# Average-case matrix discrepancy: satisfiability bounds

Antoine Maillard[*]

Inria Paris & DIENS, Ecole Normale Supérieure, PSL Research University.
Department of Mathematics, ETH Zürich, Switzerland.

**Abstract**

Given a sequence of $d \times d$ symmetric matrices $\{\mathbf{W}_i\}_{i=1}^n$, and a margin $\Delta > 0$, we investigate whether it is possible to find signs $(\varepsilon_1, \dots, \varepsilon_n) \in \{\pm 1\}^n$ such that the operator norm of the signed sum satisfies $\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \Delta$. Kunisky and Zhang (2023) recently introduced a random version of this problem, where the matrices $\{\mathbf{W}_i\}_{i=1}^n$ are drawn from the Gaussian orthogonal ensemble. This model can be seen as a random variant of the celebrated Matrix Spencer conjecture, and as a matrix-valued analog of the symmetric binary perceptron in statistical physics. In this work, we establish a satisfiability transition in this problem as $n, d \to \infty$ with $n/d^2 \to \tau > 0$. Our main results are twofold. First, we prove that the expected number of solutions with margin $\Delta = \kappa\sqrt{n}$ has a sharp threshold at a critical $\tau_1(\kappa)$: for $\tau < \tau_1(\kappa)$ the problem is typically unsatisfiable, while for $\tau > \tau_1(\kappa)$ the average number of solutions becomes exponentially large. Second, combining a second-moment method with recent results from Altschuler (2023) on margin concentration in perceptron-type problems, we identify a second threshold $\tau_2(\kappa)$, such that for $\tau > \tau_2(\kappa)$ the problem admits solutions with high probability. In particular, we establish that a system of $n = \Theta(d^2)$ Gaussian random matrices can be balanced so that the spectrum of the resulting matrix macroscopically shrinks compared to the typical semicircle law. Finally, under a technical assumption, we show that there exists values of $(\tau, \kappa)$ for which the number of solutions has large variance, implying the failure of the second moment method and uncovering a richer picture than in the vector-analog symmetric binary perceptron problem. Our proofs rely on concentration inequalities and large deviation properties for the law of correlated Gaussian matrices under spectral norm constraints.

# Contents



[*] antoine.maillard@inria.fr.

# 1 Introduction and main results

## 1.1 Setting and related literature

We start by introducing in Sections 1.1.1 and 1.1.2 two sets of independent definitions and models, both being important motivations behind the problem we study, which is discussed in Section 1.1.3.

### 1.1.1 Discrepancy theory

Computing the *discrepancy* (i.e. the best possible balancing) of a collection of sets or vectors is a classical question in mathematics and theoretical computer science. It has wide-ranging applications in combinatorics, computational geometry, experimental design, and the theory of approximation algorithms, to name a few. The reader will find in Spencer (1994), Matousek (2009), and Chen, Srivastav, and Travaglini (2014) a detailed account of the history and applications of discrepancy theory. Arguably one of the most celebrated results in this area is the following theorem, known as Spencer's "six deviations suffice" (Spencer, 1985).

**Theorem 1.1** (Spencer (1985)). *There exists $C > 0$ such that for all $n, d \geq 1$, and all $\mathbf{u}_1, \cdots, \mathbf{u}_n \in \mathbb{R}^d$ with $\|\mathbf{u}_i\|_\infty \leq 1$ for all $i \in [n]$:*

$$\mathrm{disc}(\mathbf{u}_1, \cdots, \mathbf{u}_n) \coloneqq \min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{i=1}^n \varepsilon_i \mathbf{u}_i \right\|_\infty \leq \begin{cases} C\sqrt{n \max\left(1, \log \dfrac{d}{n}\right)} & \text{if } n \leq d, \\ C\sqrt{d} & \text{if } n > d. \end{cases}$$

The second case $n > d$ can be deduced from the result for $n = d$ using classical arguments based on iterated rounding, see e.g. Spencer (1994). Interestingly, Spencer's theorem shows that one can drastically improve over a naive pick of random signings $\varepsilon_1, \cdots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{\pm 1\})$. Indeed, it is not hard to show that (taking $n = d$ for simplicity) random signings achieve $\|\sum_{i=1}^n \varepsilon_i \mathbf{u}_i\|_\infty = \Theta(\sqrt{n \log n})$ for a worst-case choice of $(\mathbf{u}_i)_{i=1}^n$, with high probability as $n \to \infty$. Spencer's theorem thus implies the existence of a signing $\varepsilon \in \{\pm 1\}^n$, which depends on the value of the $\mathbf{u}_i$'s, and whose discrepancy generically improves over the one of random signs by a logarithmic factor. While the original proof of Spencer (1985) is not constructive, there has been recently a great number of results regarding efficient algorithmic constructions of these signings. We will discuss this point further in Section 1.3.

**Matrix discrepancy** – In the present work we consider the problem of *matrix discrepancy*. Given a set of $n$ symmetric $d \times d$ matrices $\mathbf{A}_1, \cdots, \mathbf{A}_n$, we aim to characterize the following discrepancy objective ($\|\mathbf{A}\|_{\mathrm{op}} := \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2$ is the spectral, or operator, norm):

$$\mathrm{disc}(\mathbf{A}_1, \cdots, \mathbf{A}_n) := \min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{i=1}^n \varepsilon_i \mathbf{A}_i \right\|_{\mathrm{op}}. \tag{1}$$

As already mentioned, a foundational question in discrepancy is how the discrepancy objective compares to the one achieved by a *random* choice of the signings $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{\pm 1\})$: matrix discrepancy is thus intimately connected to the study of large random matrices. Matrix discrepancy has also been shown to have implications in the theory of quantum random access codes (Hopkins, Raghavendra, and Shetty, 2022; Bansal, Jiang, and Meka, 2023), in generalizations of the Kadison-Singer problem (Marcus, Spielman, and Srivastava, 2015; Kyng, Luh, and Song, 2020), as well as graph sparsification (Batson, Spielman, and Srivastava, 2014) to name a few.

The celebrated "Matrix Spencer" conjecture (Zouzias, 2012; Meka, 2014) is likely the most important open problem in matrix discrepancy. It asserts that Spencer's Theorem 1.1 can be generalized to the matrix setting, as follows.

**Conjecture 1.2** (Matrix Spencer (Zouzias, 2012; Meka, 2014))**.** *There exists $C > 0$ such that for all $n, d \geq 1$, and all $\mathbf{A}_1, \cdots, \mathbf{A}_n$ symmetric $d \times d$ matrices with $\max_{i \in [n]} \|\mathbf{A}_i\|_{\mathrm{op}} \leq 1$:*

$$\mathrm{disc}(\mathbf{A}_1, \cdots, \mathbf{A}_n) \leq C \sqrt{n \max\left(1, \log \frac{d}{n}\right)}.$$

*In particular, the discrepancy is $\mathcal{O}(\sqrt{n})$ for $d = n$.*

We stress that Spencer's theorem can be seen as the special case of Conjecture 1.2 in which all $\mathbf{A}_i$ commute with each other (and are thus diagonalizable in the same basis). Moreover, a weaker form of Conjecture 1.2, with a bound $\mathcal{O}(\sqrt{n \log d})$ on the right-hand side, can easily be shown to be achievable using a random choice of signs $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{\pm 1\})$, using the non-commutative Khintchine inequality of Lust-Piquard and Pisier (1991). Despite a recent surge in efforts (Levy, Ramadas, and Rothvoss, 2017; Hopkins, Raghavendra, and Shetty, 2022; Dadush, Jiang, and Reis, 2022), Conjecture 1.2 remains open at the time of this writing. The best-known result is a proof of Matrix Spencer if we additionally assume $\mathrm{rk}(\mathbf{A}_i) \lesssim n/\log^3 n$ (Bansal, Jiang, and Meka, 2023), and is based on the recent improvements over the non-commutative Khintchine inequality of Bandeira, Boedihardjo, and van Handel (2023). In this work, we consider an average-case version of Conjecture 1.2, introduced in Section 1.1.3.

### 1.1.2 Random vector discrepancy, and the symmetric binary perceptron

A natural question in vector discrepancy (i.e. in the setting of Theorem 1.1) is to shift our attention to average-case settings, where the $\mathbf{u}_i$ are chosen randomly rather than potentially adversarially. The typical discrepancy for random vectors is now well understood: in the regime $n = \omega(d)$ (i.e. $d = o(n)$, many more signs than dimensions), Turner, Meka, and Rigollet (2020) established the following result (with partial results preceding in Karmarkar et al. (1986) and Costello (2009)):

**Theorem 1.3** (Turner, Meka, and Rigollet (2020))**.** *Assume that $\mathbf{u}_1, \cdots, \mathbf{u}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d)$ and that $n = \omega(d)$. Then*

$$\underset{d \to \infty}{\text{p-lim}} \frac{\mathrm{disc}(\mathbf{u}_1, \cdots, \mathbf{u}_n)}{\sqrt{\frac{\pi n}{2}} 2^{-n/d}} = 1,$$

*where the limit is meant in probability.*

While Spencer's Theorem 1.1 does not apply to such random vectors (as one can easily show that $\|\mathbf{u}_i\|_\infty \sim \sqrt{2\log d} \gg 1$), the randomness makes the problem amenable to a detailed mathematical analysis with different tools.

In a complementary way, recent works have characterized very precisely the minimal discrepancy in the proportional regime $n = \Theta(d)$, as a function of the "aspect ratio" $\beta \coloneqq \lim_{d\to\infty} n/d$. This setting of random vector discrepancy is an instance of the *symmetric binary perceptron* (SBP), a random constraint satisfaction problem which was introduced in Aubin, Perkins, and Zdeborová (2019) as a variant to the classical asymmetric binary perceptron. The latter is a simple model of a neural network storing random patterns, which has a long history of study in computer science, statistical physics, and probability theory (Cover, 1965; Gardner, 1988; Gardner and Derrida, 1988; Krauth and Mézard, 1989; Sompolinsky, Tishby, and Seung, 1990; Talagrand, 1999a; Talagrand, 2010). For a *margin* $K > 0$, and in the limit $n/d \to \beta > 0$, the question of satisfiability in the SBP was introduced in Aubin, Perkins, and Zdeborová (2019) as:

$$\textit{Given } \mathbf{g}_1, \cdots, \mathbf{g}_d \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_n), \textit{ can we find } \varepsilon \in \{\pm 1\}^n \textit{ such that } \max_{i\in[d]} |\langle \mathbf{g}_i, \varepsilon\rangle| \le K\sqrt{n}?$$

Letting $(\mathbf{u}_i)_k \coloneqq (\mathbf{g}_j)_k$, so that $(\mathbf{u}_i)_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d)$ it is clear that the the SBP can be thought as an average-case version of vector discrepancy, and more precisely to the setting of Theorem 1.3 with $n = \Theta(d)$. The SBP has received significant attention in the recent literature: its relative simplicity allows studying in detail the relation between its structural properties and the performance of solving algorithms, which can then guide the theory of more complex statistical models, see e.g. Aubin, Perkins, and Zdeborová (2019), Perkins and Xu (2021), Abbe, Li, and Sly (2022), Gamarnik et al. (2022), Kizildag and Wakhare (2023), Barbier et al. (2024), El Alaoui and Gamarnik (2024), and Barbier (2024). Concretely, it is shown in Aubin, Perkins, and Zdeborová (2019), Abbe, Li, and Sly (2022), and Gamarnik et al. (2022) (among other results) that the SBP undergoes the following sharp satisfiability/unsatisfiability transition.

**Theorem 1.4** (Sharp threshold for the SBP (Aubin, Perkins, and Zdeborová, 2019; Abbe, Li, and Sly, 2022))**.** *Let $K > 0$. Define*

$$\beta_1(K) \coloneqq -\frac{\log \mathbb{P}_{z\sim\mathcal{N}(0,1)}[|z| \le K]}{\log 2}.$$

*For $\mathbf{u}_1, \cdots, \mathbf{u}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d)$, let*

$$Z_K \coloneqq \left| \left\{ \varepsilon \in \{\pm 1\}^n : \left\| \sum_{i=1}^n \varepsilon_i \mathbf{u}_i \right\|_\infty \le K\sqrt{n} \right\} \right|.$$

*Then in the limit $n, d \to \infty$ with $n/d \to \beta > 0$:*

- *(i) If $\beta < \beta_1(K)$, $\lim_{d\to\infty} \mathbb{P}[Z_K \ge 1] = 0$.*
- *(ii) If $\beta > \beta_1(K)$, $\lim_{d\to\infty} \mathbb{P}[Z_K \ge 1] = 1$.*

The analysis of Aubin, Perkins, and Zdeborová (2019) is based on the second moment method, and has been refined in subsequent works (Abbe, Li, and Sly, 2022; Gamarnik et al., 2022). In particular, $\beta_1(K)$ can be derived as the critical value of $\beta$ where the value of $\mathbb{E}[Z_K]$ transitions from being exponentially small (in $n, d$) to exponentially large.

### 1.1.3 Average-case matrix discrepancy

In this work, alongside Kunisky and Zhang (2023), we initiate the study of the average-case matrix discrepancy problem. Concretely, given $n, d \ge 1$ a *margin* $\kappa > 0$, and $\mathbf{W}_1, \cdots, \mathbf{W}_n$ random independent matrices, drawn with centered Gaussian i.i.d. elements (up to symmetry) with variance $1/d$, we seek to answer the question:

$$\textbf{(P)} : \textit{Can we find signs } (\varepsilon_1, \cdots, \varepsilon_n) \in \{\pm 1\}^n \textit{ such that } \|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \kappa\sqrt{n}\textit{?}$$

This problem can be seen as an average-case analog of Conjecture 1.2, with the simple Gaussian random matrix model serving as a natural starting point. By investigating this simplified case in great detail, we firstly aim to gain insight, in order to possibly probe Conjecture 1.2 in more structured random matrix models in the future. This would form an alternative direction to other recent works (Hopkins, Raghavendra, and Shetty, 2022; Dadush, Jiang, and Reis, 2022; Bansal, Jiang, and Meka, 2023) towards a better understanding of the Matrix Spencer conjecture.

Additionally, **(P)** can naturally be thought of as the matrix discrepancy analog to the random vector discrepancy problem (or the symmetric binary perceptron) described above: a mapping can be achieved by modifying our model to diagonal matrices $\mathbf{W}_i$ with i.i.d. elements drawn from $\mathcal{N}(0,1)$ on the diagonal, and setting $(\mathbf{g}_i)_k = (\mathbf{W}_i)_{kk}$.

Importantly, we will focus on studying **(P)** in the case of a *finite margin* $\kappa > 0$ as $n, d \to \infty$. In random vector discrepancy, the satisfiability transition for a finite margin occurs in the scale $n = \Theta(d)$, – i.e. in the symmetric binary perceptron setting – see Theorem 1.4. As we will see, in our model the critical scaling is rather $n = \Theta(d^2)$, and we will focus on this regime throughout our work. Our approach to tackle **(P)** further builds on the methods introduced by Aubin, Perkins, and Zdeborová (2019) for the SBP model. Ultimately, the main goal we pursue is to obtain a detailed understanding of the satisfiability properties of **(P)**, and to reach a counterpart to Theorem 1.4 in the context of average-case matrix discrepancy.

**Parallel work** – Shortly after a first pre-print of the present manuscript was made available online, an independent study (Wengiel, 2024) appeared, exploring as well the asymptotic discrepancy of Gaussian i.i.d. matrices. Their findings extend the approach of Turner, Meka, and Rigollet (2020) for random vector discrepancy to this context, by proving a counterpart to Theorem 1.3 for random matrix discrepancy. Their results provide a sharp characterization in the regime $n = \omega(d^2)$ (i.e. $d = o(\sqrt{n})$), and $\kappa = o(1)$. Technically, Wengiel (2024) employs the second moment method, similar to our work, but relies on a general upper bound for the density of the joint law of the spectra of two correlated Gaussian matrices. In contrast, the present work focuses on the critical regime $n = \Theta(d^2)$, where the asymptotic margin $\kappa > 0$ is finite. In this critical regime, the bound used in Wengiel (2024) becomes vacuous, and we employ here different ideas to control this joint law. Combining the results of Wengiel (2024) with ours, we achieve a nearly complete description of the asymptotic discrepancy of Gaussian i.i.d. matrices, leaving open only a fraction of the phase diagram in the critical regime $n = \Theta(d^2)$, see Fig. 1.

## 1.2 Main results

### Notations and background

We denote $[d] \coloneqq \{1, \cdots, d\}$ the set of integers from 1 to $d$, and $\mathcal{S}_d$ the set of $d \times d$ real symmetric matrices. For a function $V : \mathbb{R} \to \mathbb{R}$, and $\mathbf{S} \in \mathcal{S}_d$ with eigenvalues $(\lambda_i)_{i=1}^d$, we define $V(\mathbf{S})$ as the matrix with the same eigenvectors as $\mathbf{S}$, and eigenvalues $(V(\lambda_i))_{i=1}^d$. We denote by $\|\mathbf{S}\|_{\mathrm{op}} \coloneqq \max_{i \in [d]} |\lambda_i|$ the operator norm. For any $B \subseteq \mathbb{R}$ we denote by $\mathcal{M}_1^+(B)$ the set of real probability distributions on $B$. For a probability measure $\mu \in \mathcal{M}_1^+(\mathbb{R})$ we denote $\Sigma(\mu) \coloneqq \int \mu(\mathrm{d}x)\mu(\mathrm{d}y) \log|x-y|$ its non-commutative entropy. We say that a random matrix $\mathbf{Y} \in \mathcal{S}_d$ is generated from the *Gaussian Orthogonal Ensemble* GOE$(d)$ if

$$Y_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, (1+\delta_{ij})/d) \text{ for } i \leq j.$$

The following celebrated result dates back to Wigner (1955), and is one of the foundational results of random matrix theory. For a modern proof and further results, see Anderson, Guionnet, and Zeitouni (2010).

**Theorem 1.5** (Semicircle law)**.** *Let* $\mathbf{Y} \sim \mathrm{GOE}(d)$*, with eigenvalues* $(\lambda_i)_{i=1}^d$*, and denote* $\mu_{\mathbf{Y}} := (1/d)\sum_{i=1}^d \delta_{\lambda_i}$ *its empirical eigenvalue distribution. Then, almost surely,*

$$\begin{cases} \mu_{\mathbf{Y}} & \xrightarrow[d\to\infty]{\text{weakly}} \rho_{\mathrm{s.c.}}(\mathrm{d}x) = \dfrac{\sqrt{4-x^2}}{2\pi}\mathbb{1}\{|x| \le 2\}, \\ \|\mathbf{Y}\|_{\mathrm{op}} = \max\limits_{i\in[d]} |\lambda_i| & \xrightarrow[d\to\infty]{} 2. \end{cases}$$

$\rho_{\mathrm{s.c.}}$ in Theorem 1.5 is often called the *semicircle law*. Finally, we generically denote constants as $C > 0$ (or $C_1 > 0, C_2 > 0, \cdots$), whose value may vary from line to line. We will specify their possible dependency on parameters of the problem when relevant.

#### 1.2.1 Setting of the problem

Let $n, d \ge 1$, and $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$. For any $\kappa > 0$, we define:

$$Z_\kappa := \#\left\{\varepsilon \in \{\pm 1\}^n \text{ s.t. } \left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \le \kappa\sqrt{n}\right\}. \tag{2}$$

We refer to $n^{-1/2}\,\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}}$ as the *margin* of $\varepsilon \in \{\pm 1\}^n$. $Z_\kappa$ thus counts the number of signings $\varepsilon \in \{\pm1\}^n$ with margin at most $\kappa$.

**The case $\kappa > 2$** – Clearly $(1/\sqrt{n})\sum_{i=1}^n \mathbf{W}_i \sim \mathrm{GOE}(d)$, so that (with $\varepsilon_i = 1$ for all $i \in [n]$), for any $\kappa > 2$,

$$\mathbb{P}[Z_\kappa \ge 1] \ge \mathbb{P}_{\mathbf{W}\sim\mathrm{GOE}(d)}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa] \overset{\text{(a)}}{=} 1 - o_d(1), \tag{3}$$

using Theorem 1.5 in (a). Notice that this bound is also the one given by a random choice of $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{\pm1\})$ (independently of $\{\mathbf{W}_i\}_{i=1}^n$), as again in this case $n^{-1/2}\sum_i \varepsilon_i \mathbf{W}_i \sim \mathrm{GOE}(d)$. In what follows, we therefore focus on the (interesting) regime $\kappa \in (0, 2]$, in which a choice of signings *dependent on the* $\mathbf{W}_i$ must be made in order to get a solution with margin at most $\kappa$. Notice that this argument implies that Conjecture 1.2 holds straightforwardly for i.i.d. $\mathrm{GOE}(d)$ matrices. For this reason, our interest in this problem stems primarily from its interpretation as a constraint satisfaction problem (e.g. as a matrix-analog of the SBP) and its connection to questions in random matrix theory, particularly in the critical regime $n = \Theta(d^2)$. Nevertheless, we also view our understanding of the discrepancy of i.i.d. $\mathrm{GOE}(d)$ matrices as a foundational step toward exploring more structured random matrix models, for which Conjecture 1.2 is significantly more complex and presents an exciting avenue for future research.

