta},C,R,D,G) = \hat{\beta}E + \lim_{n\to 0}\frac{1}{n}\left(-(\mu+f'(0))\operatorname{Tr}(G+R) + \frac{1}{2}\log\det[G^{-2}(CD+R^2)] + \alpha\log\det[I+G\odot f'(C)] - \frac{\alpha}{2}\log\det[(f'(C)\odot D - \hat{\beta}I + (G^{\circ 2}-R^{\circ 2})\odot f''(C))f(C) + (I - R\odot f'(C))^2]\right), \tag{D11}$$

where $\odot$ gives the Hadamard or componentwise product between the matrices and $A^{\circ n}$ gives the Hadamard power of A, while other products and powers are matrix products and powers.

In the case where μ is not specified, we can make use of the BRST symmetry of Appendix B whose Ward identities give $D = \hat{\beta}R$ and $G = -R$. Using these relations, the effective action becomes particularly simple:

$$\mathcal{S}_{\mathrm{KR}}(\hat{\beta},C,R) = \hat{\beta}E + \frac{1}{2}\lim_{n\to 0}\frac{1}{n}(\log\det(I+\hat{\beta}CR^{-1}) - \alpha\log\det[I - \hat{\beta}f(C)(I - R\odot f'(C))^{-1}]). \tag{D12}$$

This effective action is general for arbitrary matrices C and R, and therefore arbitrary RSB order. When using a replica symmetric ansatz of $C_{ab} = \delta_{ab} + c_0(1-\delta_{ab})$ and $R_{ab} = r\delta_{ab} + r_0(1-\delta_{ab})$, the resulting function of $\hat{\beta}$, c_0, r, and r_0 is

$$\mathcal{S}_{\mathrm{KR}}(\hat{\beta},c_0,r,r_0) = \hat{\beta}E + \frac{1}{2}\left[\log\left(1+\frac{\hat{\beta}(1-c_0)}{r-r_0}\right) + \frac{\hat{\beta}c_0+r_0}{\hat{\beta}(1-c_0)+r-r_0} - \frac{r_0}{r-r_0}\right] - \frac{\alpha}{2}\left[\log\left(1 - \frac{\hat{\beta}(f(1)-f(c_0))}{1-rf'(1)+r_0f'(c_0)}\right) - \frac{\hat{\beta}f(c_0)+r_0f'(c_0)}{1-\hat{\beta}(f(1)-f(c_0))-rf'(1)+rf'(c_0)} + \frac{r_0f'(c_0)}{1-rf'(1)+r_0f'(c_0)}\right]. \tag{D13}$$

When $f(0) = 0$ as in the cases directly studied in this work, this further simplifies as $c_0 = r_0 = 0$. The effective action is then

$$\mathcal{S}_{\mathrm{KR}}(\hat{\beta}, r) = \hat{\beta}E + \frac{1}{2}\log\left(1 + \frac{\hat{\beta}}{r}\right) - \frac{\alpha}{2}\log\left(1 - \frac{\hat{\beta}f(1)}{1 - rf'(1)}\right). \tag{D14}$$

Extremizing this expression with respect to the order parameters $\hat{\beta}$ and r produces the red line of dominant minima shown in Fig. 3.

[1] V. Ros and Y. V. Fyodorov, The high-dimensional landscape paradigm: Spin-glasses, and beyond, in *Spin Glass Theory and Far Beyond* (World Scientific, New Jersey, London, Singapore, 2023), pp. 95–114.

[2] G. Biroli and J. P. Garrahan, Perspective: The glass transition, J. Chem. Phys. **138**, 12A301 (2013).

[3] F. Krzakala and J. Kurchan, Landscape analysis of constraint satisfaction problems, Phys. Rev. E **76**, 021122 (2007).

[4] G. Folena, S. Franz, and F. Ricci-Tersenghi, Rethinking mean-field glassy dynamics and its relation with the energy landscape: The surprising case of the spherical mixed p-spin model, Phys. Rev. X **10**, 031045 (2020).

[5] G. Folena and F. Zamponi, On weak ergodicity breaking in mean-field spin glasses, SciPost Phys. **15**, 109 (2023).

[6] A. El Alaoui and A. Montanari, Algorithmic thresholds in mean field spin glasses, arXiv:2009.11481.

