

Exact threshold for approximate ellipsoid fitting of random points

Antoine Maillard^{*, \diamond} , Afonso S. Bandeira^{*}

October 10, 2023

Abstract

We consider the problem (P) of exactly fitting an ellipsoid (centered at 0) to n standard Gaussian random vectors in \mathbb{R}^d , as $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha > 0$. This problem is conjectured to undergo a sharp transition: with high probability, (P) has a solution if $\alpha < 1/4$, while (P) has no solutions if $\alpha > 1/4$. So far, only a trivial bound $\alpha > 1/2$ is known to imply the absence of solutions, while the sharpest results on the positive side assume $\alpha \leq \eta$ (for $\eta > 0$ a small constant) to prove that (P) is solvable. In this work we study universality between this problem and a so-called ‘‘Gaussian equivalent’’, for which the same transition can be rigorously analyzed. Our main results are twofold. On the positive side, we prove that if $\alpha < 1/4$, there exist an ellipsoid fitting all the points up to a small error, and that the lengths of its principal axes are bounded above and below. On the other hand, for $\alpha > 1/4$, we show that achieving small fitting error is not possible if the length of the ellipsoid’s shortest axis does not approach 0 as $d \rightarrow \infty$ (and in particular there does not exist any ellipsoid fit whose shortest axis length is bounded away from 0 as $d \rightarrow \infty$). To the best of our knowledge, our work is the first rigorous result characterizing the expected phase transition in ellipsoid fitting at $\alpha = 1/4$. In a companion non-rigorous work, the first author and D. Kunisky give a general analysis of ellipsoid fitting using the replica method of statistical physics, which inspired the present work.

1 Introduction and main results

1.1 The ellipsoid fitting conjecture

We consider the *random ellipsoid fitting problem*: given n random standard Gaussian vectors in dimension d , when do they all lie on the boundary of a (centered) ellipsoid? Formally, we define an ellipsoid fit using the set \mathcal{S}_d of $d \times d$ real symmetric matrices, as follows.

Definition 1.1 (*Ellipsoid fit*)

Let $x_1, \dots, x_n \in \mathbb{R}^d$. We say that $S \in \mathcal{S}_d$ is an *ellipsoid fit* for $(x_\mu)_{\mu=1}^n$ if it satisfies:

$$(P) : \begin{cases} x_\mu^\top S x_\mu = d \text{ for all } \mu \in \{1, \dots, n\}, \\ S \succeq 0. \end{cases} \quad (1)$$

In Definition 1.1, the matrix $S \succeq 0$ defines the ellipsoid $\Sigma := \{x \in \mathbb{R}^d : x^\top S x = d\}$. Geometrically speaking, the eigenvectors of S give the directions of the principal axes of the ellipsoid, while its eigenvalues $(\lambda_i)_{i=1}^d$ are related to the lengths $(r_i)_{i=1}^d$ of its principal (semi-)axes by $r_i = \sqrt{d\lambda}^{-1/2}$.

Scaling – In what follows, we will rather refer to the rescaled quantities $r'_i = r_i/\sqrt{d}$ as the lengths of the ellipsoid axes, effectively rescaling distances so that the sphere of radius \sqrt{d} (with $S = I_d$) has all

^{*} Department of Mathematics, ETH Zürich, Switzerland.

^{\diamond} To whom correspondence shall be sent: antoine.maillard@math.ethz.ch.

(semi-)axes of length 1. In particular, the lengths of the ellipsoid's longest and shortest axis are then respectively $\lambda_{\min}(S)^{-1/2}$ and $\lambda_{\max}(S)^{-1/2}$.

We are interested in finding an ellipsoid fit to a set of *random* points $x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I}_d)$. The question of the existence of such an ellipsoid arose first in [1–3], which conjectured the following.

Conjecture 1.1 (*The ellipsoid fitting conjecture*)

Let $n, d \geq 1$, and $x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I}_d)$. Let $p(n, d) := \mathbb{P}[\exists S \in \mathcal{S}_d \text{ an ellipsoid fit for } (x_\mu)_{\mu=1}^n]$. For any $\varepsilon > 0$, the following holds:

$$\limsup_{d \rightarrow \infty} \frac{n}{d^2} \leq \frac{1 - \varepsilon}{4} \Rightarrow \lim_{d \rightarrow \infty} p(n, d) = 1, \tag{2}$$

$$\liminf_{d \rightarrow \infty} \frac{n}{d^2} \geq \frac{1 + \varepsilon}{4} \Rightarrow \lim_{d \rightarrow \infty} p(n, d) = 0. \tag{3}$$

Informally, Conjecture 1.1 predicts a sharp transition for the existence of an ellipsoid fit in the regime $n/d^2 \rightarrow \alpha > 0$ exactly at $\alpha = 1/4$.

1.2 Related works

Conjecture 1.1 was first stated and studied in the series of works [1–3], where it arose as being connected to the decomposition of a (random) data matrix M as $M = L + D$, with $L \succeq 0$ being low-rank, and D a diagonal matrix. Connections to other problems throughout theoretical computer science have since then been unveiled, such as certifying a lower bound on the discrepancy of a random matrix using a canonical semidefinite relaxation [2, 4], overcomplete independent component analysis [5], or Sum-of-Squares lower bound for the Sherrington-Kirkpatrick Hamiltonian [6]. We refer the reader to the more detailed expositions of [4, 7] on the connections of ellipsoid fitting to theoretical computer science and machine learning.

Interestingly, Conjecture 1.1 arose both from numerical evidence¹ and the remark that $d^2/4$ is known to be the statistical dimension (or squared Gaussian width) of \mathcal{S}_d^+ , the set of positive semidefinite matrices [8, 9]. As such, if one replaces eq. (1) by

$$(P_{\text{Gauss.}}) : \begin{cases} \text{Tr}[SG_\mu] = d \text{ for all } \mu \in \{1, \dots, n\}, \\ S \succeq 0, \end{cases} \tag{4}$$

in which $(G_\mu)_{\mu=1}^n$ are (independent) standard Gaussian matrices, Conjecture 1.1 provably holds for $(P_{\text{Gauss.}})$. The crucial property of $(P_{\text{Gauss.}})$ behind this result is that the affine subspace $\{S \in \mathcal{S}_d : (\text{Tr}[SG_\mu] = d)_{\mu=1}^n\}$ is randomly oriented, *uniformly* in all directions. Although this motivation for the conjecture was known, our work is (to the best of the knowledge) the first mathematically rigorous approach to leverage the connection between (P) and $(P_{\text{Gauss.}})$.

Indeed, previous progress on Conjecture 1.1 has mostly focused on proving the existence of a fitting ellipsoid using an ansatz solution: the first line of eq. (1) defines an affine subspace V of symmetric matrices of codimension n , so one can study a well-chosen $S^* \in V$, and argue that for small enough n it satisfies $S^* \succeq 0$ with high probability as $d \rightarrow \infty$. Various such constructions have been used, and we summarize in Fig. 1 the current rigorous progress on the ellipsoid fitting conjecture that arose from these approaches. Presently, the best rigorous results on Conjecture 1.1 are due to the recent works [11–13] and can be summarized as follows:

¹Given $(x_\mu)_{\mu=1}^n$, eq. (1) is a convex problem (it is an example of a semidefinite program, or SDP), and thus efficiently solvable when solutions exist.

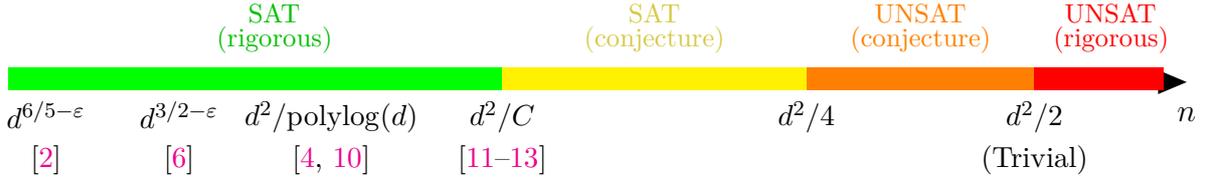


Figure 1: A summary of the current state of the ellipsoid fitting conjecture. In red, we show regions for which ellipsoid fitting is rigorously known to be unsatisfiable (UNSAT), and in orange regions which are conjectured to be. Similarly, we show in green regions rigorously known to be satisfiable (SAT), and in yellow regions which are conjectured to be so. Figure is taken from the companion work [7] to this manuscript.

Theorem 1.2 ([11–13])

Let $n, d \geq 1$, and $x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$. Let $p(n, d) := \mathbb{P}[\exists S \in \mathcal{S}_d \text{ an ellipsoid fit for } (x_\mu)_{\mu=1}^n]$. There exists a (small) universal constant $\eta > 0$ such that:

$$\limsup_{d \rightarrow \infty} \frac{n}{d^2} \leq \eta \Rightarrow \lim_{d \rightarrow \infty} p(n, d) = 1,$$

Moreover, if $n > d(d+1)/2$, then $p(n, d) = 0$.

Note that the bound $n > d(d+1)/2$ in Theorem 1.2 arises from a simple dimension counting argument, as $d(d+1)/2$ is the dimension of the space of symmetric matrices: for such values of n , not only does there not exist a solution to eq. (1), there does not exist any solution even without the constraint $S \succeq 0$!

Statistical physics approaches: heuristic and rigorous – In this work, we tackle Conjecture 1.1 using techniques inspired by the statistical physics of disordered systems. While analytical methods developed in this field were originally designed to study models known as spin glasses [14, 15], they have seen in the past decades a great number of applications in high dimensional statistics, theoretical computer science, and machine learning. Moreover, despite these techniques often being non-rigorous, a growing line of mathematics literature has emerged establishing many of their predictions. Notably, ellipsoid fitting is an example of a *semidefinite program (SDP)*² with random linear constraints, and some such SDPs have been previously analyzed with tools of statistical physics [16, 17], although the methods of these works fall short for analyzing the satisfiability transition in random ellipsoid fitting [7]. We refer the interested reader to the recent book [18] that compiles many (sometimes surprising) applications of the theory of disordered systems, as well as mathematically rigorous approaches to it.

Notably, in the companion work to our manuscript [7], non-rigorous methods of statistical physics are employed to provide a detailed picture of the satisfiability transition in random ellipsoid fitting. Besides predicting a threshold for $n \sim d^2/4$, this work gives analytical formulas for the typical shape of ellipsoid fits in the satisfiable phase (i.e. the spectral density of S), generalizes these predictions for non-Gaussian but rotationally-invariant vectors $\{x_\mu\}_{\mu=1}^n$, and also studies the performance of different explicit solutions, notably ones used in the previous literature (see Fig. 1). We emphasize that the present paper is, in contrast, mathematically rigorous.

Inspiration of our approach – Importantly, the non-rigorous analysis of [7] suggests that a quantity known as the free entropy (or free energy) in statistical physics, is universal: its value is (with high probability) the same for (P) and a variant of (P_{Gauss.}), as $d \rightarrow \infty$. Such a universality property would have major consequences, as the free entropy carries deep information about the structure of the space of solutions to the problem. Remarkably, similar phenomena have been studied numerically and theoretically in statistical learning models, in which one can effectively replace an arbitrary (and

²i.e. a combination of linear equations with a positivity constraint $S \succeq 0$.

possibly complicated) data distribution by its ‘‘Gaussian equivalent’’. Investigating this Gaussian equivalence phenomenon is the object of a recent and very active line of work, with consequences on the theory of empirical risk minimization and beyond [19–27]. Inspired by these works (in particular [21]) we provide a rigorous proof of the universality conjectured in [7], using an interpolation argument. We then leverage tools of the theory of random convex programs [8, 9], such as Gordon’s min-max inequality [28], to sharply characterize the space of solutions to $(P_{\text{Gauss.}})$. Using the aforementioned universality allows to transfer many of these conclusions to the original problem (P), yielding our main results.

1.3 Main results

Notation – We denote \mathcal{S}_d the set of $d \times d$ real symmetric matrices, while $\mathbb{S}^{d-1}(r)$ refers to the Euclidean unit sphere of radius r in \mathbb{R}^d . For $S \in \mathcal{S}_d$, $\text{Sp}(S) = \{\lambda_i\}_{i=1}^d$ is the set of eigenvalues of S . For $\gamma \in [1, \infty]$, $\|S\|_{S_\gamma} := (\sum_i |\lambda_i|^\gamma)^{1/\gamma}$ stands for the Schatten- γ norm. $B_\gamma(S, \delta)$ is the Schatten- γ ball of radius δ centered in S , and $B_\gamma(\delta)$ the ball centered at $S = 0$. We denote by $\|S\|_{\text{op}} := \|S\|_{S_\infty}$ the operator norm, and $\|S\|_F := \|S\|_{S_2}$ the Frobenius norm. For a function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, we write $\|\psi\|_L$ to denote its Lipschitz constant. Finally, we use a generic nomenclature $C, c > 0$ to denote positive constants (not depending on the dimension), that may vary from line to line. If necessary, we will make explicit the dependency of these constants on parameters of the problems.

We now state our main results, separating the conjecturally satisfiable ($\alpha = n/d^2 < 1/4$) and unsatisfiable ($\alpha > 1/4$) regimes.

1.3.1 The satisfiable phase: $\alpha < 1/4$

Our main result on the ‘‘positive’’ side of the ellipsoid fitting conjecture can be stated as follows.

Theorem 1.3 (*Satisfiable regime*)

Assume $\alpha := \limsup(n/d^2) < 1/4$ and let $r \in [1, 4/3)$. There exists $0 < \lambda_- \leq \lambda_+$, depending only on α , such that the following holds. Let

$$\Gamma_r(\varepsilon) := \left\{ S \in \mathcal{S}_d : \text{Sp}(S) \subseteq [\lambda_-, \lambda_+] \text{ and } \frac{1}{n} \sum_{\mu=1}^n \left| \sqrt{d} \left[\frac{x_\mu^\top S x_\mu}{d} - 1 \right] \right|^r \leq \varepsilon \right\}.$$

Then for any $\varepsilon > 0$, if $x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$, $\mathbb{P}[\Gamma_r(\varepsilon) \neq \emptyset] \rightarrow 1$ as $n, d \rightarrow \infty$.

Let us make a series of remarks on Theorem 1.3. First, one can alternatively formulate its conclusion as:

$$\text{p-lim}_{d \rightarrow \infty} \min_{\text{Sp}(S) \subseteq [\lambda_-, \lambda_+]} \frac{1}{n} \sum_{\mu=1}^n \left| \sqrt{d} \left[\frac{x_\mu^\top S x_\mu}{d} - 1 \right] \right|^r = 0, \quad (5)$$

where p-lim denotes limit in probability. Secondly, while our current proof limits the choice of $r \in [1, 4/3)$, it might be possible to refine our arguments to reach the same result for any $r \in [1, 2]$, see our discussion in Section 1.4. Moreover, note that by standard concentration arguments, we expect Gaussian points to be close to the sphere $\mathbb{S}^{d-1}(\sqrt{d})$, i.e. the ellipsoid defined by $S = I_d$. A detailed analysis yields, for any $r \in [1, 2]$:

$$\text{p-lim}_{d \rightarrow \infty} \frac{1}{n} \sum_{\mu=1}^n \left| \sqrt{d} \left[\frac{\|x_\mu\|^2}{d} - 1 \right] \right|^r = \mathbb{E}[|Z|^r] > 0, \quad (6)$$

where $Z \sim \mathcal{N}(0, 2)$. This follows from classical concentration arguments, see Appendix B for a detailed derivation. Theorem 1.3 therefore shows that there exists an ellipsoid whose ‘‘fitting error’’ improves by an arbitrary factor over the one achieved by the unit sphere, as long as $\alpha < 1/4$. On the other

hand, we will see that this is not possible for $\alpha > 1/4$, strongly suggesting that our results capture the phenomenon responsible for the conjectured satisfiability transition of ellipsoid fitting. In Section 1.4 we will consider possible future directions that could allow to improve our results to the existence of fitting ellipsoids with *exactly zero* error, i.e. the conjecture of eq. (2).

Finally, we notice that Theorem 1.3 is coherent with the non-rigorous analysis of [7], which predicts that for any $\alpha < 1/4$ typical solutions to ellipsoid fitting have spectral density contained in an interval of the type $[\lambda_-, \lambda_+]$ depending only on α .

1.3.2 The unsatisfiable phase: $\alpha > 1/4$

Our main result towards proving the non-existence of fitting ellipsoids for $\alpha > 1/4$ is the following.

Theorem 1.4 (*Unsatisfiable regime*)

Assume $\alpha := \liminf(n/d^2) > 1/4$. Let $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a non-decreasing differentiable function, with $\phi(0) = 0$, and such that ϕ has a unique global minimum in 0. For any $\varepsilon > 0$ and $M > 0$ we let:

$$\Gamma(\varepsilon, M) := \left\{ S \in \mathcal{S}_d : \text{Sp}(S) \subseteq [0, M] \text{ and } \frac{1}{n} \sum_{\mu=1}^n \phi \left(\sqrt{d} \left| \frac{x_\mu^\top S x_\mu}{d} - 1 \right| \right) \leq \varepsilon \right\}.$$

There exists $\varepsilon = \varepsilon(\alpha, \phi) > 0$ such that for all $M > 0$, if $x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \text{I}_d)$,

$$\lim_{d \rightarrow \infty} \mathbb{P}[\Gamma(\varepsilon, M) \neq \emptyset] = 0.$$

Theorem 1.4 shows that when $\alpha > 1/4$, ellipsoid fitting admits no solutions (even allowing a small approximation error ε) with spectrum bounded above as $d \rightarrow \infty$ (i.e. an ellipsoid whose smallest axis has length bounded away from zero).

As a simple corollary of Theorem 1.4, we get that eq. (3) holds under the assumption that if there exists ellipsoid fits, then one of them has bounded spectrum.

Corollary 1.5 (*Inexistence of fitting ellipsoids under a boundedness assumption*)

Consider the hypothesis:

(H) For any $\alpha \in (0, 1/2)$ there exists $M = M(\alpha) > 0$ such that the following holds. Let $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha > 0$, and $x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \text{I}_d)$. We denote Γ the set of ellipsoid fits for $(x_\mu)_{\mu=1}^n$. Then (with high probability), if $\Gamma \neq \emptyset$, there exists $S \in \Gamma$ such that $\|S\|_{\text{op}} \leq M(\alpha)$.

Assumption (H) implies the negative side of the ellipsoid fitting conjecture, i.e. eq. (3).

While we are not aware at the present of a proof of (H), it seems possible that this question is easier to resolve than the original ellipsoid fitting conjecture. Moreover, non-rigorous calculations [7] predict that, with high probability, *all ellipsoid fits* (when existing) have bounded operator norm, which would imply Assumption (H).

1.4 Discussion and consequences

The combination of our two main results (Theorem 1.3 and Theorem 1.4) provides strong evidence for the original ellipsoid fitting conjecture (Conjecture 1.1). Our conclusions are attained through the study of a ‘‘Gaussian equivalent’’ problem, which partly motivated Conjecture 1.1. Informally, we show that an approximate version of the ellipsoid fitting (i.e. by allowing infinitesimally small error) undergoes a sharp satisfiability transition at $\alpha = n/d^2 = 1/4$. Moreover, we also show in our proof that in the Gaussian equivalent problem, the satisfiability transition for this ‘‘approximate’’ version

corresponds to the one of the exact fitting problem (i.e. not allowing for any non-zero error). This strongly suggests that our method is indeed capturing the phenomenon responsible for the ellipsoid fitting transition.