#### 1.2.2 Asymptotics of the first moment

Define, for $\kappa \in (0, 2]$:

$$\tau_1(\kappa) := \frac{1}{\log 2}\left[-\frac{\kappa^4}{128} + \frac{\kappa^2}{8} - \frac{1}{2}\log\frac{\kappa}{2} - \frac{3}{8}\right]. \tag{4}$$

Notice that $\tau_1(2) = 0$, and $\tau_1(\kappa) \to +\infty$ as $\kappa \downarrow 0$. Our first main result is the following sharp asymptotics for the expectation of $Z_\kappa$.

**Theorem 1.6** (Asymptotics of the first moment)**.** *Let* $\kappa \in (0, 2]$*, and* $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$*. Assume* $n/d^2 \to \tau \in [0, \infty)$ *as* $n, d \to \infty$*. Then*

$$\lim_{d\to\infty} \frac{1}{d^2}\log \mathbb{E}Z_\kappa = (\tau - \tau_1(\kappa))\log 2, \tag{5}$$

*where $Z_\kappa$ is defined in eq. (2). In particular, if $\tau < \tau_1(\kappa)$, then*

$$\lim_{d\to\infty} \mathbb{P}[Z_\kappa = 0] = \lim_{d\to\infty} \mathbb{P}\left[\min_{\varepsilon\in\{\pm1\}^n} \left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} > \kappa\sqrt{n}\right] = 1.$$

In Section 2 we carry out the proof of Theorem 1.6. It relies on Proposition 2.1, which establishes the large deviations of the operator norm of a GOE($d$) matrix, in the scale $d^2$. This is a consequence of now-classical results on the large deviations of the empirical measure in so-called $\beta$-matrix models (Ben Arous and Guionnet, 1997; Anderson, Guionnet, and Zeitouni, 2010), which yield the large deviation rate function in a variational form. Further, using results of logarithmic potential theory (Saff and Totik, 2013) and the Tricomi theorem (Tricomi, 1985), we are able to then solve this variational principle, and its outcome gives eq. (4). As a byproduct, we obtain the limiting spectral measure of a matrix $\mathbf{W} \sim \mathrm{GOE}(d)$ constrained on the event $\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa$, see Theorem 2.2.

**Relation to Kunisky and Zhang (2023)** – Theorem 1.6 is a refinement of Theorem 1.13 of Kunisky and Zhang (2023), which shows that $Z_\kappa = 0$ with high probability if $\kappa \leq \delta \cdot 4^{-\tau}$, for an (unspecified) absolute constant $\delta > 0$. For instance, the bound of Kunisky and Zhang (2023) correctly predicts $\tau_1(\kappa) \sim -\log\kappa/\log 4$ for $\kappa \to 0$, but fails to capture that $\tau_1(\kappa) \to 0$ continuously as $\kappa \to 2$.

### 1.2.3 The satisfiability region

We start by introducing a function $\tau_2(\kappa)$, which will serve as a threshold for the validity of our satisfiability analysis. Let $H(p) := -p\log p - (1-p)\log(1-p)$ denote the "binary entropy" function.

**Proposition 1.7.** *For any $\eta > 0$, let $\delta_\eta \in (0,1)$ to be the unique solution to:*

$$H\left(\frac{1+\delta}{2}\right) = \frac{\eta}{1+\eta}\log 2. \tag{6}$$

*For any $\eta > 0$, let*

$$\widetilde{\tau}(\eta,\kappa) := \max\left\{(1+\eta)\tau_1(\kappa), \frac{1+\delta_\eta^2}{2(1-\delta_\eta^2)^2} + \left[\frac{\delta_\eta(1+6\delta_\eta+3\delta_\eta^2+2\delta_\eta^3)}{(1-\delta_\eta^2)^3(1-\delta_\eta)}\right]\kappa \right.$$
$$\left. + \left[\frac{2(1+\delta_\eta)^5}{(1-\delta_\eta^2)^4} - \frac{(1+3\delta_\eta^2)}{4(1-\delta_\eta^2)^3}\right]\kappa^2 + \frac{\kappa^4(1+3\delta_\eta^2)}{32(1-\delta_\eta^2)^3}\right\}. \tag{7}$$

*We define*

$$\tau_2(\kappa) := \min_{u\in[0,\kappa]}\min_{\eta>0}\widetilde{\tau}(\eta,u). \tag{8}$$

*Then $\kappa \mapsto \tau_2(\kappa)$ is a continuous and non-increasing function of $\kappa$.*

Proposition 1.7 is elementary, we prove it in Section 3.1 for completeness, along with a straightforward way to evaluate numerically $\tau_2(\kappa)$, see eq. (28). We are now ready to state our main result on the existence of solutions with a given required margin. It is based on the second moment method, building upon similar techniques to the ones used in the symmetric binary perceptron (Aubin, Perkins, and Zdeborová, 2019).

**Theorem 1.8** (Satisfiability region)**.** *Let $\kappa \in (0,2]$. Let $n, d \geq 1$, such that, as $d \to \infty$, $n/d^2 \to \tau > \tau_2(\kappa)$ defined in eq. (8). For $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, we have (recall the definition of $Z_\kappa$ in eq. (2)):*

$$\lim_{d\to\infty} \mathbb{P}[Z_\kappa \geq 1] = \lim_{d\to\infty} \mathbb{P}\left[\min_{\varepsilon\in\{\pm1\}^n} \left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n}\right] = 1.$$

In Section 3, we prove Theorem 1.8. We establish concentration of $Z_\kappa$ in Proposition 3.2, upper bounding $\mathbb{E}[Z_\kappa^2]/\mathbb{E}[Z_\kappa]^2$ by some (large) constant as $d \to \infty$. Our proof relies on a discrete analog of Laplace's method, combined with showing concentration results (via a log-Sobolev inequality) for the distribution of correlated Gaussian matrices under spectral norm constraints. We finally strengthen the result of Proposition 3.2 thanks to the general techniques on sharp transitions for integer feasibility problems developed in Altschuler (2023), and deduce Theorem 1.8. Let us emphasize that having access to two-sided bounds on the first moment asymptotics (which are here sharp, and given in Theorem 1.6) is crucial in order to develop our second moment analysis.

#### 1.2.4 Failure of the second moment method

Finally, under a technical assumption, we show that in average-case matrix discrepancy and for some values of the parameters $(\tau, \kappa)$, the number of solutions $Z_\kappa$ can be both large in expectation and have large variance, indicating the failure of the second moment method approach.

**Theorem 1.9** (Failure of the second moment method in part of the phase diagram)**.** *Assume that Hypothesis 4.1 holds. For $\kappa \in (0, 2]$, let*

$$\tau_{\text{fail.}}(\kappa) := \frac{1}{2}\left(\frac{\kappa^2}{4} - 1\right)^4. \tag{9}$$

*Then for $n, d \to \infty$ with $n/d^2 \to \tau$, if $\tau < \tau_{\text{fail.}}(\kappa)$:*

$$\liminf_{d\to\infty} \frac{1}{d^2} \log \frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} > 0. \tag{10}$$

Theorem 1.9 relies on a local analysis of what is referred to as a second moment potential. The technical Hypothesis 4.1 pertains to controlling the second derivative of this potential uniformly with respect to $d$. In Section 4, within the proof of Theorem 1.9, we explain why this assumption is plausible and outline approaches toward a potential proof, noting significant technical challenges that we defer to future work. In Section 1.3, we further discuss the implications of Theorem 1.9 in relation to our other primary results.

### 1.3 Discussion and consequences

In Figure 1, we plot a sketch of the phase diagram of the problem, as established by Theorems 1.6,1.8 and 1.9. Our results characterize a large part of the $(\kappa, \tau)$ phase diagram: we discuss in the following some important consequences, and highlight open problems and research directions arising from our analysis.

**Balancing $\Theta(d^2)$ random matrices –** An immediate consequence of Theorem 1.8 is the following corollary.

**Corollary 1.10.** *For any $\tau > 0$ large enough*[1]*, if $n, d \to \infty$ with $n/d^2 \to \tau$, and letting $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, then (with high probability) there exists $\varepsilon \in \{\pm 1\}^n$ with margin*

$$\frac{1}{\sqrt{n}} \left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\text{op}} \leq \kappa_c(\tau) < 2.$$

*Furthermore, $\kappa_c(\tau) \to 0$ as $\tau \to \infty$.*

Corollary 1.10 establishes that for $n$ large enough but still in the scale $n = \Theta(d^2)$, one can find solutions with margin arbitrarily close to 0. As far as we know, our result is the first proof

[1] Numerically, we find $\tau \geq 6$ is enough, see Fig. 1.

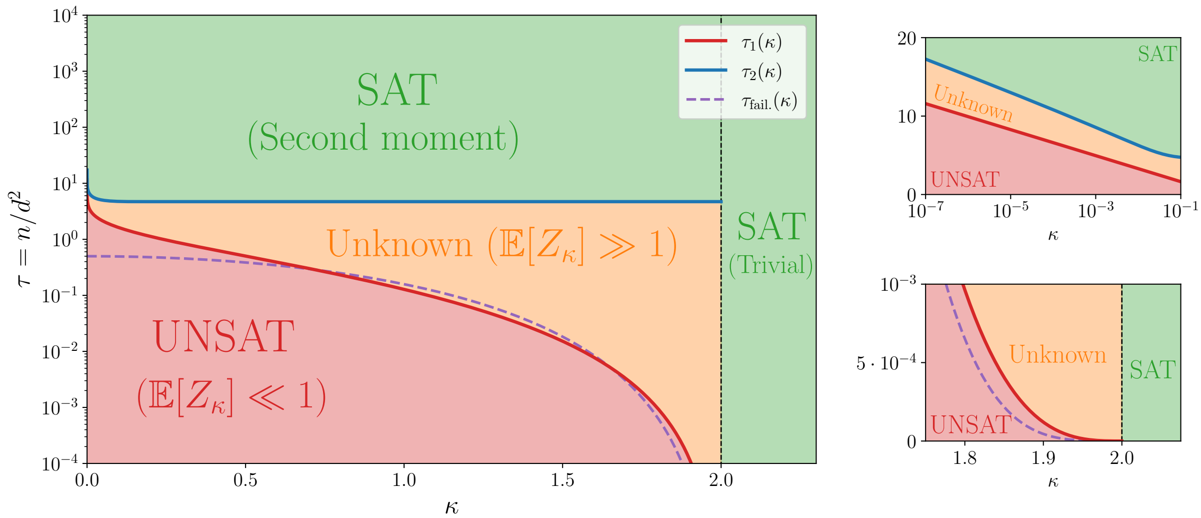


Figure 1: Sketch of the satisfiable (SAT) and unsatisfiable (UNSAT) regimes in average-case matrix discrepancy, as proven by Theorems 1.6 and 1.8. The border of the SAT region for $\kappa < 2$ is given by $\tau_2(\kappa)$, see Proposition 1.7. Numerically, we find $\tau_2(\kappa \uparrow 2) \simeq 5.67$. For $\kappa > 2$, the problem trivially admits a solution, see eq. (3). The orange region is not characterized by our results, and remains open. The dotted purple line shows $\tau_{\text{fail.}}(\kappa)$: according to Theorem 1.9, for $\tau_1(\kappa) < \tau < \tau_{\text{fail.}}(\kappa)$ the number of solutions has large expectation and variance, and the second moment method fails (see also Fig. 2). The right plots show the limits $\kappa \downarrow 0$ (top) and $\kappa \uparrow 2$ (bottom). We emphasize that $\tau_1(\kappa), \tau_2(\kappa) \to +\infty$ as $\kappa \downarrow 0$.

that a number $n = \Theta(d^2)$ of Gaussian random matrices can be balanced in a way to make their spectrum *macroscopically shrink* compared to the typical spectrum of a Gaussian matrix: it solves an open problem posed by Kunisky and Zhang (2023).

**Tightness of the second moment analysis** – Similarly to Aubin, Perkins, and Zdeborová (2019) (and subsequent works) for the symmetric binary perceptron (SBP), we use a second moment approach to characterize the feasibility of this problem. However, while this method gives a tight threshold for the SBP in Theorem 1.4 (and thus leaves no “Unknown” region in the SBP counterpart to Figure 1), the situation in average-case matrix discrepancy is quite different.

First, the threshold $\tau_2(\kappa)$ of eq. (8) is likely not optimal, as the upper bounds on $\mathbb{E}[Z_\kappa^2]$ proven in Section 3 are not expected to be tight. For this reason, the second moment ratio $\mathbb{E}[Z_\kappa^2]/\mathbb{E}[Z_\kappa]^2$ might still be bounded by a constant (cf. Proposition 3.2) even for some values $\tau \le \tau_2(\kappa)$. For instance, one might conjecture that this includes values of $\tau$ arbitrarily close to 0 when $\kappa$ approaches 2, which is not captured by our current bounds. Improving the estimates we obtain in Section 3 to obtain a sharp study of the range of parameters $(\tau, \kappa)$ with bounded second moment ratio is significantly more complex than in the SBP setting: it requires a precise understanding of the large deviation properties of the law of $(\|\mathbf{W}_1\|_{\text{op}}, \|\mathbf{W}_2\|_{\text{op}})$, where $\mathbf{W}_1, \mathbf{W}_2$ are two correlated $\text{GOE}(d)$ matrices, both conditioned on having small spectral norm (see eq. (34)). A rate function for this law can be obtained in a variational form (see e.g. Theorem 3.3 of Guionnet (2004)), however a numerical analysis of this variational formula is a very challenging problem, which we leave as a future direction.

**Open Question 1.1** (Sharp second moment)**.** Obtain the sharp limit of $(1/n)\log \mathbb{E}[Z_\kappa^2]$, the exponential scale of the second moment. This should likely rely on computing numerically $(i)$ the large-$d$ limit of the large deviations rate function of eq. (34), and $(ii)$ the asymptotic spectral density of two *correlated* $\text{GOE}(d)$ matrices, conditioned to both have spectral norm at most $\kappa$.

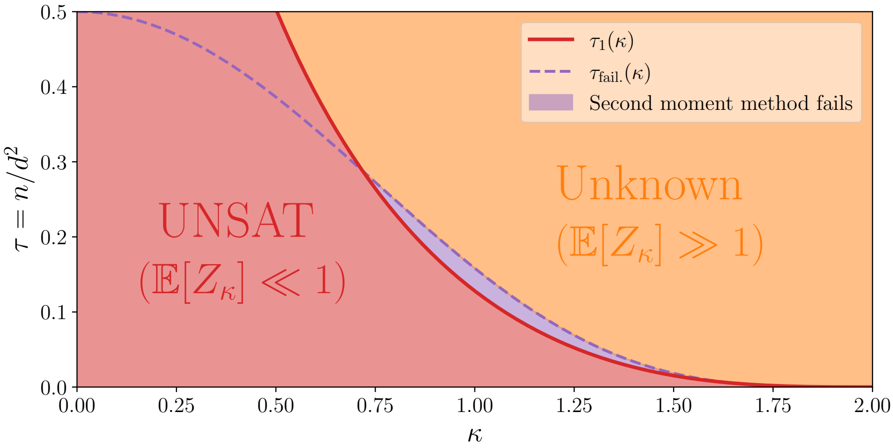


Figure 2: Illustration of Theorem 1.9. There exists a region of parameters (in purple) $\tau \in (\tau_1(\kappa), \tau_c(\kappa))$ for which the number of solutions $Z_\kappa$ to average-case matrix discrepancy satisfies both $\mathbb{E}[Z_\kappa] \gg 1$ and $\mathbb{E}[Z_\kappa^2] \gg (\mathbb{E}[Z_\kappa])^2$, so that the second moment method fails at characterizing the feasibility of the problem.

Furthermore, and very interestingly, we show in Theorem 1.9 that (assuming Hypothesis 4.1) the second moment method does not yield a sharp satisfiability threshold in average-case matrix discrepancy. Indeed, there exists a range of $\kappa \in (0, 2)$ such that $\tau_1(\kappa) < \tau_c(\kappa)$, so that if $\tau \in (\tau_1(\kappa), \tau_c(\kappa))$, then both $\lim(1/d^2) \log \mathbb{E}[Z_\kappa] > 0$ and $\lim\inf(1/d^2) \log \mathbb{E}[Z_\kappa^2]/\mathbb{E}[Z_\kappa]^2 > 0$. For such values of $(\tau, \kappa)$, the variance of $Z_\kappa$ is thus exponentially large (in $d^2$), and an approach based solely on the second moment method will fail at characterizing the feasibility of the problem. Numerically, we find that $\tau_1(\kappa) < \tau_c(\kappa)$ in the range $0.718 \lesssim \kappa \lesssim 1.652$. We summarize these findings in Fig. 2.

We stress that the result of Theorem 1.9 is in stark contrast with the symmetric binary perceptron, in which the second moment method succeeds in the entirety of the phase diagram (Theorem 1.4). In particular, this unveils a potentially richer picture for the geometry of the solution space than in the SBP. We note that, given the proof of Theorem 1.9 (see Section 4), we do not expect the second moment to be large because of spurious rare events, in which case one can often still apply the second moment method by conditioning away from these rare events (see e.g. Janson (1996), and an application in random vector discrepancy in Altschuler and Niles-Weed (2022)). As such, we believe that sharply characterizing the satisfiability threshold would require other techniques, perhaps by directly tackling the typical number of solutions $(1/d^2)\mathbb{E}[\log Z_\kappa]$ (what is called the "quenched" free energy in the statistical physics jargon). We leave such a challenging investigation for a future work. Finally, we emphasize that the failure of the second moment method might extend to other parts of the "Unknown" region of the phase diagram (beyond the purple region of Fig. 2) which are not captured by the local analysis used in Theorem 1.9: we refer to Section 4 for details.

**Sharp threshold sequence** – Theorem 7 of Altschuler (2023) (see also Lemma 3.3) directly implies the existence of a sharp threshold sequence for average-case matrix discrepancy. Using Theorems 1.6 and 1.8, we can locate it in the closure of the "Unknown region" of Fig. 1.

**Corollary 1.11** (Sharp threshold sequence)**.** *Let $n, d \geq 1$, and assume $n/d^2 \to \tau > 0$ as $d \to \infty$. There exists a sequence $\kappa_c(d, \tau)$ such that:*

*(i) For any $\varepsilon > 0$, $\mathbb{P}[Z_{\kappa_c - \varepsilon} \geq 1] = o_d(1)$, while $\mathbb{P}[Z_{\kappa_c + \varepsilon} \geq 1] = 1 - o_d(1)$.*

*(ii) $\tau_1(\kappa_c) \leq \tau \leq \tau_2(\kappa_c)$, i.e. $(\kappa_c, \tau)$ is in the closure of the "Unknown" region of Fig. 1.*

The existence of a sharp threshold sequence has been established recently in a variety of other perceptron-like problems, see for instance Talagrand (1999b), Talagrand (2011), Xu (2021), Nakajima and Sun (2023), and Altschuler (2023). Corollary 1.11 shows the existence of a sharp threshold $\kappa_c$ depending on $d$, and bounds it in an interval of size $\Theta(1)$. We further conjecture $\kappa_c(d, \tau)$ to converge to a well-defined limit, as we plan to discuss in a future work.

**Open Question 1.2** (Sharp SAT/UNSAT transition)**.** Obtain the sharp limit of $(1/d^2)\mathbb{E}\log Z_\kappa$, and from it a sharp SAT/UNSAT threshold, closing the gap in Fig. 1. In particular, compare it with the success of the second moment method (see Open Problem 1.1).

**Structure of the solution space** – Beyond satisfiability, a natural question concerns the geometric structure of the space of solutions $\{\varepsilon \in \{\pm 1\}^n : \|\sum_i \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \kappa\sqrt{n}\}$. In the symmetric binary perceptron, the geometric structure of the solution space was investigated by Aubin, Perkins, and Zdeborová (2019), Perkins and Xu (2021), and Abbe, Li, and Sly (2022), and was shown to exhibit a "frozen-1RSB" structure: that is, typically, solutions are isolated and far apart.

**Open Question 1.3** (Structure of the set of solutions)**.** Elucidate whether the solution space (in the SAT region of Fig. 1 for $\kappa < 2$) exhibits the "frozen-1RSB" property.

In the SBP it was the second moment analysis that suggested the existence of freezing (Aubin, Perkins, and Zdeborová, 2019), however in the present context two difficulties arise: $(i)$ this argument usually relies on establishing a *contiguity* property with a "planted" version of the model, and $(ii)$ even under contiguity, the second moment upper bound developed in Section 3 is unfortunately not sharp enough to determine whether freezing occurs, even withing the SAT region of Fig. 1. We discuss points $(i)$ and $(ii)$ above in further details in Appendix A. Going beyond, the failure of the second moment method (discussed above) may also signal a different geometric structure than in the SBP, at least in parts of the phase diagram.