[7] J. Kent-Dobias and J. Kurchan, How to count in hierarchical landscapes: A full solution to mean-field complexity, Phys. Rev. E **107**, 064111 (2023).

[8] M. Müller and M. Wyart, Marginal stability in structural, spin, and electron glasses, Annu. Rev. Condens. Matter Phys. **6**, 177 (2015).

[9] P. W. Anderson, Lectures on amorphous systems, in *Ill-Condensed Matter* (North-Holland Publishing Company, Amsterdam, New York, Oxford, 1984), pp. 159–261.

[10] H. J. Sommers and W. Dupont, Distribution of frozen fields in the mean-field theory of spin glasses, J. Phys. C **17**, 5785 (1984).

[11] G. Parisi, On the statistical properties of the large time zero temperature dynamics of the SK model, arXiv:cond-mat/9501045.

[12] H. Horner, Time dependent local field distribution and metastable states in the SK-spin-glass, Eur. Phys. J. B **60**, 413 (2007).

[13] S. Pankov, Low-temperature solution of the Sherrington-Kirkpatrick model, Phys. Rev. Lett. **96**, 197204 (2006).

[14] V. Erba, F. Behrens, F. Krzakala, and L. Zdeborová, Quenches in the Sherrington–Kirkpatrick model, J. Stat. Mech.: Theory Exp. (2024) 083302.

[15] A. L. Efros and B. I. Shklovskii, Coulomb interaction in disordered systems with localized electronic states, in *Electron–Electron Interactions in Disordered Systems*, Modern Problems in Condensed Matter Sciences, Vol. 10 (Elsevier, Amsterdam, Oxford, New York, 1985), pp. 409–482.

[16] B. I. Shklovskii, Half century of Efros-Shklovskii Coulomb gap. romance with Coulomb interaction and disorder, Low Temp. Phys. **50**, 1101 (2024).

[17] G. Folena and P. Urbani, Marginal stability of soft anharmonic mean field spin glasses, J. Stat. Mech.: Theory Exp. (2022) 053301.

[18] R. Monasson, Structural glass transition and the entropy of the metastable states, Phys. Rev. Lett. **75**, 2847 (1995).

[19] Y. V. Fyodorov, Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices, Phys. Rev. Lett. **92**, 240601 (2004).

[20] A. J. Bray and D. S. Dean, Statistics of critical points of Gaussian fields on large-dimensional spaces, Phys. Rev. Lett. **98**, 150201 (2007).

[21] J. Kent-Dobias, Algorithm-independent bounds on complex optimization through the statistics of marginal optima, arXiv:2407.02092 [cond-mat.dis-nn].

[22] H. Ikeda, Bose–Einstein-like condensation of deformed random matrix: a replica approach, J. Stat. Mech.: Theory Exp. (2023) 023302.

[23] J. Kent-Dobias, Arrangement of nearby minima and saddles in the mixed spherical energy landscapes, SciPost Phys. **16**, 001 (2024).

[24] Y. V. Fyodorov and P. Le Doussal, Topology trivialization and large deviations for the minimum in the simplest random optimization, J. Stat. Phys. **154**, 466 (2014).

[25] Y. V. Fyodorov and P. Le Doussal, Hessian spectrum at the global minimum of high-dimensional random landscapes, J. Phys. A: Math. Theor. **51**, 474002 (2018).

[26] D. S. Dean and S. N. Majumdar, Large deviations of extreme eigenvalues of random matrices, Phys. Rev. Lett. **97**, 160201 (2006).

[27] J. Kent-Dobias, On the topology of solutions to random continuous constraint satisfaction problems, arXiv:2409.12781 [cond-mat.dis-nn].

[28] M. Kac, On the average number of real roots of a random algebraic equation, Bull. Am. Math. Soc. **49**, 314 (1943).

[29] S. O. Rice, The distribution of the maxima of a random curve, Am. J. Math. **61**, 409 (1939).

[30] A. Annibale, A. Cavagna, I. Giardina, and G. Parisi, Supersymmetric complexity in the Sherrington-Kirkpatrick model, Phys. Rev. E **68**, 061103 (2003).

[31] A. Annibale, A. Cavagna, I. Giardina, G. Parisi, and E. Trevigne, The role of the Becchi–Rouet–Stora–Tyutin supersymmetry in the calculation of the complexity for the Sherrington–Kirkpatrick model, J. Phys. A: Math. Gen. **36**, 10937 (2003).