Our results are an example of a universality phenomenon in high-dimensional stochastic geometry: we show that the statistical dimension (or the square of the Gaussian width) of the set of positive semidefinite matrices determines the satisfiability of – a modified version of – random ellipsoid fitting, even though the affine subset $\{S \in \mathcal{S}_d : (x_\mu^\top S x_\mu = d)_{\mu=1}^n\}$ is *not* randomly oriented uniformly in all directions. In general, understanding the conditions under which universality holds in such problems of high-dimensional random geometry is an important open question. We mention [29], which proves universality between a model in which the random subspace is given by the kernel of a random i.i.d. Gaussian matrix, and a second model where the subspace is the kernel of a matrix with independent elements (not necessarily Gaussian).

1.4.1 Towards Conjecture 1.1

Unfortunately, while our main theorems characterize a satisfiability transition for ellipsoid fitting at the expected threshold, they do not formally imply Conjecture 1.1. We discuss briefly some improvements of our results that could potentially allow to bridge this gap.

On the “positive” side of the conjecture (i.e. the regime $\alpha = n/d^2 < 1/4$), Theorem 1.3 shows the existence of bounded ellipsoids that can achieve arbitrarily small error ε (where the error is taken to 0 *after* $d \rightarrow \infty$). On the other hand, a proof of eq. (2) would require to invert these limits, and take $\varepsilon \rightarrow 0$ *before* $d \rightarrow \infty$. In this regard, an important strengthening of Theorem 1.3 would be to obtain non-asymptotic bounds on $\mathbb{P}[\Gamma_r(\varepsilon) = \emptyset]$, that depend on ε .

Another potential for improvement stems from geometrical considerations: denoting $V := \{S \in \mathcal{S}_d : x_\mu^\top S x_\mu = d \text{ for all } \mu \in [n]\}$, one may use Theorem 1.3 to bound the distance of the set $\Gamma_r(\varepsilon)$ to the affine subspace V . Since $\lambda_{\min}(S) \geq \lambda_-(\alpha)$ for any $S \in \Gamma_r(\varepsilon)$, it would suffice to show that $d_{\text{op}}(\Gamma_r(\varepsilon), V) \leq \lambda_-(\alpha)$ to deduce eq. (2), the first part of Conjecture 1.1. We perform in Appendix C a naive analysis of necessary conditions for this conclusion to follow from Theorem 1.3. Unfortunately, we find that (among other considerations) these conditions would require a significantly stronger form of Theorem 1.3, by proving the conclusion for larger values of $r \in [1, 2]$ and/or a better scaling with d of the minimal error achievable (i.e. proving the conclusion of Theorem 1.3 for $\Gamma_r(\varepsilon_d)$ with $\varepsilon_d \rightarrow 0$ as $d \rightarrow \infty$).

Moreover, let us emphasize that a critical difficulty in improving our proof techniques would be to quantitatively sharpen the universality arguments we carry out, and in particular to strengthen Proposition 2.1, which shows the universality of the minimal fitting error, or “ground state” energy, for ellipsoid fitting and a simpler “Gaussian equivalent” problem. While the present form of Proposition 2.1 shows universality of this error up to a $o_d(1)$ difference, this estimate would likely have to be improved in order to carry out the aforementioned approaches. This part of our proof is greatly inspired by a recent literature on similar universality phenomena [19–27], and we are not aware of the existence of such universality results at a finer scale (or even predictions/conjectures of conditions under which they should hold).

Finally, on the “negative” side of the conjecture (i.e. for $\alpha > 1/4$), we have seen that Theorem 1.4 reduces the second part of Conjecture 1.1 (eq. (3)) to proving the boundedness of the spectrum of solutions, as emphasized in Corollary 1.5. If a proof of Assumption (H) were to become available, our results would imply eq. (3), i.e. the regime $\alpha > 1/4$ of Conjecture 1.1.

1.4.2 Further directions

Our proof method that leverages universality of the minimal fitting error is quite versatile, and we end our discussion by mentioning a few further directions and generalized results that stem from our analysis.

The dual program – First, as a semidefinite program, ellipsoid fitting admits a dual formulation, as written e.g. in [11]. While the limitations of Theorems 1.3 and 1.4, discussed above, prevent us from directly drawing conclusions on the dual, it might be possible to directly apply to it a similar universality approach. Such an application might allow to overcome some current limitations of our results, and we leave this investigation for future work.

Beyond Gaussian vectors – Secondly, while we perform our analysis for $x_1, \dots, x_n \sim \mathcal{N}(0, \mathbf{I}_d)$, it is clear from our proof that our results (both Theorems 1.3 and 1.4) hold for any i.i.d. $(x_\mu)_{\mu=1}^n$ such that the matrices $W_\mu := (x_\mu x_\mu^\top - \mathbf{I}_d)/\sqrt{d}$ satisfy a uniform pointwise normality (or uniform one-dimensional CLT) assumption, as defined in Definition 4.1, and proven for the case of Gaussian vectors in Lemma 4.8. An interesting example of a non-Gaussian distribution is given by the case of rotationally-invariant vectors with fluctuating norm, of the form

$$x_\mu \stackrel{d}{=} \sqrt{r_\mu} \omega_\mu,$$

with r_μ and ω_μ independent, and $\omega_\mu \sim \text{Unif}(\mathbb{S}^{d-1}(\sqrt{d}))$. Letting $\tau := \lim_{d \rightarrow \infty} [\sqrt{d} \text{Var}(r_1)]$, [7] conjectures that the ellipsoid fitting transition point for this model is located at $n/d^2 = \alpha_c(\tau) \in (0, 1/2)$, and gives an exact expression of $\alpha_c(\tau)$ (see Fig. 5 of [7]), showing that ellipsoid fitting becomes harder as the fluctuations of the norm increase. While pointwise normality may not hold in this setting, it is conceivable that our proof techniques can be adapted to handle these distributions, by following the calculations of [7], to obtain results akin to Theorems 1.3 and 1.4. More generally, while it is clear that some distributions can not satisfy uniform pointwise normality (see the discussion below Lemma 4.8 for examples), a more thorough investigation of the class of distributions of x_μ 's for which pointwise normality holds (and thus our proof applies) is, in our opinion, an interesting direction to explore.

A minimal nuclear norm estimator – Let us conclude by mentioning a different approach to a possible solution of the first part of Conjecture 1.1. It is conjectured in [7] (through non-rigorous methods) that the minimal nuclear norm solution, i.e.

$$\hat{S}_{\text{NN}} := \min_{\substack{S \in \mathcal{S}_d \\ \{x_\mu^\top S x_\mu = d\}_{\mu=1}^n}} \|S\|_{S_1} = \min_{\substack{S \in \mathcal{S}_d \\ \{x_\mu^\top S x_\mu = d\}_{\mu=1}^n}} \sum_{i=1}^d |\lambda_i(S)|,$$

satisfies $\hat{S}_{\text{NN}} \succeq 0$ with high probability for any $\alpha < 1/4$. Analyzing \hat{S}_{NN} , whether through the techniques of the present paper or with different methods, is another promising approach to prove eq. (2), the “positive” part of Conjecture 1.1.

1.5 Structure of the paper

In Section 2 we present the proof of our main results. The proof of some of the intermediate results are postponed to later sections: in Section 3 we study in detail the “Gaussian equivalent” problem to random ellipsoid fitting, and in Section 4 we prove an important universality property between the two models.

2 Proof of the main results

We prove here Theorems 1.3 and 1.4. The core idea of our proof can be sketched as follows:

- (i) Using rigorous methods inspired by statistical physics, we prove that a quantity known as the *asymptotic free entropy* is universal for the ellipsoid fitting problem of eq. (1) and a variant of its Gaussian counterpart of eq. (4), for any value of $\alpha = n/d^2$. The main technique we use is an interpolation method. In a suitable limit (known as the low-temperature limit in statistical physics), this implies the universality of the minimal “fitting error”.

(ii) We study the Gaussian equivalent problem using methods of random convex geometry [8, 9], leveraging in particular Gordon’s min-max inequality [28]. When $\alpha < 1/4$ we show that not only a zero error is achievable, but that one can achieve it by a matrix whose spectrum is contained in an interval of the type $[\lambda_-, \lambda_+]$, i.e. the axis of the corresponding ellipsoid have lengths bounded above and below. On the other hand, for $\alpha > 1/4$, we prove that not only is the Gaussian equivalent problem not satisfiable, but one can lower bound the minimal fitting error as long as the set of candidate matrices is contained in an operator norm ball $B_{\text{op}}(M)$ (for any constant $M > 0$).

(iii) We prove that the conclusions of (ii) transfer to the original ellipsoid fitting problem, using the universality shown in (i).

Our proof of (i) leverages an important line of work on free entropy universality [20–22], and a part of it closely follows the proof of [21], which we will point out in relevant places. Nevertheless, as our setting does not satisfy all the hypotheses of this work, and for completeness of our exposition, we chose to write the whole proof in a self-contained manner.

2.1 Reduction of the problem and Gaussian equivalent

For any $0 \leq \lambda_- \leq \lambda_+$, any function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ and any $\varepsilon > 0$ we define the set

$$\Gamma(\phi, \lambda_-, \lambda_+, \varepsilon) := \left\{ S \in \mathcal{S}_d : \text{Sp}(S) \subseteq [\lambda_-, \lambda_+] \text{ and } \frac{1}{n} \sum_{\mu=1}^n \phi \left(\sqrt{d} \left[\frac{x_\mu^\top S x_\mu}{d} - 1 \right] \right) \leq \varepsilon \right\}. \quad (7)$$

If one thinks of ϕ as an error (or loss) function, then $\Gamma(\phi, \lambda_-, \lambda_+, \varepsilon)$ represents the set of matrices with spectrum in $[\lambda_-, \lambda_+]$ that solve (P) up to an approximation error ε . Notice that for any $S \in \mathcal{S}_d$:

$$\sqrt{d} \left[\frac{x_\mu^\top S x_\mu}{d} - 1 \right] = \text{Tr}[W_\mu S] - \frac{d - \text{Tr}[S]}{\sqrt{d}}, \quad (8)$$

with $W_\mu := (x_\mu x_\mu^\top - \text{I}_d)/\sqrt{d}$. Moreover, W_μ has the same first two moments as a Gaussian matrix. Formally, we define:

Definition 2.1 (*Matrix ensembles*)

Let $d \geq 1$. We say that a random symmetric $W \in \mathcal{S}_d$ is generated according to:

- $W \sim \text{GOE}(d)$ if $W_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, [1 + \delta_{ij}]/d)$ for $i \leq j$ ³.
- $W \sim \text{Ellipse}(d)$ if $W \stackrel{\text{d}}{=} (x x^\top - \text{I}_d)/\sqrt{d}$ for $x \sim \mathcal{N}(0, \text{I}_d)$.

One checks easily that $\mathbb{E}_{\text{Ellipse}(d)}[W_{ij} W_{kl}] = \mathbb{E}_{\text{GOE}(d)}[W_{ij} W_{kl}]$ for any $i \leq j$ and $k \leq l$. This remark and eq. (8) lead to consider the following modified problem, with $W_\mu := (x_\mu x_\mu^\top - \text{I}_d)/\sqrt{d}$ and $b \in \mathbb{R}$:

$$\Gamma_b(\phi, \lambda_-, \lambda_+, \varepsilon) := \left\{ S \in \mathcal{S}_d : \text{Sp}(S) \subseteq [\lambda_-, \lambda_+] \text{ and } \frac{1}{n} \sum_{\mu=1}^n \phi(\text{Tr}[W_\mu S] - b) \leq \varepsilon \right\}. \quad (9)$$

In the rest of the proof we will focus on studying the set Γ_b of eq. (9) with⁴ $b \in \mathbb{R}$, for both $W_\mu \sim \text{Ellipse}(d)$ and $W_\mu \sim \text{GOE}(d)$ (which we call the “Gaussian equivalent” problem). At the end of our proof, we will transfer our conclusions on Γ_b back to the original solution set Γ of eq. (7).

³GOE(d) stands for *Gaussian Orthogonal Ensemble*.

⁴Furthermore, by rescaling S (and up to a change in $\lambda_-, \lambda_+, \phi$) we will reduce to the case $b \in \{-1, 0, 1\}$.

2.2 Free entropy universality

We can now state the main result concerning on the universality of the minimal error (or “ground state energy” in statistical physics jargon). This result is inspired by a rich line of work on universality of empirical risk minimization [20–22, 24].

Proposition 2.1 (*Ground state universality*)

Let $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ and $\psi : \mathbb{R} \rightarrow \mathbb{R}$ two bounded differentiable functions with bounded derivatives, and assume furthermore $\|\psi'\|_L < \infty$. Let $n, d \geq 1$ and $n, d \rightarrow \infty$ with $\alpha_1 d^2 \leq n \leq \alpha_2 d^2$ for some $0 < \alpha_1 \leq \alpha_2$, and $B \subseteq \mathcal{S}_d$ a closed set such that $B \subseteq B_{\text{op}}(C_0)$ for some $C_0 > 0$ (not depending on d). For $X_1, \dots, X_n \in \mathcal{S}_d$ we define the *ground state energy*:

$$\text{GS}_d(\{X_\mu\}) := \inf_{S \in B} \frac{1}{d^2} \sum_{\mu=1}^n \phi(\text{Tr}[X_\mu S]). \quad (10)$$

Then we have:

$$\lim_{d \rightarrow \infty} \left| \mathbb{E}_{\{W_\mu\} \overset{\text{i.i.d.}}{\sim} \text{Ellipse}(d)} \psi[\text{GS}_d(\{W_\mu\})] - \mathbb{E}_{\{G_\mu\} \overset{\text{i.i.d.}}{\sim} \text{GOE}(d)} \psi[\text{GS}_d(\{G_\mu\})] \right| = 0. \quad (11)$$

Therefore, for any $\rho \geq 0$ and $\delta > 0$:

$$\begin{cases} \lim_{d \rightarrow \infty} \mathbb{P}[\text{GS}_d(\{W_\mu\}) \geq \rho + \delta] \leq \lim_{d \rightarrow \infty} \mathbb{P}[\text{GS}_d(\{G_\mu\}) \geq \rho], \\ \lim_{d \rightarrow \infty} \mathbb{P}[\text{GS}_d(\{W_\mu\}) \leq \rho - \delta] \leq \lim_{d \rightarrow \infty} \mathbb{P}[\text{GS}_d(\{G_\mu\}) \leq \rho]. \end{cases} \quad (12)$$

A word on the proof – The main proof technique we use is Gaussian interpolation: namely we define an interpolating $U_\mu(t)$ such that $U_\mu(0) = G_\mu$ and $U_\mu(1) = W_\mu$, and show that $\text{GS}_d(\{U_\mu(t)\})$ is constant (up to negligible terms) along the interpolation path. Note that while Proposition 2.1 is very close to the results of [21], there is a technical difference with the setup of this work: for any fixed S , the random variable $\text{Tr}[WS]$ for $W \sim \text{Ellipse}(d)$ is not sub-Gaussian but only sub-exponential. As a consequence, we can not achieve a good control of the Lipschitz constant of the error (or “energy” function) of eq. (10) with respect to the Frobenius norm of S , as is required in [21]. We bypass this difficulty by controlling instead the Lipschitz constant with respect to the operator norm (see Lemma 4.1), using important empirical process bounds over the operator norm ball (see Lemma 4.3): this leads to the limitation $B \subseteq B_{\text{op}}(C_0)$. Interestingly, improving these bounds would also allow to relax the limitation $r \in [1, 4/3)$ in Theorem 1.3, as we discuss after. Having dealt with this difficulty, the rest of the interpolation argument is very similar to [21]. We show Proposition 2.1 in Section 4, deferring some arguments to Appendix D.

2.3 The Gaussian equivalent problem

We now study the Gaussian equivalent problem. We will later transfer our analysis to the original ellipsoid fitting case using Proposition 2.1. Our results are stated separately for the satisfiable and unsatisfiable regimes.

Proposition 2.2 (Regular solutions in the satisfiable regime)

Let $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha < 1/4$, and let $\{G_\mu\}_{\mu=1}^n \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)$. Let

$$V := \{S \in \mathcal{S}_d : \forall \mu \in [n], \text{Tr}[G_\mu S] = 1\}.$$

There exists $0 < \lambda_- \leq \lambda_+$ (depending only on α) such that:

$$\lim_{d \rightarrow \infty} \mathbb{P}\{\exists S \in V \text{ s.t. } \text{Sp}(S) \subseteq [\lambda_-, \lambda_+]\} = 1.$$

Proposition 2.2 shows that for $\alpha < 1/4$, there exists w.h.p. ellipsoids satisfying the ‘‘Gaussian equivalent’’ to random ellipsoid fitting, and that such solutions might also be assumed to have their axes’ lengths bounded above and below as $d \rightarrow \infty$. In the unsatisfiable regime $\alpha > 1/4$, we show on the other hand that with high probability there are no solutions to the Gaussian equivalent problem, even allowing for some error when fitting the random points.

Proposition 2.3 (No approximate solution in the unsatisfiable regime)

Let $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha > 1/4$, and let $\{G_\mu\}_{\mu=1}^n \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)$. Let $b \in \mathbb{R}$ and denote $\mathcal{C}_\mu^{(b)}(S) := |\text{Tr}(G_\mu S) - b|$. Define the affine subspace:

$$V_b := \{S \in \mathcal{S}_d : \forall \mu \in [n], \mathcal{C}_\mu^{(b)}(S) = 0\}.$$

(i) Assume that $b \neq 0$. Then there exists $c = c(\alpha, b) > 0$ and $\eta = \eta(\alpha, b) \in (0, 1)$ such that

$$\lim_{d \rightarrow \infty} \mathbb{P}\{\forall S \succeq 0 : \#\{\mu \in [n] : \mathcal{C}_\mu^{(b)}(S) > c\} \geq \eta n\} = 1.$$

(ii) Assume that $b = 0$. Then there exists $c = c(\alpha) > 0$ and $\eta = \eta(\alpha) \in (0, 1)$ such that, with probability $1 - o_d(1)$, the following holds for all $\tau \geq 0$:

$$\sup_{\substack{S \succeq 0 \\ \#\{\mu \in [n] : \mathcal{C}_\mu^{(0)}(S) > c\tau\} < \eta n}} \|S\|_F \leq \tau \sqrt{d}.$$

Propositions 2.2 and 2.3 are proven in Section 3. Our proof follows a standard approach in random geometry problems involving Gaussian distributions, by leveraging Gordon’s min-max inequality [28] and its sharpness in convex settings [30, 31]. It strengthens for our setting the results obtained for general random convex programs in [8, 9] (using either Gordon’s inequality or tools of integral geometry).

2.4 The satisfiable regime: proof of Theorem 1.3

Propositions 2.1 and 2.2 have the following consequence, taking $B := \{S : \lambda_- I_d \preceq S \preceq \lambda_+ I_d\}$, with (λ_-, λ_+) given by Proposition 2.2.