**Algorithmic discrepancy** – The past decade has seen a surge of interest in *algorithmic discrepancy*, that is the design of efficient algorithms to produce signings $\varepsilon \in \{\pm 1\}^n$ minimizing a discrepancy objective such as eq. (1). In the classical context of vector discrepancy, this line of work was initiated by Bansal (2010), and we refer to Bansal, Jiang, and Meka (2023) and Kunisky and Zhang (2023) for a more detailed description of the literature that followed. For the problem of average-case matrix discrepancy, Kunisky and Zhang (2023) analyze an online algorithm, introduced originally by Zouzias (2012), and show that is able to achieve a discrepancy $\|\sum_i \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \lesssim d \log(n + d)$ for a large class of random matrices $\mathbf{W}_i$, including the GOE distribution. In the regime $n = \Theta(d^2)$, this bound unfortunately falls short of obtaining a discrepancy lower than $2\sqrt{n}$.

**Open Question 1.4** (Efficient algorithms)**.** Is there a polynomial-time algorithm which, in the regime $n = \Theta(d^2)$ and with high probability, outputs $\varepsilon \in \{\pm 1\}^n$ such that $\|\sum_i \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \kappa\sqrt{n}$ for some $\kappa < 2$?

For the symmetric binary perceptron, which is the vector analog to our problem, it was recently shown that the phase diagram presents a large computational-to-statistical gap in which low-discrepancy solutions exist, while large classes of polynomial-time algorithms can not find them, and that this was related to the geometric structure of the solution space (Bansal and Spencer, 2020; Gamarnik et al., 2022).

## Acknowledgements

The author wants to thank Dylan Altschuler for discussions and a careful reading of a first version of this manuscript, Tim Kunisky and Afonso S. Bandeira for insightful discussions about the early stages of this work, and anonymous reviewers for helpful comments and suggestions.

## 2 Sharp asymptotics of the first moment

We carry out in this section the proof of Theorem 1.6. A direct computation from eq. (2) using the linearity of expectation yields:

$$\mathbb{E}Z_\kappa = \sum_{\varepsilon \in \{\pm 1\}^n} \mathbb{P}\left[\left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n}\right] = 2^n\, \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa], \tag{11}$$

where $\mathbf{W} \sim \mathrm{GOE}(d)$. The main result we prove is the following.

**Proposition 2.1** (Left large deviations for the operator norm of a $\mathrm{GOE}(d)$ matrix)**.** *Let $\mathbf{W} \sim \mathrm{GOE}(d)$. For any $\kappa > 0$:*

$$\lim_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] = \begin{cases} \frac{\kappa^4}{128} - \frac{\kappa^2}{8} + \frac{1}{2}\log\frac{\kappa}{2} + \frac{3}{8} & \text{if } \kappa \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Using Proposition 2.1 in eq. (11) (recall that $n/d^2 \to \tau$) yields eq. (5). The second result of Theorem 1.6 is a direct consequence of Markov's inequality combined with eq. (5). Notice that Markov's inequality even shows that $\mathbb{P}[Z_\kappa > 0]$ goes to zero exponentially fast in $d^2$ for $\tau < \tau_1(\kappa)$.

In the remainder of Section 2, we focus on proving Proposition 2.1. We note first that for $\kappa > 2$ we have $\log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] = o_d(1)$ as a consequence of Theorem 1.5. We thus focus on the case $\kappa \in (0, 2]$ in what follows.

**Sketch of proof and important related work** – The proof of Proposition 2.1 builds upon a large deviation principle (with respect to the weak topology) for the empirical spectral measure of a matrix $\mathbf{W}$ drawn from $\mathrm{GOE}(d)$, *conditioned* on $\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa$. We study the conditioned law directly, since it is a so-called $\beta$-ensemble (albeit with a singular potential enforcing the spectral norm constraint), so that the large deviation analysis follows from the general results stated in Proposition 2.3. As a direct consequence, we obtain in Corollary 2.4 the asymptotics of eq. (12) as a variational principle over probability measures supported in $[-\kappa, \kappa]$. We note that one could rely solely on the original large deviations principles for (unconditioned) $\mathrm{GOE}(d)$ matrices proven in Ben Arous and Guionnet (1997), however this requires additional care because the set of probability measures supported in $[-\kappa, \kappa]$ has empty interior under the weak topology: for completeness, we also provide in Appendix B a proof of Corollary 2.4, that appeared in an earlier version of this work, and which uses only the result of Ben Arous and Guionnet (1997). We solve the resulting variational principle using the theory of logarithmic potentials (Saff and Totik, 2013) and Tricomi's theorem (Tricomi, 1985), similarly to the alternative proof of Wigner's semicircle law obtained by Ben Arous and Guionnet (1997) from their large deviations result. Similar arguments also appeared in the theoretical physics literature (Dean and Majumdar, 2006; Vivo, Majumdar, and Bohigas, 2007; Dean and Majumdar, 2008; Majumdar and Schehr, 2014). We refer to Anderson, Guionnet, and Zeitouni (2010) and Guionnet (2022) for more background and open problems in the theory of large deviations for random matrices.

**Remark** – The following result is a byproduct of our analysis[1].

**Theorem 2.2** (Limiting spectral density of a constrained $\mathrm{GOE}(d)$ matrix)**.** *For $\kappa \in (0, 2]$, denote $\mathbb{P}_\kappa$ the law of $\mathbf{W} \sim \mathrm{GOE}(d)$ conditioned on $\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa$. If $\mathbf{W} \sim \mathbb{P}_\kappa$, then its empirical spectral density $\mu_{\mathbf{W}}$ converges weakly (as $d \to \infty$, and a.s.) to $\mu_\kappa(\mathrm{d}x) := \rho_\kappa(x)\mathrm{d}x$ given, for $x \in (-\kappa, \kappa)$, by:*

$$\rho_\kappa(x) := \frac{4 + \kappa^2 - 2x^2}{4\pi\sqrt{\kappa^2 - x^2}}. \tag{13}$$

[1] After the present work was made available as a pre-print, the author realized that another proof of Theorem 2.2 had appeared in an unpublished manuscript (Bouali, 2015).

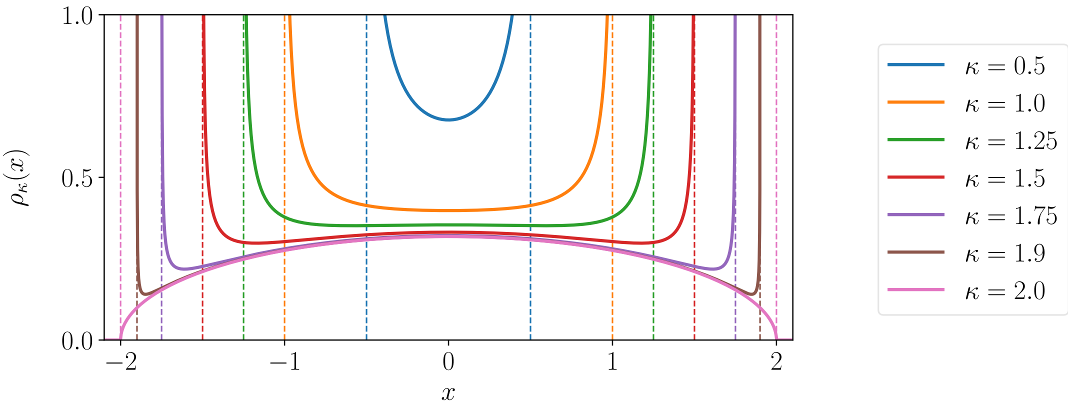


Figure 3: $\rho_\kappa(x)$ of eq. (13) for different values of $\kappa$. For $\kappa = 2$ one recovers the semicircle law.

*Notice that* $\rho_2(x) = \sqrt{4 - x^2}/(2\pi)$ *is Wigner's semicircle law, see Theorem 1.5.*

We illustrate the form of $\rho_\kappa(x)$ in Figure 3. Theorem 2.2 is proven in Section 2.3.

Let us denote $S_\kappa := \{\varepsilon \in \{\pm 1\}^n \text{ s.t. } \|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \kappa\sqrt{n}\}$, so that $Z_\kappa = \# S_\kappa$. Anticipating our satisfiability results, we note that even if $\mathbb{P}[Z_\kappa \geq 1] = 1 - o_d(1)$ as shown in Theorem 1.8 for sufficiently large $\tau = n/d^2$, Theorem 2.2 does not ensure that the asymptotic spectrum of $\sum_i \varepsilon_i \mathbf{W}_i$ (for $\varepsilon \sim \mathrm{Unif}(S_\kappa)$) equals $\rho_\kappa$. However, one can easily prove that this would be implied e.g. by the condition $\mathbb{E}[Z_\kappa]^2 = (1 + o_d(1))\mathbb{E}[Z_\kappa]^2$, which is strictly stronger than $\mathbb{P}[Z_\kappa \geq 1] = 1 - o_d(1)$. Unfortunately, the bounds established in Section 3 to prove Theorem 1.8 only yield $\mathbb{E}[Z_\kappa]^2 \leq C\mathbb{E}[Z_\kappa]^2$ in the satisfiable phase, for a possibly large $C > 1$. For this reason, we leave the establishment of the limiting spectral density of $\sum_i \varepsilon_i \mathbf{W}_i$ (for $\varepsilon \sim \mathrm{Unif}(S_\kappa)$) as an open question, with $\rho_\kappa$ a natural conjecture.

### 2.1 Large deviations results

Our proof leverages a large deviation principle for the empirical eigenvalue distribution in a large class of matrix models. Such results were initiated in the mathematics literature by Ben Arous and Guionnet (1997) in the case of GOE($d$) matrices, and have been then extended, see e.g. Guionnet (2022) for a discussion.

The space $\mathcal{M}_1^+(\mathbb{R})$ of probability measures on $\mathbb{R}$ is endowed with its usual weak topology. It is metrizable by the Dudley metric (Bogachev and Ruas, 2007):

$$d(\mu, \nu) := \left\{\left|\int f \mathrm{d}\mu - \int f \mathrm{d}\nu\right| \,:\, |f(x)| \leq 1 \text{ and } |f(x) - f(y)| \leq |x - y|, \, \forall (x, y) \in \mathbb{R}^2\right\}. \quad (14)$$

Recall that $\Sigma(\mu) := \int \mu(\mathrm{d}x)\mu(\mathrm{d}y) \log|x - y|$. The following statement is borrowed from Fan et al. (2015) (variants can be found e.g. in Ben Arous and Guionnet (1997) and Anderson, Guionnet, and Zeitouni (2010)).

**Proposition 2.3** (Fan et al. (2015))**.** *Let $B \subseteq \mathbb{R}$ an interval, and $V : B \to \mathbb{R}$ a continuous function such that $\lim_{x \to \pm\infty} \frac{V(x)}{2\log|x|} = +\infty$ (if $B$ is not bounded). For $\boldsymbol{\lambda} \in \mathbb{R}^d$, let*

$$\mu_d^{V,B}(\mathrm{d}\boldsymbol{\lambda}) := \frac{1}{Z_d^{V,B}} \prod_{i=1}^d (\mathrm{d}\lambda_i \, \mathbb{1}\{\lambda_i \in B\}) \, e^{-\frac{d}{2}\sum_{i=1}^d V(\lambda_i)} \prod_{i<j} |\lambda_i - \lambda_j|, \quad (15)$$

*with $Z_d^{V,B}$ a normalization factor, which ensures that $\int \mu_d^{V,B}(\mathrm{d}\boldsymbol{\lambda}) = 1$.*
*Denote $\nu_{\boldsymbol{\lambda}} := (1/d)\sum_{i=1}^d \delta_{\lambda_i}$ the empirical measure of $\boldsymbol{\lambda}$. Then the law of $\nu_{\boldsymbol{\lambda}}$ satisfies a large*

*deviation principle, in the scale $d^2$, with good rate function $I_{V,B}(\nu) := \mathcal{E}_V(\nu) - \inf_{\nu \in \mathcal{M}_1^+(B)}[\mathcal{E}_V(\nu)]$, where*

$$\mathcal{E}_V(\nu) := -\frac{1}{2}\Sigma(\nu) + \frac{1}{2}\int V(x)\,\nu(\mathrm{d}x). \tag{16}$$

*In particular:*

*(i) For any $M > 0$, $\{I_{V,B} \leq M\}$ is a compact subset of $\mathcal{M}_1^+(B)$.*

*(ii) For any open (respectively closed) subset $O \subseteq \mathcal{M}_1^+(B)$ (respectively $F \subseteq \mathcal{M}_1^+(B)$):*

$$\begin{cases} \liminf_{d\to\infty} \frac{1}{d^2} \log \mu_d^{V,B}[\nu_{\boldsymbol{\lambda}} \in O] \geq -\inf_{\nu \in O} I_{V,B}(\nu), \\ \limsup_{d\to\infty} \frac{1}{d^2} \log \mu_d^{V,B}[\nu_{\boldsymbol{\lambda}} \in F] \leq -\inf_{\nu \in F} I_{V,B}(\nu). \end{cases}$$

*(iii)*

$$\lim_{d\to\infty} \frac{1}{d^2} \log Z_d^{V,B} = -\inf_{\nu \in \mathcal{M}_1^+(B)} [\mathcal{E}_V(\nu)].$$

Proposition 2.3 directly applies to $\mathbb{P}_\kappa$, the law of $\mathbf{W} \sim \mathrm{GOE}(d)$ conditioned on $\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa$, with $V(x) = x^2/2$. We obtain from it the following variational formulation for the limit of the left-hand side of eq. (12).

**Corollary 2.4.** *For any $\kappa > 0$, with $\mathbf{W} \sim \mathrm{GOE}(d)$:*

$$\lim_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] = \frac{3}{8} - \inf_{\nu \in \mathcal{M}_1^+([-\kappa,\kappa])} \left[-\frac{1}{2}\Sigma(\nu) + \frac{1}{4}\int \nu(\mathrm{d}x)\, x^2\right].$$

*Proof of Corollary 2.4.* The joint law of the eigenvalues $(\lambda_1, \cdots, \lambda_d)$ of $\mathbf{W}$ is well-known thanks to the rotation invariance of the law of $\mathbf{W}$. We have (see e.g. Theorem 2.5.2 of Anderson, Guionnet, and Zeitouni (2010)):

$$\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] = \frac{\int_{[-\kappa,\kappa]^d} \prod_{i<j} |\lambda_i - \lambda_j| e^{-\frac{d}{4}\sum_{i=1}^d \lambda_i^2} \prod_{i=1}^d \mathrm{d}\lambda_i}{\int_{\mathbb{R}^d} \prod_{i<j} |\lambda_i - \lambda_j| e^{-\frac{d}{4}\sum_{i=1}^d \lambda_i^2} \prod_{i=1}^d \mathrm{d}\lambda_i}. \tag{17}$$

The denominator (or partition function) can be computed from Selberg's integrals (Mehta, 2014). Its limit is given by

$$\lim_{d\to\infty} \frac{1}{d^2} \log \int_{\mathbb{R}^d} \prod_{i<j} |\lambda_i - \lambda_j| e^{-\frac{d}{4}\sum_{i=1}^d \lambda_i^2} \prod_{i=1}^d \mathrm{d}\lambda_i = -\frac{3}{8}. \tag{18}$$

The result follows then from Proposition 2.3-$(iii)$. □

**Alternative proof –** For reasons of completeness, and as a first version of this manuscript contained it, we detail in Appendix B a proof of Corollary 2.4 that does not rely on Proposition 2.3, but only on its particular case $B = \mathbb{R}$ and $V(x) = x^2/2$, i.e. when $\mu_d^{V,B}$ is the joint law of the eigenvalues of a $\mathrm{GOE}(d)$ matrix. This setting was tackled in the seminal work of Ben Arous and Guionnet (1997).

### 2.2 Solving the variational principle: proof of Proposition 2.1

For $\mu \in \mathcal{M}_1^+(\mathbb{R})$, let

$$I(\mu) := -\frac{1}{2}\Sigma(\mu) + \frac{1}{4}\int \mu(\mathrm{d}x)\, x^2 - \frac{3}{8}. \tag{19}$$

We establish here the following equation, for $\kappa \le 2$:

$$E_\kappa := \inf_{\mu \in \mathcal{M}_1^+([-\kappa,\kappa])} I(\mu) = -\frac{\kappa^4}{128} + \frac{\kappa^2}{8} - \frac{1}{2}\log\frac{\kappa}{2} - \frac{3}{8}. \tag{20}$$

Combining eq. (20) with Corollary 2.4 yields then Proposition 2.1.

In order to characterize the minimizer of $I(\mu)$ for $\mu \in \mathcal{M}_1^+([-\kappa,\kappa])$, we rely on classical results of logarithmic potential theory, such as Theorem 1.3 of Chapter I of Saff and Totik (2013), see also Mhaskar and Saff (1985) and Theorem 2.4 of Ben Arous and Guionnet (1997). In our context, the "admissible weight function" of Saff and Totik (2013) reads

$$w(x) = \exp\{-x^2/4\}\mathbb{1}\{|x| \le \kappa\}.$$

Adapting the results of the aforementioned literature to our notations, we obtain the following theorem.

**Theorem 2.5** (Saff and Totik (2013))**.** *Let $\kappa > 0$ and $E_\kappa := \inf_{\mu \in \mathcal{M}_1^+([-\kappa,\kappa])} I(\mu)$. Then*

*(i)* $E_\kappa < \infty$.

*(ii) There exists a unique $\mu_\kappa^\star \in \mathcal{M}_1^+([-\kappa,\kappa])$ such that $I(\mu_\kappa^\star) = E_\kappa$.*

*(iii) $\mu_\kappa^\star$ is the unique measure in $\mathcal{M}_1^+([-\kappa,\kappa])$ such that for $\mu_\kappa^\star$-almost all $x$:*

$$\int \mu_\kappa^\star(\mathrm{d}y)\, \log|x-y| = \frac{x^2}{4} + \frac{1}{4}\int \mu_\kappa^\star(\mathrm{d}y)\, y^2 - \frac{3}{4} - 2E_\kappa.$$

One can further show that this theorem allows to prove that a candidate measure is the optimal one without computing $E_\kappa$.

**Lemma 2.6.** *For $\kappa \in (0,2]$, assume that $\mu \in \mathcal{M}_1^+([-\kappa,\kappa])$ and $C \in \mathbb{R}$ are such that for all $x \in (-\kappa,\kappa)$:*

$$\int \mu(\mathrm{d}y)\, \log|x-y| = \frac{x^2}{4} + C. \tag{21}$$

*Then $\mu = \mu_\kappa^\star$.*

Note that Lemma 2.6 can also be seen as a consequence of Theorem 3.3 of Chapter I of Saff and Totik (2013). We give here a short proof for the sake of completeness.

*Proof of Lemma 2.6 –.* For any $\sigma, \nu$ real signed measures, we have (recall eq. (19))

$$I(\sigma+\nu) - I(\sigma) = -\frac{1}{2}\Sigma(\nu) + \int \nu(\mathrm{d}x)\left[\frac{x^2}{4} - \int \sigma(\mathrm{d}y)\log|x-y|\right].$$

Applying this formula to $\sigma = \mu$ and $\nu = \mu_\kappa^\star - \mu$, we reach:

$$\begin{aligned} I(\mu_\kappa^\star) - I(\mu) &= -\frac{1}{2}\Sigma(\mu_\kappa^\star - \mu) + \int (\mu_\kappa^\star - \mu)(\mathrm{d}x)\left[\frac{x^2}{4} - \int \mu(\mathrm{d}y)\log|x-y|\right], \\ &\overset{\text{(a)}}{=} -\frac{1}{2}\Sigma(\mu_\kappa^\star - \mu), \end{aligned}$$

$$\overset{\text{(b)}}{\geq} 0.$$

We used eq. (21) in (a) and the fact that $\mu_\kappa^\star$ has no atom since $I(\mu_\kappa^\star) < \infty$ (Ben Arous and Guionnet, 1997), so we can restrict the integral to $x \in (-\kappa, \kappa)$. In (b) we used the following classical property of the non-commutative entropy $\Sigma(\mu)$, which can be found e.g. as Proposition II.2.2 in Faraut (2014).

**Lemma 2.7.** *Let $\mu$ be a signed measure on $\mathbb{R}$ with compact support, and such that $\int_{\mathbb{R}} \mu(\mathrm{d}x) = 0$. Let $\hat{\mu}(t) \coloneqq \int \mu(\mathrm{d}x)\, e^{itx}$ be the characteristic function of $\mu$. Then*

$$\Sigma(\mu) \coloneqq \int \mu(\mathrm{d}x)\mu(\mathrm{d}y) \log|x-y| = -\int_0^\infty \frac{|\hat{\mu}(t)|^2}{t}\mathrm{d}t.$$

*In particular, $\Sigma(\mu) \leq 0$.*

This shows that $I(\mu) \leq I(\mu_\kappa^\star) = \inf_{\mu \in \mathcal{M}_1^+([-\kappa,\kappa])} I(\mu)$. By (*ii*) of Theorem 2.5, $\mu = \mu_\kappa^\star$. $\square$

Thanks to Lemma 2.6, we can give an exact formula for $\mu_\kappa^\star$ by exhibiting a candidate measure satisfying eq. (21), which in turns implies the exact formula (20) for $E_\kappa$.

