The Landscape of the Spiked Tensor Model

GÉRARD BEN AROUS
Courant Institute

SONG MEI
Institute for Computational and Mathematical Engineering, Stanford University

ANDREA MONTANARI
Department of Electrical Engineering and Statistics, Stanford University

AND

MIHAI NICA
University of Toronto

Abstract

We consider the problem of estimating a large rank-one tensor $\boldsymbol{u}^{\otimes k} \in (\mathbb{R}^n)^{\otimes k}$, $k \geq 3$, in Gaussian noise. Earlier work characterized a critical signal-to-noise ratio $\lambda_{\text{Bayes}} = O(1)$ above which an ideal estimator achieves strictly positive correlation with the unknown vector of interest. Remarkably, no polynomial-time algorithm is known that achieved this goal unless $\lambda \geq C n^{(k-2)/4}$, and even powerful semidefinite programming relaxations appear to fail for $1 \ll \lambda \ll n^{(k-2)/4}$.

In order to elucidate this behavior, we consider the maximum likelihood estimator, which requires maximizing a degree-k homogeneous polynomial over the unit sphere in n dimensions. We compute the expected number of critical points and local maxima of this objective function and show that it is exponential in the dimensions n, and give exact formulas for the exponential growth rate. We show that (for λ larger than a constant) critical points are either very close to the unknown vector $\boldsymbol{u}$ or are confined in a band of width $\Theta(\lambda^{-1/(k-1)})$ around the maximum circle that is orthogonal to $\boldsymbol{u}$. For local maxima, this band shrinks to be of size $\Theta(\lambda^{-1/(k-2)})$. These "uninformative" local maxima are likely to cause the failure of optimization algorithms. © 2019 Wiley Periodicals, Inc.

Contents

1. Introduction 2283
2. Main Results 2286
3. Related Literature 2293
4. Proofs 2294
Appendix A. Technical Lemmas 2323

Communications on Pure and Applied Mathematics, Vol. LXXII, 2282–2330 (2019)

Appendix B. Derivation of an Explicit Formula for $S_\star$ 2326
Bibliography 2328

1 Introduction

Nonconvex formulations are the most popular approach for a number of problems across high-dimensional statistics and machine learning. Over the last few years, substantial effort has been devoted to establishing rigorous guarantees for these methods in the context of important applications. A small subset of examples include matrix completion [17, 23], phase retrieval [12, 41], high-dimensional regression with missing data [29], and two-layer neural networks [22, 44]. The general picture that emerges from theses studies, as formalized in [31], is that nonconvex losses can sometimes be "benign" and allow for nearly optimal statistical estimation using gradient-descent-type optimization algorithms. Roughly speaking, this happens when the population risk does not have flat regions, i.e., regions in which the gradient is small and the Hessian is nearly rank deficient.

In this paper we explore the flipside of this picture, namely what happens when the population risk has large "flat regions." We focus on a simple problem, tensor principal component analysis under the spiked tensor model, and show that the empirical risk can easily become extremely complex in these cases. This picture matches recent computational complexity results on the same model.

The spiked tensor model [33] captures, in a highly simplified fashion, a number of statistical estimation tasks in which we need to extract information from a noisy high-dimensional data tensor; see, e.g., [24, 27, 28, 35]. We are given a tensor $\boldsymbol{Y} \in (\mathbb{R}^n)^{\otimes k}$ of the form

$$\boldsymbol{Y} = \lambda \boldsymbol{u}^{\otimes k} + \frac{1}{\sqrt{2n}} \boldsymbol{W}\,, \tag{1.1}$$

where $\boldsymbol{W}$ is a noise tensor, and would like to estimate the unit vector $\boldsymbol{u} \in \mathbb{S}^{n-1}$. The parameter $\lambda \geq 0$ corresponds to the signal-to-noise ratio. The noise tensor $\boldsymbol{W} \in (\mathbb{R}^n)^{\otimes k}$ is distributed as

$$\boldsymbol{W} \stackrel{\mathrm{d}}{=} \sum_{\pi \in \mathfrak{S}_n} \boldsymbol{G}^{\pi}/(k!) \quad \text{where } \{G_{i_1 \cdots i_k}\}_{1 \leq i_1, \ldots, i_k \leq n} \stackrel{i.i.d.}{\sim} \mathsf{N}(0,1),$$

$\mathfrak{S}_n$ are permutations of the set $[n]$, and $(\boldsymbol{G}^{\pi})_{i_1 \cdots i_k} = G_{\pi(i_1) \cdots \pi(i_k)}$. Throughout the paper $k \geq 3$.

We say that the weak recovery problem is solvable for this model if there exists an estimator (a measurable function) $\hat{\boldsymbol{u}}$: $(\mathbb{R}^n)^{\otimes k} \to \mathbb{S}^{n-1}$ such that

$$\liminf_{n \to \infty} \mathbb{E}|\langle \hat{\boldsymbol{u}}(\boldsymbol{Y}), \boldsymbol{u} \rangle| \geq \varepsilon \tag{1.2}$$

for some $\varepsilon > 0$. It was proven in [33] that weak recovery is solvable provided $\lambda \geq \lambda_1(k)$, and in [32] that it is unsolvable for $\lambda < \lambda_0(k)$ for some constant

$0 < \lambda_0(k) < \lambda_1(k) < \infty$. In fact, for $\lambda < \lambda_0(k)$, it is altogether impossible to distinguish between the distribution (1.1) and the null model $\lambda = 0$. A sharp theshold $\lambda_{\text{Bayes}}(k)$ for the weak recovery problem was established in [26] (see also [7] for related results), and better lower bounds for the hypothesis-testing problem were proved in [36].

In light of these contributions, it is fair to say that optimal statistical estimation for the model (1.1) is well understood. In contrast, many questions are still open for what concerns computationally efficient procedures. Consider the maximum-likelihood estimator, which requires solving

$$\begin{cases} \text{maximize} & f(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes k} \rangle, \\ \text{subject to} & \boldsymbol{\sigma} \in \mathbb{S}^{n-1}. \end{cases} \tag{1.3}$$

It was shown in [33] that the maximum-likelihood estimator achieves weak recovery (cf. equation (1.2)) provided that $\lambda > \lambda_{\text{ML}}(k)$ for some constant $\lambda_{\text{ML}}(k) < \infty$.[1] However, solving the problem (1.3) (maximizing a homogeneous degree-k polynomial over the unit sphere) is NP-hard for all $k \geq 3$ [9].

Note that the population risk associated to the problem (1.3) is

$$f_0(\boldsymbol{\sigma}) \equiv \mathbb{E}\langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes k} \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^k. \tag{1.4}$$

For $k \geq 3$, the (Riemannian) gradient and Hessian of $f_0(\boldsymbol{\sigma})$ vanishes on the hyperplane orthogonal to $\boldsymbol{u}$: $\{\boldsymbol{\sigma} \in \mathbb{S}^{n-1}: \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle = 0\}$. In the intuitive language used above, the population risk has a large flat region. Since most of the volume of the sphere concentrates around this hyperplane [25], this is expected to have a dramatic impact on the optimization problem (1.3).

Polynomial-time computable estimators have been studied in a number of papers. In particular, [33] considers a spectral algorithm based on tensor unfolding and proved that it succeeds for k even provided $\lambda \geq C\, n^{(k-2)/4}$. (Here and below, we denote by C a constant that might depend on k but is independent of n.) This result was generalized in [21] to arbitrary $k \geq 3$ by using a sophisticated semidefinite programming relaxation from the sum-of-squares hierarchy. A lower-complexity spectral algorithm that succeeds under the same condition was developed in [20], and further results can be found in [2, 10]. However, no polynomial-time algorithm is known that achieves weak recovery for $1 \ll \lambda \ll n^{(k-2)/4}$, and it is possible that statistical estimation in the spiked tensor model is hard in this regime.

A large gap between known polynomial-time algorithms and statistical limits arises in the tensor completion problem, which shares many similarities with the spiked tensor model [16, 34, 43]. In the setting of tensor completion, hardness under Feige's hypothesis was proven in [6] for a certain regime of the number of observed entries.

[1] Indeed, an exact characterization of $\lambda_{\text{ML}}(k)$ should be possible using the "one-step replica symmetry breaking" formula proven in [42]. A nonrigorous analysis of the implications of this formula was carried out in [18], yielding $\lambda_{\text{ML}}(k) = \lambda_{\text{Bayes}}(k)$.

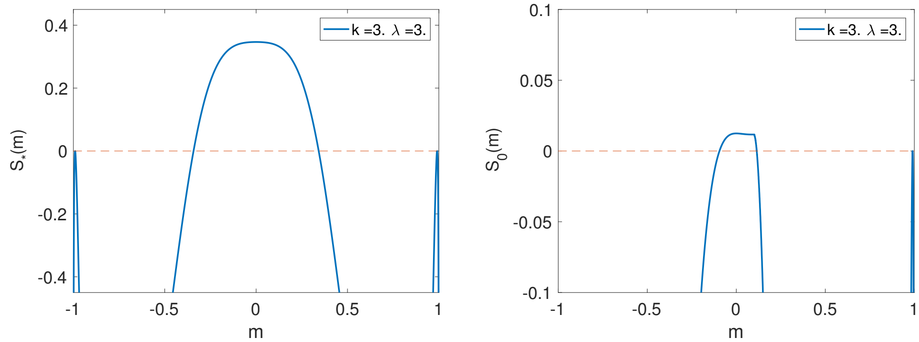

FIGURE 1.1. Complexity of the spiked tensor model of order $k = 3$ at signal-to-noise ratio $\lambda = 3$: exponential growth rate of the number of critical points $\boldsymbol{\sigma} \in \mathbb{S}^{n-1}$ as a function of the scalar product $m = \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle$. Left: complexity for the total number of critical points $S_\star(m)$. Right: complexity for local maxima $S_0(m)$.

Here we reconsider the maximum-likelihood estimator, and we explore the landscape of the optimization problem (1.3). In what regime is it hard to maximize the function $f(\cdot)$ for a typical realization of the random tensor $\boldsymbol{Y}$? In [33] a power iteration algorithm was studied that attempts to compute the maximum-likelihood estimator, and it was proven that it is successful for $\lambda \geq C n^{(k-2)/2}$. What is the origin of this threshold at $n^{(k-2)/2}$? In this paper we compute the expected number of critical points of the likelihood function $f(\boldsymbol{\sigma})$ to the leading exponential order.

Let us summarize the qualitative picture that emerges from our results. For clarity of exposition, we summarize only our results on local maxima, but similar results will be presented about generic critical points.

The expected number of local maxima grows exponentially with the dimension n. We compute the exponential growth rate, denoted by $S_0(m, x)$, as a function of the value of the cost function $x = f(\boldsymbol{\sigma})$ and of the scalar product $m = \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle$. Namely, the expected number of local maxima with $f(\boldsymbol{\sigma}) \approx x$ and $\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \approx m$ is $\exp\{n S_0(m, x) + o(n)\}$, with $S_0(m, x)$ given explicitly below. The exponent $S_0(m, x)$ and its variants $S_0(m)$, $S_\star(m, x)$, and so on, are referred to as "complexity" functions. In Figure 1.1 we plot $S_0(m) = \max_x S_0(m, x)$, which is the exponential growth rate of the number of local maxima with scalar product $\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \approx m$ for the case $k = 3$, $\lambda = 3$. (We also plot the analogous quantity for general critical points, $S_\star(m)$.)

The expected number of local maxima with scalar product $m = \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \approx 0$, i.e., lying close to the space orthogonal to the unknown vector $\boldsymbol{u}$ is exponentially large. The complexity function $S_0(m)$ decreases as $|m|$ increases, i.e., as we move away from this plane, and eventually vanishes.

For λ sufficiently large (in particular, for $\lambda > \lambda_c(k)$ given explicitly in Section 2.4), the complexity $S_0(m)$ reveals an interesting structure. It is positive in an interval $m \in (m_1(\lambda,k), m_2(\lambda,k))$, where $m_1(\lambda,k), m_2(\lambda,k) = \Theta(\lambda^{-1/(k-2)})$, and becomes nonpositive outside this interval. However, it increases again and touches 0 for $m = m_*(\lambda,k)$ close to one (for k even it also becomes 0 for $m = -m_*(\lambda,k)$ by symmetry). In other words, all the local maxima are either very close to the unknown vector $\boldsymbol{u}$ (and to the global maximum) or they are on a narrow spherical annulus orthogonal to $\boldsymbol{u}$.

It is interesting to discuss the behavior of local ascent optimization algorithms in such a landscape. While at this point the discussion is necessarily heuristic, it points at some interesting directions for future work. The expected exponential number of local maxima in the annulus $|\langle \boldsymbol{u}, \boldsymbol{\sigma}\rangle| \leq \Theta(\lambda^{-1/(k-2)})$ suggests that algorithms can converge to a local maximum that is well correlated with $\boldsymbol{u}$ only if they are initialized outside that annulus. In other words, the initialization $\boldsymbol{\sigma}_0$ must be such that $\langle \boldsymbol{u}, \boldsymbol{\sigma}_0\rangle \geq C\lambda^{-1/(k-2)}$. If no side information is available on $\boldsymbol{u}$, a random initialization will be used. This achieves $\langle \boldsymbol{u}, \boldsymbol{\sigma}_0\rangle = \Theta(n^{-1/2})$ with positive probability, and hence will escape local maxima provided $\lambda \geq Cn^{(k-2)/2}$. Remarkably, this is the same scaling as the threshold for power iteration obtained in [33]. It would be interesting to make rigorous this connection.

Let us emphasize that our results only concern the *expected number* of critical points. As is customary with random variables that fluctuate on the exponential scale, this is not necessarily close to the typical number of critical points. While we expect that several qualitative features found in this work will hold when considering the typical number, a rigorous justification is still open (see Section 3 for further discussion of this point).

The rest of the paper is organized as follows. We state formally our main results in Section 2, which also sketches the main ideas of the proofs. We will then review earlier literature in Section 3 and present proofs in Section 4.

2 Main Results

Our main results concern the number of critical points and the number of local maxima of the function $f(\boldsymbol{\sigma})$ introduced in equation (1.3), where $\boldsymbol{Y} \in (\mathbb{R}^n)^{\otimes k}$ is distributed as per equation (1.1).

Throughout, we denote by $\nabla f(\boldsymbol{\sigma})$ and $\nabla^2 f(\boldsymbol{\sigma})$ the euclidean gradient and Hessian of f at $\boldsymbol{\sigma}$, respectively, and denote by $\operatorname{grad} f(\boldsymbol{\sigma})$ and $\operatorname{Hess} f(\boldsymbol{\sigma})$ the Riemannian gradient and Hessian of f at $\boldsymbol{\sigma}$ on the unit sphere $\mathbb{S}^{n-1}$. Conceptually, the Riemannian gradient $\operatorname{grad} f(\boldsymbol{\sigma})$ is the projection of the euclidean gradient $\nabla f(\boldsymbol{\sigma})$ onto the tangent space of the sphere at $\boldsymbol{\sigma}$, and the Riemannian Hessian $\operatorname{Hess} f(\boldsymbol{\sigma})$ captures the second-order behavior of function f at $\boldsymbol{\sigma}$ on the sphere.

The completed real line is denoted by $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$. For a set $S \subseteq \mathbb{R}$, we denote by $\overline{S}$ its closure and by S^o its interior.

2.1 Complexity of Critical Points

For any Borel sets $E \subset \mathbb{R}$ and $M \in [-1, 1]$, we define $\mathrm{Crt}_{n,\star}(M, E)$ to be the number of critical points of f with function value in E and correlation in M:

$$\mathrm{Crt}_{n,\star}(M, E) := \sum_{\boldsymbol{\sigma} : \operatorname{grad} f(\boldsymbol{\sigma}) = \mathbf{0}} \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \in M\} \mathbf{1}\{f(\boldsymbol{\sigma}) \in E\}. \tag{2.1}$$

We define function $S_\star\colon [-1, 1] \times \mathbb{R} \to \overline{\mathbb{R}}$ as

$$\begin{aligned} S_\star(m, x) := {} & \frac{1}{2}(\log(k-1) + 1) + \frac{1}{2}\log(1 - m^2) - k\lambda^2 m^{2k-2}(1 - m^2) \\ & - (x - \lambda m^k)^2 + \Phi_\star\left(\sqrt{\frac{2k}{k-1}}x\right), \end{aligned} \tag{2.2}$$

where

$$\Phi_\star(x) = \begin{cases} \frac{x^2}{4} - \frac{1}{2}, & |x| \le 2, \\ \frac{x^2}{4} - \frac{1}{2} - \frac{|x|}{4} \cdot \sqrt{x^2 - 4} + \log\left\{\sqrt{\frac{x^2}{4} - 1} + \frac{|x|}{2}\right\}, & |x| > 2. \end{cases} \tag{2.3}$$

THEOREM 2.1. *For any Borel sets $M \subset [-1, 1]$ and $E \subset \mathbb{R}$, assume λ is fixed. Then, we have*

$$\limsup_{n\to\infty}\left\{\frac{1}{n}\log \mathbb{E}\big[\mathrm{Crt}_{n,\star}(M, E)\big] - \sup_{m\in\overline{M}, e\in\overline{E}} S_\star(m, e)\right\} \le 0, \tag{2.4}$$

$$\liminf_{n\to\infty}\left\{\frac{1}{n}\log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E)\} - \sup_{m\in M^o, e\in E^o} S_\star(m, e)\right\} \ge 0. \tag{2.5}$$

2.2 Complexity of Local Maxima

For any Borel set $E \subset \mathbb{R}$ and $M \in [-1, 1]$, we define $\mathrm{Crt}_{n,0}(M, E)$ to be the number of local maxima of f with function value in E and correlation in M:

$$\begin{aligned} \mathrm{Crt}_{n,0}(M, E) := \sum_{\boldsymbol{\sigma} : \operatorname{grad} f(\boldsymbol{\sigma}) = \mathbf{0}} & \mathbf{1}\{\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \in M\} \\ & \times \mathbf{1}\{f(\boldsymbol{\sigma}) \in E\}\mathbf{1}\{\operatorname{Hess} f(\boldsymbol{\sigma}) \preceq 0\}. \end{aligned} \tag{2.6}$$

We define the function $S_0\colon [-1, 1] \times \mathbb{R} \to \overline{\mathbb{R}}$ as

$$S_0(m, x) := S_\star(m, x) - L(\theta(m), t(x)), \tag{2.7}$$

where

$$L(\theta, t) = \begin{cases} L_0(\theta, t), & 2 \le t < \theta + \frac{1}{\theta},\ 1 < \theta, \\ \infty, & t < 2, \\ 0, & \text{otherwise}, \end{cases} \tag{2.8}$$

with

$$L_0(\theta, t) = \frac{1}{4}\int_{\theta+\frac{1}{\theta}}^{t} \sqrt{y^2 - 4} \cdot \mathrm{d}y - \frac{1}{2}\theta\left[t - \left(\theta + \frac{1}{\theta}\right)\right] + \frac{1}{8}\left[t^2 - \left(\theta + \frac{1}{\theta}\right)^2\right] \tag{2.9}$$

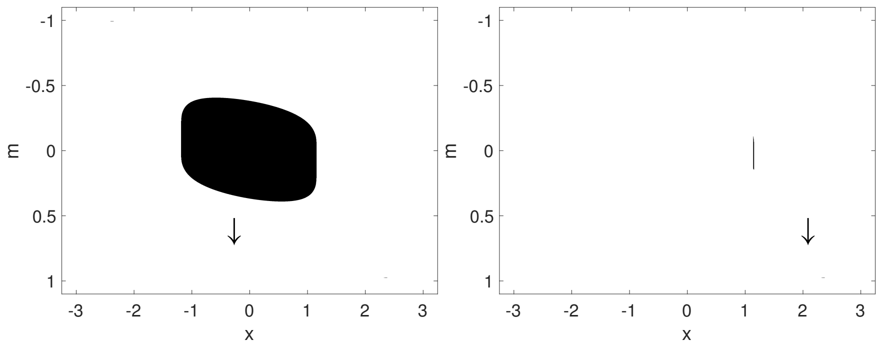

FIGURE 2.1. Spiked tensor model with $k = 3$ and $\lambda = 2.25$. The black region corresponds to nonnegative complexity: $S_\star(m, x) \geq 0$ (left) and $S_0(m, x) \geq 0$ (right). The arrows indicate the point where the complexity touches 0, in correspondence with the "good" local maxima.

and $\theta = \theta(m) = \sqrt{2k(k-1)} \cdot \lambda m^{k-2}(1-m^2)$, $t = t(x) = \sqrt{2k/(k-1)} \cdot x$. We also note that

$$\int_2^t \sqrt{y^2-4} \cdot \mathrm{d}y = t\sqrt{\frac{t^2}{4}-1} - 2\log\left(\frac{t}{2} + \sqrt{\frac{t^2}{4}-1}\right). \tag{2.10}$$

THEOREM 2.2. *For any Borel sets $M \subset [-1,1]$ and $E \subset \mathbb{R}$, assume λ is fixed. Then we have*

$$\limsup_{n\to\infty}\left\{\frac{1}{n}\log \mathbb{E}[\mathrm{Crt}_{n,0}(M,E)] - \sup_{m\in\overline{M}, e\in\overline{E}} S_0(m,e)\right\} \leq 0, \tag{2.11}$$

$$\liminf_{n\to\infty}\left\{\frac{1}{n}\log \mathbb{E}\{\mathrm{Crt}_{n,0}(M,E)\} - \sup_{m\in M^o, e\in E^o} S_0(m,e)\right\} \geq 0. \tag{2.12}$$

2.3 Evaluating the Complexity Function

The expressions for $S_\star(m, x)$ and $S_0(m, x)$ given in the previous section can be easily evaluated numerically: the figures in this section demonstrate such evaluations. Throughout this section we consider $k = 3$, but the behavior for other values of $k \geq 3$ is qualitatively similar (with the important difference that, for k even, the landscape is symmetric under change of sign of m). In Figure 2.1 we plot the region of the (m, x)-plane in which $S_\star(m, x)$ and $S_0(m, x)$ are nonnegative for $\lambda = 2.25$. By the Markov inequality, the probability of any critical point or any local maximum to be present outside these regions is exponentially small.