Corollary 2.4

Let $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha < 1/4$, and $W_1, \dots, W_n \stackrel{\text{i.i.d.}}{\sim} \text{Ellipse}(d)$. There exists $\lambda_-, \lambda_+ > 0$ depending only on α such that the following holds. If we have $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ with $\|\phi\|_\infty, \|\phi'\|_\infty < \infty$ and such that $\phi(0) = 0$, then

$$\text{p-lim}_{d \rightarrow \infty} \min_{\text{Sp}(S) \subseteq [\lambda_-, \lambda_+]} \frac{1}{n} \sum_{\mu=1}^n \phi(\text{Tr}[W_\mu S] - 1) = 0.$$

The proof of Corollary 2.4 is immediate by combining Propositions 2.1 and 2.2. We can furthermore relax some of the assumptions on ϕ in Corollary 2.4, as we now show.

Lemma 2.5

Corollary 2.4 holds for $\phi(x) = |x|^r$, for any $1 \leq r < 4/3$.

Note that the limitation $r < 4/3$ is a consequence of a limitation on the control of an empirical process that is done in Lemma 4.3 (see also the discussion in Section 4.1).

Proof of Lemma 2.5 – Let $\varepsilon > 0$ and $A > 0$, and let us denote $u_A : \mathbb{R}_+ \rightarrow [0, 1]$ a C^∞ function such that $u_A(x) = 1$ if $x \leq A$ and $u_A(x) = 0$ if $x \geq A + 1$. We denote $\phi_A(z) := |z|^r u_A(|z|)$. Then ϕ_A is bounded, with bounded derivative. Moreover, we have for any $x \in \mathbb{R}$:

$$|x|^r = \phi_A(x) + |x|^r(1 - u_A(|x|)) \leq \phi_A(x) + |x|^r \mathbf{1}\{|x| \geq A\}.$$

By Corollary 2.4, under an event of probability $1 - o_d(1)$ we can fix S with $\text{Sp}(S) \in [\lambda_-, \lambda_+]$ and such that $\sum_{\mu=1}^n \phi_A(\text{Tr}[W_\mu S] - 1) \leq n\varepsilon/2$. We pick $\gamma > 1$ such that $\gamma r \leq 4/3$, and condition on the $1 - o_d(1)$ probability event, thanks to Lemma 4.3:

$$\max_{\|R\|_{\text{op}}=1} \sum_{\mu=1}^n |\text{Tr}(W_\mu R)|^{\gamma r} \leq Cn \tag{13}$$

We have, with probability $1 - o_d(1)$:

$$\begin{aligned} \frac{1}{n} \sum_{\mu=1}^n |\text{Tr}[W_\mu S] - 1|^r &\leq \frac{\varepsilon}{2} + \frac{1}{n} \sum_{\mu=1}^n |\text{Tr}[W_\mu S] - 1|^r \mathbf{1}\{|\text{Tr}[W_\mu S] - 1| \geq A\}, \\ &\stackrel{(a)}{\leq} \frac{\varepsilon}{2} + \frac{A^{r(1-\gamma)}}{n} \sum_{\mu=1}^n |\text{Tr}[W_\mu S] - 1|^{\gamma r}, \\ &\stackrel{(b)}{\leq} \frac{\varepsilon}{2} + \frac{A^{r(1-\gamma)} 2^{\gamma r - 1}}{n} [n + C\lambda_+^{\gamma r} n], \\ &\leq \frac{\varepsilon}{2} + C(\gamma, r, \alpha) A^{r(1-\gamma)}. \end{aligned}$$

We used in (a) the following inequality, for a positive random variable X , $t \geq 0$, and any $\gamma > 1$:

$$\mathbb{E}[X \mathbf{1}\{X \geq t\}] = t \mathbb{E} \left[\frac{X}{t} \mathbf{1} \left\{ \frac{X}{t} \geq 1 \right\} \right] \leq t^{1-\gamma} \mathbb{E}[X^\gamma].$$

In (b) we used eq. (13) and $|a + b|^r \leq 2^{r-1}(|a|^r + |b|^r)$. We pick $A = [\varepsilon/(2C(\gamma, r, \alpha))]^{1/[r(1-\gamma)]}$. We have then, with probability $1 - o_d(1)$:

$$\min_{\text{Sp}(S) \subseteq [\lambda_-, \lambda_+]} \frac{1}{n} \sum_{\mu=1}^n |\text{Tr}[W_\mu S] - 1|^r \leq \varepsilon,$$

which ends the proof. □

Proof of Theorem 1.3 – Notice that Lemma 2.5 precisely shows that, for $\phi(x) = |x|^r$ and $\varepsilon > 0$, the set Γ_1 of eq. (9) is non-empty with high probability. We now use the following remark (recall the definition of Γ in eq. (7)):

Lemma 2.6

For any $(x_\mu)_{\mu=1}^n$ and $\lambda_-, \lambda_+, \varepsilon > 0, r \geq 1$, if $S \in \Gamma_1(|\cdot|^r, \lambda_-, \lambda_+, \varepsilon)$, then $\hat{S} \in \Gamma(|\cdot|^r, \lambda'_-, \lambda'_+, \varepsilon')$, with

$$\hat{S} = \frac{dS}{\sqrt{d} + \text{Tr}[S]}, \quad \lambda'_- = \frac{\lambda_-}{\lambda_+ + d^{-1/2}}, \quad \lambda'_+ = \frac{\lambda_+}{\lambda_- + d^{-1/2}}, \quad \varepsilon' = \frac{\varepsilon}{(\lambda_- + d^{-1/2})^r}.$$

Since $\lambda'_+ \leq \lambda_+/\lambda_-$, $\varepsilon' \leq \varepsilon/(\lambda_-)^r$, and $\lambda'_- \geq \lambda_-/(2\lambda_+)$ for d large enough, combining Lemmas 2.5 and 2.6 imply that for any $r \in [1, 4/3)$ and $\varepsilon > 0$:

$$\mathbb{P}[\Gamma(|\cdot|^r, a, b, \varepsilon) \neq \emptyset] \rightarrow_{d \rightarrow \infty} 1,$$

for some $0 < a \leq b$ depending only on α , which ends the proof of Theorem 1.3. \square

Proof of Lemma 2.6 – Let $S \in \Gamma_1(|\cdot|^r, \lambda_-, \lambda_+, \varepsilon)$. Defining $\hat{S} = dS/(\sqrt{d} + \text{Tr}[S])$, we have by eq. (8):

$$\left| \sqrt{d} \left[\frac{x_\mu^\top \hat{S} x_\mu}{d} - 1 \right] \right|^r = \left(\frac{d}{\sqrt{d} + \text{Tr}[S]} \right)^r |\text{Tr}[SW_\mu] - 1|^r \leq \frac{1}{(\lambda_- + d^{-1/2})^r} |\text{Tr}[SW_\mu] - 1|^r.$$

\square

2.5 The unsatisfiable regime: proof of Theorem 1.4

Propositions 2.1 and 2.3 have the following corollary.

Corollary 2.7

Let $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha > 1/4$, and $W_1, \dots, W_n \stackrel{\text{i.i.d.}}{\sim} \text{Ellipse}(d)$. Let $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a non-decreasing differentiable function, with $\phi(0) = 0$, and such that ϕ has a unique global minimum in 0. Then:

(i) Let $b \in \{-1, 1\}$. There exists $\varepsilon = \varepsilon(\alpha, \phi) > 0$ such that for all $M > 0$:

$$\lim_{d \rightarrow \infty} \mathbb{P} \left[\min_{\text{Sp}(S) \subseteq [0, M]} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_\mu S] - b|) \geq \varepsilon \right] = 1.$$

(ii) Let $b = 0$. For all $\tau > 0$, there exists $\varepsilon = \varepsilon(\tau, \alpha, \phi) > 0$ such that for all $M > 0$:

$$\lim_{d \rightarrow \infty} \mathbb{P} \left[\min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ \|S\|_F \geq \tau \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_\mu S]|) \geq \varepsilon \right] = 1.$$

Proof of Corollary 2.7 – Note that we can assume that ϕ is bounded with bounded derivative and $\phi'(0) = 0$: if not it is always possible to lower bound ϕ by such a function. $x \mapsto \phi(|x|)$ is then a bounded function on \mathbb{R} with bounded derivative. We start with (i). By Proposition 2.3, there exists $c_\alpha, \eta_\alpha > 0$ such that

$$\lim_{d \rightarrow \infty} \mathbb{P}\{\forall S \succeq 0 : \#\{\mu \in [n] : |\text{Tr}(G_\mu S) - b| \leq c_\alpha\} \leq (1 - \eta_\alpha)n\} = 1.$$

Conditioning on this event, we have

$$\inf_{S \succeq 0} \sum_{\mu=1}^n \phi(|\text{Tr}[G_\mu S] - b|) \geq n\eta_\alpha \phi(c_\alpha).$$

Using Proposition 2.1 with $B = \{S : \text{Sp}(S) \subseteq [0, M]\}$ we reach that, for all $M > 0$, with probability $1 - o_d(1)$:

$$\inf_{\text{Sp}(S) \subseteq [0, M]} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_\mu S] - b|) \geq \frac{1}{2} \eta_\alpha \phi(c_\alpha).$$

We now turn to (ii). Again by Proposition 2.3, we fix $c_\alpha, \eta_\alpha > 0$ such that for all $\tau \geq 0$:

$$\lim_{d \rightarrow \infty} \mathbb{P} \left\{ \sup_{\substack{S \succeq 0 \\ \#\{\mu \in [n] : |\text{Tr}(G_\mu S)| > c_\alpha \tau\} < \eta_\alpha n}} \|S\|_F \leq \frac{\tau}{2} \sqrt{d} \right\} = 1.$$

Stated differently:

$$\lim_{d \rightarrow \infty} \mathbb{P} \{ \forall S \succeq 0 : \|S\|_F \leq \frac{\tau}{2} \sqrt{d} \text{ or } \#\{\mu \in [n] : |\text{Tr}(G_\mu S)| > c_\alpha \tau\} \geq \eta_\alpha n \} = 1.$$

Conditioning on this event and since ϕ is non-decreasing on \mathbb{R}_+ :

$$\inf_{\substack{S \succeq 0 \\ \|S\|_F \geq \tau \sqrt{d}}} \sum_{\mu=1}^n \phi(|\text{Tr}[G_\mu S]|) \geq n \eta_\alpha \phi(c_\alpha \tau).$$

Using Proposition 2.1 with $B = \{S : 0 \preceq S \preceq M I_d \text{ and } \|S\|_F \geq \tau \sqrt{d}\}$ we reach that, for all τ, M , with probability $1 - o_d(1)$:

$$\inf_{\substack{\text{Sp}(S) \subseteq [0, M] \\ \|S\|_F \geq \tau \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_\mu S]|) \geq \frac{1}{2} \eta_\alpha \phi(c_\alpha \tau),$$

which ends the proof. \square

We now turn to the proof of Theorem 1.4.

Proof of Theorem 1.4 – Let $M > 0$. As in the proof of Corollary 2.7, we can assume without loss of generality that ϕ has bounded derivative: if it is not, it is always possible to lower bound ϕ by such a function.

Let $\delta \in (0, 1)$, and S with $\text{Sp}(S) \subseteq [0, M]$ and $|\text{Tr}[S] - d| \geq \delta \sqrt{d}$. Notice that, defining $S' := \sqrt{d} S / |d - \text{Tr}[S]| \succeq 0$, we have with $b := \text{sign}(d - \text{Tr}[S]) \in \{\pm 1\}$:

$$\text{Tr}[S' W_\mu] - b = \frac{x_\mu^\top S x_\mu - d}{|\text{Tr} S - d|},$$

and so since ϕ is non-decreasing:

$$\phi \left(\sqrt{d} \left| \frac{x_\mu^\top S x_\mu}{d} - 1 \right| \right) \geq \phi(\delta |\text{Tr}(W_\mu S') - b|).$$

Moreover, $\text{Sp}(S') \subseteq [0, M/\delta]$. Since this argument is valid for any S with $\text{Sp}(S) \subseteq [0, M]$ we get:

$$\min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ |\text{Tr}[S] - d| \geq \delta \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi \left(\sqrt{d} \left| \frac{x_\mu^\top S x_\mu}{d} - 1 \right| \right) \geq \min_{b \in \{-1, 1\}} \min_{\text{Sp}(S) \subseteq [0, M/\delta]} \frac{1}{n} \sum_{\mu=1}^n \phi(\delta |\text{Tr}[W_\mu S] - b|).$$

Using Corollary 2.7 applied to $x \mapsto \phi(\delta x)$ there exists therefore $c = c(\alpha, \delta, \phi) > 0$ such that with probability $1 - o_d(1)$:

$$\min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ |\text{Tr}[S] - d| \geq \delta \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi \left(\sqrt{d} \left| \frac{x_\mu^\top S x_\mu}{d} - 1 \right| \right) \geq c(\alpha, \delta, \phi). \quad (14)$$

Let now $S \in \mathcal{S}_d$ with $\text{Sp}(S) \subseteq [0, M]$ and $|\text{Tr}[S] - d| \leq \delta \sqrt{d}$. Then:

$$\text{Tr}[W_\mu S] = \sqrt{d} \left(\frac{x_\mu^\top S x_\mu}{d} - 1 \right) + \underbrace{r(S)}_{|r(S)| \leq \delta},$$

so that

$$\min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ |\text{Tr}[S] - d| \leq \delta \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi \left(\sqrt{d} \left| \frac{x_{\mu}^{\top} S x_{\mu}}{d} - 1 \right| \right) \geq \min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ |\text{Tr}[S] - d| \leq \delta \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_{\mu} S]|) - \|\phi'\|_{\infty} \delta. \quad (15)$$

Notice that since $\delta < 1$, for large enough d we have $|\text{Tr}[S] - d| \leq \delta \sqrt{d} \Rightarrow \text{Tr}[S] \geq d/2$. If moreover $S \succeq 0$, by Cauchy-Schwarz we have $\|S\|_F \geq \text{Tr}[S]/\sqrt{d} \geq \sqrt{d}/2$. This implies:

$$\min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ |\text{Tr}[S] - d| \leq \delta \sqrt{d}}} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_{\mu} S]|) \geq \min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ \|S\|_F \geq \sqrt{d}/2}} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_{\mu} S]|). \quad (16)$$

Using Corollary 2.7 we can obtain $\varepsilon = \varepsilon(\alpha, \phi) > 0$ such that, with probability $1 - o_d(1)$, we have:

$$\min_{\substack{\text{Sp}(S) \subseteq [0, M] \\ \|S\|_F \geq \sqrt{d}/2}} \frac{1}{n} \sum_{\mu=1}^n \phi(|\text{Tr}[W_{\mu} S]|) \geq \varepsilon. \quad (17)$$

Combining eq. (14) with all three equations (15),(16),(17), we get that for any $\delta > 0$, with probability $1 - o_d(1)$:

$$\min_{\text{Sp}(S) \subseteq [0, M]} \frac{1}{n} \sum_{\mu=1}^n \phi \left(\sqrt{d} \left[\frac{x_{\mu}^{\top} S x_{\mu}}{d} - 1 \right] \right) \geq \min[c(\alpha, \delta, \phi), \varepsilon(\alpha, \phi) - \delta \|\phi'\|_{\infty}].$$

Taking $\delta := \min(1, \varepsilon(\alpha, \phi)/(2\|\phi'\|_{\infty})) > 0$ ends the proof. \square

3 The Gaussian equivalent problem

3.1 The satisfiable regime: proof of Proposition 2.2

3.1.1 Gordon's min-max theorem

We will use the Gaussian min-max theorem of Gordon [28], as stated in [30, 31]:

Proposition 3.1 (*Gaussian min-max theorem [28, 30, 31]*)

Let $n, p \geq 1$, $W \in \mathbb{R}^{n \times p}$ an i.i.d. standard normal matrix, and $g \in \mathbb{R}^n, h \in \mathbb{R}^p$ two independent vectors with i.i.d. $\mathcal{N}(0, 1)$ coordinates. Let $\mathcal{S}_v, \mathcal{S}_u$ be two compact subsets respectively of \mathbb{R}^p and \mathbb{R}^n , and let $\psi : \mathcal{S}_v \times \mathcal{S}_u \rightarrow \mathbb{R}$ a continuous function. We define the two optimization problems:

$$\begin{cases} C(W) & := \min_{v \in \mathcal{S}_v} \max_{u \in \mathcal{S}_u} \{u^{\top} W v + \psi(v, u)\}, \\ \mathcal{C}(g, h) & := \min_{v \in \mathcal{S}_v} \max_{u \in \mathcal{S}_u} \{\|u\|_2 h^{\top} v + \|v\|_2 g^{\top} u + \psi(v, u)\}. \end{cases}$$

Then:

(i) For all $t \in \mathbb{R}$, one has

$$\mathbb{P}[C(W) < t] \leq 2\mathbb{P}[\mathcal{C}(g, h) \leq t].$$

(ii) Assume that $\mathcal{S}_v, \mathcal{S}_u$ are convex and that ψ is convex-concave on $\mathcal{S}_v \times \mathcal{S}_u$. Then for all $t \in \mathbb{R}$:

$$\mathbb{P}[C(W) > t] \leq 2\mathbb{P}[\mathcal{C}(g, h) \geq t].$$

3.1.2 Gordon's min-max inequality and random geometry

We first introduce the notion of Gaussian width of a convex cone.

Definition 3.1 (*Gaussian width*)

For $p \geq 1$, and a closed convex cone $K \subseteq \mathbb{R}^p$, we define its *Gaussian width* as

$$\omega(K) := \mathbb{E} \max_{x \in K \cap \mathbb{S}^{p-1}} \langle g, x \rangle,$$

for $g \sim \mathcal{N}(0, \mathbf{I}_p)$.

We show now a general result leveraging Gordon's min-max inequality to prove the existence of a solution to a general type of random geometry problem. Such applications are classical, and we show here that one can assume furthermore that the solution is bounded.