**Proposition 2.8.** *Let $\kappa \in (0, 2]$. Recall that $E_\kappa = \inf_{\mu \in \mathcal{M}_1^+([-\kappa,\kappa])} I(\mu)$ is reached in a unique measure $\mu_\kappa^\star$. Let $\mu_\kappa(\mathrm{d}x) \coloneqq \rho_\kappa(x)\mathrm{d}x$ be given, for $x \in (-\kappa, \kappa)$, by:*

$$\rho_\kappa(x) \coloneqq \frac{4 + \kappa^2 - 2x^2}{4\pi\sqrt{\kappa^2 - x^2}}. \tag{22}$$

*Then $\mu_\kappa = \mu_\kappa^\star$, and we have eq. (20):*

$$E_\kappa = I(\mu_\kappa^\star) = -\frac{\kappa^4}{128} + \frac{\kappa^2}{8} - \frac{1}{2}\log\frac{\kappa}{2} - \frac{3}{8}.$$

*Proof of Proposition 2.8.* By Lemma 2.6, to show that $\mu_\kappa = \mu_\kappa^\star$ it is enough to show that eq. (21) holds for $\mu_\kappa$. Since $U(x) \coloneqq \int \mu_\kappa(\mathrm{d}y) \log|x-y|$ satisfies (in the sense of distributions)

$$U'(x) = \mathrm{P.V.} \int \frac{\rho_\kappa(y)}{x-y}\mathrm{d}y,$$

it is enough to check for any $x \in (-\kappa, \kappa)$:

$$\mathrm{P.V.} \int_{-r}^{r} \frac{\rho_\kappa(y)}{x-y}\mathrm{d}y = \frac{x}{2}, \tag{23}$$

where P.V. refers to the principal value. Notice that

$$\mathrm{P.V.} \int_{-r}^{r} \frac{\rho_\kappa(y)}{x-y}\mathrm{d}y = \lim_{\varepsilon \downarrow 0} \mathrm{Re}\left\{\int_{-r}^{r} \frac{\rho_\kappa(y)}{x + i\varepsilon - y}\mathrm{d}y\right\}.$$

We compute $G_\kappa(z) \coloneqq \int_{-r}^{r} \rho_\kappa(y)/(z-y)\mathrm{d}y$ for all $z$ such that $\mathrm{Im}(z) > 0$. Changing variables to $y = \kappa\cos\theta$, and then to $\zeta = e^{i\theta}$, we have:

$$\begin{aligned}
G_\kappa(z) &= \frac{1}{4\pi}\int_0^\pi \frac{4 + \kappa^2 - 2\kappa^2\cos^2\theta}{z - \kappa\cos\theta}\,\mathrm{d}\theta, \\
&= \frac{1}{8\pi}\int_{-\pi}^\pi \frac{4 + \kappa^2 - 2\kappa^2\cos^2\theta}{z - \kappa\cos\theta}\,\mathrm{d}\theta, \\
&= \frac{1}{8i\pi}\oint_{|\zeta|=1} \frac{\kappa^2\zeta^4 - 8\zeta^2 + \kappa^2}{\zeta^2(\kappa\zeta^2 - 2\zeta z + \kappa)}\mathrm{d}\zeta.
\end{aligned}$$

The denominator has three poles: $\{0, \zeta_-, \zeta_+\}$, where $\zeta_\pm := (z \pm \sqrt{z^2 - \kappa^2})/\kappa$ (we choose the branch of the square root such that $\mathrm{Im}[\sqrt{w}] \geq 0$ for all $w \in \mathbb{C}$). Since $\mathrm{Im}(z) > 0$, it is easy to show that $|\zeta_-| < 1 < |\zeta_+|$. We then apply the residue theorem and find:

$$G_\kappa(z) = \frac{z}{2} + \frac{4 + \kappa^2 - 2z^2}{4\sqrt{z^2 - \kappa^2}}. \tag{24}$$

Taking $\lim_{\varepsilon \to 0} \mathrm{Re}[G_\kappa(x + i\varepsilon)]$ for $|x| < \kappa$ yields eq. (23), and thus $\mu_\kappa = \mu_\kappa^\star$. It remains to compute $E_\kappa = I(\mu_\kappa^\star)$. One can use the same arguments as above (based on the residue theorem) to show

$$\int \mu_\kappa^\star(\mathrm{d}x)\, x^2 = \frac{\kappa^2(8 - \kappa^2)}{16}. \tag{25}$$

By $(iii)$ of Theorem 2.5 and Lemma 2.6, we then have for all $x \in (-\kappa, \kappa)$:

$$E_\kappa = \frac{\kappa^2(8 - \kappa^2)}{128} - \frac{3}{8} + \frac{x^2}{8} - \frac{1}{2}\int_{-\kappa}^{\kappa} \rho_\kappa(y) \log|x - y|\, \mathrm{d}y. \tag{26}$$

Notice that eq. (24) is valid for all $z$ with $\mathrm{Im}(z) > 0$. In particular, we reach from it, that for all $x \geq 0$:

$$\mathrm{P.V.} \int_{-r}^{r} \frac{\rho_\kappa(y)}{x - y} \mathrm{d}y = \begin{cases} \dfrac{x}{2} & \text{if } x < \kappa, \\[2ex] \dfrac{x}{2} + \dfrac{4 + \kappa^2 - 2x^2}{4\sqrt{x^2 - \kappa^2}} & \text{if } x > \kappa. \end{cases}$$

Since this is an integrable function, we have

$$\int_{-\kappa}^{\kappa} \rho_\kappa(y) \log|x - y|\, \mathrm{d}y = \begin{cases} \dfrac{x^2}{4} + C & \text{if } x \leq \kappa, \\[2ex] \dfrac{x^2}{4} + C - \dfrac{x\sqrt{x^2 - \kappa^2}}{4} - \log\left(\dfrac{x - \sqrt{x^2 - \kappa^2}}{\kappa}\right) & \text{if } x > \kappa. \end{cases}$$

Using that $\int_{-\kappa}^{\kappa} \rho_\kappa(y) \log|x - y| \mathrm{d}y - \log x \to 0$ as $x \to \infty$ yields $C = \log(\kappa/2) - \kappa^2/8$. Eq. (26) becomes:

$$\begin{aligned} E_\kappa = I(\mu_\kappa^\star) &= \frac{\kappa^2(8 - \kappa^2)}{128} - \frac{3}{8} - \frac{1}{2}\left[\log\frac{\kappa}{2} - \frac{\kappa^2}{8}\right], \\ &= -\frac{\kappa^4}{128} + \frac{\kappa^2}{8} - \frac{1}{2}\log\frac{\kappa}{2} - \frac{3}{8}, \end{aligned}$$

which ends the proof. □

**Remark: predicting the form of $\rho_\kappa$** – In order to predict the density $\rho_\kappa$ given by eq. (22), we used an argument based on a heuristic application of Tricomi's theorem (Tricomi, 1985), which states that if eq. (23) is satisfied and $\rho_\kappa$ is supported on $[-\kappa, \kappa]$, then

$$\rho_\kappa(x) = \frac{1}{\pi\sqrt{\kappa^2 - x^2}}\left[C - \frac{1}{\pi}\mathrm{P.V.}\int_{-\kappa}^{\kappa} \frac{\sqrt{\kappa^2 - y^2}}{x - y} \times \left(\frac{y}{2}\right) \mathrm{d}y\right], \tag{27}$$

for some constant $C$ chosen to ensure $\int_{-\kappa}^{\kappa} \rho_\kappa(x) \mathrm{d}x = 1$. A careful evaluation of eq. (27) based on the residue theorem yields eq. (22).

### 2.3 Limiting spectral distribution of a norm-constrained Gaussian matrix

Theorem 2.2 follows as a direct consequence of Proposition 2.3 and Proposition 2.8. Indeed, let $\mathbb{P}_\kappa$ be the law of $\mathbf{W} \sim \mathrm{GOE}(d)$ conditioned on $\|\mathbf{W}\|_{\mathrm{op}} \le \kappa$, and denote $\mu_{\mathbf{W}}$ the empirical spectral distribution of $\mathbf{W}$. Then for any $\delta > 0$, if $B(\mu_\kappa^\star, \delta) \subseteq \mathcal{M}_1^+([-\kappa,\kappa])$ is the open ball of radius $\delta$ centered in $\mu_\kappa^\star$ for the distance of eq. (14),

$$\limsup_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}_\kappa[\mu_{\mathbf{W}} \notin B(\mu_\kappa^\star, \delta)] \le - \inf_{\mu \in B(\mu_\kappa^\star,\delta)^c} [J_\kappa(\mu)],$$

with

$$J_\kappa(\mu) \coloneqq I(\mu) - I(\mu_\kappa^\star),$$

for $I$ given in eq. (19). Since $J_\kappa$ is a good rate function (see Proposition 2.3) and has a unique minimizer (cf. $(ii)$ of Theorem 2.5), $\inf_{\mu\in B(\mu_\kappa^\star,\delta)^c}[J_\kappa(\mu)] > 0$. Therefore, by the Borel-Cantelli lemma,

$$\mathbb{P}[\limsup_{d\to\infty} d(\mu_{\mathbf{W}}, \mu_\kappa^\star) \le \delta] = 1,$$

which ends the proof by taking the limit $\delta \to 0$. $\square$

# 3 The satisfiability region

Section 3.1 is dedicated to studying the properties of the threshold $\tau_2(\kappa)$. The proof of Theorem 1.8, which is the main goal of this section, is outlined in Section 3.2, and details are given in the remainder of Section 3.

### 3.1 Properties of the satisfiability bound

We prove here Proposition 1.7, which follows from the following lemma.

**Lemma 3.1.** *Define, for any $\kappa > 0$, $\bar\tau(\kappa) \coloneqq \min_{\eta>0} \tilde\tau(\eta,\kappa)$. Then $\bar\tau(\kappa) = \tilde\tau(\eta^\star(\kappa),\kappa)$, where $\eta^\star(\kappa)$ is the unique value of $\eta > 0$ such that:*

$$(1+\eta)\tau_1(\kappa) = \frac{1+\delta_\eta^2}{2(1-\delta_\eta^2)^2} + \left[\frac{\delta_\eta(1+6\delta_\eta+3\delta_\eta^2+2\delta_\eta^3)}{(1-\delta_\eta^2)^3(1-\delta_\eta)}\right]\kappa \tag{28}$$

$$+ \left[\frac{2(1+\delta_\eta)^5}{(1-\delta_\eta^2)^4} - \frac{(1+3\delta_\eta^2)}{4(1-\delta_\eta^2)^3}\right]\kappa^2 + \frac{\kappa^4(1+3\delta_\eta^2)}{32(1-\delta_\eta^2)^3}. \tag{29}$$

*Moreover, $\kappa \mapsto \bar\tau(\kappa)$ is a continuous function of $\kappa$.*

Recall that $\tau_2(\kappa) = \min_{u\in[0,\kappa]} \bar\tau(u)$, as defined in Lemma 3.1, so that $\kappa \mapsto \tau_2(\kappa)$ is clearly continuous and non-increasing. Moreover, solving eq. (28) gives a simple way to numerically evaluate $\kappa \in [0,2] \mapsto \bar\tau(\kappa)$, which then yields the values of $\tau_2$.

*Proof of Lemma 3.1.* Let

$$\begin{cases} f_\kappa(\eta) & \coloneqq (1+\eta)\tau_1(\kappa), \\ g_\kappa(\eta) & \coloneqq \dfrac{1+\delta_\eta^2}{2(1-\delta_\eta^2)^2} + \left[\dfrac{\delta_\eta(1+6\delta_\eta+3\delta_\eta^2+2\delta_\eta^3)}{(1-\delta_\eta^2)^3(1-\delta_\eta)}\right]\kappa \\ & \quad + \left[\dfrac{2(1+\delta_\eta)^5}{(1-\delta_\eta^2)^4} - \dfrac{(1+3\delta_\eta^2)}{4(1-\delta_\eta^2)^3}\right]\kappa^2 + \dfrac{\kappa^4(1+3\delta_\eta^2)}{32(1-\delta_\eta^2)^3}. \end{cases}$$

Recall that $\delta_\eta$ is defined as the unique solution to $H[(1+\delta)/2]/\log 2 = \eta/(1+\eta)$, with $H(p) = -p\log p - (1-p)\log(1-p)$. If $G(\delta) := H[(1+\delta)/2]/\log 2$, then $G$ is smooth and strictly decreasing on $[0,1]$. So $\delta_\eta = G^{-1}[\eta/(1+\eta)]$ is a smooth and strictly decreasing function of $\eta > 0$. It is then immediate by elementary arguments (it is easy to see that each of the $\kappa$-coefficients of $g_\kappa(\eta)$ is a strictly increasing function of $\delta$, similarly to what is done e.g. in eq. (48)) that $g_\kappa(\eta)$ is also a smooth and strictly decreasing function of $\eta$.

Moreover, we have $g_\kappa(0^+) = \infty$, $f_\kappa(0^+) = \tau_1(\kappa) < \infty$, and $g_\kappa(\infty) < \infty$, $f_\kappa(\infty) = \infty$. It is then elementary to show that for any $\kappa > 0$, $\min_{\eta>0} \tilde{\tau}(\eta,\kappa) = \min_{\eta>0} \max\{f_\kappa(\eta), g_\kappa(\eta)\}$ is reached in a unique $\eta^\star(\kappa)$, such that $f_\kappa(\eta^\star(\kappa)) = g_\kappa(\eta^\star(\kappa))$, and that $\eta^\star(\kappa)$ is a continuous function of $\kappa$, which implies that $\bar{\tau}(\kappa) = f_\kappa(\eta^\star(\kappa))$ is also continuous. □

## 3.2 Reduction to a second moment upper bound

The main element of our analysis is the following upper bound.

**Proposition 3.2.** *Let $\kappa \in (0,2]$. Recall the definition of $Z_\kappa$ in eq. (2), and of $\bar{\tau}(\kappa)$ in Lemma 3.1. Assume that $\tau > \bar{\tau}(\kappa)$. Then, for $n, d \to \infty$ with $n/d^2 \to \tau$:*

$$\limsup_{d\to\infty} \frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} \leq L \cdot \left[1 - \frac{\bar{\tau}(\kappa)}{\tau}\right]^{-1/2},$$

*for an absolute constant $L > 0$.*

We will detail the proof of Proposition 3.2 in Section 3.3, deferring some intermediate results to Section 3.4 and 3.5. We first show how to deduce Theorem 1.8.

*Proof of Theorem 1.8.* Since $\tau > \tau_2(\kappa) = \min_{u\in[0,\kappa]} \bar{\tau}(u)$ (see Lemma 3.1), and $Z_{\kappa'} \leq Z_\kappa$ for any $\kappa' \leq \kappa$, we can assume without loss of generality that $\tau > \bar{\tau}(\kappa)$ in order to prove Theorem 1.8.

Because the bound on the right-hand side is strictly higher than 1, Proposition 3.2 is not strong enough to directly guarantee the existence of solutions with high probability. This is a recurring challenge in many random constraint satisfaction problems where $\mathbb{E}[Z_\kappa^2]/\mathbb{E}[Z_\kappa]^2 \to C > 1$. This occurs e.g. in the symmetric binary perceptron, the vector analog of our matrix discrepancy task, see Abbe, Li, and Sly (2022), and prevents from applying the classical second moment method to get high-probability bounds. Fortunately, in Altschuler (2023) the author develops general techniques on sharp transitions for integer feasibility problems, and applies them to show the concentration of the discrepancy $\min_{\varepsilon\in\{\pm1\}^n} \|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}}$.

**Lemma 3.3** (Theorem 7 of Altschuler (2023))**.** *Let $d \geq 1$, and $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$. Let $\mathrm{disc}(\mathbf{W}_1, \cdots, \mathbf{W}_n) := \min_{\varepsilon\in\{\pm1\}^n} \|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}}$. Assume that $n/d^2 \to \tau$ as $d \to \infty$. Then there exists $c(\tau) > 0$ such that*

$$\frac{\mathbb{E}[\mathrm{disc}(\mathbf{W}_1, \cdots, \mathbf{W}_n)]}{\sqrt{\mathrm{Var}[\mathrm{disc}(\mathbf{W}_1, \cdots, \mathbf{W}_n)]}} \geq c(\tau)\sqrt{d}.$$

Let us see how the combination of Lemma 3.3 with our second moment estimates (Proposition 3.2) ends the proof of Theorem 1.8.

Since $\bar{\tau}(\kappa)$ is a continuous function of $\kappa$ by Lemma 3.1, we choose $\delta > 0$ small enough such that $\tau > \bar{\tau}(\kappa - \delta)$. Let $X := \mathrm{disc}(\mathbf{W}_1, \cdots, \mathbf{W}_n)$. Notice that $X \leq \|\sum_{i=1}^n \mathbf{W}_i\|_{\mathrm{op}}$, so that $\mathbb{E}X \leq 2\sqrt{n}$ since $(1/\sqrt{n})\sum_{i=1}^n \mathbf{W}_i \sim \mathrm{GOE}(d)$, and $\mathbb{E}\|\mathbf{Y}\|_{\mathrm{op}} \leq 2$ for $\mathbf{Y} \sim \mathrm{GOE}(d)$ (see e.g. Exercise 7.3.5 of Vershynin (2018)). From Lemma 3.3 we thus get

$$\mathrm{Var}(X) \leq \frac{4n}{c(\tau)^2 d}, \tag{30}$$

where $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}X)^2]$. Recall that the Paley-Zygmund inequality states that for any random variable $X \geq 0$:

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Applying it to $Z_{\kappa-\delta}$ and using Proposition 3.2, we get that

$$\mathbb{P}[X \leq (\kappa - \delta)\sqrt{n}] \geq L^{-1} \cdot \left[1 - \frac{\bar{\tau}(\kappa - \delta)}{\tau}\right]^{1/2} + o_d(1). \tag{31}$$

Let us denote $C(\tau, \kappa, \delta) := L \cdot [1 - \bar{\tau}(\kappa - \delta)/\tau]^{-1/2}$. By Chebyshev's inequality and eq. (30), we further have for all $t > 0$:

$$\mathbb{P}[X \geq \mathbb{E}X - t] \geq 1 - \frac{4n}{c(\tau)^2 dt^2}.$$

In particular, if $t = c(\tau)^{-1}\sqrt{8C(\tau, \kappa, \delta)n/d}$, we have $\mathbb{P}[X \geq \mathbb{E}X - t] \geq 1 - (2C(\tau, \kappa, \delta))^{-1}$, which combined with eq. (31) implies that

$$\mathbb{E}X \leq (\kappa - \delta)\sqrt{n} + c_2(\tau, \kappa, \delta)\sqrt{\frac{n}{d}},$$

where we redefined the constant $c_2(\tau, \kappa, \delta) > 0$. Again by Chebyshev's inequality and eq. (30), this implies that for all $u > 0$:

$$\mathbb{P}[X \leq (\kappa - \delta)\sqrt{n} + c_2\sqrt{\frac{n}{d}} + u] \geq 1 - \frac{c_1(\tau)n}{du^2}.$$

Picking $u = \delta\sqrt{n} - c_2\sqrt{n/d}$, we have $u \geq (\delta/2)\sqrt{n}$ for $n, d$ large enough, and this yields:

$$\mathbb{P}[X \leq \kappa\sqrt{n}] \geq 1 - \frac{4c_1(\tau)}{\delta^2 d} \to_{d\to\infty} 1,$$

which ends the proof. □

### 3.3 Proof of the second moment upper bound

We prove here Proposition 3.2. We compute the second moment as:

$$\begin{aligned}
\mathbb{E}[Z_\kappa^2] &= \sum_{\varepsilon,\varepsilon' \in \{\pm 1\}^n} \mathbb{P}\left[\left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n} \text{ and } \left\|\sum_{i=1}^n \varepsilon_i' \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n}\right], \\
&\overset{\text{(a)}}{=} 2^n \sum_{\varepsilon \in \{\pm 1\}^n} \mathbb{P}\left[\left\|\sum_{i=1}^n \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n} \text{ and } \left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n}\right], \\
&\overset{\text{(b)}}{=} 2^n \sum_{l=0}^n \binom{n}{l} \mathbb{P}\left[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa \text{ and } \|q_l \mathbf{W} + \sqrt{1 - q_l^2}\mathbf{Z}\|_{\mathrm{op}} \leq \kappa\right].
\end{aligned} \tag{32}$$

In (a) and (b) we used the rotation invariance of the $\mathrm{GOE}(d)$ distribution. In eq. (32), we changed variables to $l := (\langle \varepsilon, \mathbf{1}_n \rangle + n)/2$ and defined the "overlap" $q_l := (1/n)\langle \varepsilon, \mathbf{1}_n \rangle = 2(l/n) - 1$. Furthermore, $\mathbf{W}, \mathbf{Z} \sim \mathrm{GOE}(d)$ independently. We get from eq. (32) that (recall as well eq. (11)):

$$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} = \frac{1}{2^n} \sum_{l=0}^n \binom{n}{l} \exp\{n G_d(q_l)\}, \tag{33}$$

where for $q \in [-1, 1]$:

$$G_d(q) \coloneqq \frac{1}{n} \log \frac{\mathbb{P}\left[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa \text{ and } \|q\mathbf{W} + \sqrt{1-q^2}\mathbf{Z}\|_{\mathrm{op}} \le \kappa\right]}{\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa]^2}. \tag{34}$$

Recall that $H(p) \coloneqq -p \log p - (1-p)\log(1-p)$. We will leverage the following lemma, which is based on standard asymptotic techniques, and whose proof is deferred to Section 3.4.