As anticipated in the introduction, we can identify two sets of local maxima:

(i) *Uninformative local maxima.* These have small x (i.e., small value of the objective) and small m (small correlation with the ground truth $\boldsymbol{u}$). They are also exponentially more numerous, and we expect them to trap descent algorithms.

(ii) *Good local maxima.* These have large x (i.e., large value of the objective) and large m (large correlation with the ground truth $\boldsymbol{u}$). Reaching such a local maximum results in accurate estimation.

Figure 2.2 shows the evolution of the two "projections" $S_0(x) = \max_m S_0(m, x)$ and $S_0(m) = \max_x S_0(m, x)$ that give the exponential growth rate of the number of local maxima as functions of the objective value $x = f(\boldsymbol{\sigma})$ and the scalar product $m = \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle$. Similar plots for the total number of critical points are found in Figure 2.3. We can identify several regimes of the signal-to-noise ratio λ:

(1) For λ small enough, we know that the landscape is qualitatively similar to the case $\lambda = 0$: local maxima are uninformative. While they are spread along the m-direction, this is purely because of random fluctuations. Local maxima with $m \approx 0$ are exponentially more numerous and have larger value.

(2) As λ crosses a threshold λ_c, the complexity develops a secondary maximum that touches 0 at $m_*(\lambda)$ close to 1. This signals a group of local maxima (or possibly only one of them) that are highly correlated with $\boldsymbol{u}$. These are good local maxima but have smaller value than the best uninformative local maxima. Maximum likelihood estimation still fails.

(3) As λ crosses a second threshold λ_s, good local maxima acquire a larger value of the objective than uninformative ones. Maximum likelihood succeeds. However, the most numerous local maxima are still uncorrelated with the signal $\boldsymbol{u}$ and are likely to trap algorithms.

The expression for threshold $\lambda_c(k)$ is explicitly calculated as in equation (2.15). For $k = 3$, $\lambda_c(3) = \sqrt{2/3} \approx 0.82$. We did not provide an analytical expression for the second threshold $\lambda_s(k)$. For $k = 3$, numerical evaluation suggests $\lambda_s(3) \approx 0.86$. The second regime for the signal-to-noise ratio $\lambda \in (\lambda_c(k), \lambda_s(k))$ is not captured in Figures 2.2 or 2.3.

Let us emphasize once more that this qualitative picture is obtained by considering the *expected number* of critical points. In order to confirm that it holds for a typical realization of $\boldsymbol{Y}$, it would be important to compute the typical number as well.

2.4 Explicit Formula for Complexity of Critical Points at a Given Location

The projection $S_\star(m) = \max_x S_\star(m, x)$, which gives the expected number of critical points at a given scalar product $m = \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle$, has a simple explicit formula in the hemisphere $m \in [0, 1]$. This is derived using elementary calculus by analyzing equation (2.2).

PROPOSITION 2.3. *The projection $S_\star(m) = \max_x S_\star(m, x)$ has the following explicit formula for $m \in [0, 1]$:*

$$S_\star(m) = \begin{cases} S_U(m), & 0 \le m < m_c, \\ S_G(m), & m \ge m_c, \end{cases} \tag{2.13}$$

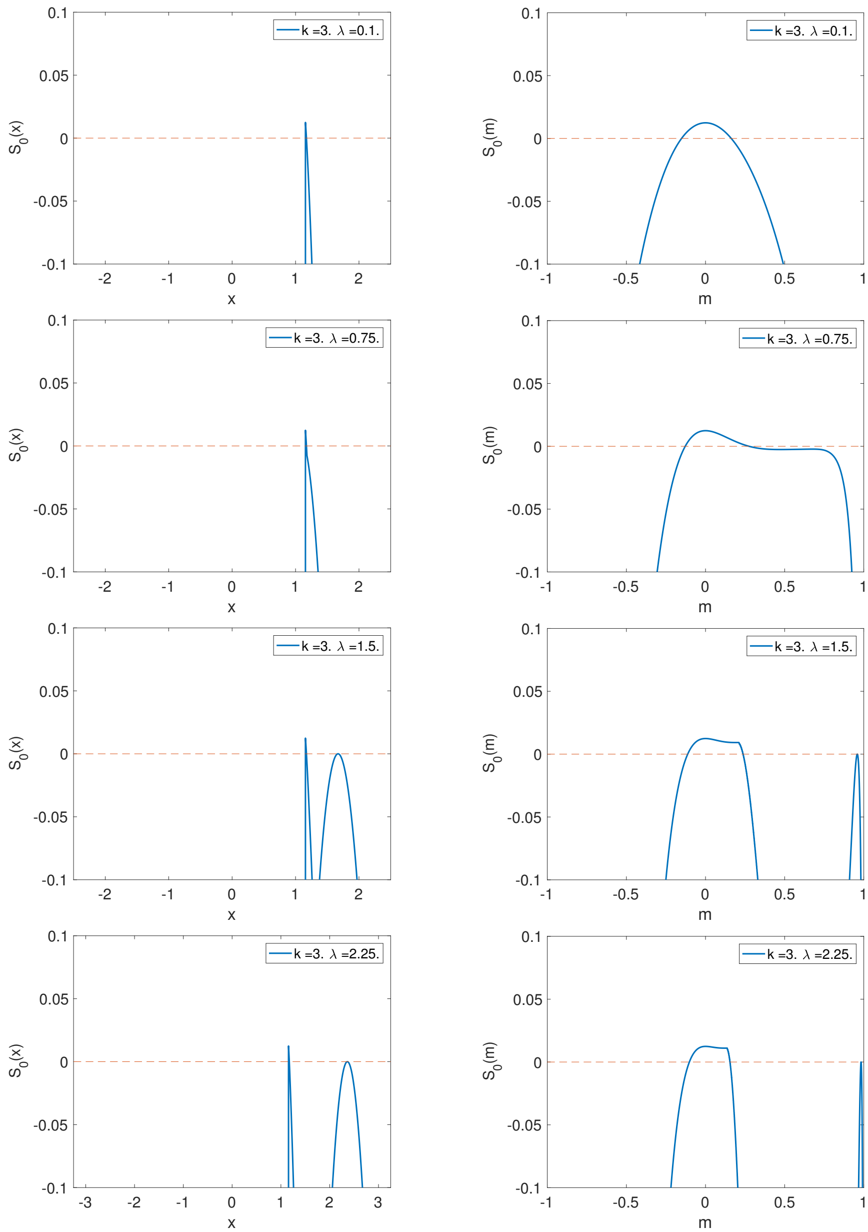

FIGURE 2.2. Complexity (exponential growth rate of the expected number of local maxima) in the spiked tensor model with $k = 3$ and (from top to bottom) $\lambda \in \{0.1, 0.75, 1.5, 2.25\}$. Left column: complexity as a function of the objective value $x = f(\boldsymbol{\sigma})$, $S_0(x) = \max_m S_0(m, x)$. Right column: complexity as a function of the scalar product $m = \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle$, $S_0(m) = \max_x S_0(m, x)$.

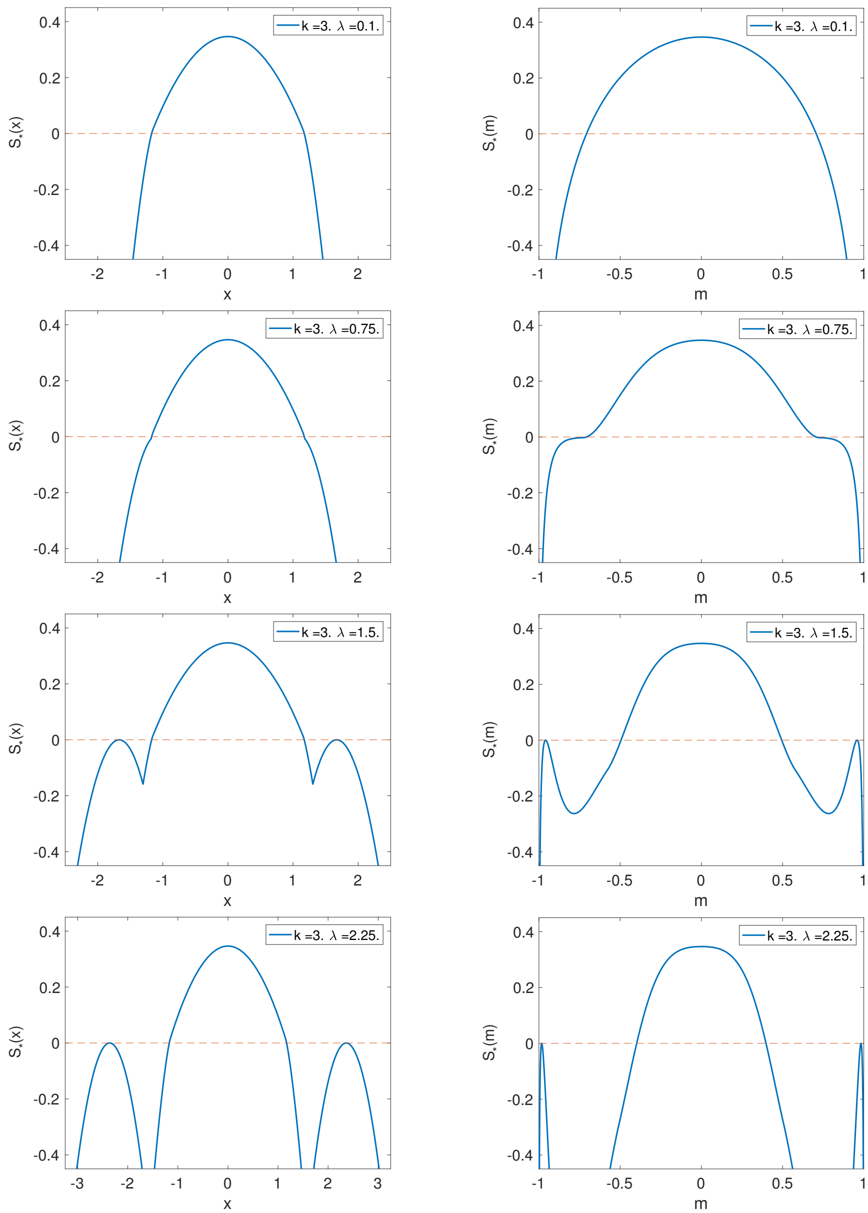

FIGURE 2.3. Complexity (exponential growth rate of the expected number of critical points) in the spiked tensor model with $k = 3$ and (from top to bottom) $\lambda \in \{0.1, 0.75, 1.5, 2.25\}$. Left column: complexity as a function of the objective value $x = f(\boldsymbol{\sigma})$, $S_\star(x) = \max_m S_\star(m, x)$. Right column: complexity as a function of the scalar product $m = \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle$, $S_\star(m) = \max_x S_\star(m, x)$.

where

$$m_c := \left(\frac{1}{\lambda}\frac{k-2}{\sqrt{2k(k-1)}}\right)^{1/k},$$

$$S_U(m) := \frac{1}{2}\log(1-m^2) - k\lambda^2 m^{2k-2}(1-m^2) + \frac{k}{k-2}\lambda^2 m^{2k} + \frac{1}{2}\log(k-1),$$

$$S_G(m) := \frac{1}{2}\log\bigl(1-m^2\bigr) - k\lambda^2 m^{(2k-2)}(1-m^2) - \left(\sqrt{\frac{1}{2}k}\,\lambda\cdot m^k\right)^2 + \sqrt{\frac{1}{2}k}\cdot\lambda\cdot m^k\cdot\sqrt{\left(1+\frac{1}{2}k\cdot\lambda^2 m^{2k}\right)} + \sinh^{-1}\left(\sqrt{\frac{1}{2}k}\,\lambda m^k\right).$$

Analysis of this formula confirms some of the qualitative observations from Section 2.3. For λ very small, namely $\lambda < (k-2)/\sqrt{2k(k-1)}$, we have that $m_c > 1$. In this case, $S_\star(m) \equiv S_U(m)$ and landscape is qualitatively similar to the case $\lambda = 0$. When $\lambda \geq (k-2)/\sqrt{2k(k-1)}$, we have that $m_c \leq 1$, and the function S_G captures the behavior of possible "good" critical points that may exist at $m > m_c$. Further analysis of the function S_G is carried out in Proposition 2.4.

PROPOSITION 2.4. *The function S_G is nonpositive, $S_G(m) \leq 0$ for all $m \in [0,1]$. Moreover, $S_G(m) = 0$ if and only if m satisfies*

$$m^{2k-4}(1-m^2) = \frac{1}{2k\lambda^2}. \tag{2.14}$$

In particular, if we set

$$\lambda_c := \sqrt{\frac{1}{2k}\frac{(k-1)^{(k-1)}}{(k-2)^{(k-2)}}}, \tag{2.15}$$

then we have that if $\lambda < \lambda_c$, then $S_G(m) < 0$, and if $\lambda \geq \lambda_c$, then S_G has a unique zero in the domain $m \in [m_c, 1]$.

The critical point λ_c identified in Proposition 2.4 represents a qualitative change in the energy landscape. When $\lambda < \lambda_c$, then $S_G < 0$ and "good" critical points are exponentially rare. On the other hand, when $\lambda \geq \lambda_c$, then S_G and has a unique zero. This is the only location in the region $m > m_c$ where critical points are not exponentially rare, and this represents the best correlation with the signal $\boldsymbol{u}$ that is achievable.

The proofs of Propositions 2.3 and 2.4 are deferred to Appendix B

2.5 Proof Ideas

The proofs of Theorems 2.1 and 2.2 rely on a representation of the expected number of critical points of a given index using the Kac-Rice formula. This approach was pioneered in [5, 15] to study the case $\lambda = 0$ of the present problem.

Evaluating the expression produced by the Kac-Rice formula requires estimating the expectation of the determinant of $\operatorname{Hess} f(\boldsymbol{\sigma})$. In the case $\lambda = 0$ considered in

[5], Hess $f(\boldsymbol{\sigma})$ is distributed as $a\boldsymbol{W}_n + b\boldsymbol{I}_n$, where $\boldsymbol{W}_n \sim \mathrm{GOE}(n)$ is a matrix form the Gaussian orthogonal ensemble. This fact, together with the explicitly known joint distribution of the eigenvalues of $\boldsymbol{W}_n$, is used in [5] to express the expected determinant in terms of the distribution of one eigenvalue and a normalization that is computed using Selberg's integral.

In the present case, Hess $f(\boldsymbol{\sigma})$ is distributed as $a\boldsymbol{W}_n + b\boldsymbol{I}_n + c\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}$, i.e., a rank one deformation of the previous matrix. Instead of an exact representation, we use the asymptotic distribution of the eigenvalue of this matrix, as well as its large-deviation properties, obtained in [30].

3 Related Literature

The complexity of random functions has been the object of a large amount of work within statistical physics, in particular in the context of mean field glasses and spin glasses. The function of interest is, typically, the Hamiltonian or energy function, and its local minima are believed to capture the long-time behavior of dynamics, as well as thermodynamic properties.

In particular, the energy function (1.3) was first studied by Crisanti and Sommers in [14] for the case $\lambda = 0$. This is referred to as the *spherical p-spin model* in the physics literature. The paper [14] uses nonrigorous methods from statistical physics to derive the complexity function, which corresponds to $S_0(x)$ in the notations used here. An alternative derivation using random matrix theory was proposed by Fyodorov [15]. Connections with thermodynamic quantities can be found in [13]. The impact of the rough energy landscape on the behavior of Langevin dynamics was studied in a number of papers; see, e.g., [11, 13].

A mathematically rigorous calculation of the expected number of critical points of any index—and the associated complexity—was first carried out in [5], again for the pure noise case $\lambda = 0$. (See also [4] for mathematically rigorous results for the complexity of some more general "pure noise" random surfaces.) As mentioned above, the expected number of critical points is not necessarily representative of typical instances. However, for the pure noise case $\lambda = 0$, it was expected that the number of critical points concentrates on its expectation. This was recently confirmed by Subag via a second-moment calculation [38]. (See also [39, 40] for additional information about the landscape geometry.)

Finally, the typical number of critical points of the spiked tensor model and variants was recently obtained in independent work [37] by using an asymptotically exact but nonrigorous generalization of the Kac-Rice formula based on the replica method [5, 15]. This computation indicates that the typical and the expected number of critical points generically do not coincide for the spiked tensor model, contrary to what happens for the pure noise case $\lambda = 0$ at low energy. By analyzing generalizations of the spiked tensor model, [37] finds different scenarios for the organization of minima on the sphere; in particular, there are cases in which the landscape is characterized by an exponential number of minima both around the spike and close to the orthogonal hyperplane.

4 Proofs

In this section we prove Theorem 2.1 and 2.2. We begin by introducing some definitions and notations in Section 4.1. We next state some useful lemmas in Section 4.2, with proofs in Sections 4.3 and 4.4. Finally, we prove Theorems 2.1 and 2.2 in Section 4.5 and 4.6.

4.1 Definitions and Notations

We will generally use lowercase letters (e.g., a, b, c) for scalars, lowercase boldface letters (e.g., $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$) for vectors, and uppercase boldface letters (e.g., $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$) for matrices. The identity matrix in n dimensions is denoted by $\boldsymbol{I}_n$, and the canonical basis in $\mathbb{R}^n$ is denoted by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Given a vector $\boldsymbol{v} \in \mathbb{R}^n$, we write $\mathsf{P}_{\boldsymbol{v}} = \boldsymbol{v}\boldsymbol{v}^\mathsf{T}/\|\boldsymbol{v}\|_2^2$ for the orthogonal projector onto the subspace spanned by $\boldsymbol{v}$, and by $\mathsf{P}_{\boldsymbol{v}}^\perp = \boldsymbol{I} - \mathsf{P}_{\boldsymbol{v}}$ the projector onto the orthogonal subspace.

For symmetric matrix $\boldsymbol{B}_n \in \mathbb{R}^{n\times n}$, we denote by $\lambda_1(\boldsymbol{B}_n) \geq \lambda_2(\boldsymbol{B}_n) \geq \cdots \geq \lambda_n(\boldsymbol{B}_n)$ the eigenvalues of $\boldsymbol{B}_n$ in decreasing order. We will also write $\lambda_{\max}(\boldsymbol{B}_n) = \lambda_1(\boldsymbol{B}_n)$ and $\lambda_{\min}(\boldsymbol{B}_n) = \lambda_n(\boldsymbol{B}_n)$ for the maximum and minimum eigenvalues.

We denote by GOE(n) the Gaussian orthogonal ensemble in n dimensions. Namely, for a symmetric random matrix $\boldsymbol{W}$ in $\mathbb{R}^{n\times n}$, we write $\boldsymbol{W} \sim \mathrm{GOE}(n)$ if the entries $(W_{ij})_{i\leq j}$ are independent, with $(W_{ij})_{1\leq i<j\leq n} \sim_{\text{iid}} \mathsf{N}(0, 1/n)$ and $(W_{ii})_{1\leq i\leq n} \sim_{\text{iid}} \mathsf{N}(0, 2/n)$.