Lemma 3.2

Let $n, p \geq 1$, and $(g_\mu)_{\mu=1}^n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{I}_p)$. We define $V := \{x \in \mathbb{R}^p : \forall \mu \in [n], \langle g_\mu, x \rangle = 1\}$. Let $K \subseteq \mathbb{R}^p$ be a closed convex cone, with Gaussian width $\omega(K)$. Assume that there exists $\varepsilon \in (0, 1)$ such that $n \leq (1 - \varepsilon) \omega(K)^2$ as $n \rightarrow \infty$. Then:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \exists x \in K \cap V \text{ s.t. } \|x\|_2 \leq \frac{2}{\sqrt{\varepsilon}} \right\} = 1. \quad (18)$$

Proof of Lemma 3.2 – Let us denote, for $A > 0$:

$$P_n(A) := \mathbb{P} \{ \exists x \in K \cap V \text{ s.t. } \|x\|_2 \leq A \},$$

and define $G \in \mathbb{R}^{n \times p}$ as the Gaussian matrix with g_μ as its μ -th row. By elementary compactness and duality arguments, we have:

$$1 - P_n(A) = \mathbb{P} \left[\min_{\substack{x \in K \\ \|x\|_2 \leq A}} \|Gx - \mathbf{1}_n\|_2 > 0 \right] = \mathbb{P} \left[\min_{\substack{x \in K \\ \|x\|_2 \leq A}} \max_{\|\lambda\|_2 \leq 1} \{-\lambda^\top \mathbf{1}_n + \lambda^\top Gx\} > 0 \right],$$

for $G \in \mathbb{R}^{n \times p}$ with i.i.d. $\mathcal{N}(0, 1)$ elements. By dominated convergence we have then

$$1 - P_n(A) = \lim_{\eta \rightarrow 0} \mathbb{P} \left[\min_{\substack{x \in K \\ \|x\|_2 \leq A}} \max_{\|\lambda\|_2 \leq 1} \{-\lambda^\top \mathbf{1}_n + \lambda^\top Gx\} > \eta \right]. \quad (19)$$

Note that in eq. (19), both x and λ belong to a convex and compact set (since K is closed and convex), and $\psi(x, \lambda) = -\lambda^\top \mathbf{1}_n$ is clearly convex-concave. We can thus apply item (ii) of Proposition 3.1:

$$1 - P_n(A) \leq 2 \lim_{\eta \rightarrow 0} \mathbb{P} \left[\min_{\substack{x \in K \\ \|x\|_2 \leq A}} \max_{\|\lambda\|_2 \leq 1} \{-\lambda^\top \mathbf{1}_n + \|\lambda\|_2 g^\top x + \|x\|_2 \lambda^\top h\} \geq \eta \right], \quad (20)$$

with $g \sim \mathcal{N}(0, \mathbf{I}_p)$ and $h \sim \mathcal{N}(0, \mathbf{I}_n)$. We then control the right-hand-side of the last equation, using that K is a cone:

$$\begin{aligned} & \min_{\substack{x \in K \\ \|x\|_2 \leq A}} \max_{\|\lambda\|_2 \leq 1} \{-\lambda^\top \mathbf{1}_n + \|\lambda\|_2 g^\top x + \|x\|_2 \lambda^\top h\} = \min_{\substack{x \in K \\ \|x\|_2 \leq A}} \max \{0, \| \|x\|_2 h - \mathbf{1}_n \|_2 + g^\top x\}, \\ & = \max \left\{ 0, \min_{\substack{x \in K \\ \|x\|_2 \leq A}} [\| \|x\|_2 h - \mathbf{1}_n \|_2 + g^\top x] \right\}, \\ & = \max \left\{ 0, \min_{v \in [0, A]} [\|vh - \mathbf{1}_n\|_2 + v \min_{x \in K \cap \mathbb{S}^{p-1}} g^\top x] \right\}. \end{aligned} \quad (21)$$

Note that $g \rightarrow \max_{x \in K \cap \mathbb{S}^{p-1}} [g^\top x]$ is 1-Lipschitz, and in particular concentrates on its average, which by definition is the Gaussian width $\omega(K)$. We use the classical result (see e.g. Theorem 3.25 of [32]):

Theorem 3.3

Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a Lipschitz function. Then for all $t \geq 0$:

$$\mathbb{P}[f(X_1, \dots, X_n) - \mathbb{E}f(X_1, \dots, X_n) \geq t] \leq \exp \left\{ -\frac{t^2}{2\|f\|_{\text{L}}^2} \right\}.$$

Therefore, for $\delta \in (0, \omega(K))$, we have⁵:

$$\mathbb{P} \left\{ \min_{x \in K \cap \mathbb{S}^{p-1}} [g^\top x] \geq -\omega(K) + \delta \right\} \leq e^{-\delta^2/2}. \quad (22)$$

From eqs. (20), (21) and (22) we have:

$$1 - P_n(A) \leq 2 \lim_{\eta \rightarrow 0} \mathbb{P} \left[\max \left\{ 0, \min_{v \in [0, A]} [\|vh - \mathbf{1}_n\|_2 + v(-\omega(K) + \delta)] \right\} \geq \eta \right] + 2e^{-\delta^2/2}. \quad (23)$$

Recall that we assumed $n \leq (1 - \varepsilon)\omega(K)^2$ and $n \rightarrow \infty$. Therefore, for $n \geq n_0(\varepsilon, \delta)$ large enough we can assume $n \leq (1 - \varepsilon/2)[\omega(K) - \delta]^2$. Let $v^* = v^*(\varepsilon) = \sqrt{(4 - \varepsilon)/\varepsilon}$ such that:

$$\left(1 - \frac{\varepsilon}{4}\right) \mathbb{E}_{X \sim \mathcal{N}(0,1)} [(v^*X - 1)^2] = (v^*)^2.$$

Thus, for $A = v^*(\varepsilon)$, we have from eq. (23):

$$1 - P_n(v^*) \leq 2 \lim_{\eta \rightarrow 0} \mathbb{P} \left[\max \left\{ 0, [\|v^*h - \mathbf{1}_n\|_2 + v^*(-\omega(K) + \delta)] \right\} \geq \eta \right] + 2e^{-\delta^2/2}. \quad (24)$$

By the law of large numbers we have ($\xrightarrow{\text{P}}$ denotes convergence in probability):

$$\frac{1}{\sqrt{n}} \|v^*h - \mathbf{1}_n\|_2 \xrightarrow{\text{P}} \sqrt{\mathbb{E}[(v^*X - 1)^2]} = \frac{v^*}{\sqrt{1 - \frac{\varepsilon}{4}}} < \frac{v^*}{\sqrt{1 - \frac{\varepsilon}{3}}}.$$

In particular, with probability $1 - o_n(1)$, we have

$$\|v^*h - \mathbf{1}_n\|_2 + v^*(-\omega(K) + \delta) \leq v^* \sqrt{n} \left[\left(1 - \frac{\varepsilon}{3}\right)^{-1/2} - \left(1 - \frac{\varepsilon}{2}\right)^{-1/2} \right] < 0.$$

Therefore, we have by eq. (24):

$$1 - P_n(v^*) \leq o_n(1) + 2e^{-\delta^2/2}.$$

Taking the limit $n \rightarrow \infty$ and then $\delta \rightarrow \infty$ finishes the proof⁶ (notice that $v^* \leq 2/\sqrt{\varepsilon}$). \square

3.1.3 The cone of positive matrices with bounded condition number

We study here the Gaussian width of the convex cone of positive semidefinite matrices with bounded condition number. We start with a classical result on the Gaussian width of \mathcal{S}_d^+ [8, 9], we refer the reader to Proposition 10.2 of [9] for a proof.

Proposition 3.4 (Gaussian width of \mathcal{S}_d^+)

The Gaussian width of \mathcal{S}_d^+ satisfies:

$$\sqrt{\frac{d(d+1)}{4} - 1} \leq \omega(\mathcal{S}_d^+) \leq \sqrt{\frac{d(d+1)}{4}}.$$

In particular, notice that $\omega(\mathcal{S}_d^+) \sim d/2$ as $d \rightarrow \infty$. We generalize this result by asking that the matrices have bounded condition number.

⁵Since $g \stackrel{\text{d}}{=} -g$, $\min_{x \in K \cap \mathbb{S}^{p-1}} [g^\top x] \stackrel{\text{d}}{=} -\max_{x \in K \cap \mathbb{S}^{p-1}} [g^\top x]$.

⁶Recall that $\delta < \omega(K)$ but $\omega(K) \rightarrow \infty$ as $n \rightarrow \infty$ by hypothesis.

Lemma 3.5 (Gaussian width of PSD matrices with bounded condition number –)

For any $\kappa \geq 1$, define $K_\kappa := \{S \in \mathcal{S}_d^+ : \lambda_{\max}(S) \leq \kappa \lambda_{\min}(S)\}$. Then K_κ is a closed convex cone, and its Gaussian width satisfies

$$\liminf_{d \rightarrow \infty} \frac{2\omega(K_\kappa)}{d} = f(\kappa),$$

where $\kappa \mapsto f(\kappa) \in [0, 1]$ is non-decreasing, with $\lim_{\kappa \rightarrow \infty} f(\kappa) = 1$.

Proof of Lemma 3.5 – The fact that K_κ is a closed convex cone is easy to verify. Moreover, for any $\kappa \leq \kappa'$ we have $K_\kappa \subseteq K_{\kappa'} \subseteq \mathcal{S}_d^+$, and $\omega(\mathcal{S}_d^+) \sim d/2$ by Proposition 3.4, so $\kappa \mapsto f(\kappa) \in [0, 1]$, and f is non-decreasing. To finish the proof, we show that $f(\kappa) \rightarrow 1$ as $\kappa \rightarrow \infty$. We let $\kappa > 1$. To identify \mathcal{S}_d with $\mathbb{R}^{d(d+1)/2}$, we use the matrix flattening function, for $M \in \mathcal{S}_d$:

$$\begin{aligned} \text{vec}(M) &:= ((\sqrt{2}M_{ab})_{1 \leq a < b \leq d}, (M_{aa})_{a=1}^d) \in \mathbb{R}^{d(d+1)/2}, \\ &= ((2 - \delta_{ab})^{1/2} M_{ab})_{a \leq b}. \end{aligned} \quad (25)$$

It is an isometry ($\langle \text{vec}(M), \text{vec}(N) \rangle = \text{Tr}[MN]$), and if $Z \sim \text{GOE}(d)$ then $\text{vec}(Z) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2/d)$. We thus reach:

$$\omega(K_\kappa) := \mathbb{E} \max_{\substack{S \in K_\kappa \\ \|S\|_F^2 = 2d}} \left\{ \frac{1}{2} \text{Tr}[ZS] \right\},$$

in which $Z \sim \text{GOE}(d)$, cf. Definition 2.1. Letting $z_1 \geq \dots \geq z_d$ be the eigenvalues of Z , by Wigner's theorem [33] $(1/d) \sum_i \delta_{z_i}$ weakly converges as $d \rightarrow \infty$ (a.s.) to $\sigma_{\text{s.c.}}(dx) = (2\pi)^{-1} \sqrt{4 - x^2} \mathbb{1}\{|x| \leq 2\} dx$. Moreover, we have (taking S having same eigenvectors as Z)⁷:

$$\omega(K_\kappa) \geq \mathbb{E} \max_{\substack{\lambda_1 \geq \dots \geq \lambda_d \geq 0 \\ \sum \lambda_i^2 \leq 2d \\ \lambda_1 \leq \kappa \lambda_d}} \left\{ \frac{1}{2} \sum_{i=1}^d \lambda_i z_i \right\}. \quad (26)$$

Let us now sketch the end of the proof. We define

$$\begin{cases} \gamma(\kappa) &:= \sqrt{\frac{2}{\int_{2\kappa^{-1}}^2 x^2 \sigma_{\text{s.c.}}(dx) + \frac{4}{\kappa^2} \int_{-2}^{2\kappa^{-1}} \sigma_{\text{s.c.}}(dx)}}, \\ \lambda_i^* &:= \gamma(\kappa) \left[z_i \mathbb{1}\{z_i \geq 2\kappa^{-1}\} + \frac{2}{\kappa} \mathbb{1}\{z_i < 2\kappa^{-1}\} \right]. \end{cases} \quad (27)$$

Let $\varepsilon > 0$. Since one can show that $z_1 \rightarrow 2$ a.s. as $d \rightarrow \infty$ [33], with high probability we have $\lambda_1^* \leq \kappa(1 + \varepsilon) \lambda_d^*$. Moreover, one checks easily that $d^{-1} \sum_{i=1}^d [\lambda_i^*]^2 \xrightarrow{\text{P}} 2$ as $d \rightarrow \infty$. Letting $\mu_i := \lambda_i^* / \sqrt{1 + \varepsilon}$, we can therefore use $\{\mu_i\}$ to lower bound $\omega(K_\kappa)$ as $d \rightarrow \infty$, with $\kappa_\varepsilon := \kappa(1 + \varepsilon)$. This yields that, for any $\varepsilon > 0$:

$$f(\kappa_\varepsilon) = \liminf_{d \rightarrow \infty} \frac{2\omega(K_{\kappa_\varepsilon})}{d} \geq \frac{1}{\sqrt{1 + \varepsilon}} \sqrt{\frac{2 \left[\int_{2\kappa^{-1}}^2 x^2 \sigma_{\text{s.c.}}(dx) + \frac{2}{\kappa} \int_{-2}^{2\kappa^{-1}} x \sigma_{\text{s.c.}}(dx) \right]^2}{\int_{2\kappa^{-1}}^2 x^2 \sigma_{\text{s.c.}}(dx) + \frac{4}{\kappa^2} \int_{-2}^{2\kappa^{-1}} \sigma_{\text{s.c.}}(dx)}}. \quad (28)$$

Taking the limits $\kappa \rightarrow \infty$ and $\varepsilon \rightarrow 0$ in eq. (28) yields $\lim_{\kappa \rightarrow \infty} f(\kappa) \geq 1$, which ends the proof. \square

3.1.4 Proof of Proposition 2.2

We can now complete the proof of Proposition 2.2. Since $\alpha = n/d^2 < 1/4$, by Lemma 3.5, we can find $\kappa = \kappa(\alpha)$ and $\varepsilon = \varepsilon(\alpha)$ such that $n \leq (1 - \varepsilon)\omega(K_\kappa)^2$ for n large enough. Therefore, by Lemma 3.2 there exists $A = A(\alpha) > 0$ such that:

$$\lim_{d \rightarrow \infty} \mathbb{P} \left\{ \exists S \in V \text{ s.t. } S \succeq 0 \text{ and } \text{Tr}[S^2] \leq Ad \text{ and } \lambda_{\max}(S) \leq \kappa \lambda_{\min}(S) \right\} = 1. \quad (29)$$

⁷Notice that we replaced $\|S\|_F^2 = 2d$ by $\|S\|_F^2 \leq 2d$ since w.h.p. one can always find $S \in K_\kappa$ such that $\text{Tr}[SZ] > 0$.

Notice that $H := n^{-1/2} \sum_{\mu=1}^n G_\mu \sim \text{GOE}(d)$, and thus $\mathbb{P}[\|H\|_{\text{op}} \leq 3] = 1 - o_d(1)$ [34]. Conditioning on this event, let $S \in V$. Then $\text{Tr}[HS] = \sqrt{n} = \sqrt{\alpha}d$, and thus by duality of $\|\cdot\|_{\text{op}}$ and $\|\cdot\|_{S_1}$:

$$\text{Tr}|S| \geq \frac{1}{\|H\|_{\text{op}}} |\text{Tr}[HS]| \geq \frac{\sqrt{\alpha}}{3}d.$$

The proof of Proposition 2.2 is then ended by noticing that:

$$\text{Sp}(S) \subseteq [\lambda_-, \lambda_+] \Leftrightarrow \begin{cases} S \succeq 0, \\ \text{Tr}[S^2] \leq A(\alpha)d, \\ \lambda_{\max}(S) \leq \kappa(\alpha)\lambda_{\min}(S), \\ \text{Tr}[S] \geq B(\alpha)d, \end{cases}$$

for some $0 < \lambda_- \leq \lambda_+$ depending only on α . \square

3.2 The unsatisfiable regime: proof of Proposition 2.3

We will show the following general result on the unsatisfiability of approximate versions of a general class of random geometry problems.

Proposition 3.6

Let $b \in \mathbb{R}$, $p, n \rightarrow \infty$, $K \subseteq \mathbb{R}^p$ a closed convex cone, with Gaussian width $\omega(K)$, $(a_\mu)_{\mu=1}^n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbb{I}_p)$, and denote $\mathcal{C}_\mu^{(b)}(x) := |a_\mu^\top x - b|$ the μ -th ‘‘constraint’’. We denote

$$V_b := \{x \in \mathbb{R}^p : \forall \mu \in [n], \mathcal{C}_\mu^{(b)}(x) = 0\}$$

a randomly-oriented affine subspace. Then we have the following:

(i) Assume that $b \neq 0$. Then

- (a) If there exists $\varepsilon > 0$ such that $n \leq (1 - \varepsilon)\omega(K)^2$ as $n \rightarrow \infty$, then $\mathbb{P}[V_b \cap K \neq \emptyset] \rightarrow_{n \rightarrow \infty} 1$.
- (b) If there exists $\varepsilon > 0$ such that $n \geq (1 + \varepsilon)\omega(K)^2$ as $n \rightarrow \infty$, then there exists $c = c(\varepsilon, b) > 0$ and $\eta = \eta(\varepsilon, b) \in (0, 1)$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\forall x \in K : \#\{\mu \in [n] : \mathcal{C}_\mu^{(b)}(x) > c\} \geq \eta n\} = 1.$$

(ii) Assume that $b = 0$. Note that V_0 is a linear subspace, and $V_0 \cap K$ is a closed convex cone.

- (a) Assume that $n \leq (1 - \varepsilon)\omega(K)^2$ for some $\varepsilon > 0$. Then as $n \rightarrow \infty$, $\mathbb{P}[V_0 \cap K \cap S^{p-1} \neq \emptyset] \rightarrow 1$.
- (b) Assume that $n \geq (1 + \varepsilon)\omega(K)^2$ for some $\varepsilon > 0$. Then there exists $c = c(\varepsilon, b) > 0$ and $\eta = \eta(\varepsilon, b) \in (0, 1)$ such that, with probability $1 - o_n(1)$, the following holds for all $\tau \geq 0$:

$$\max_{\substack{x \in K \\ \#\{\mu \in [n] : \mathcal{C}_\mu^{(0)}(x) > c\tau\} < \eta n}} \|x\|_2 \leq \tau.$$

It is clear that Proposition 3.6 ends the proof of Proposition 2.3. Indeed, seen as an element of $\mathbb{R}^{d(d+1)/2}$, $\sqrt{d/2}G_\mu \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbb{I}_{d(d+1)/2})$, see the canonical embedding in eq. (25). By Proposition 3.4, we have $\omega(\mathcal{S}_d^+)^2 = d^2/4 + o(d^2)$, and thus we can apply Proposition 3.6 with $p = d(d+1)/2$. The difference \sqrt{d} in normalization in point (ii) of Proposition 2.3 comes from the additional \sqrt{d} factor that arises when relating G_μ to a standard Gaussian.

We now focus on proving Proposition 3.6.

3.2.1 Proof of Proposition 3.6

We first show (i), assuming $b \neq 0$. Results of [8, 9] show that $V_b \cap K \neq \emptyset$ with high probability if $n \leq (1 - \epsilon)\omega(K)^2$ (see e.g. Theorem 8.1 of [9]), i.e. point (a). While the converse is also shown in these works for $n \geq (1 + \epsilon)\omega(K)^2$, here we wish to prove the stronger statement (b). Assume therefore that $n \geq (1 + \epsilon)\omega(K)^2$. By the union bound, for all $c \geq 0$ and $\eta \in (0, 1)$:

$$\begin{aligned}
& \mathbb{P}(\exists x \in K : \#\{\mu \in [n] : \mathcal{C}_\mu^{(b)}(x) > c\} < \eta n) \\
&= \mathbb{P}(\exists x \in K, \exists S \subseteq [n] : |S| > (1 - \eta)n \text{ and } \forall \mu \in S, \mathcal{C}_\mu^{(b)}(x) \leq c) \\
&\leq \sum_{\substack{S \subseteq [n] \\ |S| > (1 - \eta)n}} \mathbb{P}(\exists x \in K : \forall \mu \in S, \mathcal{C}_\mu^{(b)}(x) \leq c), \\
&\stackrel{(a)}{\leq} \left[\sum_{k < \eta n} \binom{n}{k} \right] \mathbb{P}(\exists x \in K : \|\{a_\mu^\top x - b\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq c), \\
&\stackrel{(b)}{\leq} \exp\left\{\eta n \log \frac{e}{\eta}\right\} \mathbb{P}(\exists x \in K : \|\{a_\mu^\top x - b\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq c). \tag{30}
\end{aligned}$$

We used that a_μ are i.i.d. in (a), and the bound $\sum_{i=0}^k \binom{n}{i} \leq (en/k)^k$ in (b). We now make use of the following lemma (proven later on).