**Lemma 3.4.** *Let $n \ge 1$, and $F_n : [-1, 1] \to \mathbb{R}$ such that $F_n(0) = 0$ and $F_n'(0) = 0$. Assume that there exists $(\gamma, \delta) > 0$ such that:*

- *(i)* $\limsup_{n\to\infty} \sup_{|q|\le\delta} F_n''(q) \le 1 - \gamma$.
- *(ii)* $\limsup_{n\to\infty} \sup_{|q|\ge\delta} \left[F_n(q) + H\left(\frac{1+q}{2}\right)\right] < \log 2$.

*Then (with $q_l \coloneqq 2l/n - 1 \in [-1, 1]$ for $l \in \{0, \cdots, n\}$):*

$$\limsup_{n\to\infty} \frac{1}{2^n} \sum_{l=0}^{n} \binom{n}{l} \exp\{nF_n(q_l)\} \le \frac{C}{\sqrt{\gamma}},$$

*for a global constant $C > 0$.*

From eq. (33), in order to finish the proof of Proposition 3.2, it suffices to check conditions $(i)$ and $(ii)$ of Lemma 3.4 for $G_d$ defined in eq. (34), for $\tau > \bar{\tau}(\kappa)$, $\gamma = 1 - \bar{\tau}(\kappa)/\tau$, and some $\delta > 0$.

Recall the definition of $\eta^\star(\kappa)$ in Lemma 3.1. We let $\delta \coloneqq \delta_{\eta^\star(\kappa)}$ as defined by eq. (6).

**Condition** $(ii)$ – Notice that $G_d(0) = 0$ and that $G_d$ is clearly an even function of $q$, so $G_d'(0) = 0$ (the smoothness of $G_d$ can be shown by direct computation, as we will see in eq. (36)). Furthermore, we have the trivial bound:

$$G_d(q) + H\left(\frac{1+q}{2}\right) \le H\left(\frac{1+q}{2}\right) - \frac{1}{n} \log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa].$$

Recall that $q \mapsto H[(1+q)/2]$ is even, and strictly decreasing on $[0, 1]$. Using Proposition 2.1, and the definition of $\tau_1(\kappa)$ in eq. (4), we get

$$\begin{aligned}
\limsup_{d\to\infty} \sup_{|q|\ge\delta} \left[G_d(q) + H\left(\frac{1+q}{2}\right)\right] &\le H\left(\frac{1+\delta}{2}\right) + \frac{\tau_1(\kappa)}{\tau} \log 2, \\
&\overset{\text{(a)}}{<} H\left(\frac{1+\delta}{2}\right) + \left[\frac{1}{1+\eta^\star(\kappa)}\right] \log 2, \\
&\overset{\text{(b)}}{=} \left[\frac{\eta^\star(\kappa)}{1+\eta^\star(\kappa)} + \frac{1}{1+\eta^\star(\kappa)}\right] \log 2, \\
&= \log 2,
\end{aligned}$$

using $\tau > \bar{\tau}(\kappa) = (1 + \eta^\star(\kappa))\tau_1(\kappa)$ in (a), and the definition of $\delta = \delta_{\eta^\star(\kappa)}$ in (b), cf. eq. (6). We have thus checked condition $(ii)$ of Lemma 3.4.

**Condition** $(i)$ – We will show that for any $\delta \in (0, 1)$:

$$\begin{aligned}
\limsup_{d\to\infty} \sup_{|q|\le\delta} G_d''(q) \le \frac{1}{\tau} \Bigg\{ &\frac{1+\delta^2}{2(1-\delta^2)^2} + \left[\frac{\delta(1+6\delta+3\delta^2+2\delta^3)}{(1-\delta^2)^3(1-\delta)}\right]\kappa \\
&+ \left[\frac{2(1+\delta)^5}{(1-\delta^2)^4} - \frac{(1+3\delta^2)}{4(1-\delta^2)^3}\right]\kappa^2 + \frac{\kappa^4(1+3\delta^2)}{32(1-\delta^2)^3}\Bigg\}.
\end{aligned} \tag{35}$$

Let us first show how eq. (35) finishes the proof of Proposition 3.2. We pick $\delta = \delta_{\eta^\star(\kappa)}$. By Lemma 3.1, eq. (35) can be rewritten for this value of $\delta$ as

$$\limsup_{d\to\infty} \sup_{|q|\le\delta} G_d''(q) \le \frac{\bar\tau(\kappa)}{\tau} = 1-\gamma,$$

with $\gamma := (1-\bar\tau(\kappa)/\tau)$. This implies that condition $(i)$ of Lemma 3.4 holds with this value of $\gamma$, and thus ends the proof of Proposition 3.2, as described above.

**Proof of eq.** (35) – There remains to show eq. (35). Let $q \in [0,1)$. We have ($\mathrm{d}\mathbf{W} = \prod_{i\le j} \mathrm{d}W_{ij}$ is the Lebesgue measure over the space $\mathcal{S}_d$ of symmetric matrices):

$$
\begin{aligned}
&\mathbb{P}\left[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa \text{ and } \|q\mathbf{W} + \sqrt{1-q^2}\mathbf{Z}\|_{\mathrm{op}} \le \kappa\right] \\
&= \frac{\int \mathbb{1}\{\|\mathbf{W}\|_{\mathrm{op}} \le \kappa\} e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{W}^2]} \mathbb{P}\left[\|q\mathbf{W}+\sqrt{1-q^2}\mathbf{Z}\|_{\mathrm{op}} \le \kappa\right] \mathrm{d}\mathbf{W}}{\int e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{W}^2]}\mathrm{d}\mathbf{W}}, \\
&= \frac{\int \mathbb{1}\{\|\mathbf{W}\|_{\mathrm{op}} \le \kappa\} e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{W}^2]} \left(\int \mathrm{d}\mathbf{Y} e^{-\frac{d}{4(1-q^2)}\mathrm{Tr}[(\mathbf{Y}-q\mathbf{W})^2]} \mathbb{1}\{\|\mathbf{Y}\|_{\mathrm{op}} \le \kappa\}\right) \mathrm{d}\mathbf{W}}{\left(\int e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{W}^2]}\mathrm{d}\mathbf{W}\right)\left(\int \mathrm{d}\mathbf{Y} e^{-\frac{d}{4(1-q^2)}\mathrm{Tr}[\mathbf{Y}^2]}\right)}, \\
&= \frac{\int \mathbb{1}\{\|\mathbf{W}\|_{\mathrm{op}}, \|\mathbf{Y}\|_{\mathrm{op}} \le \kappa\} e^{-\frac{d}{4(1-q^2)}(\mathrm{Tr}[\mathbf{W}^2]+\mathrm{Tr}[\mathbf{Y}^2]) + \frac{dq}{2(1-q^2)}\mathrm{Tr}[\mathbf{Y}\mathbf{W}]} \mathrm{d}\mathbf{Y}\mathrm{d}\mathbf{W}}{\left(\int \mathrm{d}\mathbf{W} e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{W}^2]}\right)^2 (1-q^2)^{d(d+1)/4}}.
\end{aligned}
\tag{36}
$$

Starting from eq. (36), we can compute the derivatives of $G_d(q)$. We will use the shorthand notation

$$\langle \cdot \rangle_{q,\kappa} := \frac{\int (\cdot) \mathbb{1}\{\|\mathbf{W}\|_{\mathrm{op}}, \|\mathbf{Y}\|_{\mathrm{op}} \le \kappa\} e^{-\frac{d}{4(1-q^2)}(\mathrm{Tr}[\mathbf{W}^2]+\mathrm{Tr}[\mathbf{Y}^2]) + \frac{dq}{2(1-q^2)}\mathrm{Tr}[\mathbf{Y}\mathbf{W}]} \mathrm{d}\mathbf{Y}\mathrm{d}\mathbf{W}}{\int \mathbb{1}\{\|\mathbf{W}\|_{\mathrm{op}}, \|\mathbf{Y}\|_{\mathrm{op}} \le \kappa\} e^{-\frac{d}{4(1-q^2)}(\mathrm{Tr}[\mathbf{W}^2]+\mathrm{Tr}[\mathbf{Y}^2]) + \frac{dq}{2(1-q^2)}\mathrm{Tr}[\mathbf{Y}\mathbf{W}]} \mathrm{d}\mathbf{Y}\mathrm{d}\mathbf{W}}, \tag{37}$$

i.e. $\langle \cdot \rangle_{q,\kappa}$ is the law of $(\mathbf{W}, \mathbf{Y})$ two correlated GOE($d$) matrices (with correlation $q$), conditioned on the event $\|\mathbf{W}\|_{\mathrm{op}}, \|\mathbf{Y}\|_{\mathrm{op}} \le \kappa$. $G_d(q)$ is the log-partition function (or "free energy" in statistical physics) of this high-dimensional probability measure, and taking derivatives will yield averages of observables under $\langle \cdot \rangle_{q,\kappa}$. We get

$$G_d'(q) = \frac{d(d+1)q}{2n(1-q^2)} + \frac{1}{2n}\left\langle -\frac{dq}{(1-q^2)^2}\mathrm{Tr}[\mathbf{W}^2+\mathbf{Y}^2] + \frac{d(1+q^2)}{(1-q^2)^2}\mathrm{Tr}[\mathbf{W}\mathbf{Y}]\right\rangle_{q,\kappa}.$$

Differentiating further, we obtain:

$$
\begin{aligned}
G_d''(q) = &\underbrace{\frac{d(d+1)(1+q^2)}{2n(1-q^2)^2}}_{=:I_1(q)} + \underbrace{\frac{1}{2n}\left\langle -\frac{d(1+3q^2)}{(1-q^2)^3}\mathrm{Tr}[\mathbf{W}^2+\mathbf{Y}^2] + \frac{2dq(3+q^2)}{(1-q^2)^3}\mathrm{Tr}[\mathbf{W}\mathbf{Y}]\right\rangle_{q,\kappa}}_{=:I_2(q)} \\
&+ \underbrace{\frac{1}{4n}\mathrm{Var}_{\langle\cdot\rangle_{q,\kappa}}\left(-\frac{dq}{(1-q^2)^2}\mathrm{Tr}[\mathbf{W}^2+\mathbf{Y}^2] + \frac{d(1+q^2)}{(1-q^2)^2}\mathrm{Tr}[\mathbf{W}\mathbf{Y}]\right)}_{=:I_3(q)}.
\end{aligned}
\tag{38}
$$

We bound successively the different terms $\{I_a\}_{a=1}^3$ in eq. (38). Since $n/d^2 \to \tau$, we have:

$$\limsup_{d\to\infty}\sup_{|q|\le\delta} I_1(q) = \frac{1}{\tau}\sup_{|q|\le\delta}\frac{(1+q^2)}{2(1-q^2)^2} = \frac{(1+\delta^2)}{2\tau(1-\delta^2)^2}. \tag{39}$$

Recall that for a real random variable $X$, we define the sub-Gaussian norm $\|X\|_{\psi_2}$ of $X$ as (Vershynin, 2018):

$$\|X\|_{\psi_2} \coloneqq \inf\{t > 0 \,:\, \mathbb{E}[\exp(X^2/t^2)] \leq 2\}.$$

To bound $I_2$ and $I_3$, we rely on the following crucial result, which we prove in Section 3.5.

**Lemma 3.5** (Concentration of moments under $\langle\cdot\rangle_{q,\kappa}$)**.** *Let $q \in (-1,1)$, $\kappa \in (0,2]$, and*

$$P(X_1, X_2) \coloneqq \sum_{p\geq 0} \sum_{i_1,\cdots,i_p\in\{1,2\}} a_{i_1\cdots i_p} X_{i_1}\cdots X_{i_p}$$

*be a polynomial in two non-commutative random variables $(X_1, X_2)$. Let $(\mathbf{W},\mathbf{Y}) \sim \langle\cdot\rangle_{q,\kappa}$ given by eq. (37). Then:*

$$\|\mathrm{Tr}\, P(\mathbf{W},\mathbf{Y}) - \langle \mathrm{Tr}\, P(\mathbf{W},\mathbf{Y})\rangle_{q,\kappa}\|_{\psi_2} \leq C\sqrt{1+q}\sum_{p\geq 0} p\cdot\kappa^{p-1} \sum_{i_1,\cdots,i_p\in\{1,2\}} |a_{i_1\cdots i_p}|,$$

*where $C > 0$ is an absolute constant. Furthermore, we have the fully explicit bound:*

$$\mathrm{Var}_{\langle\cdot\rangle_{q,\kappa}}[\mathrm{Tr}\, P(\mathbf{W},\mathbf{Y})] \leq 2(1+q)\left(\sum_{p\geq 0} p\cdot\kappa^{p-1} \sum_{i_1,\cdots,i_p\in\{1,2\}} |a_{i_1\cdots i_p}|\right)^2. \tag{40}$$

Lemma 3.5 is a consequence of a log-Sobolev inequality we prove for $\langle\cdot\rangle_{q,\kappa}$.

**Bounding** $I_2$ – Note that under the law of $\langle\cdot\rangle_{0,\kappa}$ of eq. (37) when $q=0$, $\mathbf{W}$ is distributed as a GOE($d$) matrix, conditioned to satisfy $\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa$. By Theorem 2.2, we know that $\mu_{\mathbf{W}}$ weakly converges (a.s.) to $\mu^\star_\kappa$. Since $\int \mu_{\mathbf{W}}(\mathrm{d}x)x^2 = \int \mu_{\mathbf{W}}(\mathrm{d}x)x^2 \mathbb{1}\{|x| \leq \kappa\}$, we have by the Portmanteau theorem and dominated convergence:

$$\begin{aligned}
\lim_{d\to\infty} \frac{1}{d}\langle \mathrm{Tr}\mathbf{W}^2\rangle_{0,\kappa} &= \int \mu^\star_\kappa(\mathrm{d}x)\, x^2\, \mathbb{1}\{|x|\leq\kappa\} \\
&= \int \mu^\star_\kappa(\mathrm{d}x)\, x^2, \qquad (41)\\
&\overset{\text{(a)}}{=} \frac{\kappa^2(8-\kappa^2)}{16}, \qquad (42)
\end{aligned}$$

using eq. (25) in (a). By symmetry $\mathbf{W}\to-\mathbf{W}$, we also trivially have

$$\langle \mathrm{Tr}\mathbf{W}\mathbf{Y}\rangle_{0,\kappa} = 0. \tag{43}$$

Let us denote $P_q(X,Y) \coloneqq -q(X^2+Y^2) + (1+q^2)XY$. One easily computes from eq. (37) that for any function $\varphi(\mathbf{W},\mathbf{Y})$:

$$\begin{aligned}
\frac{\partial}{\partial q}\langle\varphi(\mathbf{W},\mathbf{Y})\rangle_{q,\kappa} &= \frac{d\left[\langle\varphi(\mathbf{W},\mathbf{Y})\cdot \mathrm{Tr}[P_q(\mathbf{W},\mathbf{Y})]\rangle_{q,\kappa} - \langle\varphi(\mathbf{W},\mathbf{Y})\rangle_{q,\kappa}\langle \mathrm{Tr}[P_q(\mathbf{W},\mathbf{Y})]\rangle_{q,\kappa}\right]}{2(1-q^2)^2}, \\
&= \frac{d\left[\langle(\varphi(\mathbf{W},\mathbf{Y}) - \langle\varphi\rangle_{q,\kappa})\cdot(\mathrm{Tr}[P_q(\mathbf{W},\mathbf{Y})] - \langle\mathrm{Tr}[P_q]\rangle_{q,\kappa})\rangle_{q,\kappa}\right]}{2(1-q^2)^2}.
\end{aligned}$$

In particular:

$$\left|\frac{\partial}{\partial q}\langle\varphi(\mathbf{W},\mathbf{Y})\rangle_{q,\kappa}\right| \leq \frac{d\left[\mathrm{Var}_{\langle\cdot\rangle_{q,\kappa}}[\varphi(\mathbf{W},\mathbf{Y})]\cdot \mathrm{Var}_{\langle\cdot\rangle_{q,\kappa}}[\mathrm{Tr}P_q(\mathbf{W},\mathbf{Y})]\right]^{1/2}}{2(1-q^2)^2}. \tag{44}$$

Using eq. (44) and Lemma 3.5, we reach that for both $\varphi = \mathrm{Tr}[\mathbf{W}^2]$ and $\varphi = \mathrm{Tr}[\mathbf{W}\mathbf{Y}]$:

$$\left|\frac{\partial}{\partial q}\langle\varphi(\mathbf{W},\mathbf{Y})\rangle_{q,\kappa}\right| \leq \frac{2d\kappa(1+q)^2}{(1-q^2)^2} = \frac{2d\kappa}{(1-q)^2}. \tag{45}$$

Integrating eq. (45), and combining it with eqs. (41) and eq. (43), we get:

$$\begin{cases} \left|\frac{1}{d}\langle \mathrm{Tr}\mathbf{W}^2\rangle_{q,\kappa} - \frac{\kappa^2(8-\kappa^2)}{16}\right| & \leq \frac{2\kappa|q|}{1-q} + o_d(1), \\ \left|\frac{1}{d}\langle \mathrm{Tr}\mathbf{W}\mathbf{Y}\rangle_{q,\kappa}\right| & \leq \frac{2\kappa|q|}{1-q} + o_d(1), \end{cases} \tag{46}$$

where $o_d(1)$ is uniform in $q$. We get from eq. (46):

$$\begin{aligned} \limsup_{d\to\infty} \sup_{|q|\leq\delta} I_2(q) &\leq \frac{1}{2\tau} \max_{|q|\leq\delta} \left[\frac{4\kappa q^2(3+q^2)}{(1-q^2)^3(1-q)} - \frac{1+3q^2}{(1-q^2)^3}\left(\frac{\kappa^2(8-\kappa^2)}{16} - \frac{2\kappa|q|}{1-q}\right)\right], \\ &= \frac{1}{2\tau} \max_{q\in[0,\delta]} \left[\frac{2\kappa q(1+6q+3q^2+2q^3)}{(1-q^2)^3(1-q)} - \frac{\kappa^2(8-\kappa^2)(1+3q^2)}{16(1-q^2)^3}\right]. \end{aligned} \tag{47}$$

If

$$f_\kappa(q) \coloneqq \frac{2\kappa q(1+6q+3q^2+2q^3)}{1-q} - \frac{\kappa^2(8-\kappa^2)(1+3q^2)}{16},$$

then for all $q \in [0,1]$ and $\kappa \in [0,2]$:

$$\begin{aligned} f'_\kappa(q) &= \frac{\kappa}{8(1-q)^2}\left[16 + 3(64 - 8\kappa + \kappa^3)q + 6(8+8\kappa-\kappa^3)q^2 + (32 - 3\kappa(8-\kappa^2))q^4 - 96q^5\right], \\ &\geq \frac{\kappa}{8(1-q)^2}\left[16 + 3(64-16)q + 6(8-8)q^2 + (32 - 3\cdot 2\cdot(8))q^4 - 96q^5\right], \\ &\geq \frac{\kappa}{8(1-q)^2}\left[16 + 144q - 16q^3 - 96q^5\right], \\ &\overset{\text{(a)}}{\geq} \frac{\kappa}{8(1-q)^2}[16+32q] > 0, \end{aligned} \tag{48}$$

using $q^k \leq q$ for any $k \geq 1$ in (a). This implies that in eq. (47), the maximum is attained at $q = \delta$, and we get:

$$\limsup_{d\to\infty} \sup_{|q|\leq\delta} I_2(q) \leq \frac{1}{2\tau}\left[\frac{2\kappa\delta(1+6\delta+3\delta^2+2\delta^3)}{(1-\delta^2)^3(1-\delta)} - \frac{\kappa^2(8-\kappa^2)(1+3\delta^2)}{16(1-\delta^2)^3}\right]. \tag{49}$$

**Bounding** $I_3$ – We apply eq. (40) of Lemma 3.5 to $P_q(X,Y) = -q(X^2+Y^2) + (1+q^2)XY$, which yields:

$$\begin{aligned} I_3(q) &\leq \frac{d^2}{4n(1-q^2)^4} \cdot 2(1+q)\left(2\kappa[2q+1+q^2]\right)^2, \\ &= \frac{2d^2(1+q)^5\kappa^2}{n(1-q^2)^4}. \end{aligned}$$

So finally we get:

$$\limsup_{d\to\infty} \sup_{|q|\leq\delta} I_3(q) \leq \frac{2(1+\delta)^5\kappa^2}{\tau(1-\delta^2)^4}. \tag{50}$$

Combining eqs. (39),(49),(50) finishes the proof of eq. (35). As we discussed above, this ends the proof of Proposition 3.2. □

### 3.4 Discrete Laplace's method for a dimension-dependent exponent

We prove here Lemma 3.4. By hypothesis $(ii)$, we fix $\varepsilon > 0$ such that, for $n$ large enough:

$$\sup_{|q|\geq\delta}\left[F_n(q) + H\left(\frac{1+q}{2}\right)\right] \leq \log 2 - \varepsilon. \tag{51}$$

Recall the classical inequality:

$$\binom{n}{l} \leq e^{nH(l/n)}, \qquad \text{for } l \in \{0, \cdots, n\}. \tag{52}$$

Combining eqs. (51) and (52), we have

$$\begin{aligned}\frac{1}{2^n}\sum_{l=0}^{n}\mathbb{1}\left\{\left|l-\frac{n}{2}\right| > \frac{n\delta}{2}\right\}\binom{n}{l}\exp\{nF_n(q_l)\} &\leq \frac{1}{2^n}\sum_{l=0}^{n}\mathbb{1}\left\{\left|l-\frac{n}{2}\right| > \frac{n\delta}{2}\right\}\exp\{n(\log 2-\varepsilon)\},\\ &\leq n\exp\{-n\varepsilon\}.\end{aligned} \tag{53}$$

Let $\sigma \in (0, \gamma)$. By hypothesis $(i)$, we get that for $n$ large enough $F_n''(q) \leq (1-\gamma+\sigma)$ for all $|q| \leq \delta$. Since $F_n(0) = 0$ and $F_n'(0) = 0$, this implies $F_n(q) \leq (1-\gamma+\sigma)q^2/2$ for all $|q| \leq \delta$. Therefore,

$$\frac{1}{2^n}\sum_{l=0}^{n}\mathbb{1}\left\{\left|l-\frac{n}{2}\right| \leq \frac{n\delta}{2}\right\}\binom{n}{l}\exp\{nF_n(q_l)\} \leq \frac{1}{2^n}\sum_{l=0}^{n}\binom{n}{l}e^{\frac{n(1-\gamma+\sigma)}{2}q_l^2}. \tag{54}$$

Recall that $q_l = 2(l/n) - 1$. The right-hand side of eq. (54) can now be analyzed with standard extensions of Laplace's method. We use here the following statement, which is a consequence of the proof of Lemma 2 of Achlioptas and Moore (2002).