For a sequence of functions $f_n \colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, $n \in \mathbb{N}_+$, we say that $f_n(\boldsymbol{x})$ is *exponentially finite* on a set $\mathscr{X} \subset \mathbb{R}^d$ if

$$\limsup_{n\to\infty} \sup_{\boldsymbol{x}\in\mathscr{X}} \left|\frac{1}{n}\log f_n(\boldsymbol{x})\right| < \infty. \tag{4.1}$$

We say that $f_n(\boldsymbol{x})$ is *exponentially vanishing* on a set $\mathscr{X} \subset \mathbb{R}^d$ if

$$\lim_{n\to\infty} \sup_{\boldsymbol{x}\in\mathscr{X}} \frac{1}{n}\log f_n(\boldsymbol{x}) = -\infty. \tag{4.2}$$

We say that $f_n(\boldsymbol{x})$ is *exponentially trivial* on a set $\mathscr{X} \subset \mathbb{R}^d$ if

$$\lim_{n\to\infty} \sup_{\boldsymbol{x}\in\mathscr{X}} \left|\frac{1}{n}\log f_n(\boldsymbol{x})\right| = 0. \tag{4.3}$$

We say $f_n(\boldsymbol{x})$ is *exponentially decaying* on a set $\mathscr{X} \subset \mathbb{R}^d$, if

$$\limsup_{n\to\infty} \sup_{\boldsymbol{x}\in\mathscr{X}} \frac{1}{n}\log f_n(\boldsymbol{x}) < 0. \tag{4.4}$$

For a metric space $(\mathscr{S}, d)$, we denote the open ball at $x \in \mathscr{S}$ with radius $r > 0$ by $\mathsf{B}(x, r) = \{z \in \mathscr{S} \colon d(z, x) < r\}$. In $\mathbb{R}^d$, we will always use euclidean distance. For any $x \in \mathbb{R}$ and $\delta > 0$, the open ball in $\mathbb{R}$ is denoted by $\mathsf{B}(x, r) = (x - r, x + r)$. Let $\mathscr{P}(\mathbb{R})$ be the space of probability measures on $\mathbb{R}$. We will equip

$\mathscr{P}(\mathbb{R})$ with the Dudley distance: for two probability measures $\mu, \nu \in \mathscr{P}(\mathbb{R})$, this is defined as

$$d(\mu,\nu) = \sup\left\{\left|\int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\nu\right|; |f(x)| \vee \left|\frac{f(x)-f(y)}{x-y}\right| \le 1, \forall x \neq y\right\}.$$

The open ball $\mathsf{B}(\mu,\delta)$ contains the probability measures with Dudley distance less than δ to μ.

Suppose μ is a probability measure on $\mathbb{R}$. We denote $H_\mu(z)$ as the Stieltjes transform of μ defined by (here "conv" denotes the convex hull and $\mathbb{C}_+$ the upper half-plane)

$$\begin{aligned} H_\mu\colon \mathbb{C}_+ \cup \mathbb{R} \setminus \operatorname{conv}(\operatorname{supp}\mu) &\to \mathbb{C}, \\ z &\mapsto \int_{\mathbb{R}} \frac{1}{z-\lambda}\mu(\mathrm{d}\lambda). \end{aligned} \tag{4.5}$$

H_μ is always injective, so we can define its inverse G_μ: $G_\mu(H_\mu(z)) = z$. Denote R_μ as the R-transform defined by

$$\begin{aligned} R_\mu(w)\colon \operatorname{Image}(H_\mu) &\to \mathbb{C}, \\ w &\mapsto G_\mu(w) - 1/w. \end{aligned} \tag{4.6}$$

We denote $\sigma_{\mathrm{sc}}(\mathrm{d}\lambda) = \mathbf{1}_{|\lambda|\le 2}\sqrt{4-\lambda^2}/(2\pi)\mathrm{d}\lambda$ as the semicircular law. The Stieltjes transform for the semicircular law is

$$H_{\sigma_{\mathrm{sc}}}(z) = \frac{z - \sqrt{z^2-4}}{2}, \tag{4.7}$$

and its R-transform is

$$R_{\sigma_{\mathrm{sc}}}(w) = w. \tag{4.8}$$

4.2 Preliminary Lemmas

We start by stating a form of the Kac-Rice formula that will be a key tool for our proof. Essentially the same statement was used in [5], and we refer to [1] for general proofs and broader context.

LEMMA 4.1. *Let f be a Gaussian field on $\mathbb{S}^{n-1}$, and let $\mathscr{A} = (\mathscr{U}_\alpha, \Psi_\alpha)_{\alpha\in\mathscr{I}}$ be a finite atlas on $\mathbb{S}^{n-1}$. Set $f^\alpha = f \circ \Psi_\alpha^{-1}\colon \Psi_\alpha(U_\alpha) \subset \mathbb{R}^{n-1} \to \mathbb{R}$, and define $f_i^\alpha = \partial f^\alpha/\partial x_i$ and $f_{ij}^\alpha = \partial^2 f^\alpha/\partial x_i\,\partial x_j$. Assume that for all $\alpha \in \mathscr{I}$ and all $x, y \in \Psi_\alpha(\mathscr{U}_\alpha)$, the joint distribution of $(f_i^\alpha(x), f_{ij}^\alpha(x))_{1\le i\le j\le n}$ is nondegenerate, and*

$$\max_{i,j} |\mathrm{Var}(f_{ij}^\alpha(x)) + \mathrm{Var}(f_{ij}^\alpha(y)) - 2\,\mathrm{Cov}(f_{ij}^\alpha(x), f_{ij}^\alpha(y))| \le K_\alpha |\ln|x-y||^{-1-\beta}$$

for some $\beta > 0$ and $K_\alpha > 0$. For Borel sets $E \subset \mathbb{R}$ and $M \subset [-1,1]$, let

$$\mathrm{Crt}_{n,k}^f(M,E) = \sum_{\boldsymbol{\sigma}:\,\mathrm{grad}\,f(\boldsymbol{\sigma})=0} \mathbf{1}\{i(\mathrm{Hess}\,f(\boldsymbol{\sigma})) = k, f(\boldsymbol{\sigma}) \in E, \langle\boldsymbol{\sigma},\boldsymbol{u}\rangle \in M\}. \tag{4.9}$$

Then, using $\mathrm{d}\boldsymbol{\sigma}$ *to denote the usual surface measure on* $\mathbb{S}^{n-1}$*, and denoting by* $\varphi_{\boldsymbol{\sigma}}(\boldsymbol{x})$ *the density of* $\nabla f(\boldsymbol{\sigma})$ *at* $\boldsymbol{x}$*, we have*

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,k}^{f}(M,E)\} \\
&\quad = \int_{\langle \boldsymbol{\sigma},\boldsymbol{u}\rangle \in M} \mathbb{E}[|\det(\mathrm{Hess}\, f(\boldsymbol{\sigma}))| \\
&\qquad \times \mathbf{1}\{i(\mathrm{Hess}\, f(\boldsymbol{\sigma})) = k,\, f(\boldsymbol{\sigma}) \in E\} \mid \mathrm{grad}\, f(\boldsymbol{\sigma}) = \mathbf{0}] \\
&\qquad \cdot \varphi_{\boldsymbol{\sigma}}(\mathbf{0}) \cdot \mathrm{d}\boldsymbol{\sigma}
\end{aligned} \tag{4.10}$$

The next lemma specializes the last formula to our specific choice of $f(\cdot)$; cf. equation (1.3). Its proof can be found in Section 4.3.

LEMMA 4.2. *We have*

$$\begin{aligned}
\mathbb{E}\{\mathrm{Crt}_{n,\star}(M,E)\} = \int_M & V_n(m) \cdot \mathbb{E}\{|\det(\boldsymbol{H})| \cdot \mathbf{1}\{f \in E\}\} \\
& \times \varphi_{\boldsymbol{\sigma}}(\mathbf{0}) \cdot (1-m^2)^{-1/2} \cdot \mathrm{d}m,
\end{aligned} \tag{4.11}$$

$$\begin{aligned}
\mathbb{E}\{\mathrm{Crt}_{n,0}(M,E)\} = \int_M & V_n(m) \cdot \mathbb{E}\{|\det(\boldsymbol{H})| \cdot \mathbf{1}\{\boldsymbol{H} \preceq 0\} \cdot \mathbf{1}\{f \in E\}\} \\
& \times \varphi_{\boldsymbol{\sigma}}(\mathbf{0}) \cdot (1-m^2)^{-1/2} \cdot \mathrm{d}m,
\end{aligned} \tag{4.12}$$

where

$$V_n(m) = \mathrm{Vol}(\partial \mathsf{B}^{n-1}((1-m^2)^{1/2})) \tag{4.13}$$

is the area of the $(n-1)$*-dimensional sphere with radius* $(1-m^2)^{1/2}$*, and* $\varphi_{\boldsymbol{\sigma}}(\mathbf{0})$ *is the density of* $\boldsymbol{g}$ *at* $\mathbf{0}$*. Furthermore, the joint distribution of* $f \in \mathbb{R}$*,* $\boldsymbol{g} \in \mathbb{R}^{n-1}$*, and* $\boldsymbol{H} \in \mathbb{R}^{(n-1)\times(n-1)}$ *is given by*

$$\begin{aligned}
f &\stackrel{\mathrm{d}}{=} \lambda m^k + \frac{1}{\sqrt{2n}} Z, \\
\boldsymbol{g} &\stackrel{\mathrm{d}}{=} k\lambda m^{k-1}\sqrt{1-m^2} \cdot \boldsymbol{e}_1 + \sqrt{\frac{k}{2n}} \cdot \widetilde{\boldsymbol{g}}_{n-1}, \\
\boldsymbol{H} &\stackrel{\mathrm{d}}{=} k(k-1)\lambda m^{k-2}(1-m^2)\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}} + \sqrt{\frac{k(k-1)(n-1)}{2n}}\, \boldsymbol{W}_{n-1} \\
&\quad - k\left(\lambda m^k + \frac{1}{\sqrt{2n}} Z\right) \boldsymbol{I}_{n-1},
\end{aligned}$$

where $Z \sim \mathscr{N}(0,1)$*,* $\widetilde{\boldsymbol{g}}_{n-1} \sim \mathscr{N}(\mathbf{0}, \boldsymbol{I}_{n-1})$*, and* $\boldsymbol{W}_{n-1} \sim \mathrm{GOE}(n-1)$ *are independent.*

The next lemma provides a simplified expression. Its proof is deferred to Section 4.4.

LEMMA 4.3. *We have*

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}(M,E)\} \\
&\quad = \mathscr{C}_n \cdot \int_E \mathrm{d}x \int_M (1-m^2)^{-3/2}\mathrm{d}m \cdot \mathbb{E}\{|\det(\boldsymbol{H}_n)|\} \\
&\qquad \times \exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr]\biggr\}
\end{aligned} \tag{4.14}$$

and

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,0}(M,E)\} \\
&\quad = \mathscr{C}_n \cdot \int_E \mathrm{d}x \int_M (1-m^2)^{-3/2}\,\mathrm{d}m \cdot \mathbb{E}\{|\det(\boldsymbol{H}_n)| \cdot \mathbf{1}\{\boldsymbol{H}_n \preceq 0\}\} \\
&\qquad \times \exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr]\biggr\}
\end{aligned} \tag{4.15}$$

where, for $\boldsymbol{W}_{n-1} \sim \mathrm{GOE}(n-1)$,

$$\begin{aligned}
\boldsymbol{H}_n &= \theta_n(m)\cdot \boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{W}_{n-1} - t_n(x)\cdot \boldsymbol{I}_{n-1}, \\
t_n(x) &= \left(\frac{2kn}{(k-1)(n-1)}\right)^{1/2} \cdot x, \\
\theta_n(m) &= \left(\frac{2k(k-1)n}{(n-1)}\right)^{1/2} \cdot \lambda m^{k-2}(1-m^2), \\
\mathscr{C}_n &= 2\cdot\left(\frac{n-1}{2e}\right)^{\frac{n-1}{2}} \cdot \Gamma\left(\frac{n-1}{2}\right)^{-1} \cdot \left(\frac{n}{(k-1)e\pi}\right)^{1/2}.
\end{aligned}$$

Furthermore, $\mathscr{C}_n$ *is exponentially trivial.*

The next lemma contains a well-known fact that we will use several times in the proofs. It follows immediately from the joint distribution of eigenvalues in the GOE ensemble [3]; see, for instance, [30].

LEMMA 4.4 (Joint density of the eigenvalues of the spiked model). *Let* $\boldsymbol{X}_n = \theta\boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{W}_n$, *where* $\boldsymbol{W}_n \sim \mathrm{GOE}(n)$ *and* $\theta \geq 0$. *The density joint for the eigenvalues of* $\boldsymbol{X}_n$ *is given by*

$$\begin{aligned}
&\mathbb{P}_n^\theta(\mathrm{d}x_1,\dots,\mathrm{d}x_n) \\
&\quad = \frac{1}{Z_n^\theta}\cdot\prod_{i<j}|x_i-x_j|\cdot I_n(\theta,x_1^n)\cdot\exp\biggl\{-\frac{n}{4}\sum_{i=1}^n x_i^2\biggr\}\mathrm{d}x_1\cdots\mathrm{d}x_n,
\end{aligned} \tag{4.16}$$

where x_1^n denotes the vector $(x_1, \dots, x_n)^\mathsf{T}$, and I_n is the spherical integral defined by

$$I_n(\theta, x_1^n) := \int_{\mathscr{O}_n} \exp\left\{\frac{n\theta}{2} \cdot (\boldsymbol{U} \cdot \operatorname{diag}(x_1^n) \cdot \boldsymbol{U}^\mathsf{T})_{11}\right\} \mathrm{d}m_n(\boldsymbol{U}), \tag{4.17}$$

with m_n the Haar probability measure on $\mathscr{O}_n$, the orthogonal group of size n.

Next, we state a lemma regarding the large deviations of the largest eigenvalue of the spiked model, proven in [30].[2]

LEMMA 4.5 (Large deviation of the largest eigenvalue of the spiked model [30]). *Let $\boldsymbol{X}_n = \theta \boldsymbol{e}_1 \boldsymbol{e}_1^\mathsf{T} + \boldsymbol{W}_n$, where $\boldsymbol{W}_n \sim \mathrm{GOE}(n)$, and denote by $\lambda_{\max}(\boldsymbol{X}_n)$ the largest eigenvalue of $\boldsymbol{X}_n$. Then we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\max}(\boldsymbol{X}_n) \le t) \le -L(\theta, t),$$
$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\max}(\boldsymbol{X}_n) < t) \ge -L(\theta, t_-),$$

where $L(\theta, t)$ is as defined in equation (2.8).

For symmetric matrix $\boldsymbol{B}_n \in \mathbb{R}^{n\times n}$, let $L_{n-1}(\boldsymbol{B}_n) = 1/(n-1) \cdot \sum_{i=2}^n \delta_{\lambda_i(\boldsymbol{B}_n)}$, the empirical distribution of the $n-1$ smallest eigenvalues.

We next state three useful lemmas on the spherical integral from the papers [19, 30].

LEMMA 4.6 (Continuity of spherical integral I [19, lemma 14]). *For any $\theta, \eta > 0$, there exists a function $g_{\theta,\eta}(\delta)$: $\mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ with $\lim_{z\to 0_+} g_{\theta,\eta}(z) = 0$ such that the following holds. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ be two vectors, with $x_{\max} = \max_{i\le n} x_i$, $x_{\min} = \min_{i\le n} x_i$, $y_{\max} = \max_{i\le n} y_i$, $y_{\min} = \min_{i\le n} y_i$. Let μ_x, μ_y be their empirical distributions and define $H_{\boldsymbol{x}}(z) = (1/n)\sum_{i=1}^n 1/(z - x_i)$. If $d(\mu_x, \mu_y) \le \delta$ and $\theta \in H_{\boldsymbol{x}}([x_{\min} - \eta, x_{\max} + \eta]^c) \cap H_{\boldsymbol{y}}([y_{\min} - \eta, y_{\max} + \eta]^c)$; then for sufficiently large n*

$$\left|\frac{1}{n} \log I_n(\theta, \boldsymbol{x}) - \frac{1}{n} \log I_n(\theta, \boldsymbol{y})\right| \le g_{\theta,\eta}(\delta). \tag{4.18}$$

LEMMA 4.7 (Continuity for spherical integral II [30, prop. 2.1]). *For any $\theta, \kappa, M > 0$, there exists a function $g_{\kappa,\theta,M}$: $\mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ with $\lim_{z\to 0} g_{\kappa,\theta,M}(z) = 0$ such that the following holds. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, denote by μ'_x, μ'_y the empirical distributions of the $(n-1)$ smallest entries of $\boldsymbol{x}, \boldsymbol{y}$, and x_1, y_1 the largest elements of $\boldsymbol{x}, \boldsymbol{y}$. If $d(\mu'_x, \mu'_y) \le n^{-\kappa}$, $|x_1 - y_1| \le \delta$, and $\|\boldsymbol{x}\|_\infty, \|\boldsymbol{y}\|_\infty \le M$, then for sufficiently large n*

$$\left|\frac{1}{n} \log I_n(\theta, \boldsymbol{x}) - \frac{1}{n} \log I_n(\theta, \boldsymbol{y})\right| \le g_{\kappa,\theta,M}(\delta). \tag{4.19}$$

[2] Notice that the formula in [30] contains a typo, which is corrected here. Also, the normalization of $\boldsymbol{W}_n$ is different from the one in [30]. Here the empirical spectral distribution converges to a semicircle supported on $[-2, 2]$, while in [30] the support is $[-\sqrt{2}, \sqrt{2}]$.

LEMMA 4.8 (Limiting distribution of spherical integral [19, theorem 6]). *Let* $\theta > 0$, $\{\boldsymbol{x}(n)\}_{n\in\mathbb{N}_+}$ *be a sequence of vectors with empirical measure* L_n *converging weakly to a compactly supported measure* μ, *and limiting largest element* $x_{\max} \geq \sup\colon x \in \operatorname{supp}(\mu)\}$ *and limiting smallest element* $x_{\min} \leq \inf\{x\colon x \in \operatorname{supp}(\mu)\} < 0$. *Then the function*

$$J(\mu, x_{\max}, \theta) = \lim_{n\to\infty} \frac{1}{n} \log I_n(\theta, \boldsymbol{x}(n)) \tag{4.20}$$

is finite and well-defined (*which does not depend on* $x_{\min}$).

Moreover, letting $x \geq \sup\{x\colon x \in \operatorname{supp}(\mu)\}$, *we have*

$$J(\mu, x, \theta) = \frac{\theta \cdot v(x,\theta)}{2} - \frac{1}{2}\int_{\mathbb{R}} \log(1 + \theta \cdot v(x,\theta) - \theta \cdot \lambda)\mu(\mathrm{d}\lambda), \tag{4.21}$$

where

$$v(x,\theta) = \begin{cases} R_\mu(\theta) & \textit{if } H_\mu(x) \geq \theta, \\ x - 1/\theta & \textit{otherwise}. \end{cases} \tag{4.22}$$

See Section 4.1 *for the definitions of the Stieltjes transform* $H_\mu(x)$ *and the R-transform* $R_\mu(x)$.

Setting $\mu = \sigma_{\mathrm{sc}}$ in the above lemma, with some simple calculations we get the following expression for $J(\sigma_{\mathrm{sc}}, x, \theta)$:

LEMMA 4.9. *Since* $\sup\{x\colon x \in \operatorname{supp}(\sigma_{\mathrm{sc}})\} = 2$, $J(\sigma_{\mathrm{sc}}, x, \theta)$ *is defined as* $x \geq 2$. *We have*

$$J(\sigma_{\mathrm{sc}}, x, \theta) = \begin{cases} \theta^2/4 & \text{if } 0 < \theta \leq 1,\ x \in [2, \rho(\theta)], \\ \frac{1}{2}\cdot[\theta x - 1 - \log(\theta) - \Phi_\star(x)] & \text{if } \theta \geq 1,\ x \geq 2 \text{ or} \\ & 0 < \theta \leq 1,\ x > \rho(\theta). \end{cases} \tag{4.23}$$

See equation (2.3) *for the definition of* $\Phi_\star(x)$.