[32] A. Annibale, G. Gualdi, and A. Cavagna, Coexistence of supersymmetric and supersymmetry-breaking states in spherical spin-glasses, J. Phys. A: Math. Gen. **37**, 11311 (2004).

[33] A. Crisanti, H. Horner, and H.-J. Sommers, The spherical p-spin interaction spin-glass model, Z. Phys. B: Condens. Matter **92**, 257 (1993).

[34] J. Kurchan, Supersymmetry in spin glass dynamics, J. Phys. I **2**, 1333 (1992).

[35] E. Subag, TAP approach for multi-species spherical spin glasses I: General theory, arXiv:2111.07132 [math.PR].
[36] E. Subag, TAP approach for multispecies spherical spin glasses II: The free energy of the pure models, Ann. Probab. **51**, 1004 (2023).
[37] E. Bates and Y. Sohn, Crisanti–Sommers formula and simultaneous symmetry breaking in multi-species spherical spin glasses, Commun. Math. Phys. **394**, 1101 (2022).
[38] E. Bates and Y. Sohn, Free energy in multi-species mixed p-spin spherical models, Electron. J. Probab. **27**, 1 (2022).
[39] B. Huang and M. Sellke, Strong topological trivialization of multi-species spherical spin glasses, arXiv:2308.09677 [math.PR].
[40] B. Huang and M. Sellke, Algorithmic threshold for multi-species spherical spin glasses, arXiv:2303.12172 [math.PR].
[41] B. Huang and M. Sellke, Optimization algorithms for multi-species spherical spin glasses, J. Stat. Phys. **191**, 29 (2024).
[42] Y. V. Fyodorov, A spin glass model for reconstructing nonlinearly encrypted signals corrupted by noise, J. Stat. Phys. **175**, 789 (2019).
[43] Y. V. Fyodorov and R. Tublin, Counting stationary points of the loss function in the simplest constrained least-square optimization, Acta Phys. Pol. B **51**, 1663 (2020).
[44] Y. V. Fyodorov and R. Tublin, Optimization landscape in the simplest constrained random least-square problem, J. Phys. A: Math. Theor. **55**, 244008 (2022).
[45] P. Vivo, Random linear systems with quadratic constraints: from random matrix theory to replicas and back, arXiv:2401.03209 [cond-mat.stat-mech].
[46] R. Tublin, A few results in random matrix theory and random optimization, Ph.D. thesis, King's College London (2022).
[47] P. Urbani, A continuous constraint satisfaction problem for the rigidity transition in confluent tissues, J. Phys. A: Math. Theor. **56**, 115003 (2023).
[48] P. J. Kamali and P. Urbani, Dynamical mean field theory for models of confluent tissues and beyond, SciPost Phys. **15**, 219 (2023).
[49] P. J. Kamali and P. Urbani, Stochastic gradient descent outperforms gradient descent in recovering a high-dimensional signal in a glassy energy landscape, arXiv:2309.04788 [cs.LG].
[50] A. Montanari and E. Subag, Solving overparametrized systems of random equations: I. model and algorithms for approximate solutions, arXiv:2306.13326 [math.PR].
[51] A. Montanari and E. Subag, On Smale's 17th problem over the reals, arXiv:2405.01735 [cs.DS].
[52] P. Urbani, Statistical physics of complex systems: glasses, spin glasses, continuous constraint satisfaction problems, high-dimensional inference and neural networks, arXiv:2405.06384 [cond-mat.dis-nn].
[53] M. Müller, L. Leuzzi, and A. Crisanti, Marginal states in mean-field glasses, Phys. Rev. B **74**, 134431 (2006).
[54] A. El Alaoui, A. Montanari, and M. Sellke, Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Los Alamitos, California, Washington, 2022), pp. 323–334.
[55] A. Montanari, Optimization of the Sherrington–Kirkpatrick Hamiltonian, SIAM J. Comput. FOCS19 (2021).
[56] B. S. DeWitt, *Supermanifolds*, 2nd ed., Cambridge monographs on mathematical physics (Cambridge University Press, Cambridge, New York, 1992).
[57] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford University Press, Oxford, New York, 2002).
[58] G. Livan, M. Novaes, and P. Vivo, *Introduction to Random Matrices*, SpringerBriefs in Mathematical Physics, Vol. 26 (Springer International Publishing, Cham, Switzerland, 2018).