**Lemma 3.6** (Achlioptas and Moore (2002))**.** *There exists $B, C > 0$ such that the following holds. Let $G$ a real analytic positive function on $[0,1]$, and define for $\alpha \in [0,1]$:*

$$g(\alpha) \coloneqq \frac{G(\alpha)}{\alpha^\alpha(1-\alpha)^{1-\alpha}}.$$

*If there exists $\alpha_{\max} \in (0,1)$ a strict global maximum of $g$ in $[0,1]$ such that $g''(\alpha_{\max}) < 0$, then for sufficiently large $n$:*

$$B \cdot \frac{g(\alpha_{\max})^{n+1/2}}{\sqrt{-g''(\alpha_{\max})}} \leq \sum_{l=0}^{n}\binom{n}{l}G(l/n)^n \leq C \cdot \frac{g(\alpha_{\max})^{n+1/2}}{\sqrt{\alpha_{\max}(1-\alpha_{\max})(-g''(\alpha_{\max}))}}.$$

**Remark** – Lemma 3.6 is stated in Achlioptas and Moore (2002) as

$$C_1 \cdot g(\alpha_{\max})^n \leq \sum_{l=0}^{n}\binom{n}{l}G(l/n)^n \leq C_2 \cdot g(\alpha_{\max})^n,$$

where the constants $C_1, C_2$ might depend on $\alpha_{\max}$ and $g(\alpha_{\max})$. Their proof (see Appendix A of Achlioptas and Moore (2002)) reveals the dependency of $C_1, C_2$ on $\alpha_{\max}$ and $g''(\alpha_{\max})$, which we make explicit here.

We apply Lemma 3.6 in eq. (54), with

$$G(x) \coloneqq \frac{1}{2}e^{\frac{(1-\gamma+\sigma)(2x-1)^2}{2}}$$

Let

$$g(x) := \frac{G(x)}{x^x(1-x)^{1-x}} = \frac{e^{\frac{(1-\gamma+\sigma)(2x-1)^2}{2}}}{2x^x(1-x)^{1-x}}.$$

It is clear that $g(x) = g(1-x)$ for all $x \in [0,1]$, and moreover

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}(\log g)(x) = -(1-\gamma+\sigma)(1-2x) + \operatorname{arctanh}(1-2x).$$

Since $\operatorname{arctanh}(u) \geq u$ for all $u \in [0,1)$, we get that for all $x \in (0,1/2]$:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\log g)(x) \geq 2(\gamma-\sigma)(1-2x).$$

Combining this with the symmetry $g(x) = g(1-x)$, we obtain that $g$ has a strict global maximum in $x = 1/2$, and we compute $g(1/2) = 1$. Moreover, we get by direct computation that $g''(1/2) = -4(\gamma-\sigma) < 0$. All in all, we reach that for $n$ large enough:

$$\frac{1}{2^n}\sum_{l=0}^{n}\binom{n}{l} e^{-\frac{n(1-\gamma)}{2}q_l^2} \leq \frac{C}{\sqrt{\gamma-\sigma}}. \tag{55}$$

Combining eqs. (53),(54) and (55), we get:

$$\limsup_{n\to\infty}\frac{1}{2^n}\sum_{l=0}^{n}\binom{n}{l}\exp\{nF_n(q_l)\} \leq \limsup_{n\to\infty}[ne^{-n\varepsilon} + C(\gamma-\sigma)^{-1/2}] = C(\gamma-\sigma)^{-1/2}.$$

Letting $\sigma \downarrow 0$ ends the proof of Lemma 3.4. $\square$

## 3.5 Log-Sobolev inequality for the conditioned law of two correlated $\mathrm{GOE}(d)$

In this section, we start by reminders on log-Sobolev inequalities, before proving such a property for the law of eq. (37), and finally proving Lemma 3.5.

### 3.5.1 Log-Sobolev inequalities and concentration of measure

**Definition 3.1.** Let $d \geq 1$. A probability measure $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ is said to satisfy the *Logarithmic Sobolev Inequality* (LSI) with constant $c > 0$ if, for any differentiable function $f$ in $L^2(\mu)$, we have

$$\int f^2 \log\frac{f^2}{\int f^2\,\mathrm{d}\mu}\mathrm{d}\mu \leq 2c\int \|\nabla f\|_2^2\,\mathrm{d}\mu. \tag{56}$$

We refer the reader to Guionnet (2009) and Anderson, Guionnet, and Zeitouni (2010) for more on the theory of log-Sobolev inequalities and their applications to concentration results in random matrix theory. A particularly useful consequence of the LSI is the following.

**Lemma 3.7** (Herbst)**.** *Assume that $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ satisfies the LSI with constant $c$. Let $G : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function, with Lipschitz constant $\|G\|_L$. Then for all $\lambda \in \mathbb{R}$:*

$$\mathbb{E}_\mu\left[e^{\lambda[G-\mathbb{E}G]}\right] \leq \exp\left\{\frac{c\|G\|_L^2\lambda^2}{2}\right\}.$$

*Therefore, for all $\delta > 0$:*

$$\mu(|G-\mathbb{E}G| \geq \delta) \leq 2\exp\left\{-\frac{\delta^2}{2c\|G\|_L^2}\right\}.$$

*In particular,* $\mathrm{Var}(G) \leq c\|G\|_L^2$. ***FIXME:*** *Is this the right place?*

Finally, we will use that a necessary condition for a measure to satisfy the LSI is the so-called *Bakry-Emery (BE)* condition.

**Theorem 3.8** (Theorem 4.4.17 of Anderson, Guionnet, and Zeitouni (2010))**.** *Let $d \geq 1$ and $\Phi : \mathbb{R}^d \to \mathbb{R}$ a $\mathcal{C}^2$ function. Assume that $\Phi$ satisfies the Bakry-Emery condition:*

$$\mathrm{Hess}\, \Phi(x) \succeq \frac{1}{c} \mathrm{I}_d,$$

*for all $x \in \mathbb{R}^d$, for some $c > 0$. Then the measure*

$$\mu_\Phi(\mathrm{d}x) := \frac{1}{\mathcal{Z}} e^{-\Phi(x)} \mathrm{d}x$$

*satisfies the LSI with constant $c$.*

### 3.5.2 A log-Sobolev inequality for the law of eq. (37)

We show the following lemma.

**Lemma 3.9.** *For any $q \in (-1, 1)$ and $\kappa > 0$, the law $\langle \cdot \rangle_{q,\kappa}$ of eq. (37) satisfies the LSI with constant $2(1 + q)/d$.*

*Proof of Lemma 3.9.* Let $\varepsilon > 0$. We denote $\phi_\varepsilon(x) := e^{-x^2/(2\varepsilon)}/\sqrt{2\pi\varepsilon}$. We define

$$V_\varepsilon(x) := -\log \left( \int_{-\kappa}^{\kappa} \mathrm{d}y \, e^{-y^2/2} \, \phi_\varepsilon(x - y) \right).$$

**Reminders on log-concavity** – A real positive integrable function $p$ is said to be *strongly log-concave* with variance parameter $\sigma^2$ (denoted $\mathrm{SLC}(\sigma^2)$) if $p(x) = \phi_{\sigma^2}(x) \cdot e^{\varphi(x)}$, for some concave function $\varphi : \mathbb{R} \to [-\infty, \infty)$. We refer the reader to Saumard and Wellner (2014) for properties of log-concave and strongly log-concave functions and probability distributions. It is clear that $x \mapsto e^{-x^2/2} \mathbb{1}\{|x| \leq \kappa\}$ is SLC(1), and that $\phi_\varepsilon$ is $\mathrm{SLC}(\varepsilon)$. By Theorem 3.7 of Saumard and Wellner (2014), if $f$ is $\mathrm{SLC}(\sigma_1^2)$ and $g$ is $\mathrm{SLC}(\sigma_2^2)$, their convolution $f \star g$ is $\mathrm{SLC}(\sigma_1^2 + \sigma_2^2)$. Therefore, $e^{-V_\varepsilon}$ is $\mathrm{SLC}(1+\varepsilon)$, which implies (since $V_\varepsilon$ is smooth) that $V_\varepsilon''(x) \geq (1+\varepsilon)^{-1}$ for all $x \in \mathbb{R}$. Since $V_\varepsilon$ is even, and

$$V_\varepsilon(0) \geq -\log \int_{-\kappa}^{\kappa} \mathrm{d}y \, \phi_\varepsilon(y) \geq 0,$$

we get that $V_\varepsilon(x) \geq x^2/[2(1+\varepsilon)]$ for all $x \in \mathbb{R}$. We define $\mu_\varepsilon$ as:

$$\mu_\varepsilon(\mathrm{d}\mathbf{Y}, \mathrm{d}\mathbf{W}) := \frac{e^{-\frac{d}{2(1-q^2)}(\mathrm{Tr} V_\varepsilon(\mathbf{W}) + \mathrm{Tr} V_\varepsilon(\mathbf{Y})) + \frac{dq}{2(1-q^2)} \mathrm{Tr}[\mathbf{Y}\mathbf{W}]} \mathrm{d}\mathbf{Y} \mathrm{d}\mathbf{W}}{\int e^{-\frac{d}{2(1-q^2)}(\mathrm{Tr} V_\varepsilon(\mathbf{W}) + \mathrm{Tr} V_\varepsilon(\mathbf{Y})) + \frac{dq}{2(1-q^2)} \mathrm{Tr}[\mathbf{Y}\mathbf{W}]} \mathrm{d}\mathbf{Y} \mathrm{d}\mathbf{W}}. \tag{57}$$

Recall $\mathcal{S}_d$ is the set of symmetric $d \times d$ matrices. Since $x \mapsto V_\varepsilon(x) - x^2/(2[1+\varepsilon])$ is convex, by Klein's lemma (cf. Lemma 4.4.12 of Anderson, Guionnet, and Zeitouni (2010) or Lemma 6.4 of Guionnet (2009)), the function $\mathbf{W} \mapsto \mathrm{Tr} V_\varepsilon(\mathbf{W}) - \mathrm{Tr}[\mathbf{W}^2]/(2[1+\varepsilon])$ is also convex. Thus, for all $\mathbf{W} \in \mathcal{S}_d$:

$$\mathrm{Hess}\, V_\varepsilon(\mathbf{W}) \succeq \frac{1}{1+\varepsilon} \mathrm{I}_{\mathcal{S}_d}.$$

All in all we get for any $\mathbf{W}$ and $\mathbf{Y}$:

$$\mathrm{Hess} \left[ \frac{d}{2(1-q^2)} (\mathrm{Tr} V_\varepsilon(\mathbf{W}) + \mathrm{Tr} V_\varepsilon(\mathbf{Y})) - \frac{dq}{2(1-q^2)} \mathrm{Tr}[\mathbf{Y}\mathbf{W}] \right] \succeq \frac{d}{2(1-q^2)} \begin{pmatrix} \frac{\mathrm{I}_{\mathcal{S}_d}}{1+\varepsilon} & -q\mathrm{I}_{\mathcal{S}_d} \\ -q\mathrm{I}_{\mathcal{S}_d} & \frac{\mathrm{I}_{\mathcal{S}_d}}{1+\varepsilon} \end{pmatrix},$$

which means

$$\lambda_{\min}\left(\mathrm{Hess}\left[\frac{d}{2(1-q^2)}(\mathrm{Tr}V_\varepsilon(\mathbf{W})+\mathrm{Tr}V_\varepsilon(\mathbf{Y}))-\frac{dq}{2(1-q^2)}\mathrm{Tr}[\mathbf{YW}]\right]\right)\geq\frac{d(1-q-\varepsilon q)}{2(1+\varepsilon)(1-q^2)}.$$

Therefore, by Theorem 3.8, $\mu_\varepsilon$ satisfies the LSI with constant

$$\frac{2(1+\varepsilon)(1-q^2)}{d(1-q-\varepsilon q)}=\frac{2(1+q)}{d}+o_{\varepsilon\to0}(1).$$

Finally, notice that we have $V_\varepsilon(x)\to_{\varepsilon\to0}V(x)$ pointwise, with $V(x)$ defined as:

$$V(x):=\begin{cases}\frac{x^2}{2} & \text{if } |x|<\kappa,\\ \frac{\kappa^2}{2}+\log 2 & \text{if } |x|=\kappa,\\ +\infty & \text{if } |x|>\kappa.\end{cases}$$

Since $V_\varepsilon(x)\geq x^2/4$ for $\varepsilon\leq1/2$, we get by dominated convergence and the Portmanteau theorem that $\mu_\varepsilon\to_{\varepsilon\to0}\mu_0$ weakly, where $\mu_0$ is defined as in eq. (57), replacing $V_\varepsilon$ by $V$. Because the set $\{\|\mathbf{W}\|_{\mathrm{op}}=\kappa\}$ has Lebesgue measure zero, we further have that $\mu_0=\langle\cdot\rangle_{q,\kappa}$. Since $\mu_\varepsilon$ satisfies the LSI with constant $2(1+q)/d+o_{\varepsilon\to0}(1)$, and weakly converges to $\langle\cdot\rangle_{q,\kappa}$ as $\varepsilon\downarrow0$, we deduce that $\langle\cdot\rangle_{q,\kappa}$ satisfies the LSI[1] with constant $2(1+q)/d$. □

### 3.5.3 Proof of Lemma 3.5

Let $P(X_1,X_2)=\sum_{p\geq0}\sum_{i_1,\cdots,i_p\in\{1,2\}}a_{i_1\cdots i_p}X_{i_1}\cdots X_{i_p}$. We make use of the following elementary result.

**Lemma 3.10** (Lemma 6.2 of Guionnet (2009)). *Let $Q$ be a polynomial in two non-commutative variables. Then, for any $\kappa>0$, the function*

$$(\mathbf{W},\mathbf{Y})\in B_{\mathrm{op}}(\kappa)\times B_{\mathrm{op}}(\kappa)\mapsto\mathrm{Tr}[Q(\mathbf{W},\mathbf{Y})]$$

*is Lipschitz with respect to the Euclidean norm, with Lipschitz norm bounded by $\sqrt{d}C(Q,\kappa)$ for some constant $C(Q,\kappa)>0$. If $Q$ is a monomial of degree $p$, one can take $C(Q,\kappa)=p\kappa^{p-1}$.*

Notice that $\mathrm{supp}(\langle\cdot\rangle_{q,\kappa})\subseteq B_{\mathrm{op}}(\kappa)\times B_{\mathrm{op}}(\kappa)$. By Lemma 3.10, $f:(\mathbf{W},\mathbf{Y})\mapsto\mathrm{Tr}P(\mathbf{W},\mathbf{Y})$ is thus Lipschitz on the support of $\langle\cdot\rangle_{q,\kappa}$, with Lipschitz constant

$$\|f\|_L\leq\sqrt{d}\sum_{p\geq0}p\cdot\kappa^{p-1}\sum_{i_1,\cdots,i_p\in\{1,2\}}|a_{i_1\cdots i_p}|.$$

Combining Lemmas 3.9 and 3.7 finishes the proof of Lemma 3.5. Indeed, notice that if $X$ is a random variable such that $\mathbb{E}[X]=0$ and $\mathbb{E}[e^{\lambda X}]\leq e^{\lambda^2K^2/2}$ for some $K>0$ and all $\lambda\in\mathbb{R}$, then $\|X\|_{\psi_2}\leq CK$, and moreover by Taylor expansion close to $\lambda=0$, we get $\mathrm{Var}(X)\leq K^2$. □

# 4 Limitations of the second moment method

We prove here Theorem 1.9, and discuss as well the technical hypothesis on which our statement relies. Let $\kappa\in(0,2]$ and $\tau>0$. We start again from the second moment computation detailed in Section 3.3, and more precisely from eq. (33), which we recall here:

$$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2}=\frac{1}{2^n}\sum_{l=0}^{n}\binom{n}{l}\exp\{nG_d(q_l)\},\tag{58}$$

[1]By taking the limit of eq. (56) for well-behaved functions $f$ using weak convergence, and extending to all differentiable and square integrable functions by density. See e.g. the proof of Theorem 4.4.17 in Anderson, Guionnet, and Zeitouni (2010) for details.

where for $q \in [-1,1]$:

$$G_d(q) := \frac{1}{n} \log \frac{\mathbb{P}\left[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa \text{ and } \|q\mathbf{W} + \sqrt{1-q^2}\mathbf{Z}\|_{\mathrm{op}} \le \kappa\right]}{\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa]^2}.$$

Using eq. (36) we further have, for any $q \in (-1,1)$:

$$G_d(q) = \frac{1}{n} \log \frac{\int \mathbb{1}\{\|\mathbf{W}_1\|_{\mathrm{op}}, \|\mathbf{W}_2\|_{\mathrm{op}} \le \kappa\} e^{-\frac{d}{4(1-q^2)}(\mathrm{Tr}[\mathbf{W}_1^2]+\mathrm{Tr}[\mathbf{W}_2^2]) + \frac{dq}{2(1-q^2)}\mathrm{Tr}[\mathbf{W}_1\mathbf{W}_2]} \mathrm{d}\mathbf{W}_1 \mathrm{d}\mathbf{W}_2}{\left(\int \mathrm{d}\mathbf{W}\, \mathbb{1}\{\|\mathbf{W}\|_{\mathrm{op}} \le \kappa\}\, e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{W}^2]}\right)^2 (1-q^2)^{d(d+1)/4}}. \tag{59}$$

We can compute explicitly the limit of $G_d''(q)$ for $q$ close to 0 as follows.

**Lemma 4.1.** *Recall the definition of $\tau_{\mathrm{fail.}}(\kappa)$ in eq. (9). We have*

$$\lim_{d\to\infty} G_d''(0) = \frac{\tau_{\mathrm{fail.}}(\kappa)}{\tau}.$$

Lemma 4.1 relies on the limiting spectral distribution theorem we established in Theorem 2.2, and is proven below. First, we establish how Lemma 4.1, alongside the following technical hypothesis, ends the proof of Theorem 1.9.