4.3 Proof of Lemma 4.2

We rewrite the objective function as

$$f(\boldsymbol{\sigma}) = \langle \boldsymbol{Y}, \boldsymbol{\sigma}^{\otimes k}\rangle = \lambda \cdot \langle \boldsymbol{u}, \boldsymbol{\sigma}\rangle^k + h(\boldsymbol{\sigma}), \tag{4.24}$$

where

$$h(\boldsymbol{\sigma}) = \frac{1}{\sqrt{2n}}\langle \boldsymbol{W}, \boldsymbol{\sigma}^{\otimes k}\rangle = \frac{1}{\sqrt{2n}} \sum_{i_1,\dots,i_k=1}^{n} G_{i_1\cdots i_k}\sigma_{i_1}\cdots\sigma_{i_k}. \tag{4.25}$$

The euclidean gradient and Hessian of the f give

$$\nabla f(\boldsymbol{\sigma}) = k\lambda \, \langle \boldsymbol{u}, \boldsymbol{\sigma}\rangle^{k-1} \cdot \boldsymbol{u} + \nabla h(\boldsymbol{\sigma}), \tag{4.26}$$

$$\nabla^2 f(\boldsymbol{\sigma}) = k(k-1)\lambda \cdot \langle \boldsymbol{u}, \boldsymbol{\sigma}\rangle^{k-2} \cdot \boldsymbol{u}\boldsymbol{u}^\mathsf{T} + \nabla^2 h(\boldsymbol{\sigma}), \tag{4.27}$$

where

$$\nabla h(\boldsymbol{\sigma})_i = \frac{k}{\sqrt{2n}} \cdot \sum_{i_1,\dots,i_{k-1}=1}^{n} W_{i i_1 \cdots i_{k-1}} \sigma_{i_1} \cdots \sigma_{i_{k-1}}$$
$$(4.28) \qquad = \frac{k}{\sqrt{2n}} \cdot \frac{1}{k!} \cdot \sum_{\pi \in \mathscr{P}_n} \sum_{i_1,\dots,i_{k-1}=1}^{n} (G^{\pi})_{i i_1 \cdots i_{k-1}} \sigma_{i_1} \cdots \sigma_{i_{k-1}}$$

and

$$\nabla^2 h(\boldsymbol{\sigma})_{ij} = \frac{k(k-1)}{\sqrt{2n}} \cdot \sum_{i_1,\dots,i_{k-2}=1}^{n} W_{ij i_1 \cdots i_{k-2}} \sigma_{i_1} \cdots \sigma_{i_{k-2}}$$
$$(4.29) \qquad = \frac{k(k-1)}{\sqrt{2n}} \cdot \frac{1}{k!} \cdot \sum_{\pi \in \mathscr{P}_n} \sum_{i_1,\dots,i_{k-2}=1}^{n} (G^{\pi})_{ij i_1 \cdots i_{k-2}} \sigma_{i_1} \cdots \sigma_{i_{k-2}}.$$

We will denote by $T_{\boldsymbol{\sigma}} \mathbb{S}^{n-1}$ the tangent space of the unit sphere $\mathbb{S}^{n-1}$ at the point $\boldsymbol{\sigma}$, which we will identify isometrically with the euclidean subspace of $\mathbb{R}^n$ orthogonal to $\boldsymbol{\sigma}$. The Riemannian gradient and Hessian of f on the manifold $\mathbb{S}^{n-1}$, restricted on the tangent space, are given by

$$(4.30) \qquad \operatorname{grad} f(\boldsymbol{\sigma}) = \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \nabla f(\boldsymbol{\sigma}) = k\lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^{k-1} \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \boldsymbol{u} + \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \nabla h(\boldsymbol{\sigma}),$$

$$\operatorname{Hess} f(\boldsymbol{\sigma}) = \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \nabla^2 f(\boldsymbol{\sigma}) \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} - \langle \boldsymbol{\sigma}, \nabla f(\boldsymbol{\sigma}) \rangle \cdot \mathsf{P}_{\boldsymbol{\sigma}}^{\perp}$$
$$(4.31) \qquad = k(k-1)\lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^{k-2} \cdot (\mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \boldsymbol{u})(\mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \boldsymbol{u})^{\mathsf{T}} - k\lambda \langle \boldsymbol{u}, \boldsymbol{\sigma} \rangle^{k} \cdot \mathsf{P}_{\boldsymbol{\sigma}}^{\perp}$$
$$+ \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \nabla^2 h(\boldsymbol{\sigma}) \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} - \langle \boldsymbol{\sigma}, \nabla h(\boldsymbol{\sigma}) \rangle \cdot \mathsf{P}_{\boldsymbol{\sigma}}^{\perp}.$$

Taking $\boldsymbol{\sigma} = \boldsymbol{e}_n$ and $\boldsymbol{u} = m\boldsymbol{e}_n + \sqrt{1-m^2}\, \boldsymbol{e}_1$, we have (and identifying $T_{\boldsymbol{\sigma}} \mathbb{S}^{n-1}$ with $\mathbb{R}^{n-1}$)

$$(4.32) \qquad f(\boldsymbol{\sigma}) \stackrel{\mathrm{d}}{=} \lambda m^k + \frac{1}{\sqrt{2n}} Z, \quad Z \sim \mathscr{N}(0,1),$$

$$(4.33) \qquad \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \nabla f(\boldsymbol{\sigma}) \big|_{T_{\boldsymbol{\sigma}} \mathbb{S}^{n-1}} \stackrel{\mathrm{d}}{=} k\lambda m^{k-1} \sqrt{1-m^2}\, \boldsymbol{e}_1 + \sqrt{\frac{k}{2n}} \boldsymbol{g}_{n-1}, \quad \boldsymbol{g}_{n-1} \sim \mathscr{N}(\boldsymbol{0}, \boldsymbol{I}_{n-1}),$$

$$\mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \nabla^2 f(\boldsymbol{\sigma}) \mathsf{P}_{\boldsymbol{\sigma}}^{\perp} \big|_{T_{\boldsymbol{\sigma}} \mathbb{S}^{n-1}}$$
$$(4.34) \qquad \stackrel{\mathrm{d}}{=} k(k-1)\lambda m^{k-2} (1-m^2) \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathsf{T}} + \sqrt{\frac{k(k-1)(n-1)}{2n}} \boldsymbol{W}_{n-1},$$
$$\boldsymbol{W}_{n-1} \sim \mathrm{GOE}(n-1).$$

Thus, the Riemannian Hessian restricted to the tangent space is distributed as

$$\begin{aligned}\operatorname{Hess} f(\boldsymbol{\sigma})\big|_{T_{\boldsymbol{\sigma}}\mathbb{S}^{n-1}} &\\ \stackrel{\mathrm{d}}{=} k(k-1)\lambda m^{k-2}(1-m^2)\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}} &\\ + \sqrt{\frac{k(k-1)(n-1)}{2n}}\boldsymbol{W}_{n-1} - k\left(\lambda m^k + \frac{1}{\sqrt{2n}}Z\right)\boldsymbol{I}_{n-1}. &\end{aligned} \tag{4.35}$$

Furthermore, note that $\operatorname{grad} f(\boldsymbol{\sigma})$ and $\operatorname{Hess} f(\boldsymbol{\sigma})$ are independent.

Plug these quantities into equation (4.10) and use rotational invariance to get equation (4.12). Summing equation (4.10) over k gives equation (4.11).

4.4 Proof of Lemma 4.3

In equation (4.12), the determinant of the Hessian is given by

$$\begin{aligned}|\det(\boldsymbol{H}_n)| &= (k(k-1)(n-1)/2n)^{(n-1)/2}\\ &\quad \times \det\big(\theta_n(m)\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}} + \boldsymbol{W}_{n-1} - t_n(f)\boldsymbol{I}_{n-1}\big).\end{aligned} \tag{4.36}$$

We denote the density of f by $p_f(x)$; we then have

$$p_f(x) = \sqrt{n/\pi}\cdot\exp\{-n(x-\lambda m^k)^2\}. \tag{4.37}$$

The inner expectation yields

$$\begin{aligned}&\mathbb{E}\{|\det(\boldsymbol{H}_n)|\cdot\mathbf{1}\{\boldsymbol{H}_n \preceq \mathbf{0}\}\cdot\mathbf{1}\{f\in E\}\}\\ &\quad= (k(k-1)(n-1)/(2n))^{(n-1)/2}\\ &\qquad\times\int_E \mathbb{E}\{|\det(\boldsymbol{H}_n)|\cdot\mathbf{1}\{H_n\preceq 0\}\}\,p_f(x)\mathrm{d}x\\ &\quad= (k(k-1)(n-1)/(2n))^{(n-1)/2}(n/\pi)^{1/2}\\ &\qquad\times\int_E \mathbb{E}\{|\det(\boldsymbol{H}_n)|\cdot\mathbf{1}\{\boldsymbol{H}_n\preceq 0\}\}\exp\{-n(x-\lambda m^k)^2\}\mathrm{d}x.\end{aligned} \tag{4.38}$$

We also have

$$V_n(m) = 2\pi^{(n-1)/2}/\Gamma((n-1)/2)\cdot(1-m^2)^{(n-2)/2}, \tag{4.39}$$

$$\varphi_{\boldsymbol{\sigma}}(\mathbf{0}) = (n/(\pi k))^{(n-1)/2}\cdot\exp\{-nk\lambda^2 m^{2k-2}(1-m^2)\}. \tag{4.40}$$

Plug these into equation (4.12) and we have the form of equation (4.15) with prefactor

$$\begin{aligned}\mathscr{C}_n &= (k(k-1)(n-1)/(2n))^{(n-1)/2}(n/\pi)^{1/2}\\ &\quad\times 2\pi^{(n-1)/2}/\Gamma((n-1)/2)\\ &\quad\times (n/(\pi k))^{(n-1)/2}\times(1/(k-1)e)^{n/2}\\ &= 2((n-1)/(2e))^{(n-1)/2}/\Gamma((n-1)/2)\times(n/(k-1)e\pi)^{1/2}.\end{aligned} \tag{4.41}$$

Expand the Γ function in $\mathscr{C}_n$ using Stirling's formula, and it is easy to see that $\mathscr{C}_n$ is exponentially trivial.

Equation (4.14) follows essentially by the same calculation.

4.5 Proof of Theorem 2.1

Throughout the proof, we will use the following notations:

$$
\begin{aligned}
\boldsymbol{J}_n &= \theta \cdot \boldsymbol{e}_1 \boldsymbol{e}_1^\top + \boldsymbol{W}_n - t \cdot \boldsymbol{I}_n, \\
\boldsymbol{X}_n &= \theta \cdot \boldsymbol{e}_1 \boldsymbol{e}_1^\top + \boldsymbol{W}_n, \\
\boldsymbol{H}_n &= \theta_n(m) \cdot \boldsymbol{e}_1 \boldsymbol{e}_1^\top + \boldsymbol{W}_{n-1} - t_n(x) \cdot \boldsymbol{I}_{n-1}, \\
\theta(m) &= \sqrt{2k(k-1)} \cdot \lambda m^{k-2}(1-m^2), \\
t(x) &= \sqrt{2k/(k-1)} \cdot x, \\
\theta_n(m) &= \sqrt{2k(k-1)n/(n-1)} \cdot \lambda m^{k-2}(1-m^2), \\
t_n(x) &= \sqrt{2kn/((k-1)(n-1))} \cdot x.
\end{aligned} \tag{4.42}
$$

In order to prove Theorem 2.1, we will establish the following key proposition, whose proof follows on page 2307.

PROPOSITION 4.10. *The following statements hold*:

(a) Exponential tightness:

$$\lim_{z\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}([-1,1], (-\infty,-z] \cup [z,\infty))\} = -\infty. \tag{4.43}$$

(b) Upper bound: *For any fixed large $U_0 > 0$ and $T_0 > 0$, let $\widehat{\mathscr{U}_0} \subset [-U_0, U_0]$ and $\overline{\mathscr{T}_0} \subset [-T_0, T_0]$ be two compact sets, and define $\overline{\mathscr{E}_0} := \widehat{\mathscr{U}_0} \times \overline{\mathscr{T}_0}$. Then we have* (*for $\Phi_\star$ defined as in equation* (2.3))

$$\limsup_{n\to\infty} \sup_{(\theta,t)\in\overline{\mathscr{E}_0}} \frac{1}{n} \log \mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \le \sup_{t\in\overline{\mathscr{T}_0}} \Phi_\star(t). \tag{4.44}$$

(c) Lower bound: *For any fixed $\delta > 0$, θ_0, and t_0, define $\mathscr{U}_0^\delta = (\theta_0 - \delta, \theta_0 + \delta)$, $\mathscr{T}_0^\delta = (t_0 - \delta, t_0 + \delta)$, and $\mathscr{E}_0^\delta := \mathscr{U}_0^\delta \times \mathscr{T}_0^\delta$. Then we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} \mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \mathrm{d}\theta \, \mathrm{d}t \ge \Phi_\star(t_0). \tag{4.45}$$

Using this proposition, we can prove Theorem 2.1.

PROOF OF THEOREM 2.1. Because of the exponential tightness property, we only need to consider the case when the set E is bounded. We will prove first the upper bound of equation (2.4), and then the lower bound; cf. equation (2.5).

Step 1. UPPER BOUND.

First, letting $E_0 = (x_0 - \delta_0, x_0 + \delta_0)$, we claim that

$$\lim_{\delta_0\to 0_+} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E_0)\} \le \sup_{m\in\overline{M}} S_\star(m, x_0). \tag{4.46}$$

Assuming this claim holds, to prove equation (2.4), we consider a general compactly supported set E. Fix an $\varepsilon > 0$; for each $x \in E$, there exists a radius δ_x such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, (x-\delta_x, x+\delta_x))\} \le \sup_{m\in\overline{M}} S_\star(m,x) + \varepsilon. \tag{4.47}$$

Then $\{(x-\delta_x, x+\delta_x): x \in E\}$ is an open cover of $\overline{E}$. Due to the compactness of $\overline{E}$, there exists a finite number of intervals $\{(x_i - \delta_{x_i}, x_i + \delta_{x_i})\}_{i=1}^m$ that form a cover of $\overline{E}$ and such that the above equation holds. Therefore

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E)\} \\ &\quad \le \max_{i\in[m]} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, (x_i - \delta_{x_i}, x_i + \delta_{x_i}))\} \\ &\quad \le \sup_{m\in\overline{M}, x\in\overline{E}} S_\star(m,x) + \varepsilon. \end{aligned} \tag{4.48}$$

Equation (2.4) holds by choosing an arbitrarily small ε.

Therefore, we just need to prove equation (4.46). For $x \in \mathbb{R}$, $S \subseteq \mathbb{R}$, define $d(x,S) = \inf\{|x-y|: y \in S\}$. For a given small $\delta > 0$, define

$$\begin{aligned} \overline{M}_\delta &:= \{m: d(m, \overline{M}) \le \delta\}, \\ \overline{E}_\delta &:= \{x: d(x, \overline{E}_0) \le \delta\}, \\ \widehat{\mathscr{U}}_\delta &:= \{\theta: \theta = \sqrt{2k(k-1)} \cdot \lambda m^{k-2}(1-m^2),\ m \in \overline{M}_\delta\}, \\ \overline{\mathscr{T}}_\delta &:= \{t: t = \sqrt{2k/(k-1)} \cdot x,\ x \in \overline{E}_\delta\}, \\ \overline{\mathscr{E}}_\delta &:= \widehat{\mathscr{U}}_\delta \times \overline{\mathscr{T}}_\delta. \end{aligned} \tag{4.49}$$

Since E_0 is bounded, we can define finite constants U_0, T_0 such that

$$\widehat{\mathscr{U}}_\delta \subset [-U_0, U_0] \quad \text{and} \quad \overline{\mathscr{T}}_\delta \subset [-T_0, T_0].$$

For any $\delta > 0$, there exists N_δ large enough such that $t_n(x) \in \overline{\mathscr{T}}_\delta$ and $\theta_n(m) \in \widehat{\mathscr{U}}_\delta$ for all $n \ge N_\delta$ and $(m,x) \in \overline{M} \times \overline{E}_0$. According to Proposition 4.10(b), there exists $N_{\varepsilon,\delta} \ge N_\delta$ such that for all $n \ge N_{\varepsilon,\delta}$,

$$\begin{aligned} &\sup_{m\in\overline{M}, x\in\overline{E}_0} \mathbb{E}\{|\det(\boldsymbol{H}_n)|\} \\ &\quad = \sup_{m\in\overline{M}, x\in\overline{E}_0} \mathbb{E}\{\big|\det\big(\theta_n(m) \cdot \boldsymbol{e}_1\boldsymbol{e}_1^\mathsf{T} + \boldsymbol{W}_{n-1} - t_n(x) \cdot \boldsymbol{I}_{n-1}\big)\big|\} \\ &\quad \le \sup_{(\theta,t)\in\overline{\mathscr{E}}_\delta} \mathbb{E}\{\big|\det\big(\theta \cdot \boldsymbol{e}_1\boldsymbol{e}_1^\mathsf{T} + \boldsymbol{W}_{n-1} - t \cdot \boldsymbol{I}_{n-1}\big)\big|\} \\ &\quad \le \exp\big\{(n-1)\big[\sup_{t\in\overline{\mathscr{T}}_\delta} \Phi_\star(t) + \varepsilon\big]\big\}. \end{aligned} \tag{4.50}$$

According to the expression for the expected number of critical points in Lemma 4.3, equation (4.14),

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E_0)\} \\
&\quad \le \sup_{m\in\bar{M}, x\in\bar{E}_0} \mathbb{E}\{|\det(\boldsymbol{H}_n)|\} \cdot \mathscr{C}_n \cdot \int_{E_0} \mathrm{d}x \int_M (1-m^2)^{-3/2}\,\mathrm{d}m \\
&\qquad \times \exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr]\biggr\} \\
&\quad \le \sup_{m\in\bar{M}, x\in\bar{E}_0} 4\mathscr{C}_n R_0 \times \exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr]\biggr\} \\
&\qquad \times \exp\biggl\{(n-3)\biggl[\frac{1}{2}\log(1-m^2)\biggr] + (n-1)\cdot \sup_{t\in\bar{\mathscr{T}}_\delta}[\Phi_\star(t)+\varepsilon]\biggr\}.
\end{aligned}$$

Note that the prefactor $2\mathscr{C}_n R_0$ is exponentially trivial. We have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E_0)\} \\
&\quad \le \sup_{m\in\bar{M}, x\in\bar{E}_0} \biggl\{\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr\} \\
&\qquad + \sup_{t\in\bar{\mathscr{T}}_\delta} \Phi_\star(t) + \varepsilon.
\end{aligned} \tag{4.51}$$

Letting $\varepsilon, \delta \to 0_+$ and using the continuity of $\Phi_\star(t)$ and compactness of $\bar{\mathscr{E}}_0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E_0)\} \\
&\quad \le \sup_{m\in\bar{M}, x\in\bar{E}_0} \biggl\{\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr\} + \sup_{t\in\bar{\mathscr{T}}_0} \Phi_\star(t).
\end{aligned} \tag{4.52}$$

Note that $E_0 = (x_0 - \delta_0, x_0 + \delta_0)$. Letting $\delta_0 \to 0$ and using the continuity of $\Phi_\star(t)$, we have proved equation (4.46).

Step 2. LOWER BOUND. For any Borel sets $M \subset [-1, 1]$ and $E \subset \mathbb{R}$, and for any $\varepsilon > 0$, there exists $(m_0, x_0) \in M^o \times E^o$ such that

$$S_\star(m_0, x_0) \geq \sup_{(m,x)\in M^o\times E^o} S_\star(m, x) - \varepsilon. \tag{4.53}$$

Denote $\theta_0 = \theta(m_0)$ and $t_0 = t(x_0)$. For a given small $\delta > 0$, define

$$\begin{aligned}
M_0^\delta &:= (m_0 - \delta, m_0 + \delta),\\
E_0^\delta &:= (x_0 - \delta, x_0 + \delta),\\
\mathscr{B}_0^\delta &:= M_0^\delta \times E_0^\delta,\\
\mathscr{U}_n^\delta &:= \{\theta\colon \theta = \sqrt{2k(k-1)n/(n-1)} \cdot \lambda m^{k-2}(1-m^2),\ m \in M_0^\delta\},\\
\mathscr{T}_n^\delta &:= \{t\colon t = \sqrt{2kn/((k-1)(n-1))} \cdot x,\ x \in E_0^\delta\},\\
\mathscr{E}_n^\delta &:= \mathscr{U}_n^\delta \times \mathscr{T}_n^\delta.
\end{aligned} \tag{4.54}$$

We fix δ sufficiently small so that $M_0^\delta \subset M^o$ and $E_0^\delta \subset E^o$. It is easy to see that $\mathscr{U}_n^\delta$ and $\mathscr{T}_n^\delta$ are open sets and $\theta_0 \in \mathscr{U}_n^\delta$, $t_0 \in \mathscr{T}_n^\delta$ are inner points.