Lemma 3.7

Recall that $n \geq (1 + \epsilon)\omega(K)^2$. There exists $c_1, c_2 > 0$ and $\eta_0 \in (0, 1)$ depending only on ϵ such that for any $\eta \in (0, \eta_0)$:

$$\mathbb{P}(\exists x \in K : \|\{a_\mu^\top x - b\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq c_1 \cdot b) \leq 2 \exp\{-nc_2\}.$$

Applying Lemma 3.7 in eq. (30), we can consider $\eta = \eta(\epsilon, b) \in (0, 1/2)$ small enough, such that $\eta < \eta_0$ and $\eta \log(e/\eta) \leq c_2/2$. This yields then that

$$\mathbb{P}(\exists x \in K : \#\{\mu \in [n] : \mathcal{C}_\mu^{(b)}(x) > c_1 \cdot b\} < \eta n) \leq 2 \exp\{-nc_2/2\} \rightarrow 0,$$

and ends the proof. We now prove (ii) of Proposition 3.6, assuming $b = 0$. (a) is here a simple consequence of the usual Gordon's "escape through a mesh" theorem [28], so we focus on (b), assuming $n \geq (1 + \epsilon)\omega(K)^2$. Note that since K is a cone, we have for all $c \geq 0$, $\eta \in (0, 1)$:

$$\begin{aligned}
& \sup_{\substack{x \in K \\ \#\{\mu \in [n] : \mathcal{C}_\mu^{(0)}(x) > c\} < \eta n}} \|x\|_2 \\
&= \sup\{v \geq 0 : \exists x \in K \cap \mathbb{S}^{p-1} \text{ s.t. } \#\{\mu \in [n] : \mathcal{C}_\mu^{(0)}(x) > c/v\} < \eta n\}. \tag{31}
\end{aligned}$$

We now use the following counterpart to Lemma 3.7 in the case $b = 0$, also proven later:

Lemma 3.8

Recall that $n \geq (1 + \epsilon)\omega(K)^2$. There exists $c_1, c_2 > 0$ and $\eta_0 \in (0, 1)$ depending only on ϵ such that for any $\eta \in (0, \eta_0)$:

$$\mathbb{P}(\exists x \in K \cap \mathbb{S}^{p-1} : \|\{a_\mu^\top x\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq c_1) \leq 2 \exp\{-nc_2\}.$$

Repeating the same reasoning as in the $b \neq 0$ case, we can then find $c = c(\epsilon) > 0$ and $\eta = \eta(\epsilon) \in (0, 1/2)$ such that with probability $1 - o_n(1)$:

$$\forall x \in K \cap \mathbb{S}^{p-1} : \#\{\mu \in [n] : \mathcal{C}_\mu^{(0)}(x) > c\} \geq \eta n.$$

Plugging this in eq. (31) yields that with probability $1 - o_n(1)$, for all $\tau \geq 0$:

$$\max_{\substack{x \in K \\ \#\{\mu \in [n] : \mathcal{C}_\mu^{(0)}(x) > c\tau\} < \eta n}} \|x\|_2 \leq \tau,$$

which ends the proof. \square

We now prove the two Lemmas 3.7 and 3.8.

3.2.2 Proofs of Lemma 3.7 and 3.8

Proof of Lemma 3.7 – Note that for all $t \geq 0$ and $\eta \in (0, 1)$:

$$\mathbb{P}\left(\exists x \in K : \|\{a_\mu^\top x - b\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq t\right) = \mathbb{P}[\exists x \in K : \|Gx - b\mathbf{1}_m\|_\infty \leq t], \quad (32)$$

with $m := n(1 - \eta)$, $G \in \mathbb{R}^{m \times p}$ an i.i.d. $\mathcal{N}(0, 1)$ matrix, and $\mathbf{1}_m$ the all-ones vector. We thus have:

$$\begin{aligned} \mathbb{P}(\exists x \in K : \|\{a_\mu^\top x - b\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq t) &\stackrel{(a)}{=} \lim_{A \rightarrow \infty} \mathbb{P}[\exists x \in K : \|x\|_2 \leq A \text{ and } \|Gx - b\mathbf{1}_m\|_\infty \leq t], \\ &\stackrel{(b)}{=} \lim_{A \rightarrow \infty} \mathbb{P}\left[\min_{\substack{x \in K \\ \|x\|_2 \leq A}} \|Gx - b\mathbf{1}_m\|_\infty \leq t\right], \end{aligned} \quad (33)$$

where (a) follows from dominated convergence, and (b) uses that the minimum is now over a compact set since K is closed. Since $\|\cdot\|_\infty$ and $\|\cdot\|_1$ are dual norms, we have for all $x \in K$:

$$\|Gx - b\mathbf{1}_m\|_\infty = \max_{\substack{\lambda \in \mathbb{R}^m \\ \|\lambda\|_1 \leq 1}} [-b\lambda^\top \mathbf{1}_m + \lambda^\top Gx].$$

We can now use the Gaussian min-max inequality (Proposition 3.1), which, together with eq. (33), implies:

$$\mathbb{P}(\exists x \in K : \|\{a_\mu^\top x - b\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq t) \leq \lim_{A \rightarrow \infty} 2\mathbb{P}[\gamma_A(g, h) \leq t] \stackrel{(a)}{\leq} 2\mathbb{P}[\gamma(g, h) \leq t].$$

Here we defined:

$$\gamma(g, h) := \inf_{x \in K} \max_{\substack{\lambda \in \mathbb{R}^m \\ \|\lambda\|_1 \leq 1}} [-b\lambda^\top \mathbf{1}_m + \|\lambda\|_2 g^\top x + \|x\|_2 h^\top \lambda], \quad (34)$$

and γ_A is defined by restricting the infimum to $\|x\|_2 \leq A$. Moreover, $g \sim \mathcal{N}(0, \mathbf{I}_p)$, $h \sim \mathcal{N}(0, \mathbf{I}_m)$. The inequality (a) holds since $\gamma(g, h) \leq \gamma_A(g, h)$ for all $A > 0$. To conclude the proof, it therefore suffices to show:

$$\mathbb{P}[\gamma(g, h) \leq c_1 b] \leq 2 \exp\{-nc_2\}, \quad (35)$$

for c_1, c_2 small enough (depending on ε, b). We use again that $g \rightarrow \max_{x \in K \cap \mathbb{S}^{p-1}} [g^\top x]$ is 1-Lipschitz, and in particular concentrates on the Gaussian width by Theorem 3.3. Using that g is distributed as $-g$, this implies that for any $u > 0$:

$$\mathbb{P}\left\{\min_{x \in K \cap \mathbb{S}^{p-1}} [g^\top x] \leq -\omega(K) - u\right\} \leq e^{-u^2/2}.$$

Since $\omega(K) \leq \sqrt{n/(1 + \varepsilon)}$ by hypothesis, we can fix $\delta = \delta(\varepsilon) > 0$ and $\eta_0 = \eta_0(\varepsilon) > 0$ such that for n large enough we have for $\eta < \eta_0$: $\omega(K) + \delta\sqrt{m} \leq \sqrt{m/(1 + \varepsilon/2)}$ (recall that $m = (1 - \eta)n$). Thus, since K is a cone, and using the max-min inequality, we have with probability at least $1 - e^{-n(1 - \eta_0)\delta^2/2}$:

$$\gamma(g, h) \geq \inf_{v \geq 0} \max_{\substack{\lambda \in \mathbb{R}^m \\ \|\lambda\|_1 \leq 1}} [-b\lambda^\top \mathbf{1}_m + v(h^\top \lambda - \|\lambda\|_2(\omega(K) + \delta\sqrt{m}))]. \quad (36)$$

Let us now assume that $b > 0$, and let $X \sim \mathcal{N}(0, 1)$, and $D := \mathbb{E}[|X|] = \sqrt{2/\pi}$. We pick $\sigma = \sigma(\varepsilon) \in (0, 1)$ (its choice will be constrained later on), and define A_ε by:

$$\mathbb{E}[X^2 \mathbf{1}\{X \leq A_\varepsilon\}] = \sigma(\varepsilon). \quad (37)$$

Finally, we define $\lambda^* = \lambda^*(h) \in \mathbb{R}^m$ by

$$\lambda_\mu^* := \frac{1}{Dm} h_\mu \mathbf{1}\{h_\mu \leq A_\varepsilon\}. \quad (38)$$

It is a simple exercise based on Hoeffding's and Bernstein's inequalities [34] to check that for any $u > 0$ (recall that $m = (1 - \eta)n \geq (1 - \eta_0(\varepsilon))n$):

$$\begin{cases} \mathbb{P}[\|\lambda^*\|_1 \leq 1] & \geq 1 - \exp\{-C_1 n\}, \\ \mathbb{P}\left[h^\top \lambda^* - \frac{\sigma(\varepsilon)^2}{D} \leq -u\right] & \leq \exp\{-C_2 n \min(u^2, u)\}, \\ \mathbb{P}\left[\|\lambda^*\|_2^2 - \frac{\sigma(\varepsilon)^2}{mD^2} \geq \frac{u}{n}\right] & \leq \exp\{-C_3 n \min(u^2, u)\}, \\ \mathbb{P}[\mathbf{1}_m^\top \lambda^* \geq -C_4] & \leq \exp\{-C_5 n\}, \end{cases} \quad (39)$$

for some positive constants $(C_a)_{a=1}^5$, all depending on ε . In particular, the first three lines of eq. (39) imply that for all $u \in (0, 1)$, with probability at least $1 - 3 \exp\{-C(\varepsilon)nu^2\}$:

$$\|\lambda^*\|_1 \leq 1 \text{ and } \frac{h^\top \lambda^*}{\|\lambda^*\|_2 \omega(K) + \delta\sqrt{m}} \geq \frac{\sigma(\varepsilon)^2/D - u}{\sqrt{\sigma(\varepsilon)^2/D^2 + u}} \sqrt{1 + \frac{\varepsilon}{2}}, \quad (40)$$

in which we used that $\omega(K) + \delta\sqrt{m} \leq \sqrt{m}(1 + \varepsilon/2)^{-1/2}$, and $m/n \leq 1$. We can choose $\sigma(\varepsilon) \in (0, 1)$ sufficiently close to 1, and $u(\varepsilon) \in (0, 1)$ sufficiently close to 0 such that the right-hand side of eq. (40) is greater than 1. Combining it with the last equation of eq. (39) and the lower bound of eq. (36), we get that (with new constants c_1, c_2 depending on ε), with probability at least $1 - 2 \exp\{-c_2(\varepsilon)n\}$:

$$\gamma(g, h) \geq bc_1(\varepsilon),$$

which implies eq. (35) and ends the proof. The case $b < 0$ is treated in a similar way, constraining $h_\mu \geq -A_\varepsilon$ rather than $h_\mu \leq A_\varepsilon$ in eq. (38). \square

Proof of Lemma 3.8 – Let $t \geq 0$. Repeating the same arguments as in the proof of Lemma 3.7 one obtains that

$$\mathbb{P}(\exists x \in K \cap \mathbb{S}^{p-1} : \|\{a_\mu^\top x\}_{\mu=1}^{(1-\eta)n}\|_\infty \leq t) \leq 2\mathbb{P}\left[\underbrace{\min_{x \in K \cap \mathbb{S}^{p-1}} \max_{\substack{\lambda \in \mathbb{R}^m \\ \|\lambda\|_1 \leq 1}} \{\|\lambda\|_2 g^\top x + h^\top \lambda\} \leq t}_{=:\gamma(g, h)}\right].$$

Again, we can fix $\eta_0(\varepsilon) > 0$ and $\delta(\varepsilon) > 0$ such that for $\eta < \eta_0$ and n large enough we have $\omega(K) + \delta\sqrt{m} \leq \sqrt{m}[1 + \varepsilon/2]^{-1/2}$. By the max-min inequality and the concentration of the Gaussian width, this implies that with probability at least $1 - e^{-n(1-\eta_0)\delta^2}$:

$$\gamma(g, h) \geq \max_{\substack{\lambda \in \mathbb{R}^m \\ \|\lambda\|_1 \leq 1}} [h^\top \lambda - \|\lambda\|_2 (\omega(K) + \delta\sqrt{m})]. \quad (41)$$

Defining again $D := \sqrt{2/\pi}$ so that $D = \mathbb{E}[|X|]$ for $X \sim \mathcal{N}(0, 1)$, we now define λ^* as:

$$\lambda_\mu^* := \frac{1}{Dm} h_\mu.$$

We have the counterpart to eq. (39) for this case: for any $u > 0$ and $\tau > 0$,

$$\begin{cases} \mathbb{P}[\|\lambda^*\|_1 \leq 1 + \tau] & \geq 1 - \exp\{-C_1 n \tau^2\}, \\ \mathbb{P}\left[h^\top \lambda^* - \frac{1}{D} \leq -u\right] & \leq \exp\{-C_2 n \min(u^2, u)\}, \\ \mathbb{P}\left[\|\lambda^*\|_2^2 - \frac{1}{mD^2} \geq \frac{u}{n}\right] & \leq \exp\{-C_3 n \min(u^2, u)\}, \end{cases} \quad (42)$$

for some $(C_a)_{a=1}^3$ depending on ε . We can fix $u = u(\varepsilon) > 0$ such that

$$D^{-1} - u - \sqrt{\frac{u + D^{-2}}{1 + \varepsilon/2}} =: 2c_1(\varepsilon) > 0,$$

since the limit as $u \rightarrow 0$ of the left-hand side is strictly positive. Letting $\tau = 1$, we finally get that with probability at least $1 - 3\exp(-c_2(\varepsilon)n)$ we can lower bound (using $\lambda = \lambda^*/2$ such that $\|\lambda\|_1 \leq 1$)

$$\begin{aligned} \gamma(g, h) &\geq \frac{D^{-1} - u}{2} - \frac{1}{2} \sqrt{\frac{u}{n} + \frac{1}{mD^2}} (\omega(K) + \delta\sqrt{m}), \\ &\geq \frac{D^{-1} - u}{2} - \frac{1}{2} \sqrt{\frac{u + D^{-2}}{1 + \varepsilon/2}}, \\ &\geq c_1(\varepsilon). \end{aligned}$$

This ends the proof. □

4 Universality: proof of Proposition 2.1

This section is devoted to the proof of Proposition 2.1. We first show in Section 4.1 a critical result on the Lipschitz constant of the “error” function appearing in eq. (10). This requires controlling a random process on the operator norm sphere, which is also useful in the proof of Theorem 1.3, see Section 2.4. We leverage this control to show in Section 4.2 a general result on the universality of a quantity known as the asymptotic free entropy of the model, both for matrices arising from ellipsoid fitting and its Gaussian equivalent. This result follows from an interpolation argument. Finally, we apply these results in the so-called “low-temperature” limit in Section 4.3 to deduce Proposition 2.1. As we mentioned, while we can not directly apply the results of [21], Sections 4.2 and 4.3 closely follows their approach. We defer to Appendix D some technicalities, as well as some parts of the proof that more directly follow the arguments of [21].

4.1 Lipschitz constant of the energy function, and bounding random processes

We show here the following result on the behavior of the error (or “energy”) function, under both models $X_\mu \sim \text{Ellipse}(d)$ and $X_\mu \sim \text{GOE}(d)$.

Lemma 4.1 (Lipschitz constant of the energy)

Let $n, d \geq 1$ and $n, d \rightarrow \infty$ with $\alpha_1 d^2 \leq n \leq \alpha_2 d^2$ for some $0 < \alpha_1 < \alpha_2$. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ such that $\|\phi'\|_\infty < \infty$, and $X_1, \dots, X_n \in \mathcal{S}_d$ be generated i.i.d. according to either $\text{GOE}(d)$ or $\text{Ellipse}(d)$. For $S \in \mathcal{S}_d$, we define the *energy*:

$$E_{\{X_\mu\}}(S) := \frac{1}{d^2} \sum_{\mu=1}^n \phi[\text{Tr}(X_\mu S)].$$

Then the following holds for some $C > 0$ (depending only on α):

$$\mathbb{P} \left[\sup_{S_1, S_2 \in \mathcal{S}_d} \frac{|E_{\{X_\mu\}}(S_1) - E_{\{X_\mu\}}(S_2)|}{\|S_1 - S_2\|_{\text{op}}} \leq C \|\phi'\|_\infty \right] \geq 1 - 2e^{-n}.$$

In other words, the energy function has a bounded Lipschitz constant (as $d \rightarrow \infty$) with respect to the operator norm. Note that this is strongly stronger than what a naive use of the triangular inequality and of the duality $\|\cdot\|_{\text{op}} \leftrightarrow \|\cdot\|_{S_1}$ yields:

$$\frac{|E_{\{X_\mu\}}(S_1) - E_{\{X_\mu\}}(S_2)|}{\|S_1 - S_2\|_{\text{op}}} \leq \frac{\|\phi'\|_\infty}{d^2} \sum_{\mu=1}^n \text{Tr}|X_\mu|,$$

since $\text{Tr}|X_\mu| \gtrsim d$ for $X_\mu \sim \text{GOE}(d)$, and $\text{Tr}|X_\mu| \gtrsim \sqrt{d}$ for $X_\mu \sim \text{Ellipse}(d)$. Instead, the proof of Lemma 4.1 is based on the following bounds for random processes, for which we separate the $\text{GOE}(d)$ and $\text{Ellipse}(d)$ setting. Lemma 4.2 is a consequence of elementary concentration results, and is proven in Appendix D.1, while Lemma 4.3 is proven in the following.