**Hypothesis 4.1.** We assume that, for $G_d$ given in eq. (59):

$$\lim_{\varepsilon\downarrow 0} \limsup_{d\to\infty} \sup_{|q|\le\varepsilon} |G_d''(q) - G_d''(0)| = 0.$$

**Discussion of Hypothesis 4.1** – Hypothesis 4.1 states that $G_d''(q)$ is continuous in $q = 0$, uniformly in $d$ as $d \to \infty$. We make two important remarks related to this hypothesis:

- First, note that $G_d$ in eq. (59) can be interpreted as a large deviations rate function for the spectral norms of two $q$-correlated GOE($d$) matrices $(\|\mathbf{W}_1\|_{\mathrm{op}}, \|\mathbf{W}_2\|_{\mathrm{op}})$, on the scale $d^2$. In general, the large deviations rate function (in the weak topology) for the joint law of the spectral measures of two matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ drawn from a $\beta$-ensemble with an interacting potential proportional to $\mathrm{Tr}[\mathbf{W}_1\mathbf{W}_2]$ has been established in Guionnet (2004)[Theorem 3.3]. From these results, one can obtain the existence of the limit of $G_d(q)$ as $d \to \infty$ (which we call $G(q)$), as well as a variational formula for it. Under this framework, Hypothesis 4.1 would follow if $G_d''$ was shown to converge uniformly to $G''$ in a neighborhood of $q = 0$.

- A sufficient condition for Hypothesis 4.1 to hold is to establish that $\sup_{|q|\le\varepsilon} |G_d^{(3)}(q)| \le C(\varepsilon)$ as $d \to \infty$. The third derivative of $G_d$ can be explicitly calculated from eq. (59), similar to our computation of the second derivative in Section 3. As discussed there, bounding the second derivative required demonstrating that $\mathrm{Var}_{\langle\cdot\rangle}[\mathrm{Tr}P[\mathbf{W}_1,\mathbf{W}_2]] = \mathcal{O}(1)$, where $P$ is a polynomial independent of $d$, and $\mathbf{W}_1, \mathbf{W}_2$ are two correlated GOE($d$) matrices conditioned to have spectral norm at most $\kappa$ (see eq. (37)). We established this bound in Lemma 3.5 using classical concentration techniques. Note that $\langle \mathrm{Tr}[P(\mathbf{W}_1,\mathbf{W}_2)]\rangle = \Theta(d)$, so we established strong concentration properties for this quantity. For the third derivative, however, we now essentially need to show that

$$\left\langle (\mathrm{Tr}P[\mathbf{W}_1,\mathbf{W}_2] - \langle \mathrm{Tr}P[\mathbf{W}_1,\mathbf{W}_2]\rangle)^3 \right\rangle = \mathcal{O}\left(\frac{1}{d}\right). \tag{60}$$

  Unfortunately, the bound established in Lemma 3.5 only yields eq. (60) with a right-hand side of $\mathcal{O}(1)$. Achieving a sharper bound would require an even more precise control over the statistics of $\mathrm{Tr} P[\mathbf{W}_1, \mathbf{W}_2]$ under the law $\langle \cdot \rangle$ of eq. (37), which we have not been able to achieve at present and leave as a direction for future work[1].

We come back to the proof of Theorem 1.9. If we assume that $\tau < \tau_{\mathrm{fail.}}(\kappa)$, by Lemma 4.1 and Hypothesis 4.1, there exists $\delta > 0$ and $\varepsilon > 0$ (depending on $\tau, \kappa$) such that, for $d$ large enough:

$$\inf_{|q| \leq \varepsilon} G_d''(q) \geq (1 + \delta). \tag{61}$$

Recall that $H(p) \coloneqq -p \log p - (1-p) \log(1-p)$ is the "binary entropy" function. We define $S(q) \coloneqq H[(1+q)/2] - \log 2$, and

$$\Phi_d(q) \coloneqq G_d(q) + S(q). \tag{62}$$

Since $S$ is a smooth function of $q$, and $S''(0) = -1$, from eq. (61) there exists new constants $(\varepsilon, \delta) > 0$ such that

$$\inf_{|q| \leq \varepsilon} \Phi_d''(q) \geq \delta. \tag{63}$$

Notice that $\Phi_d(0) = 0$ and $\Phi_d'(0) = 0$ (since $\Phi_d$ is an even function of $q$), so eq. (63) implies that for $d$ large enough:

$$\inf_{|q| \leq \varepsilon} \left[ \Phi_d(q) - \frac{\delta q^2}{2} \right] \geq 0. \tag{64}$$

Using the classical inequality that for any $l \in \{0, \cdots, n\}$:

$$\binom{n}{l} \geq \frac{1}{n+1} 2^{n H(l/n)},$$

we obtain from eq. (58):

$$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} \geq \frac{1}{n+1} \sum_{l=0}^{n} \exp\{n \Phi_d(q_l)\} \overset{\text{(a)}}{\geq} \frac{1}{n+1} \sum_{\substack{0 \leq l \leq n \\ |q_l| \leq \varepsilon}} \exp\left\{ \frac{n \delta q_l^2}{2} \right\},$$

where $q_l = 2(l/n) - 1$, and we used eq. (64) in (a). Choosing $l \in \{0, \cdots, n\}$ such that $\varepsilon/2 \leq |q_l| \leq \varepsilon$, we reach (recall that $n/d^2 \to \tau$):

$$\liminf_{d \to \infty} \frac{1}{d^2} \log \frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} \geq \frac{\delta \tau \varepsilon^2}{8} > 0,$$

which ends the proof of Theorem 1.9. $\square$

**Remark** – Notice that a statement akin to Theorem 1.9 might still hold even if $\Phi_d''(0) < 0$ for large $d$, as long as $\Phi_d$ reaches its global maximum in a value $q$ which is far from 0 as $d \to \infty$, as our argument can then be easily adapted to this setting. As such, we do not know if $\tau_{\mathrm{fail.}}(\kappa)$ (which comes out of our local analysis around $q = 0$) is a sharp threshold for $\mathbb{E}[Z_\kappa^2] \gg \mathbb{E}[Z_\kappa]^2$.

*Proof of Lemma 4.1 –.* We start from eq. (38), which for $q = 0$ gives:

$$G_d''(0) = \frac{d(d+1)}{2n} - \frac{d}{n} \mathbb{E}[\mathrm{Tr} \mathbf{W}^2] + \frac{d^2}{4n} \mathrm{Var}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']]. \tag{65}$$

[1] A bound $o_d(1)$ for the quantity of eq. (60) can be shown if the joint density of $(\mathbf{W}_1, \mathbf{W}_2)$ is proportional to $\exp\{-d \mathrm{Tr} V[\mathbf{W}_1, \mathbf{W}_2]\}$, for some strongly convex and polynomial potentials $V$, as a consequence of a central limit theorem for $\mathrm{Tr} P[\mathbf{W}_1, \mathbf{W}_2]$, see Guionnet (2009)[Chapter 9]. However even achieving this bound here appears non-trivial because our potential is singular as a consequence of the constraint $\mathbb{1}\{\|\mathbf{W}_a\|_{\mathrm{op}} \leq \kappa\}$ (see e.g. Borot and Guionnet (2013) for the case of single-matrix models).

In eq. (65), $\mathbf{W}$ and $\mathbf{W}'$ are sampled independently according to the law $\mathbb{P}_\kappa$ of $\mathbf{Z} \sim \mathrm{GOE}(d)$ conditioned on $\|\mathbf{Z}\|_{\mathrm{op}} \leq \kappa$, i.e. for any test function $\varphi$:

$$\mathbb{E}_{\mathbb{P}_\kappa}[\varphi(\mathbf{Z})] = \frac{\int \varphi(\mathbf{Z})\, \mathbb{1}\{\|\mathbf{Z}\|_{\mathrm{op}} \leq \kappa\} e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{Z}^2]} \mathrm{d}\mathbf{Z}}{\int \mathbb{1}\{\|\mathbf{Z}\|_{\mathrm{op}} \leq \kappa\} e^{-\frac{d}{4}\mathrm{Tr}[\mathbf{Z}^2]} \mathrm{d}\mathbf{Z}}. \tag{66}$$

We know that for $\mathbf{W} \sim \mathbb{P}_\kappa$, the empirical spectral distribution $\mu_{\mathbf{W}}$ weakly converges (a.s.) to $\mu_\kappa^\star$ given by Theorem 2.2. Since $\int \mu_{\mathbf{W}}(\mathrm{d}x) x^2 = \int \mu_{\mathbf{W}}(\mathrm{d}x) x^2 \mathbb{1}\{|x| \leq \kappa\}$, we have by the Portmanteau theorem and dominated convergence:

$$\lim_{d\to\infty} \frac{1}{d} \mathbb{E}[\mathrm{Tr} \mathbf{W}^2] = \int \mu_\kappa^\star(\mathrm{d}x)\, x^2\, \mathbb{1}\{|x| \leq \kappa\} = \int \mu_\kappa^\star(\mathrm{d}x)\, x^2. \tag{67}$$

We now focus on the last term of eq. (65). Notice that $\mathbb{E}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']] = \mathrm{Tr}[(\mathbb{E}\mathbf{W})^2] = 0$, since $\mathbb{E}\mathbf{W} = 0$ because $\mathbb{P}_\kappa$ is symmetric under $\mathbf{W} \leftrightarrow -\mathbf{W}$. Moreover, for any orthogonal matrix $\mathbf{O} \in \mathcal{O}(d)$, $\mathbf{W} \overset{\mathrm{d}}{=} \mathbf{O}\mathbf{W}\mathbf{O}^\top$ (as is directly seen from eq. (66)), so that we further have:

$$\mathrm{Var}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']] = \mathbb{E}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']^2] = \mathbb{E}_{\mathbf{O},\boldsymbol{\Lambda},\boldsymbol{\Lambda}'}[\mathrm{Tr}[\mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^\top\boldsymbol{\Lambda}']^2]. \tag{68}$$

In eq. (68), $\boldsymbol{\Lambda} = \mathrm{Diag}(\{\lambda_i\})$ is a diagonal matrix containing the eigenvalues of $\mathbf{W}$ (and similarly for $\boldsymbol{\Lambda}'$), and $\mathbf{O}$ is an orthogonal matrix sampled from the Haar measure on $\mathcal{O}(d)$, independently of $\mathbf{W}, \mathbf{W}'$. Thus:

$$\mathrm{Var}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']] = \sum_{i,j,k,l} \mathbb{E}[\lambda_i\lambda_k]\mathbb{E}[\lambda_j\lambda_l]\mathbb{E}[O_{ij}^2 O_{kl}^2]. \tag{69}$$

The terms involving $\lambda_i$ eq. (69) can be computed using the permutation invariance of the law of $\{\lambda_i\}$ as well as the invariance under $\boldsymbol{\Lambda} \leftrightarrow -\boldsymbol{\Lambda}$. Concretely, for all $i \in [d]$:

$$\mathbb{E}[\lambda_i^2] = \mathbb{E}[\lambda_1^2] = \frac{1}{d}\sum_{j=1}^d \mathbb{E}[\lambda_j^2] = \frac{1}{d}\mathbb{E}[\mathrm{Tr}\mathbf{W}^2], \tag{70}$$

and for $i \neq j$:

$$\mathbb{E}[\lambda_i\lambda_j] = \mathbb{E}[\lambda_1\lambda_2] = \frac{1}{d-1}\mathbb{E}\left[\lambda_1 \sum_{k\geq 2}\lambda_k\right] = \frac{1}{d(d-1)}\mathbb{E}[(\mathrm{Tr}\mathbf{W})^2 - \mathrm{Tr}(\mathbf{W}^2)]. \tag{71}$$

The first moments of the matrix elements of a Haar-sampled orthogonal matrix are elementary (see e.g. Banica, Collins, and Schlenker (2011) for general results):

$$\mathbb{E}[O_{ij}^2 O_{kl}^2] = \begin{cases} \dfrac{3}{d(d+2)} & (i = k \text{ and } j = l), \\ \dfrac{1}{d(d+2)} & (i = k \text{ and } j \neq l, \text{ or } i \neq k \text{ and } j = l), \\ \dfrac{d+1}{d(d-1)(d+2)} & (i \neq k \text{ and } j \neq l). \end{cases} \tag{72}$$

Using eq. (72) in eq. (69), separating cases in the sum, we get:

$$\begin{aligned} \mathrm{Var}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']] &= \frac{3}{d(d+2)} \cdot d^2 \cdot \mathbb{E}[\lambda_1^2]^2 + \frac{1}{d(d+1)} \cdot 2d^2(d-1) \cdot \mathbb{E}[\lambda_1^2]\mathbb{E}[\lambda_1\lambda_2] \\ &+ \frac{d+1}{d(d-1)(d+2)} \cdot d^2(d-1)^2 \cdot \mathbb{E}[\lambda_1\lambda_2]^2, \\ &= [1 + o_d(1)] \left(3\mathbb{E}[\lambda_1^2]^2 + 2d\mathbb{E}[\lambda_1\lambda_2]\mathbb{E}[\lambda_1^2] + d^2\mathbb{E}[\lambda_1\lambda_2]^2\right). \end{aligned} \tag{73}$$

From eqs. (67) and (70), we have $\mathbb{E}[\lambda_1^2] \to \mathbb{E}_{\mu_\kappa^\star}[X^2]$ as $d \to \infty$. Furthermore, by Lemma 3.5, $\mathbb{E}[(\mathrm{Tr}\mathbf{W})^2] = \mathrm{Var}[\mathrm{Tr}\mathbf{W}] = \mathcal{O}(1)$ as $d \to \infty$, so eq. (71) gives that $d\mathbb{E}[\lambda_1\lambda_2] \to -\mathbb{E}_{\mu_\kappa^\star}[X^2]$ as $d \to \infty$. Plugging these limits in eq. (73) we get:

$$\mathrm{Var}[\mathrm{Tr}[\mathbf{W}\mathbf{W}']] = 2\left(\int \mu_\kappa^\star(\mathrm{d}x)\, x^2\right)^2 + o_{d\to\infty}(1). \tag{74}$$

Finally, combining eqs. (65), (67) and (74) we obtain (recall $n/d^2 \to \tau$):

$$\lim_{d\to\infty} G_d''(0) = \frac{1}{\tau}\left[\frac{1}{2} - \int \mu_\kappa^\star(\mathrm{d}x)\, x^2 + \frac{1}{2}\left(-\frac{1}{2}\int \mu_\kappa^\star(\mathrm{d}x)\, x^2\right)^2\right]. \tag{75}$$

The integral in eq. (75) was already computed in eq. (25): plugging its value in eq. (75) shows that $\lim_{d\to\infty} G_d''(0) = \tau_{\mathrm{fail.}}(\kappa)/\tau$, which ends the proof of Lemma 4.1. □

# References


Abbe, Emmanuel, Shuangping Li, and Allan Sly (2022). "Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, pp. 327–338.

Achlioptas, Dimitris and Cristopher Moore (2002). "The asymptotic order of the random k-SAT threshold". In: *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.* IEEE, pp. 779–788.

Altschuler, Dylan J (2023). "Zero-One Laws for Random Feasibility Problems". In: *arXiv preprint arXiv:2309.13133*.

Altschuler, Dylan J and Jonathan Niles-Weed (2022). "The discrepancy of random rectangular matrices". In: *Random Structures & Algorithms* 60.4, pp. 551–593.

Anderson, Greg W, Alice Guionnet, and Ofer Zeitouni (2010). *An introduction to random matrices.* Cambridge university press.

Aubin, Benjamin, Will Perkins, and Lenka Zdeborová (2019). "Storage capacity in symmetric binary perceptrons". In: *Journal of Physics A: Mathematical and Theoretical* 52.29, p. 294003.

Bandeira, Afonso, March Boedihardjo, and Ramon van Handel (2023). "Matrix concentration inequalities and free probability". In: *Inventiones mathematicae* 234.1, pp. 419–487.

Banica, Teodor, Benoit Collins, and Jean-Marc Schlenker (2011). "On polynomial integrals over the orthogonal group". In: *Journal of Combinatorial Theory, Series A* 118.3, pp. 778–795.

Bansal, Nikhil (2010). "Constructive algorithms for discrepancy minimization". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, pp. 3–10.

Bansal, Nikhil, Haotian Jiang, and Raghu Meka (2023). "Resolving matrix Spencer conjecture up to poly-logarithmic rank". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pp. 1814–1819.

Bansal, Nikhil and Joel Spencer (2020). "On-line balancing of random inputs". In: *Random Structures & Algorithms* 57.4, pp. 879–891.

Barbier, Damien (2024). "How to escape atypical regions in the symmetric binary perceptron: a journey through connected-solutions states". In: *arXiv preprint arXiv:2408.04479*.

Barbier, Damien et al. (2024). "On the atypical solutions of the symmetric binary perceptron". In: *Journal of Physics A: Mathematical and Theoretical* 57.19, p. 195202.

Batson, Joshua, Daniel A Spielman, and Nikhil Srivastava (2014). "Twice-ramanujan sparsifiers". In: *SIAM Review* 56.2, pp. 315–334.

Ben Arous, Gérard and Alice Guionnet (1997). "Large deviations for Wigner's law and Voiculescu's non-commutative entropy". In: *Probability theory and related fields* 108, pp. 517–542.

Bogachev, Vladimir Igorevich and Maria Aparecida Soares Ruas (2007). *Measure theory.* Vol. 2. Springer.

Borot, Gaëtan and Alice Guionnet (2013). “Asymptotic expansion of $\beta$ matrix models in the one-cut regime”. In: *Communications in Mathematical Physics* 317, pp. 447–483.
Bouali, Mohamed (2015). “Constrained Eigenvalues Density of Invariant Random Matrices Ensembles”. In: *arXiv preprint arXiv:1509.03745*.
Chen, William, Anand Srivastav, and Giancarlo Travaglini (2014). *A panorama of discrepancy theory*. Vol. 2107. Springer.
Costello, Kevin P (2009). “Balancing gaussian vectors”. In: *Israel Journal of Mathematics* 172.1, pp. 145–156.
Cover, Thomas M (1965). “Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition”. In: *IEEE transactions on electronic computers*, pp. 326–334.
Dadush, Daniel, Haotian Jiang, and Victor Reis (2022). “A new framework for matrix discrepancy: partial coloring bounds via mirror descent”. In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 649–658.
Dean, David S and Satya N Majumdar (2006). “Large deviations of extreme eigenvalues of random matrices”. In: *Physical review letters* 97.16, p. 160201.
— (2008). “Extreme value statistics of eigenvalues of Gaussian random matrices”. In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 77.4, p. 041108.
El Alaoui, Ahmed and David Gamarnik (2024). “Hardness of sampling solutions from the Symmetric Binary Perceptron”. In: *arXiv preprint arXiv:2407.16627*.
Fan, Chenjie et al. (2015). “Convergence of eigenvalues to the support of the limiting measure in critical $\beta$ matrix models”. In: *Random Matrices: Theory and Applications* 4.03, p. 1550013.
Faraut, Jacques (2014). “Logarithmic potential theory, orthogonal polynomials, and random matrices”. In: *Modern methods in multivariate statistics, Lecture Notes of CIMPA-FECYT-UNESCO-ANR. Hermann.*
Gamarnik, David et al. (2022). “Algorithms and barriers in the symmetric binary perceptron model”. In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, pp. 576–587.
Gardner, Elizabeth (1988). “The space of interactions in neural network models”. In: *Journal of physics A: Mathematical and general* 21.1, p. 257.
Gardner, Elizabeth and Bernard Derrida (1988). “Optimal storage properties of neural network models”. In: *Journal of Physics A: Mathematical and general* 21.1, p. 271.
Guionnet, Alice (2004). “First order asymptotics of matrix integrals; a rigorous approach towards the understanding of matrix models”. In: *Communications in mathematical physics* 244.3, pp. 527–569.
— (2009). *Large Random Matrices: Lectures on Macroscopic Asymptotics: École d’Été de Probabilités de Saint-Flour 2006*. Vol. 36. Springer.
— (2022). “Rare events in random matrix theory”. In: *Proc. Int. Cong. Math.* Vol. 2, pp. 1008–1052.
Hopkins, Samuel B, Prasad Raghavendra, and Abhishek Shetty (2022). “Matrix discrepancy from quantum communication”. In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 637–648.
Janson, Svante (1996). “The second moment method, conditioning and approximation”. In: *Random discrete structures (Minneapolis, MN, 1993)*. Vol. 76. IMA Vol. Math. Appl. Springer, New York, pp. 175–183.
Karmarkar, Narendra et al. (1986). “Probabilistic analysis of optimum partitioning”. In: *Journal of Applied probability* 23.3, pp. 626–645.
Kizildag, Eren C and Tanay Wakhare (2023). “Symmetric Perceptron with Random Labels”. In: *Fourteenth International Conference on Sampling Theory and Applications*.
Krauth, Werner and Marc Mézard (1989). “Storage capacity of memory networks with binary couplings”. In: *Journal de Physique* 50.20, pp. 3057–3066.