For this choice of δ and ε, according to Proposition 4.10(c), for any $\varepsilon_0 > 0$, we can find $N_{\varepsilon,\varepsilon_0,\delta}$ and $\delta_0 > 0$ such that as $n \geq N_{\varepsilon,\varepsilon_0,\delta}$,

$$\mathscr{E}_0^{\delta_0} := (\theta_0 - \delta_0, \theta_0 + \delta_0) \times (t_0 - \delta_0, t_0 + \delta_0) \subset \mathscr{E}_n^\delta, \tag{4.55}$$

and

$$\begin{aligned}
&\int_{(\theta,t)\in\mathscr{E}_0^{\delta_0}} \mathbb{E}\{\big|\det\big(\theta \cdot \boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{W}_{n-1} - t \cdot \boldsymbol{I}_{n-1}\big)\big|\}\mathrm{d}\theta\,\mathrm{d}t\\
&\quad \geq \exp\{(n-1)[\Phi_\star(t_0) - \varepsilon_0]\}.
\end{aligned}$$

According to the expression for the expected number of critical points as in equation (4.15) in Lemma 4.3,

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}(M, E)\}\\
&\quad \geq \mathbb{E}\big\{\mathrm{Crt}_{n,\star}\big(M_0^\delta, E_0^\delta\big)\big\}\\
&\quad \geq \mathscr{C}_n \cdot \int_{\mathscr{B}_0^\delta} \mathbb{E}\{|\det(H_n)|\}\mathrm{d}x\,\mathrm{d}m\\
&\qquad\qquad \times \inf_{(m,x)\in\mathscr{B}_0^\delta} \exp\bigg\{(n-3)\cdot\bigg[\frac{1}{2}\log(1-m^2)\bigg]\\
&\qquad\qquad + n\bigg[\frac{1}{2}(\log(k-1)+1)\\
&\qquad\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\bigg]\bigg\} \geq
\end{aligned} \tag{4.56}$$

$$\begin{aligned}
&\geq \mathscr{C}_n \cdot \int_{\mathscr{E}_0^{\delta_0}} \mathbb{E}\{|\det(\theta \cdot \boldsymbol{e}_1 \boldsymbol{e}_1^\top + W_{n-1} - t \cdot \boldsymbol{I}_{n-1})|\} \\
&\quad \times \frac{n-1}{2k\lambda n|(k-2)\cdot m(\theta)^{k-3} - k \cdot m(\theta)^{k-1}|} \mathrm{d}\theta \, \mathrm{d}t \\
&\quad \times \inf_{(m,x)\in\mathscr{B}_0^\delta} \exp\left\{n\left[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \right.\right. \\
&\qquad \left.\left. - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\right]\right\},
\end{aligned}$$

which gives

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}(M,E)\} \\
&\quad \geq \frac{\mathscr{C}_n}{8k^2\lambda} \cdot \exp\left\{(n-1)\cdot[\Phi_\star(t_0) - \varepsilon_0]\right\} \\
&\qquad \times \inf_{(m,x)\in\mathscr{B}_0^\delta} \exp\left\{n\left[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \right.\right. \\
&\qquad\qquad \left.\left. - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\right]\right\}.
\end{aligned} \tag{4.57}$$

Note that the preconstant $\mathscr{C}_n/8k^2\lambda$ is exponentially trivial on a compact set. We have

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M,E)\} \\
&\quad \geq \inf_{(m,x)\in\mathscr{B}_0^\delta} \left\{\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \right. \\
&\qquad \left. - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\right\} + \Phi_\star(t_0) - \varepsilon_0.
\end{aligned} \tag{4.58}$$

Letting $\varepsilon_0, \delta \to 0_+$, we have

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,\star}(M,E)\} &\geq S_\star(m_0, x_0) \\
&\geq \sup_{m\in M^o, x\in E^o} S_\star(m,x) - \varepsilon.
\end{aligned} \tag{4.59}$$

Letting $\varepsilon \to 0_+$ gives the desired result. □

In the following we prove Proposition 4.10.

Proof of Proposition 4.10(a): Exponential Tightness

We need to upper-bound $\mathbb{E}\{\mathrm{Crt}_{n,\star}([-1,1],(-\infty,-z]\cup[z,\infty))\}$. Starting from equation (4.14), we have a crude upper bound

$$
\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}([-1,1],(-\infty,z]\cup[z,\infty))\}\\
&\quad\le 4\mathscr{C}_n\cdot\int_z^\infty \mathrm{d}x\cdot\mathbb{E}\{[4x+2k\lambda+\|W_{n-1}\|_{\mathrm{op}}]^n\}\\
&\qquad\times\exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1)-(x-\lambda)^2\biggr]\biggr\}.
\end{aligned}
\tag{4.60}
$$

We let $D_n=4\mathscr{C}_n\cdot\exp\{n[1/2\cdot(\log(k-1)+1)]\}$. It is easy to check that D_n is exponentially finite.

Taking $z\ge\max(2k\lambda,1)$ (note that we consider $k\ge 2$) and letting $Y_n=\|W_{n-1}\|_{\mathrm{op}}$, we have

$$
\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}([-1,1],(-\infty,z]\cup[z,\infty))\}\\
&\quad\le D_n\cdot\int_z^\infty \mathbb{E}\{(5x+Y_n)^n\}\cdot\exp\{-nx^2/4\}\mathrm{d}x\\
&\quad\le D_n\mathbb{E}\{(1+Y_n)^n\}\int_z^\infty (5x)^n\cdot\exp\{-nx^2/4\}\mathrm{d}x.
\end{aligned}
\tag{4.61}
$$

The operator norm of a GOE matrix has sub-Gaussian tails (cf. Lemma A.2). This immediately implies

$$
\mathbb{E}\{(1+Y_n)^n\}\le\mathbb{E}\{e^{nY_n}\}\le C^n
\tag{4.62}
$$

for some universal constant C, whence

$$
\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,\star}([-1,1],(-\infty,z]\cup[z,\infty))\}\\
&\quad\le D_nC^n\int_z^\infty (5x)^n\cdot\exp\{-nx^2/4\}\mathrm{d}x,
\end{aligned}
\tag{4.63}
$$

and the claim in equation (4.43) follows by Lemma A.3.

Proof of Proposition 4.10(b): Upper Bound

Recall that $\boldsymbol{J}_n=\theta\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}+\boldsymbol{W}_n-t\boldsymbol{I}_n$ and $\boldsymbol{X}_n=\theta\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{T}}+\boldsymbol{W}_n$. Let $\sigma_{\mathrm{sc}}(\mathrm{d}\lambda)=\mathbf{1}_{|\lambda|\le 2}\sqrt{4-\lambda^2}/(2\pi)\mathrm{d}\lambda$ be the semicircle law, and denote by $\mathsf{B}(\sigma_{\mathrm{sc}},\delta)$ the ball of radius δ around $\sigma_{\mathrm{sc}}(\mathrm{d}\lambda)$, with the Dudley metric defined in Section 4.1. Let $\mathsf{B}_R(\sigma_{\mathrm{sc}},\delta)$ be the set of probability measures in $\mathsf{B}(\sigma_{\mathrm{sc}},\delta)$ with support in $[-R,R]$. For μ a probability measure on $\mathbb{R}$, define (for all x such that the integral on the right-hand side is well-defined)

$$
\Phi(\mu,x)=\int_{\mathbb{R}}\log|\lambda-x|\cdot\mu(\mathrm{d}\lambda).
\tag{4.64}
$$

We will often make use of the following fact: for any event A, we have (defining $L_n = 1/n \cdot \sum_{i=1}^n \delta_{x_i}$ the empirical measure of the numbers $\{x_i\}_{i=1}^n$):

$$\begin{aligned}
&\mathbb{E}\{|\det(\boldsymbol{J}_n)|; A\} \\
&\quad = \int_{\mathbb{R}^n} \prod_{i=1}^n |x_i - t| \cdot \mathbf{1}_A \cdot \mathbb{P}_n^\theta(\mathrm{d}x_1, \dots, \mathrm{d}x_n) \\
&\quad = \frac{1}{Z_n^\theta} \int_{\mathbb{R}^n} \prod_{i=1}^n |x_i - t| \cdot I_n(\theta, x_1^n) \cdot \mathbf{1}_A^X \\
&\qquad\qquad \cdot \prod_{i<j} |x_i - x_j| \cdot \prod_{i=1}^n \exp\{-n x_i^2/4\} \mathrm{d}x_i \\
&\quad = \frac{Z_n^0}{Z_n^\theta} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \cdot \mathbf{1}_A \cdot \mathbb{P}_n^0(\mathrm{d}x_1, \dots, \mathrm{d}x_n),
\end{aligned} \tag{4.65}$$

where

$$Z_n^\theta = Z_n^0 \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \cdot \mathbb{P}_n^0(\mathrm{d}x_1, \dots, \mathrm{d}x_n). \tag{4.66}$$

We have the upper bound

$$\begin{aligned}
\mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \le & \underbrace{\mathbb{E}\{|\det(\boldsymbol{J}_n)|; L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)\}}_{E_1} \\
& + \underbrace{\mathbb{E}\{|\det(\boldsymbol{J}_n)|; L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)\}}_{E_2}
\end{aligned} \tag{4.67}$$

where $\delta > 0$ is a fixed arbitrary small number.

According to Lemma A.4, $E_2 \le B_n^3/A_n^2$ as a function of (θ, t) is exponentially vanishing on any compact set. Hence, we just need to consider the term E_1:

$$\begin{aligned}
E_1 &= \frac{Z_n^0}{Z_n^\theta} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \cdot \mathbf{1}\{L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)\} \cdot \mathrm{d}\mathbb{P}_n^0 \\
&\le \exp\Big\{n \cdot \sup_{\mu \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)} \Phi(\mu, t)\Big\} \cdot \frac{Z_n^0}{Z_n^\theta} \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \mathrm{d}\mathbb{P}_n^0 \\
&= \exp\Big\{n \cdot \sup_{\mu \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)} \Phi(\mu, t)\Big\}.
\end{aligned} \tag{4.68}$$

Defining $\Phi_\eta(\mu, t) = \int_{\mathbb{R}} \log(|t - \lambda| \vee \eta) \mu(\mathrm{d}\lambda)$, it is easy to verify that $\Phi_\eta(\mu, t)$ is continuous in $(\mu, t) \in M_1([-R_0, R_0]) \times \overline{\mathscr{T}_0}$ for each η. Since $\Phi(\mu, t) =$

$\inf_{\eta>0}\{\Phi_\eta(\mu,t)\}$, it holds that $\Phi(\mu,t)$ is upper-semicontinuous on the same domain. Further, a direct calculation yields $\Phi(\sigma_{\mathrm{sc}},t)=\Phi_\star(t)$. Therefore,

$$\begin{aligned}\limsup_{\delta\to 0}\limsup_{n\to\infty}\sup_{(\theta,t)\in\overline{\mathscr{E}}_0}\frac{1}{n}\log E_1 &\le \limsup_{\delta\to 0}\sup_{t\in\overline{\mathscr{T}}_0,\mu\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta)}\Phi(\mu,t)\\ &\le \sup_{t\in\overline{\mathscr{T}}_0}\Phi_\star(t).\end{aligned} \tag{4.69}$$

Consequently, we have

$$\limsup_{n\to\infty}\sup_{(\theta,t)\in\overline{\mathscr{E}}_0}\frac{1}{n}\log\mathbb{E}\{|\det(\boldsymbol{J}_n)|\}\le\sup_{t\in\overline{\mathscr{T}}_0}\Phi_\star(t). \tag{4.70}$$

Proof of Proposition 4.10(c): Lower Bound

Since $t\mapsto\Phi_\star(t)$ is continuous, we only need to prove the lower bound for (θ_0,t_0) in a dense subset of $\mathbb{R}^2$. We consider two cases for t_0:

Case 1. $t_0\in(-\infty,-2)\cup(2,\infty)$. In this case, the proof is easier, since t_0 is separated from the support of the semicircle law. We only consider the subcase $t_0>2$ and $\theta_0>1$, which is more difficult. The proof for $t_0>2$ and $\theta_0<1$ follows by a very similar argument.

Case 2. $t_0\in(-2,2)$. This case is more challenging since t_0 is inside the support of the semicircle law. We will distinguish two subcases. In subcase 2.1, $t_0\in(-2,2)$ and $\theta_0>1$, and in subcase 2.2 $t_0\in(-2,2)$ and $\theta_0<1$. We use the estimate of the spherical integral in [19] and [30].

Case 1: $t_0\in(-\infty,-2)\cup(2,\infty)$. As mentioned, we consider $t_0>2$ and $\theta_0>1$ here. The other cases are similar.

Let $\rho(\theta)=\theta+1/\theta$. Let $\delta_0\in(0,\delta)$ be such that $t_0>2+2\delta_0$. We can then choose $\varepsilon_0\in(0,\delta)$ such that $\rho(\theta_0+2\varepsilon_0)-\rho(\theta_0-2\varepsilon_0)\le\delta_0$ and $\rho(\theta_0-2\varepsilon_0)>2$. Let $\mathscr{T}_2(\delta_0,\varepsilon_0)=[t_0-\delta_0,t_0+\delta_0]\setminus[\rho(\theta_0-2\varepsilon_0),\rho(\theta_0+2\varepsilon_0)]$, and $\mathscr{T}_1(\delta_0,\varepsilon_0)=[\rho(\theta_0-\varepsilon_0),\rho(\theta_0+\varepsilon_0)]\cup[t_0-2\delta_0,t_0+2\delta_0]^c$. We have $d(\mathscr{T}_1(\delta_0,\varepsilon_0),\mathscr{T}_2(\delta_0,\varepsilon_0))>0$, and the eigenvalues of the spiked matrix $\boldsymbol{X}_n$ belongs to $\mathscr{T}_1(\delta_0,\varepsilon_0)$ with probability converging to 1 as $n\to\infty$.

Thus, for $t\in\mathscr{T}_2(\delta_0,\varepsilon_0)$, $\theta\in\mathscr{U}_0^{\varepsilon_0}=(\theta_0-\varepsilon_0,\theta_0+\varepsilon_0)$, we have the following lower bound, holding for any $\delta'>0$ (here $L_n(\boldsymbol{X}_n)$ denotes the empirical spectral distribution of the matrix $\boldsymbol{X}_n$):

$$
\begin{aligned}
&\mathbb{E}\{|\det(\boldsymbol{J}_n)|\}\\
&\quad = \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}}\\
&\quad \geq \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n)\\
&\qquad \times \mathbf{1}\{L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'), \operatorname{supp}(L_n) \in \mathscr{T}_1(\delta_0, \varepsilon_0)\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}}\\
&\quad \geq \biggl\{\frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \mathbf{1}\{L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'),\\
&\qquad\qquad \operatorname{supp}(L_n) \in \mathscr{T}_1(\delta_0, \varepsilon_0)\} \mathrm{d}\mathbb{P}_n^{\mathbf{0}}\biggr\}\\
&\qquad \times \exp\biggl\{n\Bigl[\inf_{\substack{\mu \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'),\\ \operatorname{supp}(\mu) \in \mathscr{T}_1(\delta_0, \varepsilon_0)}} \Phi(\mu, t)\Bigr]\biggr\},\\
&\quad \geq \biggl\{\mathbb{P}(\operatorname{supp}(L_n(\boldsymbol{X}_n)) \subseteq \mathscr{T}_1(\delta_0, \varepsilon_0))\\
&\qquad - \underbrace{\frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \cdot \mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta')\} \mathrm{d}\mathbb{P}_n^{\mathbf{0}}}_{G_1}\biggr\}\\
&\qquad \times \exp\biggl\{n\Bigl[\inf_{\substack{\mu \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'),\\ \operatorname{supp}(\mu) \in \mathscr{T}_1(\delta_0, \varepsilon_0)}} \Phi(\mu, t)\Bigr]\biggr\}.
\end{aligned} \tag{4.71}
$$

According to Lemma A.4, $G_1 = B_n^2/A_n^2$ is exponentially vanishing on compact sets, so we can drop this term. We also know that $\mathbb{P}(\operatorname{supp}(L_n(\boldsymbol{X}_n)) \subseteq \mathscr{T}_1(\delta_0, \varepsilon_0))$ is exponentially trivial on compact sets.

This gives

$$
\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^{\delta}} \mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \mathrm{d}\theta\, \mathrm{d}t\\
&\quad \geq \liminf_{n\to\infty} \frac{1}{n} \log \int_{\theta\in\mathscr{U}_0^{\varepsilon_0}, t\in\mathscr{T}_2(\delta_0,\varepsilon_0)} \mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \mathrm{d}\theta\, \mathrm{d}t\\
&\quad \geq \liminf_{\delta'\to 0_+} \inf_{\substack{t\in\mathscr{T}_2(\delta_0,\varepsilon_0), \mu\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta')\\ \operatorname{supp}(\mu)\in\mathscr{T}_1(\delta_0,\varepsilon_0)}} \Phi(\mu, t) = \inf_{t\in\mathscr{T}_2(\delta_0,\varepsilon_0)} \Phi_\star(t).
\end{aligned} \tag{4.72}
$$

The last equality holds because $\Phi(\mu, t)$ is continuous with respect to (μ, t) on $\{(\mu, t)\colon \mu \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'), \operatorname{supp}(\mu) \in \mathscr{T}_1(\delta_0, \varepsilon_0), t \in \mathscr{T}_2(\delta_0, \varepsilon_0)\}$.

Since $\Phi_\star(t)$ is continuous, letting first $\varepsilon_0 \to 0_+$ and then $\delta_0 \to 0_+$, we have

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} \mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \mathrm{d}\theta\, \mathrm{d}t \\ \geq \limsup_{\delta_0\to 0_+} \limsup_{\varepsilon_0\to 0_+} \inf_{t\in\mathscr{T}_2(\delta_0,\varepsilon_0)} \Phi_\star(t) = \Phi_\star(t_0).\end{aligned} \tag{4.73}$$

Case 2.1: We next consider the case of $t_0 \in (-2, 2)$ and $\theta_0 > 1$. We further assume $t_0 > 0$, as the case $t_0 < 0$ follows by a similar argument. Define

$$H_1 = \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \cdot \mathrm{d}\mathbb{P}_n^0(x_1^n), \tag{4.74}$$

$$H_2 = \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \cdot \mathrm{d}\mathbb{P}_n^0(x_1^n). \tag{4.75}$$

We have $\mathbb{E}\{|\det(\boldsymbol{J}_n)|\} = H_1/H_2$. Let $\rho(\theta) = \theta + 1/\theta$. Since $\Phi_\star(t_0) = t_0^2/4 - 1/2$ for $t_0 \in (-2, 2)$, it suffices to show that

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} H_1\, \mathrm{d}\theta\, \mathrm{d}t \\ \geq t_0^2/4 + \Phi_\star(\rho(\theta_0)) - \rho(\theta_0)^2/4 + J(\sigma_{\mathrm{sc}}, \rho(\theta_0), \theta_0),\end{aligned} \tag{4.76}$$

$$\begin{aligned}\limsup_{\delta\to 0_+} \limsup_{n\to\infty} \sup_{\theta\in\mathscr{U}_0^\delta} \frac{1}{n} \log H_2 \\ \leq 1/2 + \Phi_\star(\rho(\theta_0)) - \rho(\theta_0)^2/4 + J(\sigma_{\mathrm{sc}}, \rho(\theta_0), \theta_0),\end{aligned} \tag{4.77}$$

with $J(\cdot)$ defined as per Lemma 4.8.

By [30, prop. 3.1], for fixed $\theta > 1$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log H_2 \leq 1/2 + \Phi_\star(\rho(\theta)) - \rho(\theta)^2/4 + J(\sigma_{\mathrm{sc}}, \rho(\theta), \theta). \tag{4.78}$$

Therefore, equation (4.77) is implied by the convexity of $1/n \cdot \log H_2$ as a function of θ.