Lemma 4.2 (Bounding random processes, GOE(d) setting)

Let $(G_\mu)_{\mu=1}^n \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)$. Let $r \in [1, 2]$. There exists $C > 0$ such that for all $t > 0$:

$$\mathbb{P} \left[\max_{\|S\|_F^2=d} \left(\sum_{\mu=1}^n |\text{Tr}[G_\mu S]|^r \right)^{1/r} \geq (C+t)n^{1/r} \right] \leq \exp(-nt^2/2).$$

Lemma 4.3 (Bounding random processes, Ellipse(d) setting)

Let $(W_\mu)_{\mu=1}^n \stackrel{\text{i.i.d.}}{\sim} \text{Ellipse}(d)$. Let $r \in [1, 4/3]$. We assume that $\alpha_1 d^2 \leq n \leq \alpha_2 d^2$, for some $0 < \alpha_1 < \alpha_2$. There are constants $C_1, C_2 > 0$ (that might depend on α_1, α_2) such that for all $t > 0$:

$$\mathbb{P} \left[\max_{\|S\|_{\text{op}}=1} \sum_{\mu=1}^n |\text{Tr}(W_\mu S)|^r \geq n(C_1 + t) \right] \leq 2 \exp \left\{ -C_2 \min(nt^{\frac{2}{r}}, n^{\frac{1}{4} + \frac{1}{r}t^{\frac{1}{r}}}) \right\}.$$

Proof of Lemma 4.1 – We finish here the proof, assuming Lemmas 4.2 and 4.3. By the mean value theorem and the duality $\|\cdot\|_{\text{op}} \leftrightarrow \|\cdot\|_{S_1}$:

$$|E_{\{X_\mu\}}(S_1) - E_{\{X_\mu\}}(S_2)| \leq \left| \sup_{S \in \mathcal{S}_d} \langle \nabla_S E_{\{X_\mu\}}(S), S_1 - S_2 \rangle \right| \leq \|S_1 - S_2\|_{\text{op}} \sup_{S \in \mathcal{S}_d} \|\nabla_S E_{\{X_\mu\}}(S)\|_{S_1}.$$

Again using the duality $\|\cdot\|_{\text{op}} \leftrightarrow \|\cdot\|_{S_1}$:

$$\|\nabla_S E_{\{X_\mu\}}(S)\|_{S_1} = \frac{1}{d^2} \left\| \sum_{\mu=1}^n X_\mu \phi'[\text{Tr}(X_\mu S)] \right\|_{S_1},$$

$$\begin{aligned}
&= \frac{1}{d^2} \sup_{\|R\|_{\text{op}}=1} \sum_{\mu=1}^n \text{Tr}[X_\mu R] \phi'[\text{Tr}(X_\mu S)], \\
&\leq \frac{\|\phi'\|_\infty}{d^2} \sup_{\|R\|_{\text{op}}=1} \sum_{\mu=1}^n |\text{Tr}[X_\mu R]|.
\end{aligned}$$

Using Lemmas 4.2 and 4.3 in the case $r = 1$, we reach the sought statement. \square

Remark I– Note that a naive argument using that $\text{Tr}[W_\mu S]$ is a sub-exponential random variable yields Lemma 4.3 for $r = 1$, which is already enough to deduce Lemma 4.1. However, Lemma 4.3 is also used later in the proof of Theorem 1.3, see Section 2.4. Since $\text{Tr}[W_\mu S]$ is the sum of many independent random variables, we can leverage its two-tailed behavior (by Bernstein’s inequality) to prove Lemma 4.3 for all $r \leq 4/3$, yielding the limitation $r < 4/3$ in Theorem 1.3. It is possible that a finer analysis could lead to a proof of Lemma 4.3 for the case $4/3 \leq r \leq 2$, which would in turn imply Theorem 1.3 for $r < 2$.

Remark II– We give an informal argument as to why we can not hope to extend Lemma 4.2 nor 4.3 for $r > 2$. Indeed, in the $\text{GOE}(d)$ setting, the choice $S = \sqrt{d}G_\mu/\|G_\mu\|_F$ (for some $\mu \in [n]$) yields that the objective function is at least $\sqrt{d}\|G_\mu\|_F \gtrsim \sqrt{n}$. In the $\text{Ellipse}(d)$ setting, assume that $r > 2$. If $1/q + 1/r = 1$, then by the dualities $\ell^p \leftrightarrow \ell^q$ and $\|\cdot\|_{\text{op}} \leftrightarrow \|\cdot\|_{S_1}$:

$$\max_{\|S\|_{\text{op}}=1} \left(\sum_{\mu=1}^n |\text{Tr}(W_\mu S)|^r \right)^{1/r} = \max_{\|\lambda\|_q=1} \left\| \sum_{\mu=1}^n \lambda_\mu W_\mu \right\|_{S_1}. \quad (43)$$

Let us lower bound the right-hand side of eq. (43). Let $\beta \in (0, 1)$, $p = \beta n$, and $\lambda_1 = \dots = \lambda_p = p^{-1/q} > \lambda_{p+1} = \dots = \lambda_n = 0$. Then $\|\lambda\|_q = 1$. Moreover,

$$\left\| \sum_{\mu=1}^n \lambda_\mu W_\mu \right\|_{S_1} = p^{1-1/q} d^{-1/2} \left\| \frac{1}{p} \sum_{\mu=1}^p x_\mu x_\mu^\top - \text{Id} \right\|_{S_1}.$$

By classical results of concentration of Wishart matrices [34], we know that since $p \lesssim d^2$ then

$$\left\| \frac{1}{p} \sum_{\mu=1}^p x_\mu x_\mu^\top - \text{Id} \right\|_{S_1} \gtrsim \frac{d^{3/2}}{\sqrt{p}} \gtrsim \sqrt{\frac{d}{\beta}},$$

where \gtrsim might hide constants that depend on α . We then reach:

$$\left\| \sum_{\mu=1}^n \lambda_\mu W_\mu \right\|_{S_1} \gtrsim \beta^{1/2-1/q} n^{1-1/q}.$$

Since $r > 2$, one has $1/2 - 1/q < 0$. Letting β going to 0, this shows that any bound of the type

$$\max_{\|\lambda\|_q=1} \left\| \sum_{\mu=1}^n \lambda_\mu W_\mu \right\|_{S_1} \leq C(\alpha) n^{1-1/q}$$

can not hold.

Proof of Lemma 4.3 – Throughout this proof, constants might depend on α_1, α_2 . Let us define, for any $S \in \mathcal{S}_d$:

$$\begin{cases} Y(S) & := \sum_{\mu=1}^n |\text{Tr}(W_\mu S)|^r, \\ X(S) & := Y(S) - \mathbb{E}Y(S). \end{cases} \quad (44)$$

We fix $\varepsilon \in (0, 1)$. It follows of classical covering number bounds [32] that if $T := \{S \in \mathcal{S}_d : \|S\|_{\text{op}} = 1\}$, then

$$\log \mathcal{N}(T, \|\cdot\|_{\text{op}}, \varepsilon) \leq \frac{d(d+1)}{2} \log \frac{3}{\varepsilon}. \quad (45)$$

Let us fix N a minimal ε -net of T for $\|\cdot\|_{\text{op}}$. If we let $S^* := \arg \max_{\|S\|_{\text{op}}=1} Y(S)$, and $S_0 \in N$ with $\|S^* - S_0\|_{\text{op}} \leq \varepsilon$, then we have by Minkowski's inequality (recall $r \geq 1$):

$$\begin{aligned} \|\{\text{Tr}[W_\mu S^*]\}_{\mu=1}^n\|_r - \|\{\text{Tr}[W_\mu S_0]\}_{\mu=1}^n\|_r &\leq \|\{\text{Tr}[W_\mu (S_0 - S^*)]\}_{\mu=1}^n\|_r \\ &\leq \varepsilon \max_{\|S\|_{\text{op}}=1} \|\{\text{Tr}[W_\mu S]\}_{\mu=1}^n\|_r. \end{aligned}$$

This implies

$$\max_{\|S\|_{\text{op}}=1} \sum_{\mu=1}^n |\text{Tr}(W_\mu S)|^r \leq \frac{1}{(1-\varepsilon)^r} \max_{S \in N} \sum_{\mu=1}^n |\text{Tr}(W_\mu S)|^r. \quad (46)$$

We combine the covering number upper bound of eq. (45) and the relation of eq. (46) with the following lemma, proven later on, which bounds the deviation probability of the process for a given S .

Lemma 4.4 (Tail bound at a fixed point)

With the notations of eq. (44), we have, for any $S \in T$:

(i) $\mathbb{E}[Y(S)] \leq C_1 n$.

(ii) For all $t \geq 0$:

$$\mathbb{P}[X(S) \geq nt] \leq 2 \exp \left\{ -C_2 \min(nt^2, nt^{\frac{2}{r}}, n^{\frac{1}{4} + \frac{1}{r}} t^{\frac{1}{r}}) \right\}.$$

Note that the above constants may depend on r .

Picking $\varepsilon = 1/2$ and performing a union bound over N , we reach using eqs. (45),(46) and Lemma 4.4:

$$\mathbb{P}[\sup_{S \in T} Y(S) \geq n(C_1 + t)] \leq 2 \exp \left\{ C_3 n - C_2 \min(nt^2, nt^{\frac{2}{r}}, n^{\frac{1}{4} + \frac{1}{r}} t^{\frac{1}{r}}) \right\}.$$

We thus have for any $t \geq 1$:

$$\mathbb{P}[\sup_{S \in T} Y(S) \geq n(C_1 + t)] \leq 2 \begin{cases} \exp \left\{ C_3 n - C_2 n t^{\frac{2}{r}} \right\} & \text{if } t \leq n^{1 - \frac{3r}{4}}, \\ \exp \left\{ C_3 n - C_2 n^{\frac{1}{4} + \frac{1}{r}} t^{\frac{1}{r}} \right\} & \text{if } t \geq n^{1 - \frac{3r}{4}}. \end{cases}$$

Note that $n^{1/4+1/r} \geq n$ since $r \leq 4/3$. Therefore, for (new) constants C_1, C_2 we have for all $t > 0$:

$$\mathbb{P}[\sup_{S \in T} Y(S) \geq n(C_1 + t)] \leq 2 \exp \left\{ -C_2 \min(nt^{\frac{2}{r}}, n^{\frac{1}{4} + \frac{1}{r}} t^{\frac{1}{r}}) \right\},$$

which ends the proof. □

We now tackle Lemma 4.4.

Proof of Lemma 4.4 – We start with (i), fixing $S \in T$. We have $\mathbb{E}[Y(S)] = n\mathbb{E}[|\text{Tr}(W_1 S)|^r]$. Let $Z := \text{Tr}(W_1 S) \stackrel{d}{=} (x^\top S x - \text{Tr}[S])/\sqrt{d}$, with $x \sim \mathcal{N}(0, \mathbf{I}_d)$. Since $\|S\|_{\text{op}} = 1$, we have by Hanson-Wright's inequality [34]:

$$\begin{aligned} \mathbb{P}[|Z| \geq t] &\leq 2 \exp \left\{ -C \min \left(\frac{dt^2}{\|S\|_F^2}, \sqrt{dt} \right) \right\}, \\ &\leq 2 \exp \left\{ -C \min \left(t^2, \sqrt{dt} \right) \right\}, \end{aligned} \quad (47)$$

where $C > 0$ is an absolute constant, and we used that $\|S\|_F \leq \sqrt{d}\|S\|_{\text{op}}$. Separating the sub-Gaussian and the sub-exponential parts of the tail we have, we have for all $p \geq 1$:

$$\begin{aligned}
\mathbb{E}[|Z|^p] &= p \int_0^\infty du u^{p-1} \mathbb{P}[|Z| \geq u], \\
&\leq 2p \left[\int_0^{\sqrt{d}} du u^{p-1} e^{-Cu^2} + \int_{\sqrt{d}}^\infty du u^{p-1} e^{-C\sqrt{d}u} \right], \\
&\leq 2p \left[\int_0^\infty du u^{p-1} e^{-Cu^2} + e^{-Cd} \int_0^\infty du (u + \sqrt{d})^{p-1} e^{-C\sqrt{d}u} \right], \\
&\stackrel{(a)}{\leq} 2p \left[\frac{1}{2C^{p/2}} \Gamma(p/2) + e^{-Cd} \max(1, 2^{p-2}) \int_0^\infty du [u^{p-1} + d^{(p-1)/2}] e^{-C\sqrt{d}u} \right], \\
&\leq 2p \left[\frac{\Gamma(p/2)}{2C^{p/2}} + e^{-Cd} \max(1, 2^{p-2}) \left(\frac{\Gamma(p)}{C^p d^{p/2}} + \frac{d^{(p-2)/2}}{C} \right) \right].
\end{aligned}$$

We used in (a) that $(a+b)^x \leq \max(1, 2^{x-1})(a^x + b^x)$ for all $a, b, x \geq 0$. Using Minkowski's inequality, we reach that for all $p \geq 1$:

$$\mathbb{E}[|Z|^p]^{1/p} \leq C_1 \sqrt{p} + C_2 e^{-\frac{C_3 d}{p}} \left(\frac{p}{\sqrt{d}} + d^{\frac{1}{2} - \frac{1}{p}} \right), \quad (48)$$

for some positive constants $(C_a)_{a=1}^3$ independent of p and d . Informally, the sub-Gaussian tail dominates the first moments of Z since the sub-exponential tail only kicks in at the scale $\mathcal{O}(\sqrt{d})$. Eq. (48) implies claim (i) of Lemma 4.4 by taking $p = r$ (since the second term goes to 0 as $d \rightarrow \infty$ for any fixed p).

We turn to (ii). We make use of classical tail bounds for sub-Weibull random variables, recalled in Lemma A.1. Denoting $Z_\mu := \text{Tr}(W_\mu S)$, we have $X(S) = \sum_{\mu=1}^n \{|Z_\mu|^r - \mathbb{E}[|Z_\mu|^r]\}$. We decompose $X(S)$ in two parts, i.e. $X(S) = X_1(S) + X_2(S)$, with

$$\begin{cases} X_1(S) &:= \sum_{\mu=1}^n \left[\min(|Z_\mu|, \sqrt{d})^r - \mathbb{E}\{\min(|Z_\mu|, \sqrt{d})^r\} \right], \\ X_2(S) &:= \sum_{\mu=1}^n \left([|Z_\mu|^r - d^{r/2}] \mathbb{1}\{|Z_\mu| > \sqrt{d}\} - \mathbb{E}\left\{ [|Z_\mu|^r - d^{r/2}] \mathbb{1}\{|Z_\mu| > \sqrt{d}\} \right\} \right). \end{cases} \quad (49)$$

We will successively bound $X_1(S), X_2(S)$. To lighten the notations, we do not write their dependency on S in what follows. Observe that $(Z_\mu)_{\mu=1}^n$ are i.i.d. random variables, and that they satisfy the tail bound of eq. (47).

Bounding X_1 – Denoting $T_\mu := \min(|Z_\mu|, \sqrt{d})$, we have $\mathbb{P}[T_\mu \geq t] \leq 2 \exp\{-C_2 t^2\}$ by the tail bound of eq. (47). Moreover, $\mathbb{E}[T_\mu^r] \leq \mathbb{E}[|Z_\mu|^r] \leq C_1$ (depending only on r) by eq. (48). Therefore, for every $t > C_1$, we have

$$\mathbb{P}[|T_\mu^r - \mathbb{E}[T_\mu^r]| \geq t] = \mathbb{P}[T_\mu^r \geq \mathbb{E}[T_\mu^r] + t] \leq 2e^{-C_2(t + \mathbb{E}[T_\mu^r])^{2/r}} \leq 2e^{-C_2 t^{2/r}}.$$

This implies that $\mathbb{P}[|T_\mu^r - \mathbb{E}[T_\mu^r]| \geq t] \leq 2e^{-Ct^{2/r}}$ for all $t \geq 0$ and some (new) constant $C > 0$, depending only on r . We can thus apply (i) of Lemma A.1 for $q = 2/r \in [1, 2]$ and $a_i = 1/n$ (so $\|a\|_2^2 = \|a\|_{q^*}^q = n^{-1}$), which yields that for all $t \geq 0$

$$\mathbb{P}[|X_1| \geq nt] = \mathbb{P}\left[\frac{1}{n} \left| \sum_{\mu=1}^n \{T_\mu^r - \mathbb{E}[T_\mu^r]\} \right| \geq t \right] \leq 2 \exp\left\{ -Cn \min(t^2, t^{2/r}) \right\}. \quad (50)$$

Bounding X_2 – We proceed similarly, using (ii) of Lemma A.1. Letting

$$U_\mu := d^{r/2}[|Z_\mu|^r - d^{r/2}]\mathbb{1}\{|Z_\mu| > \sqrt{d}\},$$

then $U_\mu \geq 0$, and moreover, by the Cauchy-Schwarz inequality:

$$\begin{aligned} \mathbb{E}[U_\mu] &\leq d^{r/2}\sqrt{\mathbb{E}|Z_\mu|^{2r}}\sqrt{\mathbb{P}\{|Z_\mu| > \sqrt{d}\}}, \\ &\leq C_1e^{-C_2d}, \end{aligned}$$

for some $C_1, C_2 > 0$ depending only on r , using the moments and tail bound of eqs. (47) and (48). Repeating the argument used on T_μ above (using this time the second part of the tail of eq. (47)), we then reach that for all $t \geq 0$:

$$\mathbb{P}[|U_\mu - \mathbb{E}[U_\mu]| \geq t] \leq 2e^{-Ct^{1/r}}.$$

We can then apply (ii) of Lemma A.1 with $q = 1/r \in [1/2, 1]$ to reach:

$$\begin{aligned} \mathbb{P}[|X_2| \geq nt] &= \mathbb{P}\left[\frac{1}{n}\left|\sum_{\mu=1}^n\{U_\mu - \mathbb{E}[U_\mu]\}\right| \geq td^{r/2}\right] \leq 2\exp\left\{-C\min(nd^r t^2, d^{1/2}(nt)^{1/r})\right\}, \\ &\leq 2\exp\left\{-C\min(n^{1+r/2}t^2, n^{1/4+1/r}t^{1/r})\right\}, \end{aligned} \quad (51)$$

using that $\alpha_1 d^2 \leq n \leq \alpha_2 d^2$.

We conclude the proof of Lemma 4.4 by combining eqs. (50) and eq. (51), along with the union bound $\mathbb{P}[|X| \geq nt] \leq \mathbb{P}[|X_1| \geq nt/2] + \mathbb{P}[|X_2| \geq nt/2]$. \square

4.2 Free entropy universality for matrix models

In this section we state and prove a general universality theorem for the asymptotic free entropy in a large class of matrix models, under a “uniform one-dimensional central limit theorem” assumption (or pointwise normality). We first need to define such an assumption.

Definition 4.1 (*Uniform pointwise normality*)

Let $d \geq 1$, and ρ a probability distribution on \mathcal{S}_d . We say that ρ satisfies a *one-dimensional CLT with respect to the set $A_d \subseteq \mathcal{S}_d$* if:

(i) The mean and covariance of ρ are matching the $\text{GOE}(d)$ distribution, i.e. for $W \sim \rho$ and $G \sim \text{GOE}(d)$, we have $\mathbb{E}[W] = \mathbb{E}[G] = 0$ and for all $i \leq j$ and $k \leq l$: $\mathbb{E}[W_{ij}W_{kl}] = \mathbb{E}[G_{ij}G_{kl}] = \delta_{ik}\delta_{jl}(1 + \delta_{ijkl})/d$.

(ii) For any bounded Lipschitz function φ , we have:

$$\lim_{d \rightarrow \infty} \sup_{S \in A_d} \left| \mathbb{E}_{W \sim \rho}[\varphi(\text{Tr}[WS])] - \mathbb{E}_{G \sim \text{GOE}(d)}[\varphi(\text{Tr}[GS])] \right| = 0. \quad (52)$$

We can now state the universality theorem for the free entropy. Its proof is in great part an adaptation of the proof arguments for Theorem 1 and Lemma 1 in [21] (see also [20, 22, 24]). We sketch the ideas of its proof in the following, deferring some technicalities and adaptations of the arguments of [21] to appendices.

Theorem 4.5 (Free entropy universality for matrix models)

Let $n, d \geq 1$ and $n, d \rightarrow \infty$ with $\alpha_1 d^2 \leq n \leq \alpha_2 d^2$ for some $0 < \alpha_1 \leq \alpha_2$. We are given:

- (i) P_0 a probability distribution on \mathcal{S}_d , such that $\text{supp}(P_0) \subseteq B_2(C_0\sqrt{d})$, for a constant $C_0 > 0$.
- (ii) $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ a bounded differentiable function with bounded derivative.
- (iii) A series of symmetric convex sets A_d such that $\text{supp}(P_0) \subseteq A_d$.
- (iv) ρ a probability distribution on \mathcal{S}_d , which satisfies a one-dimensional CLT with respect to A_d as per Definition 4.1.