Kunisky, Dmitriy and Peiyuan Zhang (2023). "Online algorithms and lower bounds for average-case matrix discrepancy". In: *arXiv preprint arXiv:2307.10055*.
Kyng, Rasmus, Kyle Luh, and Zhao Song (2020). "Four deviations suffice for rank 1 matrices". In: *Advances in Mathematics* 375, p. 107366.
Levy, Avi, Harishchandra Ramadas, and Thomas Rothvoss (2017). "Deterministic discrepancy minimization via the multiplicative weight update method". In: *International Conference on Integer Programming and Combinatorial Optimization*. Springer, pp. 380–391.
Lust-Piquard, Françoise and Gilles Pisier (1991). "Non commutative Khintchine and Paley inequalities". In: *Arkiv för matematik* 29, pp. 241–260.
Majumdar, Satya N and Grégory Schehr (2014). "Top eigenvalue of a random matrix: large deviations and third order phase transition". In: *Journal of Statistical Mechanics: Theory and Experiment* 2014.1, P01012.
Marcus, Adam W, Daniel A Spielman, and Nikhil Srivastava (2015). "Interlacing families II: Mixed characteristic polynomials and the Kadison—Singer problem". In: *Annals of Mathematics*, pp. 327–350.
Matousek, Jiri (2009). *Geometric discrepancy: An illustrated guide*. Vol. 18. Springer Science & Business Media.
Mehta, Madan Lal (2014). *Random Matrices: Revised and Enlarged Second Edition*. Elsevier.
Meka, Raghu (2014). *Discrepancy and beating the union bound*. https://windowsontheory.org/2014/02/07/discrepancy-and-beating-the-union-bound.
Mhaskar, HN and Edward B. Saff (1985). "Where does the sup norm of a weighted polynomial live? A Generalization of Incomplete Polynomials". In: *Constructive Approximation* 1, pp. 71–91.
Nakajima, Shuta and Nike Sun (2023). "Sharp threshold sequence and universality for ising perceptron models". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, pp. 638–674.
Perkins, Will and Changji Xu (2021). "Frozen 1-RSB structure of the symmetric Ising perceptron". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1579–1588.
Saff, Edward B. and Vilmos Totik (2013). *Logarithmic potentials with external fields*. Vol. 316. Springer Science & Business Media.
Saumard, Adrien and Jon A Wellner (2014). "Log-concavity and strong log-concavity: a review". In: *Statistics surveys* 8, p. 45.
Sompolinsky, Haim, Naftali Tishby, and H Sebastian Seung (1990). "Learning from examples in large neural networks". In: *Physical Review Letters* 65.13, p. 1683.
Spencer, Joel (1985). "Six standard deviations suffice". In: *Transactions of the American mathematical society* 289.2, pp. 679–706.
— (1994). *Ten lectures on the probabilistic method*. SIAM.
Talagrand, Michel (1999a). "Intersecting random half cubes". In: *Random Structures & Algorithms* 15.3-4, pp. 436–449.
— (1999b). "Self averaging and the space of interactions in neural networks". In: *Random Structures & Algorithms* 14.3, pp. 199–213.
— (2010). *Mean field models for spin glasses: Volume I: Basic examples*. Vol. 54. Springer Science & Business Media.
— (2011). *Mean Field Models for Spin Glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*. Vol. 55. Springer Science & Business Media.
Tricomi, Francesco Giacomo (1985). *Integral equations*. Vol. 5. Courier corporation.
Turner, Paxton, Raghu Meka, and Philippe Rigollet (2020). "Balancing Gaussian vectors in high dimension". In: *Conference on Learning Theory*. PMLR, pp. 3455–3486.
Vershynin, Roman (2018). *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press.

Vivo, Pierpaolo, Satya N Majumdar, and Oriol Bohigas (2007). “Large deviations of the maximum eigenvalue in Wishart random matrices”. In: *Journal of Physics A: Mathematical and Theoretical* 40.16, p. 4317.
Wengiel, Nicola (2024). “Asymptotic Discrepancy of Gaussian Orthogonal Ensemble Matrices”. In: *arXiv preprint arXiv:2410.23915*.
Wigner, Eugene P (1955). “Characteristic Vectors of Bordered Matrices With Infinite Dimensions”. In: *Annals of Mathematics* 62.3, pp. 548–564.
Xu, Changji (2021). “Sharp threshold for the Ising perceptron model”. In: *The Annals of Probability* 49.5, pp. 2399–2415.
Zouzias, Anastasios (2012). “A matrix hyperbolic cosine algorithm and applications”. In: *International Colloquium on Automata, Languages, and Programming*. Springer, pp. 846–858.

# A Contiguity with a planted model, and freezing of solutions

We discuss briefly here Open Problem 1.3, and more specifically how freezing could be established, or at least conjectured, from the second moment computation. For the symmetric binary perceptron (SBP), freezing was conjectured in this way in Aubin, Perkins, and Zdeborová (2019), before being established rigorously in Perkins and Xu (2021) and Abbe, Li, and Sly (2022).

Let us first fix some notations. For the remainder of this section we assume that the margin $\kappa \in (0, 2)$ is given and fixed. **FIXME:** Use $\Sigma_n$ here! We denote $\Sigma_d := \{\pm 1\}^d$, and $\mathcal{S}_d$ the set of symmetric $d \times d$ matrices. We let $\mathbf{H} := (\mathbf{W}_1, \cdots, \mathbf{W}_n)$, and denote $S(\mathbf{H}) := \{\varepsilon \in \{\pm 1\}^n \text{ s.t. } \|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \kappa\sqrt{n}\}$. Recall that $Z_\kappa = |S(\mathbf{H})|$, see eq. (2).

**Step 1: contiguity** – Freezing of solutions is a property of $\mathrm{Unif}(S(\mathbf{H}))$, the uniform measure over the set of solutions. The classical approach to establish properties of this measure is to use a contiguity argument with a *planted* version of the problem, which is often easier to study. Concretely, we define two models.

- **Random model** ($\mathbb{P}_{\mathrm{ra.}}$): draw first $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, conditioned on $S(\mathbf{H}) \neq \emptyset$. Draw then $\varepsilon \sim \mathrm{Unif}(S(\mathbf{H}))$.
- **Planted model** ($\mathbb{P}_{\mathrm{pl.}}$): draw first $\varepsilon \sim \mathrm{Unif}(\Sigma_d)$. Draw then $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, conditioned on satisfying $\|\sum_i \varepsilon_i \mathbf{W}_i\|_{\mathrm{op}} \leq \kappa\sqrt{n}$.

$\mathbb{P}_{\mathrm{ra.}}$ and $\mathbb{P}_{\mathrm{pl.}}$ define two probability measures over $\mathcal{S}_d^n \times \Sigma_d$. In the planted distribution, we often denote $\varepsilon = \varepsilon^\star$. In many problems, $\mathbb{P}_{\mathrm{pl.}}$ turns out to be easier to study than $\mathbb{P}_{\mathrm{ra.}}$, and moreover we have the general identity, for any $(\varepsilon, \mathbf{H})$ such that $\varepsilon \in S(\mathbf{H})$:

$$\frac{\mathrm{d}\mathbb{P}_{\mathrm{pl.}}}{\mathrm{d}\mathbb{P}_{\mathrm{ra.}}}(\varepsilon, \mathbf{H}) = \frac{Z_\kappa(\mathbf{H})}{\mathbb{E}_0[Z_\kappa(\mathbf{H}')]} \cdot \mathbb{P}_0(S(\mathbf{H}') \neq 0), \tag{76}$$

where $\mathbb{P}_0$ is the law of $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$. When $\mathbb{P}_0(S(\mathbf{H}') \neq 0) \to 1$ as $d \to \infty$, contiguity of $\mathbb{P}_{\mathrm{pl.}}$ and $\mathbb{P}_{\mathrm{ra.}}$ thus follows from concentration of $Z_\kappa$ on its average, and in particular
e refer to the previously-cited literature on the SBP for further references on planting and contiguity in constraint satisfaction problems.

**Step 2: the planted solutions is isolated** – Once contiguity with the planted model is established, a key point is to notice that of $q = 1 - 2l/n \in [-1, 1]$: **FIXME:** Do I use $\langle \cdot \rangle$ for scalar product?

$$\begin{aligned}
\mathbb{P}_{\mathrm{pl.}}(\exists \varepsilon \in S(\mathbf{H}) \backslash \{\varepsilon^\star\} : \langle \varepsilon, \varepsilon^\star \rangle = nq) &\leq \mathbb{E}_{\mathrm{pl.}}\left[\#\{\varepsilon \in S(\mathbf{H}) \backslash \{\varepsilon^\star\} : \langle \varepsilon, \varepsilon^\star \rangle = nq\}\right], \\
&\overset{(a)}{\leq} \binom{n}{l} \mathbb{P}_{\mathrm{pl.}}\left[\left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\right],
\end{aligned}$$

where in (a) we used the invariance of the law of $\mathrm{GOE}(d)$, and the probability is the same for all $\varepsilon \in \Sigma_n$ such that $\langle \varepsilon, \varepsilon^\star \rangle = nq$. **TODO:** Finish

# B Large deviations for the spectral norm of a Gaussian matrix

For completeness, we provide here an alternative proof of Corollary 2.4, which appeared in an earlier version of this manuscript, and which relies only on the the main result of Ben Arous and

Guionnet (1997), that is a large deviation principle for the empirical eigenvalue distribution of $\mathrm{GOE}(d)$ matrices[1]. Recall that $\Sigma(\mu) := \int \mu(\mathrm{d}x)\mu(\mathrm{d}y) \log|x-y|$.

**Proposition B.1** (Ben Arous and Guionnet (1997))**.** *Let* $\mathbf{W} \sim \mathrm{GOE}(d)$*. We denote* $\mu_{\mathbf{W}} := (1/d)\sum_{i=1}^d \delta_{\lambda_i(\mathbf{W})}$ *its empirical spectral distribution. For* $\mu \in \mathcal{M}_1^+(\mathbb{R})$*, recall eq. (19):*

$$I(\mu) := -\frac{1}{2}\Sigma(\mu) + \frac{1}{4}\int \mu(\mathrm{d}x)\, x^2 - \frac{3}{8}.$$

*Then:*

*(i)* $I : \mathcal{M}_1^+(\mathbb{R}) \to [0,\infty]$ *is a strictly convex function and a good rate function, i.e.* $\{I_1 \leq M\}$ *is a compact subset of* $\mathcal{M}_1^+(\mathbb{R})$ *for any* $M > 0$.

*(ii) The law of* $\mu_{\mathbf{W}}$ *satisfies a large deviation principle, in the scale* $d^2$*, with rate function* $I$*, that is for any open (respectively closed) subset* $O \subseteq \mathcal{M}_1^+(\mathbb{R})$ *(respectively* $F \subseteq \mathcal{M}_1^+(\mathbb{R})$*):*

$$\begin{cases} \liminf_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}[\mu_{\mathbf{W}} \in O] \geq -\inf_{\mu\in O} I(\mu), \\ \limsup_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}[\mu_{\mathbf{W}} \in F] \leq -\inf_{\mu\in F} I(\mu). \end{cases}$$

**Remark** – Ben Arous and Guionnet (1997) also prove large deviations results for the empirical measure of $\mathbf{W}$ sampled under more general distributions, with density proportional to $\exp\{-d\mathrm{Tr}[V(\mathbf{W})]/2\}$, for a continuous potential $V$ growing fast enough at infinity. Here we essentially adapt these results to the discontinuous potential

$$V(x) = \frac{x^2}{2} + \mathbb{1}\{|x| > \kappa\} \times \infty.$$

## B.1 Large deviation upper bound

Let $Q := \{\mu \in \mathcal{M}_1^+(\mathbb{R}) \, : \, \mu([-\kappa,\kappa]) = 1\}$. By the Portmanteau theorem, $Q$ is sequentially closed under weak convergence, and thus closed since the weak topology on $\mathcal{M}_1^+(\mathbb{R})$ is metrizable. We apply Proposition B.1 to get:

$$\begin{aligned} \limsup_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] &= \limsup_{d\to\infty} \frac{1}{d^2} \log \mathbb{P}[\mu_{\mathbf{W}} \in Q], \\ &\leq -\inf_{\mu\in Q} I(\mu), \\ &= -\inf_{\mu\in\mathcal{M}_1^+([-\kappa,\kappa])} I(\mu). \end{aligned} \tag{77}$$

The last equality follows since $I(\mu_{|[-\kappa,\kappa]}) = I(\mu)$ for all $\mu \in Q$. This proves the upper bound for $(1/d^2)\log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa]$ in Corollary 2.4

## B.2 Large deviation lower bound

We focus now on the lower bound for $(1/d^2)\log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa]$ in Corollary 2.4.

Unfortunately the large deviation statement of Proposition B.1 is not enough to obtain the corresponding lower bound to eq. (77), because the set of probability measures supported in $[-\kappa,\kappa]$ has empty interior under the weak topology[2]. Instead, we come back to the joint law of

[1]Notice that Ben Arous and Guionnet (1997) use a convention where $\mathrm{GOE}(d)$ matrices have off-diagonal entries with variance $1/(2d)$. We state their result adapted to our conventions.

[2]For any $\mu \in \mathcal{M}_1^+(\mathbb{R})$, there is a sequence $\mu_n$ weakly converging to $\mu$ while $\mathrm{supp}(\mu_n) \nsubseteq [-\kappa,\kappa]$.

the eigenvalues of a GOE($d$) matrix, and restrict the integration domain to a small neighborhood of the quantiles of $\mu^\star_\kappa$. This strategy is similar to the one used in the proof of the large deviation lower bound in Ben Arous and Guionnet (1997).

We use the results of Section 2.2, which are independent of Corollary 2.4, and characterize the minimizing measure of $I(\mu)$. In particular, by Theorem 2.5, we know that $I(\mu)$ has a unique minimizer in $\mathcal{M}_1^+([-\kappa,\kappa])$, which we denote $\mu^\star_\kappa$, and it has a density $\rho_\kappa$ given in eq. (22). Let $\delta \in (0,\kappa)$. In what follows, we let $\nu_\delta := \mu^\star_{\kappa-\delta}$. We define the quantiles of $\nu_\delta$ as

$$-(\kappa-\delta) = x_0^{(d)} < x_1^{(d)} < \cdots < x_d^{(d)} < x_{d+1}^{(d)} = \kappa - \delta,$$

with, for all $i \in \{0, \cdots, d\}$:

$$\int_{x_i^{(d)}}^{x_{i+1}^{(d)}} \nu_\delta(u)\,\mathrm{d}u = \frac{1}{d+1}.$$

We will drop the subscript and write $x_i$ for $x_i^{(d)}$ to lighten notations.

Clearly, the empirical measure $(1/d)\sum_{i=1}^d \delta_{x_i}$ weakly converges to $\nu_\delta$ as $d \to \infty$. Notice that $\{(\lambda_i)_{i=1}^d : |\lambda_i - x_i| \leq \delta, \ \forall i \in [1,d]\} \subseteq [-\kappa,\kappa]^d$, so that from eqs. (17) and (18):

$$\begin{aligned}
&e^{-\frac{3d^2}{8}+o(d^2)}\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] \qquad (78)\\
&\geq \int_{[-\delta,\delta]^d} \prod_{i<j} |u_i - u_j + x_i - x_j| e^{-\frac{d}{4}\sum_{i=1}^d (u_i+x_i)^2} \prod_{i=1}^d \mathrm{d}u_i,\\
&\overset{(a)}{\geq} \int_{[-\delta,\delta]^d\cap\Delta_d} \prod_{i<j} (u_j - u_i + x_j - x_i) e^{-\frac{d}{4}\sum_{i=1}^d (u_i+x_i)^2} \prod_{i=1}^d \mathrm{d}u_i,\\
&\overset{(b)}{\geq} \int_{[-\delta,\delta]^d\cap\Delta_d} \prod_{i+1<j} (x_j - x_i) \prod_{i=1}^{d-1} [x_{i+1}-x_i]^{1/2}[u_{i+1}-u_i]^{1/2} e^{-\frac{d}{4}\sum_{i=1}^d (\delta+|x_i|)^2} \prod_{i=1}^d \mathrm{d}u_i,
\end{aligned}$$

where we defined in (a) the set $\Delta_d := \{u_1 < \cdots < u_d\}$, and used in (b) that $u_i \leq u_j$ and $x_i \leq x_j$, as well as the inequality $A + B \geq \sqrt{AB}$ for $A, B \geq 0$. The integral on the variables $u_i$ can be lower-bounded as follows:

$$\begin{aligned}
\int_{[-\delta,\delta]^d\cap\Delta_d} \prod_{i=1}^{d-1}\sqrt{u_{i+1}-u_i}\prod_{i=1}^d \mathrm{d}u_i &= \delta^{(3d-1)/2}\int_{[-1,1]^d\cap\Delta_d}\prod_{i=1}^{d-1}\sqrt{u_{i+1}-u_i}\prod_{i=1}^d \mathrm{d}u_i,\\
&\geq \delta^{(3d-1)/2}\prod_{i=1}^d \int_{-1+\frac{2(i-1)}{d}}^{-1+\frac{2i-1}{d}} \mathrm{d}u_i \left(\prod_{i=1}^{d-1}\sqrt{u_{i+1}-u_i}\right),\\
&\geq \left(\frac{\delta}{d}\right)^{(3d-1)/2}. \qquad (79)
\end{aligned}$$

Combining eqs. (78) and (79):

$$\begin{aligned}
\liminf_{d\to\infty} \frac{1}{d^2}\log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] \geq \frac{3}{8} - \frac{\delta^2}{4} + \liminf_{d\to\infty}\Bigg[&-\frac{\delta}{2d}\sum_{i=1}^d |x_i| - \frac{1}{4d}\sum_{i=1}^d x_i^2\\
&+\frac{1}{2d^2}\sum_{i=1}^{d-1}\log(x_{i+1}-x_i) + \frac{1}{d^2}\sum_{i,j=1}^d \log(x_j - x_i)\mathbb{1}\{j > i+1\}\Bigg].
\end{aligned}$$

By the weak convergence described above, and since $\sum_{i=1}^d x_i^2 = \sum_{i=1}^d x_i^2 \mathbb{1}\{|x_i| \leq \kappa\}$, we get by the Portmanteau theorem:

$$\liminf_{d\to\infty}\frac{1}{d^2}\log\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] \geq \frac{3}{8} - \frac{\delta^2}{4} - \frac{\delta}{2}\int |x|\nu_\delta(\mathrm{d}x) - \frac{1}{4}\int x^2 \nu_\delta(\mathrm{d}x) \qquad (80)$$

$$+\liminf_{d\to\infty}\left[\frac{1}{2d^2}\sum_{i=1}^{d-1}\log(x_{i+1}-x_i)+\frac{1}{d^2}\sum_{i,j=1}^{d}\log(x_j-x_i)\mathbb{1}\{j>i+1\}\right].$$

Finally, notice that

$$\begin{aligned}
\Sigma(\nu_\delta) &= 2\sum_{0\le i,j\le d}\int_{x_i}^{x_{i+1}}\nu_\delta(\mathrm{d}x)\int_{x_j}^{x_{j+1}}\nu_\delta(\mathrm{d}y)\,\log(y-x)\,\mathbb{1}\{x<y\},\\
&\le \sum_{i=0}^{d}\int_{x_i}^{x_{i+1}}\nu_\delta(\mathrm{d}x)\int_{x_i}^{x_{i+1}}\nu_\delta(\mathrm{d}y)\,\log|y-x| + 2\sum_{0\le i<j\le d}\frac{\log(x_{j+1}-x_i)}{(d+1)^2},\\
&\le \frac{1}{(d+1)^2}\left[\sum_{i=0}^{d}\log(x_{i+1}-x_i)+2\sum_{0\le i<j\le d}\log(x_{j+1}-x_i)\right],\\
&\overset{\text{(a)}}{\le} \frac{1}{(d+1)^2}\left[\sum_{i=1}^{d-1}\log(x_{i+1}-x_i)+2\sum_{i,j=1}^{d}\log(x_j-x_i)\mathbb{1}\{j>i+1\}\right]+\frac{2d+1}{(d+1)^2}\log 2\kappa.
\end{aligned}$$

We used in (a) that $|x_i-x_j|\le 2(\kappa-\delta)\le 2\kappa$ for all $i,j$. Using this in eq. (80) gives:

$$\begin{aligned}
\liminf_{d\to\infty}\frac{1}{d^2}\log\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}}\le\kappa] &\ge \frac{3}{8}-\frac{\delta^2}{4}-\frac{\delta}{2}\int|x|\nu_\delta(\mathrm{d}x)-\frac{1}{4}\int x^2\nu_\delta(\mathrm{d}x)+\frac{1}{2}\Sigma(\nu_\delta),\\
&\ge -\frac{\delta}{2}\int|x|\nu_\delta(\mathrm{d}x)-\frac{\delta^2}{4}-I(\nu_\delta).
\end{aligned}$$

Recall that $\nu_\delta=\mu^\star_{\kappa-\delta}$, so that taking the limit $\delta\to 0$, we get:

$$\liminf_{d\to\infty}\frac{1}{d^2}\log\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}}\le\kappa]\ge -I(\mu^\star_\kappa)=-\inf_{\mu\in\mathcal{M}_1^+([-\kappa,\kappa])}I(\mu). \tag{81}$$

Combining eqs. (77) and (81) ends the proof of Corollary 2.4. □