To prove equation (4.76), first we choose $\delta_0 \in (0, \delta)$ and $\varepsilon_0 > 0$ small enough such that $\rho(\theta_0 - \delta_0) - \varepsilon_0 > t_0 + 2\delta_0$. For any fixed $\theta \in (\theta_0 - \delta_0, \theta_0 + \delta_0)$, we have

$$\begin{aligned}&\int_{\mathscr{T}_0^\delta} H_1\, \mathrm{d}t \\ &\quad = \frac{1}{Z_n^0} \int_{\mathscr{T}_0^\delta} \mathrm{d}t \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \cdot \prod_{i=1}^n |t - x_i| \\ &\quad\quad \times \prod_{1\le i<j\le n} |x_i - x_j| \cdot \exp\left\{-\frac{n}{4} \sum_{i=1}^n x_i^2\right\} \prod_{i=1}^n \mathrm{d}x_i =\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{Z_n^0} \int_{x_0 \in \mathcal{T}_0^\delta} \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \cdot \prod_{0 \le i < j \le n} |x_i - x_j| \\
&\quad \times \exp\left\{-\frac{n}{4} \sum_{i=0}^n x_i^2\right\} \cdot \prod_{i=0}^n \mathrm{d}x_i \cdot \exp\left\{\frac{n}{4} x_0^2\right\} \\
&\ge \frac{1}{Z_n^0} \int_{x_n \in \mathsf{B}(\rho(\theta), \varepsilon_0)} \int_{x_0 \in \mathcal{T}_0^{\delta_0}} \int_{x_1^{n-1} \in \mathbb{R}^{n-1}} \prod_{i=0}^{n-1} |x_n - x_i| \exp\left\{-\frac{n}{4} x_n^2\right\} \mathrm{d}x_n \\
&\quad \times I_n(\theta, x_1^n) \cdot \mathbf{1}\{L_n(x_0^{n-1}) \in \mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}}, n^{-1/4})\} \prod_{0 \le i < j \le n-1} |x_i - x_j| \\
&\quad \times \exp\left\{-\frac{n}{4} \sum_{i=0}^{n-1} x_i^2\right\} \prod_{i=0}^{n-1} \mathrm{d}x_i \times \exp\left\{\frac{n}{4} (t_0 - \delta_0)^2\}\right\}.
\end{aligned}
$$

Note that for n sufficiently large, $L_n(x_0^{n-1}) \in \mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}}, n^{-1/4})$ implies that $L_{n-1}(x_1^{n-1}) \in \mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}}, 2n^{-1/4})$. Therefore, for any $\theta \in (\theta_0 - \delta_0, \theta_0 + \delta_0)$, we have

$$
\begin{aligned}
&\int_{\mathcal{T}_0^\delta} H_1 \, \mathrm{d}t \\
&\quad \ge \underbrace{(\rho(\theta_0 - \delta_0) - t_0 - \delta_0 - \varepsilon_0) \cdot \exp\{-\rho(\theta_0 + \delta_0)^2/4\} \times 2\varepsilon_0}_{A_1} \\
&\qquad \times \underbrace{\exp\left\{\frac{n}{4}(t_0 - \delta_0)^2\}\right\}}_{A_2} \\
&\qquad \times \underbrace{\inf_{\substack{L_{n-1}(x_1^{n-1}) \in \mathsf{B}_{2+\delta_0}(\sigma_{\mathrm{sc}}, 2n^{-1/4}), \\ x_n \in \mathsf{B}(\rho(\theta), \varepsilon_0 + 2\delta_0)}} \exp\left\{(n-1)[\Phi(L_{n-1}(x_1^{n-1}), x_n) - \frac{1}{4} x_n^2]\right\}}_{A_3} \\
&\qquad \times \underbrace{\inf_{\substack{L_{n-1}(x_1^{n-1}) \in \mathsf{B}_{2+\delta_0}(\sigma_{\mathrm{sc}}, 2n^{-1/4}), \\ x_n \in \mathsf{B}(\rho(\theta), \varepsilon_0 + 2\delta_0)}} I_n(\theta, x_1^n)}_{A_4} \\
&\qquad \times \underbrace{\int_{x_0 \in \mathcal{T}_0^\delta} \int_{x_1^{n-1} \in [-2-\varepsilon_0, 2+\varepsilon_0]^{n-1}} \mathbf{1}\{L_n(x_0^{n-1}) \in \mathsf{B}(\sigma_{\mathrm{sc}}, n^{-1/4})\} \mathbb{P}_n^0(\mathrm{d}x_0^{n-1})}_{A_5}. \qquad (4.79)
\end{aligned}
$$

The term A_1 is strictly positive and does not depend on n. Therefore it is exponentially trivial.

Since $\Phi(\mu, t)$ is continuous on the set $\{(\mu, t) \colon \mu \in \mathsf{B}_{2+\delta_0}(\sigma_{\mathrm{sc}}, \delta'), t \in \mathsf{B}(\rho(\theta), \varepsilon_0 + 2\delta_0)\}$, the term A_3 is lower-bounded as follows:

$$
\liminf_{n \to \infty} \frac{1}{n} \log A_3 \ge \inf_{x \in \mathsf{B}(\rho(\theta), \varepsilon_0 + 2\delta_0)} \left[\Phi_\star(x) - \frac{1}{4} x^2\right]. \qquad (4.80)
$$

For the term A_4, using the continuity of the spherical integrals in Lemma 4.7 and 4.8, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log A_4 \geq J(\sigma_{\rm sc},\rho(\theta),\theta) - g_\theta(2\varepsilon_0+4\delta_0), \tag{4.81}$$

where $g_\theta(\cdot) = g_{1/4,\theta,\rho(\theta)+1}(\cdot)$

For the term A_5, we have

$$\begin{aligned} A_5 \geq{}& \mathbb{E}_{{\rm GOE},n}\left[\frac{1}{n}\#\{\lambda_i : \lambda_i \in \mathscr{T}_0^\delta\}\right] \\ &- \mathbb{P}_{{\rm GOE},n}\big(\max_{i\in[n]}|\lambda_i| \geq 2+\delta_0\big) - \mathbb{P}_{{\rm GOE},n}\big(L_n \notin \mathsf{B}(\sigma_{\rm sc}, n^{-1/4})\big). \end{aligned}\tag{4.82}$$

The first term is exponentially trivial, the second term is exponentially decaying, and the third term is exponentially vanishing. Therefore, A_5 is exponentially trivial.

Putting the various terms together we get, for any $\theta \in (\theta_0-\delta_0, \theta_0+\delta_0)$ and $t_0 > 0$,

$$\begin{aligned} &\liminf_{n\to\infty}\frac{1}{n}\log\int_{\mathscr{T}_0^\delta} H_1\,{\rm d}t \\ &\quad\geq \frac{1}{4}(t_0-\delta_0)^2 + J(\sigma_{\rm sc},\rho(\theta),\theta) - g_\theta(2\varepsilon_0+4\delta_0) \\ &\qquad + \inf_{x\in\mathsf{B}(\rho(\theta),\varepsilon_0+2\delta_0)}[\Phi_\star(x) - 1/4\cdot x^2]. \end{aligned}\tag{4.83}$$

For any fixed $\theta \in (\theta_0-\delta_0, \theta_0+\delta_0)$, letting $\varepsilon_0, \delta_0 \to 0$ and using the continuity of $\Phi_\star(x)$ and $J(\sigma_{\rm sc}, x, \theta)$ in variable x (see equations (2.3) and (4.23)), we have

$$\begin{aligned} &\liminf_{n\to\infty}\frac{1}{n}\log\int_{\mathscr{T}_0^\delta} H_1\,{\rm d}t \\ &\quad\geq \frac{1}{4}t_0^2 + J(\sigma_{\rm sc},\rho(\theta),\theta) + \Phi_\star(\rho(\theta)) - 1/4\cdot\rho(\theta)^2. \end{aligned}\tag{4.84}$$

Note that $\{1/n\cdot\log\int_{\mathscr{T}_0^\delta} H_1\,{\rm d}t\}_{n\in\mathbb{N}_+}$ are convex functions and are uniformly bounded in θ. Therefore, according to Lemma A.1, the above inequality holds uniformly for $\theta \in (\theta_0-\delta_0, \theta_0+\delta_0)$. That is,

$$\begin{aligned} &\liminf_{n\to\infty}\frac{1}{n}\log\int_{\mathscr{E}_0^\delta} H_1\,{\rm d}\theta\,{\rm d}t \\ &\quad\geq \liminf_{n\to\infty}\inf_{\theta\in\mathscr{U}_0^{\delta_0}}\frac{1}{n}\log\int_{\mathscr{E}_0^\delta} H_1\,{\rm d}t \\ &\quad\geq \inf_{\theta\in\mathscr{U}_0^{\delta_0}}\left[\frac{1}{4}t_0^2 + J(\sigma_{\rm sc},\rho(\theta),\theta) + \Phi_\star(\rho(\theta)) - 1/4\cdot\rho(\theta)^2\right]. \end{aligned}\tag{4.85}$$

Letting $\delta_0 \to 0$ gives the desired result.

Case 2.2: $t_0 \in (-2, 2)$ and $\theta_0 < 1$. We further assume $t_0 > 0$, as the case $t_0 < 0$ can be treated analogously. For any fixed small $\varepsilon_0, \delta' > 0$, we have the lower bound

$$\begin{aligned}
&\mathbb{E}\{|\det(\boldsymbol{J}_n)|\} \\
&\quad = \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}} \\
&\quad \geq \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \cdot \mathbf{1}\{L_n \in \mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}}, \delta')\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}} \\
&\quad \geq \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \cdot \Big\{ \underbrace{\int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot \mathbf{1}\{L_n \in \mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}}, \delta')\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}}}_{F_1} \\
&\qquad\qquad \times \underbrace{\inf_{L_n \in \mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}}, \delta')} I_n(\theta, x_1^n)}_{F_2} \Big\}.
\end{aligned}$$

For the term F_1, we have

$$\begin{aligned}
F_1 \geq &\underbrace{\int_{[-2-\varepsilon_0, 2+\varepsilon_0]^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}}}_{F_3} \\
&- \underbrace{\int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot \mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta')\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}}}_{F_4}.
\end{aligned}$$

According to Lemma A.4, $F_4 = B_n^1$ is exponentially vanishing on compact sets. For the term F_3, letting $0 < \delta_0 < \delta$, we have

$$\begin{aligned}
&\int_{t \in \mathscr{T}_0^{\delta}} \mathrm{d}t \int_{[-2-\varepsilon_0, 2+\varepsilon_0]^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}} \\
&\quad = \frac{1}{Z_n^{\mathbf{0}}} \int_{t \in \mathscr{T}_0^{\delta}} \int_{[-2-\varepsilon_0, 2+\varepsilon_0]^n} \prod_{t=1}^{n} |t - x_i| \\
&\qquad \times \prod_{1 \leq i < j \leq n} |x_i - x_j| \cdot \exp\Big\{-\frac{n}{4} \sum_{i=1}^{n} x_i^2\Big\} \cdot \prod_{i=1}^{n} \mathrm{d}x_i \cdot \mathrm{d}t \\
&\quad = \frac{1}{Z_n^{\mathbf{0}}} \int_{x_0 \in \mathscr{T}_0^{\delta}} \int_{[-2-\varepsilon_0, 2+\varepsilon_0]^n} \prod_{0 \leq i < j \leq n} |x_i - x_j| \\
&\qquad \times \exp\Big\{-\frac{n}{4} \sum_{i=0}^{n} x_i^2\Big\} \cdot \prod_{i=0}^{n} \mathrm{d}x_i \cdot \exp\Big\{\frac{n}{4} x_0^2\Big\},
\end{aligned}$$

which gives

$$
\begin{aligned}
&\int_{t\in\mathscr{T}_0^\delta} \mathrm{d}t \int_{[-2-\varepsilon_0,2+\varepsilon_0]^n} \exp\{n\cdot\Phi(L_n,t)\}\cdot \mathrm{d}\mathbb{P}_n^0 \\
&\quad \geq \frac{1}{Z_n^0}\left(1+\frac{1}{n}\right)^{\frac{(n+1)(n+2)}{4}} \int_{y_0\in\sqrt{\frac{n}{n+1}}\mathscr{T}_0^{\delta_0}} \int_{[-2-\varepsilon_0/2,2+\varepsilon_0/2]^n} \prod_{0\leq i<j\leq n} |y_i-y_j| \\
&\qquad \times \exp\left\{-\frac{n+1}{4}\sum_{i=0}^n y_i^2\right\}\cdot \prod_{i=0}^n \mathrm{d}y_i \times \exp\left\{\frac{n}{4}(t_0-\delta_0)^2\right\} \\
&\quad = \frac{Z_{n+1}^0}{Z_n^0}\left(1+\frac{1}{n}\right)^{\frac{(n+1)(n+2)}{4}} \\
&\qquad \times \mathbb{E}_{\mathrm{GOE}}^{n+1}\left[\frac{1}{n+1}\#\left\{\lambda_i : \lambda_i \in \sqrt{\frac{n}{n+1}}\mathscr{T}_0^{\delta_0}\right\}\right. \\
&\qquad\qquad \left.\times \mathbf{1}\{\max|\lambda_i|\leq 2+\varepsilon_0/2\}\right] \times \exp\left\{\frac{n}{4}(t_0-\delta_0)^2\right\}
\end{aligned}
$$

Using Selberg's integral formula, we have

$$
\lim_{n\to\infty}\frac{1}{n}\log\left\{\frac{Z_{n+1}^0}{Z_n^0}\left(1+\frac{1}{n}\right)^{\frac{(n+1)(n+2)}{4}}\right\} = -\frac{1}{2}. \tag{4.86}
$$

Similar to the method dealing with the term A_5 in equation (4.79), we have

$$
\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log \mathbb{E}_{\mathrm{GOE}}^{n+1}&\left[\frac{1}{n+1}\#\left\{\lambda_i : \lambda_i \in \sqrt{\frac{n}{n+1}}\mathscr{T}_0^{\delta_0}\right\}\right. \\
&\qquad \left.\times \mathbf{1}\{\max|\lambda_i|\leq 2+\varepsilon_0/2\}\right] = 0.
\end{aligned}
$$

Now we turn to look at the term F_2. For any fixed $\theta\in\mathscr{U}_0^\delta$, there is a margin between θ and 1, so we can find η small enough so that

$$
\theta \in \bigcup_{\mu\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta')} H_\mu([-2-\varepsilon_0-\eta, 2+\varepsilon_0+\eta]^c)
$$

as ε_0,δ' are small enough. Due to the continuity of the spherical integral (cf. Lemmas 4.6 and 4.8), there exists $g_{\theta,\eta}(\delta)>0$ as $\delta>0$ and $\lim_{\delta\to 0} g_{\theta,\eta}(\delta)=0$ such that for all n large enough,

$$
\frac{1}{n}\log \inf_{L_n\in\mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}},\delta')} I_n(\theta,x_1^n) \geq J(\sigma_{\mathrm{sc}},2+\varepsilon_0,\theta)-g_{\theta,\eta}(\delta'). \tag{4.87}
$$

Using the right-continuity of function $J(\sigma_{\mathrm{sc}},x,\theta)$ with respect to x at $x=2$, we have

$$
\liminf_{\varepsilon_0,\delta'\to 0_+}\liminf_{n\to\infty}\frac{1}{n}\log \inf_{L_n\in\mathsf{B}_{2+\varepsilon_0}(\sigma_{\mathrm{sc}},\delta')} I_n(\theta,x_1^n) \geq J(\sigma_{\mathrm{sc}},2,\theta). \tag{4.88}
$$

Therefore for any fixed $\theta \in \mathscr{U}_0^\delta$,

$$\begin{aligned}(4.89)\quad &\liminf_{n\to\infty} \frac{1}{n} \log \int_{t\in\mathscr{T}_0^\delta} \int_{\mathbb{R}^n} \exp\{n\cdot\Phi(L_n,t)\}\cdot I_n(\theta,x_1^n)\cdot \mathrm{d}\mathbb{P}_n^0 \\ &\geq \limsup_{\delta_0\to 0} \left\{ J(\sigma_{\mathrm{sc}},2,\theta) + \frac{1}{4}(t_0-\delta_0)^2 - \frac{1}{2} \right\} = J(\sigma_{\mathrm{sc}},2,\theta) + \Phi_\star(t_0).\end{aligned}$$

Since $1/n\cdot\log \int_{t\in\mathscr{T}_0^\delta} \int_{\mathbb{R}^n} \exp\{n\cdot\Phi(L_n,t)\}\cdot I_n(\theta,x_1^n)\cdot\mathrm{d}\mathbb{P}_n^0$ is convex in θ, according to Lemma A.1, we have

$$\begin{aligned}(4.90)\quad &\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} \int_{\mathbb{R}^n} \exp\{n\cdot\Phi(L_n,t)\}\cdot I_n(\theta,x_1^n)\cdot\mathrm{d}\mathbb{P}_n^0 \\ &\geq J(\sigma_{\mathrm{sc}},2,\theta_0) + \Phi_\star(t_0).\end{aligned}$$

By [30, prop. 3.1], for fixed $\theta \in (\theta_0-\delta, \theta_0+\delta)$, we have

$$(4.91)\quad \limsup_{n\to\infty} \frac{1}{n} \log(Z_n^\theta/Z_n^0) \;\leq\; J(\sigma_{\mathrm{sc}},2,\theta).$$

By the convexity of $\sup_{t\in\mathscr{T}_0^\delta} \frac{1}{n}\log(Z_n^\theta/Z_n^0)$ as a function of θ, we have

$$(4.92)\quad \limsup_{\delta\to 0+} \limsup_{n\to\infty} \sup_{\theta\in\mathscr{U}_0^\delta} \frac{1}{n} \log(Z_n^\theta/Z_n^0) \;\leq\; J(\sigma_{\mathrm{sc}},2,\theta_0).$$

Therefore, as $t_0 \in (-2,2)$ and $\theta_0 < 1$, we have

$$(4.93)\quad \liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} \mathbb{E}\{|\det(\boldsymbol{J}_n)|\}\mathrm{d}\theta\,\mathrm{d}t \geq \Phi_\star(t_0).$$

4.6 Proof of Theorem 2.2

PROPOSITION 4.11. *The following statements hold*:

(a) *Exponential tightness.*

$$(4.94)\quad \lim_{z\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,0}([-1,1],(-\infty,-z]\cup[z,\infty))\} = -\infty.$$

(b) *Upper bound. For any fixed large $U_0 > 0$ and $T_0 > 0$, denote $\widehat{\mathscr{U}_0} \subset [-U_0,U_0]$ and $\overline{\mathscr{T}_0} \subset [-T_0,T_0]$ to be two compact sets, and denote $\overline{\mathscr{E}_0} := \widehat{\mathscr{U}_0} \times \overline{\mathscr{T}_0}$. Then we have*

$$\begin{aligned}(4.95)\quad &\limsup_{n\to\infty} \sup_{(\theta,t)\in\overline{\mathscr{E}_0}} \frac{1}{n} \log \mathbb{E}\{|\det(\boldsymbol{J}_n)|\cdot \mathbf{1}\{\boldsymbol{J}_n \preceq 0\}\} \\ &\leq \sup_{(\theta,t)\in\overline{\mathscr{E}_0}} [\Phi_\star(t) - L(\theta,t)]\end{aligned}$$

(c) *Lower bound. For any fixed* $\delta > 0$, θ_0, *and* t_0, *denote* $\mathscr{U}_0^\delta := (\theta_0-\delta, \theta_0+\delta)$, $\mathscr{T}_0^\delta := (t_0-\delta, t_0+\delta)$, *and* $\mathscr{E}_0^\delta := \mathscr{U}_0^\delta \times \mathscr{T}_0^\delta$. *Then we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} \mathbb{E}\{|\det(\boldsymbol{J}_n)| \cdot \mathbf{1}\{\boldsymbol{J}_n \preceq 0\}\} \mathrm{d}\theta\, \mathrm{d}t \geq \Phi_\star(t_0) - L(\theta_0, t_0). \tag{4.96}$$

Assume this proposition holds, we are in a good position to prove Theorem 2.2.

PROOF. Because of the exponential tightness property, we only need to consider the case when E is bounded.

Step 1. UPPER BOUND. Denoting $E_0 = (x_0 - \delta_0, x_0 + \delta_0)$. Using the same argument as in the proof of the upper bound in Theorem 2.1, we just need to show that

$$\lim_{\delta_0\to 0_+} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,0}(M, E_0)\} \leq \sup_{m\in\overline{M}} S_0(m, x_0). \tag{4.97}$$

For any small $\delta > 0$, define

$$\begin{aligned}
\overline{M}_\delta &= \{m\colon d(m, M) \leq \delta\},\\
\overline{E}_\delta &= \{x\colon d(x, E_0) \leq \delta\},\\
\widehat{\mathscr{U}_\delta} &= \{\theta\colon \theta = \sqrt{2k(k-1)} \cdot \lambda m^{k-2}(1-m^2),\ m \in \overline{M}_\delta\},\\
\overline{\overline{\mathscr{T}_\delta}} &= \{t\colon t = \sqrt{2k/(k-1)} \cdot x,\ x \in \overline{E}_\delta\},\\
\overline{\overline{\mathscr{E}_\delta}} &= \widehat{\mathscr{U}_\delta} \times \overline{\overline{\mathscr{T}_\delta}}.
\end{aligned} \tag{4.98}$$

Since E_0 is bounded, we can define finite constants $R_0 = \sup\{|x|\colon x \in E_0\}$, $U_0 = 2\sup\{|\sqrt{2k/(k-1)} \cdot x|\colon x \in E_0\}$ and

$$T_0 = 2\sup\{|\sqrt{2k(k-1)} \cdot \lambda m^{k-2}(1-m^2)|\colon m \in M\}.$$

Therefore, as δ is sufficiently small, we have $\widehat{\mathscr{U}_\delta} \subset [-U_0, U_0]$ and $\overline{\overline{\mathscr{T}_\delta}} \subset [-T_0, T_0]$.

We only prove the case for (M, E_0) such that $\sup_{(\theta,t)\in\overline{\overline{\mathscr{E}_0}}}[\Phi_\star(t) - L(\theta, t)] > -\infty$. For (M, E_0) such that $\sup_{(\theta,t)\in\overline{\overline{\mathscr{E}_0}}}[\Phi_\star(t) - L(\theta, t)] = -\infty$, we can prove it using similar arguments.