For $W_1, \dots, W_n \in \mathcal{S}_d$ we define the free entropy:

$$F_d(\{W_\mu\}) := \frac{1}{d^2} \log \int P_0(dS) \exp \left\{ - \sum_{\mu=1}^n \phi(\text{Tr}[W_\mu S]) \right\}. \quad (53)$$

Then for any bounded differentiable function ψ with bounded Lipschitz derivative we have

$$\lim_{d \rightarrow \infty} \left| \mathbb{E}_{\{W_\mu\} \stackrel{\text{i.i.d.}}{\sim} \rho} \psi[F_d(\{W_\mu\})] - \mathbb{E}_{\{G_\mu\} \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)} \psi[F_d(\{G_\mu\})] \right| = 0. \quad (54)$$

Remark I – One could straightforwardly weaken the hypothesis $\text{supp}(P_0) \subseteq A$ in Theorem 4.5 to the weaker condition $d^{-2} \log P_0(A^c) \rightarrow -\infty$ as $d \rightarrow \infty$.

Remark II – Note that our setup differs slightly from the one of [21], as we consider distributions P_0 with possibly continuous support, and (more importantly) for a fixed $S \in \mathcal{S}_d$, the projections $\{\text{Tr}[W_\mu S]\}_{\mu=1}^n$ are not sub-Gaussian when $W_\mu \sim \text{Ellipse}(d)$, but only sub-exponential. Nevertheless, we will see that the approach of [21] can in large part be adapted to prove Theorem 4.5, thanks to the results we showed in Section 4.1.

Sketch of proof of Theorem 4.5 – Since $\text{supp}(P_0) \subseteq A_d$, the integral in eq. (53) can be restricted to $S \in A_d$. We make use of an interpolation argument to show the universality of the free entropy. We define, for $t \in [0, \pi/2]$ and $\mu \in [n]$:

$$\begin{cases} U_\mu(t) & := \cos(t)W_\mu + \sin(t)G_\mu, \\ \tilde{U}_\mu(t) & := \frac{\partial U_\mu(t)}{\partial t} = -\sin(t)W_\mu + \cos(t)G_\mu. \end{cases} \quad (55)$$

Note that $\{U_\mu\}_{\mu=1}^n$ are still i.i.d., and are smooth functions of t . Moreover, if W_μ was also a $\text{GOE}(d)$ matrix, then $U_\mu(t), \tilde{U}_\mu(t)$ would be independent $\text{GOE}(d)$ matrices. By the fundamental theorem of calculus:

$$|\mathbb{E}\psi[F_d(W)] - \mathbb{E}\psi[F_d(G)]| = \left| \int_0^{\pi/2} \frac{\partial}{\partial t} \{\mathbb{E}\psi[F_d(U(t))]\} dt \right| \stackrel{(a)}{\leq} \int_0^{\pi/2} \left| \mathbb{E} \frac{\partial \psi[F_d(U(t))]}{\partial t} \right| dt, \quad (56)$$

where (a) follows by dominated convergence since $\psi[F_d(U(t))]$ is continuously differentiable on $[0, \pi/2]$, and the triangular inequality. We will deduce Theorem 4.5 if we can show the following two lemmas:

Lemma 4.6 (Domination)

Under the hypotheses of Theorem 4.5:

$$\int_0^{\pi/2} \sup_{d \geq 1} \left| \mathbb{E} \frac{\partial \psi[F_d(U(t))]}{\partial t} \right| dt < \infty.$$

Lemma 4.7 (Pointwise limit)

Under the hypotheses of Theorem 4.5, for any $t \in (0, \pi/2)$:

$$\lim_{d \rightarrow \infty} \mathbb{E} \frac{\partial \psi[F_d(U(t))]}{\partial t} = 0.$$

Indeed, plugging Lemmas 4.6 and 4.7 in eq. (56) and taking the $d \rightarrow \infty$ limit using dominated convergence ends the proof of Theorem 4.5. We therefore focus on proving these two lemmas in the following.

As it will be useful, we state the result of the elementary computation of the derivative:

$$\frac{\partial \psi[F_d(U(t))]}{\partial t} = -\frac{\psi'[F_d(U(t))]}{d^2} \sum_{\mu=1}^n \frac{\int P_0(dS) e^{-\sum_{\nu} \phi(\text{Tr}[U_{\nu}(t)S])} \left(\text{Tr}[S\tilde{U}_{\mu}(t)] \phi'(\text{Tr}[U_{\mu}(t)S]) \right)}{\int P_0(dS) e^{-\sum_{\nu} \phi(\text{Tr}[U_{\nu}(t)S])}}. \quad (57)$$

Because $\{G_{\mu}, W_{\mu}\}$ are i.i.d. we get further:

$$\mathbb{E} \frac{\partial \psi[F_d(U(t))]}{\partial t} = -\frac{n}{d^2} \mathbb{E} \left[\psi'[F_d(U(t))] \frac{\int P_0(dS) e^{-\sum_{\nu} \phi(\text{Tr}[U_{\nu}(t)S])} \left(\text{Tr}[S\tilde{U}_1(t)] \phi'(\text{Tr}[U_1(t)S]) \right)}{\int P_0(dS) e^{-\sum_{\nu} \phi(\text{Tr}[U_{\nu}(t)S])}} \right]. \quad (58)$$

Note that if W_{μ} was also a $\text{GOE}(d)$ matrix, for any t , $U_{\mu}(t)$ and $\tilde{U}_{\mu}(t)$ would be independent $\text{GOE}(d)$ matrices. The main idea behind the interpolation is that the matrix W_{μ} will only appear through some one-dimensional projection with a matrix S . We will then use Definition 4.1 to argue that one can effectively replace W_{μ} by a $\text{GOE}(d)$ matrix, which by the argument above would mean that we can consider the case of independent $\text{GOE}(d)$ matrices $U_{\mu}(t)$ and $\tilde{U}_{\mu}(t)$. In this case, the RHS of eq. (58) would be 0, since there is only a single term involving $\tilde{U}_1(t)$, which has zero mean: this crucial idea is the intuition behind Lemma 4.7.

The details of the proofs of Lemmas 4.6 and 4.7 are fairly technical and substantially follow the ones of their counterparts in [21]. For this reason, we defer them to Appendix D.2.

4.3 Proof of Proposition 2.1

4.3.1 Consequences of universality for ellipsoid fitting

We investigate here the consequences of Theorem 4.5 for the ellipsoid fitting problem. It follows by the Berry-Esseen central limit theorem [35] that the distribution $\text{Ellipse}(d)$ satisfies uniform pointwise normality on a large set of matrices (in the sense of Definition 4.1).

Lemma 4.8 (One-dimensional CLT for the ellipse problem)

Let $d \geq 1$ and $W \sim \text{Ellipse}(d)$. Fix any $\eta \in (0, 1/2)$. Let $A_d := \{S \in \mathcal{S}_d : \text{Tr}[|S|^3] \leq d^{3/2-\eta}\}$. Then A_d is convex and symmetric, and the law of W satisfies a one-dimensional CLT with respect to A_d , in the sense of Definition 4.1.

Remark – This lemma makes crucial use of the Gaussian nature of the vectors, and more specifically it relies on their rotation invariance and the first moments of their norm, as is clear from the proof. On the other hand, for vectors sampled from other distributions, such as $x \sim \text{Unif}(\{\pm 1\}^d)$ or $x \sim \text{Unif}(\mathbb{S}^{d-1})$, it is easy to see that Lemma 4.8 can not hold: indeed, $S = I_d$ is such that $\text{Tr}[SW] = 0$ deterministically, while $\text{Tr}[SG] = \text{Tr}[G] \sim \mathcal{N}(0, 2)$ for $G \sim \text{GOE}(d)$. This is consistent, as in the example of these two distributions there always exists an ellipsoid fit, which is the sphere itself, and therefore Theorem 1.3 can not possibly hold.

Proof of Lemma 4.8 – Note that A_d is a centered ball for the S_3 -norm, and is therefore convex and symmetric. The proof of the first and second moments condition of Definition 4.1 is immediate via a

simple calculation. We focus on proving condition (ii) of Definition 4.1. Fix $S \in A_d$, with eigenvalues $(\lambda_i)_{i=1}^d$. With $W \sim \text{Ellipse}(d)$ and $G \sim \text{GOE}(d)$, let

$$\begin{cases} X & := \text{Tr}[SW], \\ Y & := \text{Tr}[SG]. \end{cases}$$

It is trivial to see that $Y \sim \mathcal{N}(0, 2\text{Tr}[S^2]/d)$, so that $Y \stackrel{d}{=} d^{-1/2} \sum_{i=1}^d \lambda_i z_i$ for $z_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2)$. Moreover,

$$X \stackrel{d}{=} \frac{1}{\sqrt{d}} \sum_{i=1}^d \lambda_i (x_i^2 - 1),$$

with $x_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. We use the Berry-Esseen central limit theorem, and in particular the formulation of Chapter 11 of [35] – itself a simple consequence of the Lindeberg exchange method.

Lemma 4.9 (Corollary 11.59 of [35])

There exists a universal constant $C > 0$ such that the following holds. Let $p \geq 1$ and X_1, \dots, X_p and Y_1, \dots, Y_p be independent random variables, such that $\mathbb{E}[X_i] = \mathbb{E}[Y_i]$ and $\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2]$ for all $i \in [p]$. Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ a Lipschitz function with Lipschitz constant $\|\varphi\|_L$. Then

$$\left| \mathbb{E} \varphi \left(\sum_{i=1}^p X_i \right) - \mathbb{E} \varphi \left(\sum_{i=1}^p Y_i \right) \right| \leq C \|\varphi\|_L \left[\sum_{i=1}^p \left(\mathbb{E} |X_i|^3 + \mathbb{E} |Y_i|^3 \right) \right]^{1/3}.$$

Lemma 4.9 yields:

$$|\mathbb{E} \varphi(X) - \mathbb{E} \varphi(Y)| \leq C \|\varphi\|_L \left[B \frac{\text{Tr}[|S|^3]}{d^{3/2}} \right]^{1/3},$$

with $B = \mathbb{E}[|z^2 - 1|^3] + 2^{3/2} \mathbb{E}|z|^3$ for $z \sim \mathcal{N}(0, 1)$. Using the definition of A_d , we reach:

$$\sup_{S \in A_d} |\mathbb{E} \varphi(X) - \mathbb{E} \varphi(Y)| \leq C \|\varphi\|_L \sup_{S \in A_d} \left[\frac{1}{\sqrt{d}} \frac{\text{Tr}[|S|^3]}{d} \right]^{1/3} \leq C \|\varphi\|_L d^{-\eta/3} \rightarrow 0.$$

This ends the proof. □

We can now state the main result of this section, a corollary of Theorem 4.5 and Lemma 4.8.

Corollary 4.10 (Universality for ellipsoid fitting)

Let $n, d \geq 1$ and $n, d \rightarrow \infty$ with $\alpha_1 d^2 \leq n \leq \alpha_2 d^2$ for some $0 < \alpha_1 \leq \alpha_2$. Let P_0 be a probability distribution such that $\text{supp}(P_0) \subseteq B_{\text{op}}(C_0)$ for some constant $C_0 > 0$. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ a bounded differentiable function with bounded derivative. For $X_1, \dots, X_n \in \mathcal{S}_d$ we define the free entropy:

$$F_d(\{X_\mu\}) := \frac{1}{d^2} \log \int P_0(dS) \exp \left\{ - \sum_{\mu=1}^n \phi(\text{Tr}[X_\mu S]) \right\}.$$

Then for any ψ such that $\|\psi\|_\infty, \|\psi'\|_\infty, \|\psi'\|_L < \infty$ we have

$$\lim_{d \rightarrow \infty} \left| \mathbb{E}_{\{W_\mu\} \stackrel{\text{i.i.d.}}{\sim} \text{Ellipse}(d)} \psi[F_d(\{W_\mu\})] - \mathbb{E}_{\{G_\mu\} \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)} \psi[F_d(\{G_\mu\})] \right| = 0. \quad (59)$$

We notice that the only requirement for Corollary 4.10 to hold is $\text{supp}(P_0) \subseteq B_F(C\sqrt{d}) \cap B_3(d^{3/2-\eta})$ for some $C > 0$ and $\eta > 0$, a weaker requirement than $\text{supp}(P_0) \subseteq B_{\text{op}}(C_0)$.

Proof of Corollary 4.10 – Since $B_{\text{op}}(C_0) \subseteq B_2(C_0\sqrt{d})$, hypothesis (i) of Theorem 4.5 is satisfied. Condition (ii) is satisfied by hypothesis. Since $B_{\text{op}}(C_0) \subseteq B_3(C_0 d^{1/3}) \subseteq B_3(d^{1/2-\eta})$ for any $\eta \in$

$(0, 1/6)$, condition (iii) of Theorem 4.5 is satisfied with $A_d = B_3(d^{1/2-\eta})$. Finally, Lemma 4.8 verifies condition (iv) for this choice of A_d . All in all we can apply Theorem 4.5, from which the conclusion follows. \square

4.3.2 Proof of Proposition 2.1

We are now ready to prove Proposition 2.1, taking a “small-temperature” limit. Such arguments are classical in rigorous statistical mechanics, see e.g. Appendix A of [21]. Notice that the restriction $B \subseteq B_{\text{op}}(C_0)$ will be critical because we proved an upper bound on the Lipschitz constant of the energy for the operator norm, cf. Lemma 4.1. Recall the definition of the energy function:

$$E_{\{X_\mu\}}(S) := \frac{1}{d^2} \sum_{\mu=1}^n \phi[\text{Tr}(X_\mu S)].$$

We fix $\eta \in (0, 1)$, and $\mathcal{N}_\eta \subseteq B$ a minimal η -net of B for $\|\cdot\|_{\text{op}}$. By Since $B \subseteq B_{\text{op}}(C_0)$ and $\dim(\mathcal{S}_d) = d(d+1)/2$, it follows by standard covering number upper bounds [32, 34] that

$$\log |\mathcal{N}_\eta| = \log \mathcal{N}(B, \|\cdot\|_{\text{op}}, \eta) \leq \log \mathcal{N}\left(B_{\text{op}}(C_0), \|\cdot\|_{\text{op}}, \frac{\eta}{2}\right) \leq d^2 \log \frac{K}{\eta}, \quad (60)$$

for some $K > 0$ depending on C_0 . Recall the definition of $\text{GS}_d(\{X_\mu\})$ in eq. (10). We define:

$$\text{GS}_d(\eta, \{X_\mu\}) := \inf_{S \in \mathcal{N}_\eta} E_{\{X_\mu\}}(S).$$

We will show the two lemmas:

Lemma 4.11

For any $\eta > 0$ and any ψ such that $\|\psi\|_\infty, \|\psi'\|_\infty, \|\psi'\|_L < \infty$:

$$\lim_{d \rightarrow \infty} \left| \mathbb{E}_{\{W_\mu\} \stackrel{\text{i.i.d.}}{\sim} \text{Ellipse}(d)} \psi[\text{GS}_d(\eta, \{W_\mu\})] - \mathbb{E}_{\{G_\mu\} \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)} \psi[\text{GS}_d(\eta, \{G_\mu\})] \right| = 0.$$

Lemma 4.12

Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \rho$, with $\rho \in \{\text{GOE}(d), \text{Ellipse}(d)\}$. Then, with probability at least $1 - 2e^{-n}$:

$$|\text{GS}_d(\eta, \{X_\mu\}) - \text{GS}_d(\{X_\mu\})| \leq C \|\phi'\|_\infty \cdot \eta.$$

These results are proven in the following, let us first see how they end the proof of Proposition 2.1. We fix $\eta \in (0, 1)$. We have

$$\begin{aligned} |\mathbb{E}\psi[\text{GS}_d(\{W_\mu\})] - \mathbb{E}\psi[\text{GS}_d(\{G_\mu\})]| &\leq |\mathbb{E}\psi[\text{GS}_d(\eta, \{W_\mu\})] - \mathbb{E}\psi[\text{GS}_d(\eta, \{G_\mu\})]| \\ &\quad + \|\psi\|_L \sum_{X \in \{W, G\}} \mathbb{E}|\text{GS}_d(\eta, \{X_\mu\}) - \text{GS}_d(\{X_\mu\})|. \end{aligned}$$

The first term goes to 0 as $d \rightarrow \infty$ by Lemma 4.11. Finally, using Lemma 4.12 and the Cauchy-Schwarz inequality, we have:

$$\begin{aligned} \mathbb{E}|\text{GS}_d(\eta, \{X_\mu\}) - \text{GS}_d(\{X_\mu\})| &\leq e^{-n/2} \sqrt{2\mathbb{E}|\text{GS}_d(\eta, \{X_\mu\}) - \text{GS}_d(\{X_\mu\})|^2} + C\|\phi'\|_\infty \eta, \\ &\stackrel{(a)}{\leq} 4\alpha\|\phi\|_\infty e^{-n/2} + C\|\phi'\|_\infty \eta, \end{aligned}$$

using in (a) that $|E(S)| \leq \alpha\|\phi\|_\infty$. Letting $d \rightarrow \infty$, we get

$$\lim_{d \rightarrow \infty} |\mathbb{E}\psi[\text{GS}_d(\{W_\mu\})] - \mathbb{E}\psi[\text{GS}_d(\{G_\mu\})]| \leq C\|\phi'\|_\infty \|\psi\|_L \cdot \eta.$$

Taking the limit $\eta \rightarrow 0$ ends the proof of eq. (11). The claim of eq. (12) can be obtained easily by picking ψ approximating an indicator function, see e.g. Section A.1.3 of [21] for a detail of this argument. \square

Proof of Lemma 4.11 – We define, for $\beta > 0$:

$$F_d(\eta, \beta, \{X_\mu\}) := \frac{1}{d^2\beta} \log \frac{1}{|\mathcal{N}_\eta|} \sum_{S \in \mathcal{N}_\eta} \exp \left\{ -\beta \sum_{\mu=1}^n \phi(\text{Tr}[X_\mu S]) \right\}.$$

Using Corollary 4.10 with P_0 being the uniform distribution over \mathcal{N}_η , we have, for any $\beta > 0$:

$$\lim_{d \rightarrow \infty} |\mathbb{E}\psi[F_d(\eta, \beta, \{W_\mu\})] - \mathbb{E}\psi[F_d(\eta, \beta, \{G_\mu\})]| = 0. \quad (61)$$

Moreover, for any fixed d, η , we have $\text{GS}_d(\eta, \{X_\mu\}) = \lim_{\beta \rightarrow \infty} F_d(\eta, \beta, \{X_\mu\})$. Thus:

$$|\text{GS}_d(\eta, \{X_\mu\}) - F_d(\eta, \beta, \{X_\mu\})| \leq \int_\beta^\infty \left| \frac{\partial F_d(\eta, s, \{X_\mu\})}{\partial s} \right| ds. \quad (62)$$