According to Proposition 4.11(b), for any $\varepsilon > 0$ and $\delta > 0$, there exists $N_{\varepsilon,\delta}$ large enough such that $t_n(x) \in \overline{\overline{\mathscr{T}_\delta}}$ and $\theta_n(m) \in \widehat{\mathscr{U}_\delta}$ for all $(m, x) \in M \times E_0$, and for all $n \geq N_{\varepsilon,\delta}$,

$$\begin{aligned}
&\sup_{m\in\overline{M}, x\in\overline{E}_0} \mathbb{E}\{|\det(\theta_n(m) \cdot \boldsymbol{e}_1\boldsymbol{e}_1^\mathsf{T} + W_{n-1} - t_n(x) \cdot \boldsymbol{I}_{n-1})| \cdot \mathbf{1}\{H_n \preceq 0\}\}\\
&\quad\leq \sup_{(\theta,t)\in\overline{\overline{\mathscr{E}_\delta}}} \mathbb{E}\{|\det(\boldsymbol{J}_{n-1})| \cdot \mathbf{1}\{\boldsymbol{J}_{n-1} \preceq 0\}\}\\
&\quad\leq \exp\{(n-1)[\sup_{(\theta,t)\in\overline{\overline{\mathscr{E}_\delta}}} \Phi_\star(t) - L(\theta, t) + \varepsilon]\}
\end{aligned} \tag{4.99}$$

Therefore, using equation (4.14) in Lemma 4.3, we have

$$\begin{aligned}
&\mathbb{E}\{\mathrm{Crt}_{n,0}(M,E_0)\} \\
&\quad\leq \sup_{m\in M, x\in E_0} 2\mathscr{C}_n\cdot R_0 \times \exp\Big\{n\Big[\frac{1}{2}(\log(k-1)+1) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2)-(x-\lambda m^k)^2\Big]\Big\} \\
&\qquad \times \exp\Big\{(n-3)\Big[\frac{1}{2}\log(1-m^2)\Big]+(n-1)\cdot \sup_{(\theta,t)\in\overline{\mathscr{E}}_\delta}[\Phi_\star(t)-L(\theta,t)+\varepsilon]\Big\}.
\end{aligned}$$

Note that the preconstant $2\mathscr{C}_n R_0$ is exponentially trivial. We have

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\{\mathrm{Crt}_{n,0}(M,E_0)\} \\
&\quad\leq \sup_{m\in\overline{M}, x\in\overline{E}_0}\Big\{\frac{1}{2}(\log(k-1)+1)+\frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2)-(x-\lambda m^k)^2\Big\} \\
&\qquad + \sup_{(\theta,t)\in\overline{\mathscr{E}}_\delta}\Big\{\Phi_\star(t)-L(\theta,t)+\varepsilon\Big\}.
\end{aligned}\tag{4.100}$$

Letting $\varepsilon,\delta\to 0_+$ and using the upper semicontinuity of $\Phi_\star(t)-L(\theta,t)$ and compactness of $\overline{\mathscr{E}}_0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\{\mathrm{Crt}_{n,0}(M,E_0)\} \\
&\quad\leq \sup_{m\in\overline{M}, x\in\overline{E}_0}\Big\{\frac{1}{2}(\log(k-1)+1)+\frac{1}{2}\log(1-m^2) \\
&\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2)-(x-\lambda m^k)^2\Big\} \\
&\qquad + \sup_{(\theta,t)\in\overline{\mathscr{E}}_0}\Big\{\Phi_\star(t)-L(\theta,t)\Big\}.
\end{aligned}\tag{4.101}$$

Note that we took $E_0=(x_0-\delta_0,x_0+\delta_0)$; letting $\delta_0\to 0$ and using the upper semicontinuity of $\Phi_\star(t)-L(\theta,t)$ gives equation (4.97).

Step 2. LOWER BOUND. It suffices to consider the case when

$$\sup_{(m,x)\in M^o\times E^o} S_0(m,x)>-\infty;$$

otherwise, the inequality holds trivially.

For any Borel sets $M \subset [-1,1]$ and $E \subset \mathbb{R}$, and for any $\varepsilon > 0$, there exists $(m_0, x_0) \in M^o \times E^o$ such that

$$S_0(m_0, x_0) \geq \sup_{(m,x)\in M^o \times E^o} S_0(m, x) - \varepsilon. \tag{4.102}$$

For this choice of (m_0, x_0), denote $\theta_0 = \theta(m_0)$ and $t_0 = t(x_0)$. For a given small $\delta > 0$, define

$$\begin{aligned} M_0^\delta &:= (m_0 - \delta, m_0 + \delta), \\ E_0^\delta &:= (x_0 - \delta, x_0 + \delta), \\ \mathscr{B}_0^\delta &:= M_0^\delta \times E_0^\delta, \\ \mathscr{U}_n^\delta &:= \{\theta \colon \theta = \sqrt{2k(k-1)n/(n-1)} \cdot \lambda m^{k-2}(1-m^2),\ m \in M_0^\delta\}, \\ \mathscr{T}_n^\delta &:= \{t \colon t = \sqrt{2kn/((k-1)(n-1))} \cdot x, x \in E_0^\delta\}, \\ \mathscr{E}_n^\delta &:= \mathscr{U}_n^\delta \times \mathscr{T}_n^\delta. \end{aligned} \tag{4.103}$$

We fix δ sufficiently small so that $M_n^\delta \subset M^o$ and $E_n^\delta \subset E^o$.

For this choice of δ and ε, according to Proposition 4.11(c), for any $\varepsilon_0 > 0$, we can find $N_{\varepsilon,\varepsilon_0,\delta}$ and $\delta_0 > 0$ such that as $n \geq N_{\varepsilon,\varepsilon_0,\delta}$,

$$\mathscr{E}_0^{\delta_0} := (\theta_0 - \delta_0, \theta_0 + \delta_0) \times (t_0 - \delta_0, t_0 + \delta_0) \subset \mathscr{E}_n^\delta, \tag{4.104}$$

and

$$\begin{aligned} &\int_{(\theta,t)\in\mathscr{E}_0^{\delta_0}} \mathbb{E}\{|\det(\boldsymbol{J}_{n-1})| \cdot \mathbf{1}\{\boldsymbol{J}_{n-1} \preceq 0\}\} \mathrm{d}\theta\, \mathrm{d}t \\ &\quad \geq \exp\{(n-1)[\Phi(t_0) - L(\theta_0, t_0) - \varepsilon_0]\}. \end{aligned} \tag{4.105}$$

According to the expression for the expected number of critical points as in equation (4.15) in Lemma 4.3,

$$\begin{aligned} &\mathbb{E}\{\mathrm{Crt}_{n,0}(M, E)\} \\ &\quad \geq \mathbb{E}\{\mathrm{Crt}_{n,0}(M_0^\delta, E_0^\delta)\} \\ &\quad \geq \mathscr{C}_n \cdot \int_{\mathscr{B}_0^\delta} \mathbb{E}\{|\det(H_n)| \cdot \mathbf{1}\{H_n \preceq 0\}\} \mathrm{d}x\, \mathrm{d}m \\ &\qquad \times \inf_{(m,x)\in\mathscr{B}_0^\delta} \exp\bigg\{(n-3) \cdot \bigg[\frac{1}{2}\log(1-m^2)\bigg] \\ &\qquad\quad + n\bigg[\frac{1}{2}(\log(k-1)+1) - k\lambda^2 m^{2k-2}(1-m^2) - (x - \lambda m^k)^2\bigg]\bigg\} \geq \end{aligned}$$

$$
\begin{aligned}
&\geq \mathscr{C}_n \cdot \int_{\mathscr{E}_0^{\delta_0}} \mathbb{E}\{|\det(\boldsymbol{J}_{n-1})| \cdot \mathbf{1}\{\boldsymbol{J}_{n-1} \preceq 0\}\} \\
&\qquad\qquad \times \frac{n-1}{2k\lambda n[(k-2)\cdot m(\theta)^{k-3} - k\cdot m(\theta)^{k-1}]} \,\mathrm{d}\theta \,\mathrm{d}t \\
&\quad \times \inf_{(m,x)\in\mathscr{B}_0^\delta} \exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr]\biggr\} \\
&\geq \frac{\mathscr{C}_n}{8k^2\lambda} \cdot \exp\{(n-1)\cdot[\Phi_0(t_0) - L(\theta_0, t_0) - \varepsilon_0]\} \\
&\quad \times \inf_{(m,x)\in\mathscr{B}_0^\delta} \exp\biggl\{n\biggl[\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr]\biggr\}.
\end{aligned}
$$

Note that the preconstant $\mathscr{C}_n/8k^2$ is exponentially trivial on a compact set. We have

$$
\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,0}(M, E)\} \\
&\quad \geq \inf_{(m,x)\in\mathscr{B}_0^\delta} \biggl\{\frac{1}{2}(\log(k-1)+1) + \frac{1}{2}\log(1-m^2) \\
&\qquad\qquad\qquad - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2\biggr\} \\
&\qquad + \Phi_\star(t_0) - L(\theta_0, t_0) - \varepsilon_0.
\end{aligned}
$$

Letting $\varepsilon_0, \delta \to 0_+$, we have

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{\mathrm{Crt}_{n,0}(M, E)\} &\geq S_0(m_0, x_0) \\
&\geq \sup_{m\in M^o, x\in E^o} S_0(m, x) - \varepsilon.
\end{aligned}
\tag{4.106}
$$

Letting $\varepsilon \to 0_+$ gives the desired result. □

For Proposition 4.11, the exponential tightness is trivial since we have the exponential tightness of the expected number of critical points. In the following, we will prove the upper bound and the lower bound.

Part (1). Upper Bound

We decompose

$$\begin{aligned}&\mathbb{E}\{|\det(\boldsymbol{J}_n)|\cdot\mathbf{1}\{\boldsymbol{J}_n\preceq 0\}\}\\&\quad\le\underbrace{\mathbb{E}\{|\det(\boldsymbol{J}_n)|\cdot\mathbf{1}\{\boldsymbol{J}_n\preceq 0\};L_n\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta)\}}_{F_1}+\underbrace{\mathbb{E}\{|\det(\boldsymbol{J}_n)|,L_n\notin\mathsf{B}(\sigma_{\mathrm{sc}},\delta)\}}_{E_2}\end{aligned}$$

where $\delta>0$ is an arbitrary small number.

According to Lemma A.4, $E_2=B_n^3/A_n^2$ as a function of (θ,t) is exponentially vanishing on compact set. We just need to consider the term F_1.

For the term F_1, we have

$$\begin{aligned}F_1&=\frac{Z_n^0}{Z_n^\theta}\int_{\mathbb{R}^n}\exp\bigl\{[n\cdot\Phi(L_n,t)\}\cdot\mathbf{1}\{\max_{i\in[n]}\{x_i\}\le t,L_n\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta)\bigr\}I_n(\theta,x_1^n)\cdot\mathrm{d}\mathbb{P}_n^0\\&\le\exp\bigl\{n\cdot\sup_{\mu\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta)}\Phi(\mu,t)\bigr\}\cdot\frac{Z_n^0}{Z_n^\theta}\int_{\mathbb{R}^n}\mathbf{1}\{\max_{i\in[n]}\{x_i\}\le t\}I_n(\theta,x_1^n)\cdot\mathrm{d}\mathbb{P}_n^0\\&=\exp\bigl\{n\cdot\sup_{\mu\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta)}\Phi(\mu,t)\bigr\}\cdot\mathbb{P}(\lambda_{\max}(X_n)\le t).\end{aligned}$$

According to Lemma 4.5, and noting that $\mathbb{P}(\lambda_{\max}(X_n)\le t)$ is a coordinate-wise monotone function with respect to (θ,t), we have

$$\limsup_{n\to\infty}\sup_{(\theta,t)\in\bar{\mathscr{E}}_0}\frac{1}{n}\log\mathbb{P}(\lambda_{\max}(X_n)\le t)\le-\inf_{(\theta,t)\in\bar{\mathscr{E}}_0}L(\theta,t).\tag{4.107}$$

Consequently,

$$\begin{aligned}&\lim_{\delta\to 0+}\limsup_{n\to\infty}\sup_{(\theta,t)\in\bar{\mathscr{E}}_0}\frac{1}{n}\log F_1\\&\quad\le\lim_{\delta\to 0+}\sup_{(\theta,t)\in\bar{\mathscr{E}}_0,\mu\in\mathsf{B}(\sigma_{\mathrm{sc}},\delta)}[\Phi(\mu,t)-L(\theta,t)]\le\sup_{(\theta,t)\in\bar{\mathscr{E}}_0}[\Phi_\star(t)-L(\theta,t)].\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{(\theta,t)\in\bar{\mathscr{E}}_0}\frac{1}{n}\log\mathbb{E}\{|\det(\boldsymbol{J}_n)|\cdot\mathbf{1}\{\boldsymbol{J}_n\preceq 0\}\}\\&\quad\le\sup_{(\theta,t)\in\bar{\mathscr{E}}_0}[\Phi_\star(t)-L(\theta,t)].\end{aligned}$$

Part (2). Lower Bound

For the lower bound, since $\Phi_\star(t)-L(\theta,t)$ is upper-semicontinuous, we only need to prove it for those (θ_0,t_0) in a dense subset of $\mathbb{R}^2$. Since $t_0\in(-\infty,2)$, we have $L(\theta,t_0)=\infty$ for any θ. So we only need to consider the case when $t_0>2$.

Fix $t_0 > 2$, choose $\delta_0 > 0$ and $\varepsilon_0 > 0$ such that $\delta_0 < \delta$, and $2 < t_0 - \delta_0 - \varepsilon_0 < t_0 - \delta_0$. We have

$$\begin{aligned}
&\mathbb{E}\{|\det(\boldsymbol{J}_n)| \cdot \mathbf{1}\{\boldsymbol{J}_n \preceq 0\}\} \\
&\quad = \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot \mathbf{1}\{x_{\max} \le t\} \cdot I_n(\theta, x_1^n) \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}}(x_1^n) \\
&\quad \ge \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} \exp\{n \cdot \Phi(L_n, t)\} \cdot I_n(\theta, x_1^n) \\
&\qquad \times \mathbf{1}\{x_{\max} \le \min\{t, t_0 - \delta_0 - \varepsilon_0\}, L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta')\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}} \\
&\quad \ge \frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \mathbf{1}\{x_{\max} \le \min\{t, t_0 - \delta_0 - \varepsilon_0\}, L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta')\} \cdot \mathrm{d}\mathbb{P}_n^{\mathbf{0}} \\
&\qquad \times \exp\Big\{n\Big[\inf_{L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'), x_{\max} \le \min\{t, t_0 - \delta_0 - \varepsilon_0\}} \Phi(L_n, t)\Big]\Big\} \\
&\quad \ge \Bigg\{\mathbb{P}(\lambda_{\max}(X_n) \le \min\{t, t_0 - \delta_0 - \varepsilon_0\}) \\
&\qquad\qquad - \underbrace{\frac{Z_n^{\mathbf{0}}}{Z_n^{\theta}} \int_{\mathbb{R}^n} I_n(\theta, x_1^n) \mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta')\}\} \mathrm{d}\mathbb{P}_n^{\mathbf{0}}}_{G_2} \Bigg\} \\
&\qquad \times \exp\Big\{n\Big[\inf_{L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'), x_{\max} \le \min\{t, t_0 - \delta_0 - \varepsilon_0\}} \Phi(L_n, t)\Big]\Big\}
\end{aligned}$$

According to Lemma A.4, $G_2 = B_n^2 / A_n^2$ is exponentially vanishing on a compact set, so we can drop this term.

According to Lemma 4.5, note that $\mathbb{P}(\lambda_{\max}(X_n) \le t)$ is a coordinate-wise monotone function with respect to (θ, t). Further, $L(\theta, t)$ is continuous for $t > 2$. Therefore, we have

$$\begin{aligned}
&\liminf_{n \to \infty} \inf_{(\theta, t) \in \mathscr{E}_0^{\delta_0}} \frac{1}{n} \log \mathbb{P}(\lambda_{\max}(X_n) < \min\{t, t_0 - \delta_0 - \varepsilon_0\}) \\
&\qquad \ge -L(\theta_0 + \delta_0, t_0 - \delta_0 - \varepsilon_0).
\end{aligned} \tag{4.108}$$

This gives

$$\begin{aligned}
&\liminf_{n \to \infty} \frac{1}{n} \log \int_{(\theta, t) \in \mathscr{E}_0^{\delta}} \mathbb{E}\{|\det(\boldsymbol{J}_n)| \cdot \mathbf{1}\{\boldsymbol{J}_n \preceq 0\}\} \mathrm{d}\theta \, \mathrm{d}t \\
&\quad \ge \lim_{\delta' \to 0+} \inf_{\substack{t \in \mathscr{T}_0^{\delta_0}, L_n \in \mathsf{B}(\sigma_{\mathrm{sc}}, \delta'), \\ \lambda_{\max} \le t_0 - \delta_0 - \varepsilon_0}} \Phi(L_n, t) - L(\theta_0 + \delta_0, t_0 - \delta_0 - \varepsilon_0) \\
&\quad = \Phi_\star(t_0 + \delta_0/2) - L(\theta_0 + \delta_0, t_0 - \delta_0 - \varepsilon_0).
\end{aligned} \tag{4.109}$$

Since $\Phi_\star(t) - L(\theta, t)$ is continuous with respect to (θ, t) on $\mathbb{R} \times (2, \infty)$, letting first $\varepsilon_0 \to 0_+$ and then $\delta_0 \to 0_+$, we have

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log \int_{(\theta,t)\in\mathscr{E}_0^\delta} \mathbb{E}\{|\det(\boldsymbol{J}_n)| \cdot \mathbf{1}\{\boldsymbol{J}_n \preceq 0\}\} \mathrm{d}\theta \, \mathrm{d}t \\
&\quad \geq \limsup_{\delta_0\to 0_+} \limsup_{\varepsilon_0 \to 0_+} \Phi_\star(t_0 + \delta_0/2) - L(\theta_0 + \delta_0, t_0 - \delta_0 - \varepsilon_0) \\
&\quad = \Phi_\star(t_0) - L(\theta_0, t_0).
\end{aligned}$$

Appendix A Technical Lemmas

LEMMA A.1. *Let $\{f_n(x)\}_{n\in\mathbb{N}_+}$ be a series of real-valued functions defined on the same compact interval $[a, b]$. Suppose that each of the $f_n(x)$ are convex, and the $\{f_n(x)\}_{n\in\mathbb{N}_+}$ are uniformly bounded. Then we have*

$$\liminf_{n\to\infty} \inf_{x\in[a,b]} f_n(x) = \inf_{x\in[a,b]} \liminf_{n\to\infty} f_n(x). \tag{A.1}$$

PROOF. It is obvious that the left-hand side is smaller or equal to the right-hand side. It suffices to prove that the left-hand side is bigger or equal to the right-hand side.

We prove by contradiction. Assume the left-hand side is smaller than the right-hand side by a margin ε. We call the right-hand side $f_\star$. Then we have an increasing sequence $n_k \in \mathbb{N}_+$ such that

$$f_n(x_{n_k}) \leq f_\star - \varepsilon. \tag{A.2}$$

The sequence x_{n_k} has an accumulation point $x_\star \in [a, b]$. Since the $f_n(x)$ are uniformly bounded, let $\sup_{x\in[a,b],n\in\mathbb{N}_+} f_n(x) - f_\star \leq U$. Consider the interval $\mathscr{I} = [x_\star - (b - x_\star)\varepsilon/(2U + \varepsilon), x_\star + (x_\star - a)\varepsilon/(2U + \varepsilon)]$. Since $\lim_{k\to\infty} x_{n_k} = x_\star$, then there exists K large enough such that as $k \geq K$, we have $x_{n_k} \in \mathscr{I}$. For any $x \in \mathscr{I} \cap [x_\star, b]$, because of the convexity of f_n, we have

$$f_n(x_\star) \leq (x - x_\star)/(x - a) \cdot f_n(a) + (x_\star - a)/(x - a) \cdot f_n(x). \tag{A.3}$$

For k such that $x_{n_k} \in \mathscr{I} \cap [x_\star, b]$, we have

$$\begin{aligned}
f_{n_k}(x_\star) &\leq \frac{x_{n_k} - x_\star}{x_{n_k} - a} \cdot (f_\star + U) + \frac{x_\star - a}{x_{n_k} - a} \cdot (f_\star - \varepsilon) \\
&\leq f_\star + \frac{(x_{n_k} - x_\star)U - (x_\star - a)\varepsilon}{x_{n_k} - a} \\
&\leq f_\star + \frac{U(x_\star - a)\varepsilon/(2U + \varepsilon) - (x_\star - a)\varepsilon}{(x_\star - a)\varepsilon/(2U + \varepsilon) + x_\star - a} \\
&\leq f_\star + \varepsilon \frac{U/(2U + \varepsilon) - 1}{\varepsilon/(2U + \varepsilon) + 1} \leq f_\star - \varepsilon/2.
\end{aligned}$$

Similarly, for k such that $x_{n_k} \in \mathscr{I} \cap [a, x_\star]$, we also have $f_{n_k}(x_\star) \le f_\star - \varepsilon/2$. Therefore

$$\liminf_{n\to\infty} f_n(x_\star) \le \liminf_{k\to\infty} f_{n_k}(x_\star) \le f_\star - \varepsilon/2, \tag{A.4}$$

which contradicts the definition of $f_\star$. □

The following lemma is from [8, lemma 6.3].

LEMMA A.2 (Concentration of the operator norm of the GOE matrix). *Let $\boldsymbol{W}_N \sim$ GOE(n). Then there exists a constant t_0 such that, for all $t \ge t_0$ and all n large enough, we have*

$$\mathbb{P}(\|\boldsymbol{W}_n\|_{\mathrm{op}} \ge t) \le \exp\{-nt^2/9\}. \tag{A.5}$$

LEMMA A.3. *We have*

$$\lim_{z\to\infty} \lim_{n\to\infty} \frac{1}{n} \log \int_z^\infty x^n \exp\{-nx^2\} \mathrm{d}x = -\infty. \tag{A.6}$$

PROOF. For large enough x, we have $x^2/2 \le x^2 - \log x$. Therefore, for large enough z, the following holds:

$$\begin{aligned} &\int_z^\infty x^n \exp\{-nx^2\}\mathrm{d}x \\ &\quad = \int_z^\infty \exp\{-n(x^2 - \log x)\}\mathrm{d}x \le \int_z^\infty \exp\{-nx^2/2\}\mathrm{d}x \\ &\quad \le \int_z^\infty x \cdot \exp\{-nx^2/2\}\mathrm{d}x = \frac{1}{n}\exp\{-nz^2/2\}. \end{aligned} \tag{A.7}$$

This proves the claim. □

LEMMA A.4. *For the following quantities as functions of (θ, t), we have that A_n^1, A_n^2, and A_n^3 are exponentially finite on any compact set, and B_n^1, B_n^2, and B_n^3 are exponentially vanishing on any compact set*:

$$\begin{aligned} A_n^1 &= \int_{\mathbb{R}^n} \exp\{n\Phi(L_n, t)\}\mathrm{d}\mathbb{P}_n^{\mathbf{0}}, \\ B_n^1 &= \int_{\mathbb{R}^n} \exp\{n\Phi(L_n, t)\}\mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)\}\mathrm{d}\mathbb{P}_n^{\mathbf{0}}, \\ A_n^2 &= \int_{\mathbb{R}^n} I_n(\theta, x_1^n)\mathrm{d}\mathbb{P}_n^{\mathbf{0}}, \\ B_n^2 &= \int_{\mathbb{R}^n} I_n(\theta, x_1^n)\mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)\}\mathrm{d}\mathbb{P}_n^{\mathbf{0}}, \\ A_n^3 &= \int_{\mathbb{R}^n} \exp\{n\Phi(L_n, t)\}I_n(\theta, x_1^n)\mathrm{d}\mathbb{P}_n^{\mathbf{0}} \\ B_n^3 &= \int_{\mathbb{R}^n} \exp\{n\Phi(L_n, t)\}I_n(\theta, x_1^n)\mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}}, \delta)\}\mathrm{d}\mathbb{P}_n^{\mathbf{0}}. \end{aligned} \tag{A.8}$$

PROOF. We prove B_n^3 as an example.

$$B_n^3 \le \int_{\mathbb{R}^n} \exp\{n\Phi(L_n,t)\} I_n(\theta, x_1^n) \mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}},\delta)\} d\mathbb{P}_n^0$$
$$= \underbrace{\int_{\mathbb{R}^n} \exp\{n\Phi(L_n,t)\} I_n(\theta, x_1^n) \mathbf{1}\{\max_{i\in[n]} |x_i| \ge R\} d\mathbb{P}_n^0}_{E_1}$$
$$+ \underbrace{\int_{\mathbb{R}^n} \exp\{n\Phi(L_n,t)\} I_n(\theta, x_1^n) \mathbf{1}\{L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}},\delta), \max_{i\in[n]} |x_i| \le R\} d\mathbb{P}_n^0}_{E_2}.$$

Step 1. BOUND FOR E_1

Let $\widehat{\mathscr{U}_0} = [-U_0, U_0]$, $\overline{\mathscr{T}_0} = [-T_0, T_0]$, and $\overline{\mathscr{E}_0} = \widehat{\mathscr{U}_0} \times \overline{\mathscr{T}_0}$. Then

$$\begin{aligned} &\lim_{R\to\infty} \limsup_{n\to\infty} \sup_{(\theta,t)\in\overline{\mathscr{E}}_0} \frac{1}{n} \log E_1 \\ &\le \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\{(\|W_n\|_{\mathrm{op}} + T_0)^n \\ &\qquad \times \exp\{U_0 \|W_n\|_{\mathrm{op}}\}; \|W_n\|_{\mathrm{op}} \ge R\} = -\infty. \end{aligned} \tag{A.9}$$

For any L, we choose an $R > 0$ large enough such that

$$\limsup_{n\to\infty} \sup_{(\theta,t)\in\overline{\mathscr{E}}_0} \frac{1}{n} \log E_1 \le L. \tag{A.10}$$

Step 2. BOUND FOR E_2: USE THE LDP OF THE EMPIRICAL DISTRIBUTION OF EIGENVALUES OF THE GOE MATRIX

To bound E_2, we resort to the large-deviation result for L_n:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n^0(L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}},\delta)) = -\infty. \tag{A.11}$$

Therefore, we have the upper bound

$$\sup_{(\theta,t)\in\overline{\mathscr{E}}_0} E_2 \le \mathbb{P}(L_n \notin \mathsf{B}(\sigma_{\mathrm{sc}},\delta)) \cdot \exp\{n[\log(R_0 + T_0) + U_0 R_0]\}, \tag{A.12}$$

which gives

$$\lim_{n\to\infty} \sup_{(\theta,t)\in\overline{\mathscr{E}}_0} \frac{1}{n} \log E_2 = -\infty. \tag{A.13}$$

Therefore, we have

$$\limsup_{n\to\infty} \sup_{(\theta,t)\in\overline{\mathscr{E}}_0} \frac{1}{n} \log B_3^n \le L. \tag{A.14}$$

Sending $L \to -\infty$ gives the desired result. □

Appendix B Derivation of an Explicit Formula for $S_\star$

PROOF OF PROPOSITION 2.3. The function $S_\star(m,x)$ can be written as

$$S_\star(m,x) = \frac{1}{2}\log(k-1) + \frac{1}{2}\log(1-m^2) - k\lambda^2 m^{2k-2}(1-m^2) - (x-\lambda m^k)^2$$
$$+ \frac{2k}{k-1}\frac{x^2}{4} - \frac{1}{2}\int_2^{\sqrt{\frac{2k}{k-1}}|x|} \sqrt{y^2-4}\,\mathrm{d}y \cdot \mathbf{1}\left\{|x| \geq \sqrt{\frac{2(k-1)}{k}}\right\}.$$

Isolating the dependence on x, to maximize $S_\star(m,x)$ we must do the optimization problem:

$$\text{(B.1)} \qquad -\frac{1}{2}\min_u\left\{\frac{k-2}{2k}u^2 - 4(\lambda m^k)\sqrt{\frac{k-1}{2k}}u + \int_2^u \sqrt{y^2-4}\mathrm{d}y \cdot 1_{\{|u|\geq 2\}}\right\}$$

where we have made the substitution $u = \sqrt{(2k/(k-1)}x$. This is exactly the setting of Lemma B.1 with $a = \frac{k-2}{2k}$ and $b = 4\lambda m^k\sqrt{\frac{k-1}{2k}}$. The consideration $b \gtrless 4a$ leads to the definition of m_c and the two separate solutions S_U, S_G. When $b < 4a$, the formula for S_U is as follows using the solution to the maximization problem in this region: $-b^2/4a = -(4(k-1)/(k-2))\lambda^2 m^{2k}$ and simplifying the resulting expression.

In the other region, when $b > 4a$, using our a, b values we compute that the maximizing u is

$$u^* = \frac{k}{\sqrt{k-1}}\sqrt{\frac{1}{2}k(\lambda m^k)^2+1} - \sqrt{\frac{k}{2(k-1)}}(k-2)\lambda m^k.$$

The min value is (after some simplifying)

$$\begin{aligned}
&-\frac{1}{2}bu^* - 2\log\left(\left(\frac{1}{2}-a\right)u^* + \frac{1}{2}b\right)\\
&= -2\sqrt{\frac{1}{2}k}\,\lambda m^k\sqrt{\frac{1}{2}k(\lambda m^k)^2+1} + (k-2)(\lambda m^k)^2\\
&\quad - 2\log\left(\sqrt{\frac{1}{2}k(\lambda m^k)^2+1} + \sqrt{\frac{1}{2}k}\,\lambda m^k\right) + \log(k-1)\\
&= -\sqrt{\frac{1}{2}k}\lambda m^k 2\sqrt{\frac{1}{2}k(\lambda m^k)^2+1} + (k-2)(\lambda m^k)^2\\
&\quad - 2\sinh^{-1}\left(\sqrt{\frac{1}{2}k}\,\lambda m^k\right)g + \log(k-1)
\end{aligned}$$

(we have used $\sinh^{-1}(x) = \log(x+\sqrt{x^2+1})$). Plugging back into the formula for $S_\star(m,x)$ now gives the desired result for S_G. □

LEMMA B.1. *Let $a > 0$ and $b > 0$ be parameters. Let*

$$g(x) = ax^2 - bx + \int_2^{|x|} \sqrt{y^2 - 4}\,\mathrm{d}y \cdot 1_{\{|x|>2\}}.$$

Then

$$\arg\min_x\{g(x)\} = \begin{cases} \frac{b}{2a}, & 0 < b < 4a, \\ x^*, & b > 4a, \end{cases}$$

where

$$x^* := \frac{2ab - \sqrt{b^2 + 4 - 16a^2}}{4a^2 - 1}.$$

Moreover,

$$\min_x\{g(x)\} = \begin{cases} -\frac{b^2}{4a}, & 0 < b < 4a, \\ -\frac{1}{2}bx^* - 2\log\left(\left(\frac{1}{2} - a\right)x^* + \frac{1}{2}b\right), & b > 4a. \end{cases}$$

PROOF. We notice that g' is monotone increasing, and hence g has a unique minimum, which occurs when

$$b - 2ax = \operatorname{sgn}(x)\sqrt{|x|^2 - 4} \cdot \mathbf{1}\{|x| > 2\}.$$

If $-4a < b < 4a$, then this occurs at $b/2a$. Otherwise, since we consider only the case $b > 0$, we have a solution with $x > 2$ and have $b - 2ax = \sqrt{x^2 - 4}$. The quadratic formula then gives the formula for arg min. To see the formula for the minimum, we use the closed form for the integral:

$$\int_2^x \sqrt{y^2 - 4}\,\mathrm{d}y = \frac{1}{2}x\sqrt{x^2 - 4} - 2\log\left(\frac{x}{2} + \frac{1}{2}\sqrt{x^2 - 4}\right).$$

Substituting the identity $b-2ax = \sqrt{x^2 - 4}$ then gives the formula for $\min(g(x))$. □

PROOF OF PROPOSITION 2.4. For any value of $\alpha \in (0, 1)$, define

$$f_\alpha(x) := \frac{1}{2}\ln(1 - \alpha) - \frac{2x^2}{\alpha} + x^2 + x\sqrt{1 + x^2} + \sinh^{-1}(x).$$

It can be verified by computing the derivative that this function has exactly one maximum at $x_\alpha := \frac{1}{2}\frac{\alpha}{\sqrt{1-\alpha}}$ and that $f_\alpha(x_\alpha) = 0$. In particular, $f_\alpha(x) \le 0$ for all x. Now notice that we may write

$$S_G(m) = f_{m^2}\left(\sqrt{\frac{1}{2}k}(\lambda m^k)\right).$$

This shows $S_G(m) \le 0$. The consideration about the zeros of f_c shows that S_G has a zero only when $\sqrt{\frac{1}{2}k}(\lambda m^k) = \frac{1}{2}\frac{m^2}{\sqrt{1-m^2}}$, which is equivalent to equation (2.14). Elementary calculus reveals that the polynomial $m^{2k-4}(1 - m^2)$ achieves a maximum value of $\frac{(k-2)^{k-2}}{(k-1)^{(k-1)}}$, and this observation yields the desired properties for λ_c. □

Acknowledgment. S. M. was supported by an Office of Technology Licensing Stanford Graduate Fellowship. A. M. was partially supported by National Science Foundation grants DMS-1613091, CCF-1714305, and IIS-1741162. M. N. was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada.

Bibliography

[1] Adler, R. J.; Taylor, J. E. *Random fields and geometry*. Springer, New York, 2009.

[2] Anandkumar, A.; Deng, Y.; Ge, R.; Mobahi, H. Homotopy analysis for tensor PCA. Preprint, 2016. arXiv:1610.09322 [stat.ML]

[3] Anderson, G. W.; Guionnet, A.; Zeitouni, O. *An introduction to random matrices*. Cambridge University Press, 2010.

[4] Auffinger, A.; Ben Arous, G. Complexity of random smooth functions on the high-dimensional sphere. *Ann. Probab.* **41** (2013), no. 6, 4214–4247. doi:10.1214/13-AOP862

[5] Auffinger, A.; Ben Arous, G.; Černỳ, J. Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.* **66** (2013), no. 2, 165–201. doi:10.1002/cpa.21422

[6] Barak, B.; Moitra, A. Noisy tensor completion via the sum-of-squares hierarchy. *Conference on Learning Theory* (2016), 417–445.

[7] Barbier, J.; Macris, N.; Miolane, L. The layered structure of tensor estimation and its mutual information. Preprint, 2017. arXiv:1709.10368 [cs.IT]

[8] Ben Arous, G.; Dembo, A.; Guionnet, A. Aging of spherical spin glasses. *Probab. Theory Related Fields* **120** (2001), no. 1, 1–67. doi:10.1007/PL00008774

[9] Bhattiprolu, V.; Ghosh, M.; Lee, E.; Guruswami, V.; Tulsiani, M. Multiplicative approximations for polynomial optimization over the unit sphere. Preprint, 2016. arXiv:1611.05998 [cs.DS]

[10] Bhattiprolu, V.; Guruswami, V.; Lee, E. Certifying random polynomials over the unit sphere via sum of squares hierarchy. Preprint, 2016. arXiv:1605.00903 [cs.CC]

[11] Bouchaud, J.-P.; Cugliandolo, L. F.; Kurchan, J.; Mezard, M. Out of equilibrium dynamics in spin-glasses and other glassy systems. *Spin glasses and random fields*, 161–223. World Scientific, Singapore, 1998.

[12] Chen, Y.; Candès, E. J. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Comm. Pure Appl. Math.* **70** (2017), no. 5, 822–883. doi:10.1002/cpa.21638

[13] Crisanti, A.; Horner, H.; Sommers, H.-J. The spherical p-spin interaction spin-glass model. *Zeitschrift für Physik B Condensed Matter* **92** (1993), no. 2, 257–271.

[14] Crisanti, A.; Sommers, H.-J. Thouless-Anderson-Palmer approach to the spherical p-spin spin glass model. *Journal de Physique I* **5** (1995), no. 7, 805–813.

[15] Fyodorov, Y. V. Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices. *Phys. Rev. Lett.* **92** (2004), no. 24, 240601, 4 pp. doi:10.1103/PhysRevLett.92.240601

[16] Gandy, S.; Recht, B.; Yamada, I. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems* **27** (2011), no. 2, 025010, 19 pp. doi:10.1088/0266-5611/27/2/025010

[17] Ge, R.; Lee, J. D.; Ma, T. Matrix completion has no spurious local minimum. *Advances in Neural Information Processing Systems* (2016), 2973–2981.

[18] Gillin, P.; Sherrington, D. $p > 2$ spin glasses with first-order ferromagnetic transitions. *J. Phys. A* **33** (2000), no. 16, 3081–3091. doi:10.1088/0305-4470/33/16/302

[19] Guionnet, A.; Maıda, M. A Fourier view on the R-transform and related asymptotics of spherical integrals. *J. Funct. Anal.* **222** (2005), no. 2, 435–490. doi:10.1016/j.jfa.2004.09.015

[20] Hopkins, S. B.; Schramm, T.; Shi, J.; Steurer, D. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, 178–191. ACM, New York, 2016.

[21] Hopkins, S. B.; Shi, J.; Steurer, D. Tensor principal component analysis via sum-of-square proofs. *Conference on Learning Theory* (2015), 956–1006.

[22] Janzamin, M.; Sedghi, H.; Anandkumar, A. Beating the perils of non-convexity: Guaranteed training of neural networks using tensor methods. Preprint, 2015. arXiv:1506.08473 [cs.LG]

[23] Keshavan, R. H.; Montanari, A.; Oh, S. Matrix completion from a few entries. *IEEE Trans. Inform. Theory* **56** (2010), no. 6, 2980–2998. doi:10.1109/TIT.2010.2046205

[24] Kreimer, N.; Stanton, A.; Sacchi, M. D. Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction. *Geophysics* **78** (2013), no. 6, V273–V284.

[25] Ledoux, M. *The concentration of measure phenomenon*. American Mathematical Society, Providence, R.I., 2005.

[26] Lesieur, T.; Miolane, L.; Lelarge, M.; Krzakala, F.; Zdeborová, L. Statistical and computational phase transitions in spiked tensor estimation. Preprint, 2017. arXiv:1701.08010 [math.ST]

[27] Li, N.; Li, B. Tensor completion for on-board compression of hyperspectral images. *2010 IEEE International Conference on Image Processing*, 517–520. IEEE, 2010.

[28] Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (2013), no. 1, 208–220.

[29] Loh, P.-L.; Wainwright, M. J. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** (2012), no. 3, 1637–1664. doi:10.1214/12-AOS1018

[30] Maida, M. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.* **12** (2007), 1131–1150. doi:10.1214/EJP.v12-438

[31] Mei, S.; Bai, Y.; Montanari, A. The landscape of empirical risk for non-convex losses. *Ann. Statist.* **46** (2018), no. 6A, 2747–2774. doi:10.1214/17-AOS1637

[32] Montanari, A.; Reichman, D.; Zeitouni, O. On the limitation of spectral methods: from the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inform. Theory* **63** (2017), no. 3, 1572–1579. doi:10.1109/TIT.2016.2637959

[33] Montanari, A.; Richard, E. A statistical model for tensor PCA. *Advances in Neural Information Processing Systems* (2014), 2897–2905.

[34] Montanari, A.; Sun, N. Spectral algorithms for tensor completion. *Comm. Pure Appl. Math.* **71** (2018), no. 11, 2381–2425. doi:10.1002/cpa.21748

[35] Mørup, M. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1** (2011), no. 1, 24–40. doi:10.1002/widm.1

[36] Perry, A.; Wein, A. S.; Bandeira, A. S. Statistical limits of spiked tensor models. Preprint, 2016. arXiv:1612.07728 [math.PR]

[37] Ros, V.; Arous, G. B.; Biroli, G.; Cammarota, C. Complex energy landscapes in spiked-tensor and simple glassy models: ruggedness, arrangements of local minima and phase transitions. Preprint, 2018. arXiv:1804.02686 [cond-mat.dis-nn]

[38] Subag, E. The complexity of spherical p-spin models—a second moment approach. *Ann. Probab.* **45** (2017), no. 5, 3385–3450. doi:10.1214/16-AOP1139

[39] Subag, E. The geometry of the Gibbs measure of pure spherical spin glasses. *Invent. Math.* **210** (2017), no. 1, 135–209. doi:10.1007/s00222-017-0726-4

[40] Subag, E.; Zeitouni, O. The extremal process of critical points of the pure p-spin spherical spin glass model. *Probab. Theory Related Fields* **168** (2017), no. 3-4, 773–820. doi:10.1007/s00440-016-0724-2

[41] Sun, J.; Qu, Q.; Wright, J. A geometric analysis of phase retrieval. *Found. Comput. Math.* **18** (2018), no. 5, 1131–1198. doi:10.1007/s10208-017-9365-9

[42] Talagrand, M. Free energy of the spherical mean field model. *Probab. Theory Related Fields* **134** (2006), no. 3, 339–382. doi:10.1007/s00440-005-0433-8

[43] Yuan, M.; Zhang, C.-H. On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** (2016), no. 4, 1031–1068. doi:10.1007/s10208-015-9269-5

[44] Zhong, K.; Song, Z.; Jain, P.; Bartlett, P. L.; Dhillon, I. S. Recovery guarantees for one-hidden-layer neural networks. Preprint, 2017. arXiv:1706.03175 [cs.LG]

GÉRARD BEN AROUS
Courant Institute
251 Mercer St.
New York, NY 10012
USA
E-mail: benarous@cims.nyu.edu

SONG MEI
Stanford University
475 Via Ortega
Stanford, CA 94305
USA
E-mail: songmei@stanford.edu

ANDREA MONTANARI
Stanford University
350, Serra Mall
Stanford, CA 94305
USA
E-mail: montanari@stanford.edu

MIHAI NICA
University of Toronto
40 St. George St.
Toronto, ON
CANADA
E-mail: mnica@math.utoronto.ca

Received November 2017.