Defining the ‘‘Gibbs’’ measure for $S \in \mathcal{N}_\eta$:

$$\mathbb{P}_\beta(S) := \frac{\exp \left\{ -\beta \sum_{\mu=1}^n \phi(\text{Tr}[X_\mu S]) \right\}}{\sum_{S' \in \mathcal{N}_\eta} \exp \left\{ -\beta \sum_{\mu=1}^n \phi(\text{Tr}[X_\mu S']) \right\}},$$

it is easy to check that

$$\begin{aligned} \left| \frac{\partial F_d(\eta, s, \{X_\mu\})}{\partial s} \right| &= \frac{1}{s^2 d^2} \left| \sum_{S \in \mathcal{N}_\eta} \mathbb{P}_s(S) \log \mathbb{P}_s(S) + \log |\mathcal{N}_\eta| \right|, \\ &\stackrel{(a)}{\leq} \frac{1}{s^2 d^2} \log |\mathcal{N}_\eta|, \\ &\stackrel{(b)}{\leq} \frac{1}{s^2} \log \frac{K}{\eta}, \end{aligned}$$

where (a) follows from the fact that, the uniform distribution over \mathcal{N}_η maximizes the entropy, and (b) is eq. (60). Plugging the result back in eq. (62) we get:

$$\begin{aligned} |\text{GS}_d(\eta, \{X_\mu\}) - F_d(\eta, \beta, \{X_\mu\})| &\leq \log \left(\frac{K}{\eta} \right) \int_\beta^\infty \frac{ds}{s^2}, \\ &\leq \frac{1}{\beta} \log \left(\frac{K}{\eta} \right). \end{aligned} \quad (63)$$

Combining eqs. (61) and (63) we get, for any $\beta > 0$:

$$\begin{aligned} \limsup_{d \rightarrow \infty} |\mathbb{E}\psi[\text{GS}_d(\eta, \{W_\mu\})] - \mathbb{E}\psi[\text{GS}_d(\eta, \{G_\mu\})]| &\leq \|\psi\|_L \sum_{X \in \{W, G\}} \mathbb{E} |\text{GS}_d(\eta, \{X_\mu\}) - F_d(\eta, \beta, \{X_\mu\})|, \\ &\leq \frac{2\|\psi\|_L}{\beta} \log \left(\frac{K}{\eta} \right). \end{aligned}$$

Taking the limit $\beta \rightarrow \infty$ ends the proof of Lemma 4.11. \square

Proof of Lemma 4.12 – Note that $\text{GS}_d(\eta, \{X_\mu\}) \geq \text{GS}_d(\{X_\mu\})$ since $\mathcal{N}_\eta \subseteq B$. The other side of this inequality is a direct consequence of Lemma 4.1. Indeed, assuming that $E(S)$ is $C\|\phi'\|_\infty$ -Lipschitz with respect to the operator norm, let us fix $S^* \in B$ such that $E(S^*) = \text{GS}_d(\{X_\mu\})$ (since B is closed and bounded it is compact, therefore this minimizer exists). Letting $S \in \mathcal{N}_\eta$ such that $\|S^* - S\|_{\text{op}} \leq \eta$, we have

$$\begin{aligned} \text{GS}_d(\{X_\mu\}) &= E(S^*), \\ &\geq E(S) - |E(S^*) - E(S)|, \\ &\geq \text{GS}_d(\eta, \{X_\mu\}) - C\|\phi'\|_\infty \cdot \eta, \end{aligned}$$

which ends the proof. \square

Acknowledgements

The authors are grateful to March Boedihardjo, Tim Kunisky, Petar Nizić-Nikolac, and Joel Tropp for insightful discussions and suggestions.

References

- [1] James James Francis Saunderson. *Subspace identification via convex optimization*. PhD thesis, Massachusetts Institute of Technology, 2011.
- [2] James Saunderson, Venkat Chandrasekaran, Pablo A Parrilo, and Alan S Willsky. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1395–1416, 2012.
- [3] James Saunderson, Pablo A Parrilo, and Alan S Willsky. Diagonal and low-rank decompositions and fitting ellipsoids to random points. In *52nd IEEE Conference on Decision and Control*, pages 6031–6036. IEEE, 2013.
- [4] Aaron Potechin, Paxton M Turner, Prayaag Venkat, and Alexander S Wein. Near-optimal fitting of ellipsoids to random points. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 4235–4295. PMLR, 2023.
- [5] Anastasia Podosinnikova, Amelia Perry, Alexander S Wein, Francis Bach, Alexandre d’Aspremont, and David Sontag. Overcomplete independent component analysis via SDP. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2583–2592. PMLR, 2019.
- [6] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for Sherrington-Kirkpatrick via planted affine planes. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 954–965. IEEE, 2020.
- [7] Antoine Maillard and Dmitriy Kunisky. Fitting an ellipsoid to random points: predictions using the replica method. *arXiv preprint arXiv:2310.01169*, 2023.
- [8] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12:805–849, 2012.
- [9] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference: A Journal of the IMA*, 3(3):224–294, 2014.
- [10] Daniel Kane and Ilias Diakonikolas. A nearly tight bound for fitting an ellipsoid to gaussian random points. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3014–3028. PMLR, 2023.
- [11] Afonso S Bandeira, Antoine Maillard, Shahar Mendelson, and Elliot Paquette. Fitting an ellipsoid to a quadratic number of random points. *arXiv preprint arXiv:2307.01181*, 2023.
- [12] Jun-Ting Hsieh, Pravesh K Kothari, Aaron Potechin, and Jeff Xu. Ellipsoid fitting up to a constant. *arXiv preprint arXiv:2307.05954*, 2023.
- [13] Madhur Tulsiani and June Wu. Ellipsoid fitting up to constant via empirical covariance estimation. *arXiv preprint arXiv:2307.10941*, 2023.

- [14] Philip W Anderson. Spin glass vi: Spin glass as cornucopia. *Physics Today*, 42(9):9–11, 1989.
- [15] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
- [16] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 814–827, 2016.
- [17] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi. Phase transitions in semidefinite relaxations. *Proceedings of the National Academy of Sciences*, 113(16):E2218–E2223, 2016.
- [18] Patrick Charbonneau, Enzo Marinari, Giorgio Parisi, Federico Ricci-Tersenghi, Gabriele Sicuro, Francesco Zamponi, and Marc Mezard. *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking after 40 Years*. World Scientific, 2023.
- [19] Sebastian Goldt, Bruno Loureiro, Galen Reeves, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. The gaussian equivalence of generative models for learning with shallow neural networks. In *Mathematical and Scientific Machine Learning*, pages 426–471. PMLR, 2022.
- [20] Hong Hu and Yue M Lu. Universality laws for high-dimensional learning with random features. *IEEE Transactions on Information Theory*, 69(3):1932–1964, 2022.
- [21] Andrea Montanari and Basil N Saeed. Universality of empirical risk minimization. In *Conference on Learning Theory*, pages 4310–4312. PMLR, 2022.
- [22] Federica Gerace, Florent Krzakala, Bruno Loureiro, Ludovic Stephan, and Lenka Zdeborová. Gaussian universality of linear classifiers with random labels in high-dimension. *arXiv preprint arXiv:2205.13303*, 2022.
- [23] Radoslaw Adamczak, Alexander E Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constructive Approximation*, 34:61–88, 2011.
- [24] Yatin Dandi, Ludovic Stephan, Florent Krzakala, Bruno Loureiro, and Lenka Zdeborová. Universality laws for gaussian mixtures in generalized linear models. *arXiv preprint arXiv:2302.08933*, 2023.
- [25] Bruno Loureiro, Cedric Gerbelot, Hugo Cui, Sebastian Goldt, Florent Krzakala, Marc Mezard, and Lenka Zdeborová. Learning curves of generic features maps for realistic datasets with a teacher-student model. *Advances in Neural Information Processing Systems*, 34:18137–18151, 2021.
- [26] Oussama Dhifallah and Yue M Lu. A precise performance analysis of learning with random features. *arXiv preprint arXiv:2008.11904*, 2020.
- [27] Dominik Schröder, Hugo Cui, Daniil Dmitriev, and Bruno Loureiro. Deterministic equivalent and error universality of deep random features learning. *arXiv preprint arXiv:2302.00401*, 2023.
- [28] Yehoram Gordon. On Milman’s inequality and random subspaces which escape through a mesh in \mathbb{R}^n . In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 1986–87*, pages 84–106. Springer, 1988.
- [29] Samet Oymak and Joel A Tropp. Universality laws for randomized dimension reduction, with applications. *Information and Inference: A Journal of the IMA*, 7(3):337–446, 2018.

- [30] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory*, pages 1683–1709. PMLR, 2015.
- [31] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. Precise error analysis of regularized m -estimators in high dimensions. *IEEE Transactions on Information Theory*, 64(8):5592–5628, 2018.
- [32] Ramon Van Handel. Probability in high dimension. *Lecture Notes (Princeton University)*, 2014.
- [33] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Cambridge university press, 2010.
- [34] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [35] Ryan O’Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
- [36] Michel Talagrand. The supremum of some canonical processes. *American Journal of Mathematics*, 116(2):283–325, 1994.
- [37] Paweł Hitczenko, Stephen J Montgomery-Smith, and Krzysztof Oleszkiewicz. Moment inequalities for sums of certain independent symmetric random variables. *Studia Math*, 123(1):15–42, 1997.
- [38] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

A A classical concentration inequality

We will make use of the following elementary concentration inequality, a generalization of Bernstein's inequality for ψ_q tails [23, 36, 37].

Lemma A.1 (Tail of sum of independent sub-Weibull random variables [23] –)

Let $q \in (0, 2]$, and W_1, \dots, W_n be i.i.d. centered random variables satisfying $\mathbb{P}[|W_1| \geq t] \leq 2e^{-Ct^q}$.

(i) If $q \in [1, 2]$, then for all $a \in \mathbb{R}^n$ and all $t \geq 0$:

$$\mathbb{P} \left[\left| \sum_{\mu=1}^n a_\mu W_\mu \right| \geq t \right] \leq 2 \exp \left\{ -c \min \left(\frac{t^2}{\|a\|_2^2}, \frac{t^q}{\|a\|_{q^*}^q} \right) \right\},$$

where $q^* \in [2, +\infty]$ with $1/q + 1/q^* = 1$.

(ii) If $q \in [1/2, 1]$, then for all $t > 0$

$$\mathbb{P} \left[\left| \frac{1}{n} \sum_{\mu=1}^n W_\mu \right| \geq t \right] \leq 2 \exp \left\{ -c_q \min(nt^2, (nt)^q) \right\}.$$

This lemma is stated in [23], see Lemmas 3.6 and 3.7 – and eq. (3.7) – and is a consequence of the same result for symmetric Weibull random variables [37].

B Fitting error of the sphere

We show here eq. (6). By Bernstein's inequality [34], we have for all $\mu \in [n]$ and $u \geq 0$:

$$\mathbb{P} \left[\left| \frac{\|x_\mu\|^2}{d} - 1 \right| \geq u \right] \leq 2 \exp \left(-Cd \min(u, u^2) \right).$$

As a consequence, if $X_\mu := \sqrt{d}(\|x_\mu\|^2/d - 1)$, then $\|X_\mu\|_{\psi_1} \leq C$. Let $Y_\mu := |X_\mu|^r - \mathbb{E}[|X_\mu|^r]$.

By the central limit theorem, $X_\mu \xrightarrow{d} \mathcal{N}(0, 2)$ as $d \rightarrow \infty$. One shows easily that $\sup_{d \geq 1} \mathbb{E}[|X_\mu|^{2+\varepsilon}] < \infty$ (e.g. for $\varepsilon = 2$), and thus (since $r \leq 2$) $|X_\mu|^r$ is uniformly integrable as $d \rightarrow \infty$. This implies (cf. Theorem 3.5 of [38]) that $\mathbb{E}[|X_\mu|^r] \rightarrow \mathbb{E}[|Z|^r]$ with $Z \sim \mathcal{N}(0, 2)$. Notice that $\mathbb{E}[|Z|^r] = 2^r \Gamma((r+1)/2) / \sqrt{\pi}$.

Let $q := 1/r \in [1/2, 1]$. Since $\|X_\mu\|_{\psi_1} \leq C$, we have $\| |X_\mu|^r \|_{\psi_q} \leq C$, and thus $\|Y_\mu\|_{\psi_q} \leq C'$. We can then use Lemma A.1, and we get:

$$\mathbb{P} \left[\left| \frac{1}{n} \sum_{\mu=1}^n (|X_\mu|^r - \mathbb{E}[|X_\mu|^r]) \right| \geq t \right] \leq 2 \exp \left\{ -c_r \min(nt^2, (nt)^{1/r}) \right\}.$$

Combining the above, we get that for any $\varepsilon > 0$, we have with probability $1 - o_d(1)$:

$$\mathbb{E}[|Z|^r] - \varepsilon \leq \frac{1}{n} \sum_{\mu=1}^n \left| \sqrt{d} \left[\frac{\|x_\mu\|^2}{d} - 1 \right] \right|^r \leq \mathbb{E}[|Z|^r] + \varepsilon.$$

C Towards exact ellipsoid fitting

We show here the following proposition.

Proposition C.1 (From approximate to exact ellipsoid fitting)

Let $n, d \rightarrow \infty$ with $n/d^2 \rightarrow \alpha \in (0, 1/2)$. Let $\{X_\mu\}_{\mu=1}^n$ be symmetric random matrices, and $H_{\mu\nu} := \text{Tr}[X_\mu X_\nu]$. Assume that:

- (i) With probability $1 - o_d(1)$, $\|H^{-1}\|_{\text{op}} \leq Cn^{-1/2}$ for some $C = C(\alpha) > 0$.
- (ii) There exists $\lambda_- > 0$ (depending only on α) such that

$$\text{p-lim}_{n \rightarrow \infty} \min_{S \succeq \lambda_- I_d} \frac{1}{\sqrt{n}} \sum_{\mu=1}^n |\text{Tr}(X_\mu S) - 1|^2 = 0. \quad (64)$$

Then, with probability $1 - o_d(1)$ there exists $S \succeq 0$ such that $\text{Tr}[X_\mu S] = 1$ for all $\mu \in [n]$.

Let us emphasize that given our current proof of Theorem 1.3 (cf. Section 2.4), if the assumptions of Proposition C.1 hold for $X_\mu \sim \text{Ellipse}(d)$ and $\alpha < 1/4$, then the first part of Conjecture 1.1 will hold.

However, the proof of Proposition C.1 is rather naive, as it crudely bounds the operator norm distance of the minimizer of eq. (64) to a subspace V (the affine subspace of solutions to the linear constraints $\text{Tr}[X_\mu S] = 1$) by its distance in Frobenius norm. For these reasons, the assumptions of Proposition C.1 may be far from being optimal.

Remark I – Note that the condition (i) is clearly satisfied if $X_\mu \stackrel{\text{i.i.d.}}{\sim} \text{GOE}(d)$ and $\alpha \in (0, 1/2)$, since H is then distributed as a Wishart matrix. On the other hand, while we expect it to hold as well for $X_\mu \stackrel{\text{i.i.d.}}{\sim} \text{Ellipse}(d)$, this condition is (to the best of our knowledge) not known unless α is small enough: interestingly, this was one of the limitations in the recent works [11–13] that proved that ellipsoid fitting is feasible for α sufficiently small.

Remark II – Note that it is sufficient for eq. (64) to hold that there exists $r \in [1, 2]$ such that:

$$\text{p-lim}_{n \rightarrow \infty} \min_{S \succeq \lambda_- I_d} \frac{1}{n^{r/4}} \sum_{\mu=1}^n |\text{Tr}(X_\mu S) - 1|^r = 0.$$

At the moment, our proof of Theorem 1.3 (cf. Lemma 2.5) only implies (for $r < 4/3$) a similar statement with a prefactor $1/n$ rather than the required $1/n^{r/4}$. It would thus need to be improved to show that eq. (64) holds for the ellipsoid fitting setting.

Proof of Proposition C.1 – Let $V := \{S \in \mathcal{S}_d : \text{Tr}[X_\mu S] = 1, \forall \mu \in [n]\}$ be the affine space of solutions to the constraints. Let $\varepsilon > 0$, and $\hat{S} \succeq \lambda_- I_d$ such that

$$\frac{1}{\sqrt{n}} \sum_{\mu=1}^n |\text{Tr}(X_\mu \hat{S}) - 1|^2 \leq \varepsilon.$$

Note that for all $M \in \mathcal{S}_d$, if $\|M\|_{\text{op}} \leq \lambda_-$, then $\hat{S} + M \succeq 0$. In particular, $\|M\|_{\text{F}} \leq \lambda_- \Rightarrow \hat{S} + M \succeq 0$. In order to conclude it thus suffices to show that $d_{\text{F}}(\hat{S}, V) \leq \lambda_-$. The following lemma is an elementary geometrical result:

Lemma C.2 (Euclidean distance to an affine subspace –)

Let $d \geq 1$ and $1 \leq r \leq d$ two integers. Let $(a_k, b_k)_{k=1}^r \in (\mathbb{R}^d \times \mathbb{R})^r$, with $(a_k)_{k=1}^r$ linearly independent. We define $G := \{x \in \mathbb{R}^d : a_k^\top x + b_k = 0, \forall k \in [r]\}$. Then, for any $y \in \mathbb{R}^d$:

$$d_2(y, G)^2 = v^\top H^{-1} v,$$

in which $v_k := a_k^\top y + b_k$, and $H_{kk'} = \langle a_k, a_{k'} \rangle$ is the Gram matrix of the $\{a_k\}$.

We now condition on the event of condition (i). For any $\varepsilon > 0$, applying Lemma C.2 yields (with probability $1 - o_d(1)$):

$$d_{\mathbb{F}}(\hat{S}, V)^2 \leq C(\alpha)\varepsilon,$$

so that picking $\varepsilon \leq \lambda_-^2/C(\alpha)$ implies the result. \square

D Additional proofs for Proposition 2.1

D.1 Proof of Lemma 4.2

The lemma is a direct corollary of the following elementary result.

Proposition D.1 (Bounding a Gaussian process on the sphere)

Let $n, p \geq 1$. Let $G \in \mathbb{R}^{n \times p}$ with $G_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. There is $A > 0$ such that for all $\delta > 0$ and $r \geq 1$:

$$\mathbb{P}\left[\max_{\|x\|_2=1} \|Gx\|_r \geq [A(\sqrt{n} + \sqrt{p}) + \delta\sqrt{n}] n^{\max(1/r-1/2, 0)}\right] \leq \exp\left\{-\frac{n\delta^2}{2}\right\}.$$

Proof of Proposition D.1 – Notice that $G \rightarrow \max_{\|x\|_2=1} \|Gx\|_r$ is Lipschitz:

$$\begin{aligned} \left| \max_{\|x\|_2=1} \|G_1x\|_r - \max_{\|x\|_2=1} \|G_2x\|_r \right| &\leq \max_{\|x\|_2=1} \|(G_1 - G_2)x\|_r, \\ &\stackrel{(a)}{\leq} n^{\max(\frac{1}{r}-\frac{1}{2}, 0)} \max_{\|x\|_2=1} \|(G_1 - G_2)x\|_2, \\ &\leq n^{\max(\frac{1}{r}-\frac{1}{2}, 0)} \|G_1 - G_2\|_{\mathbb{F}}, \end{aligned}$$

where we used Hölder's inequality in the form $\|y\|_r \leq n^{\frac{1}{r}-\frac{1}{2}}\|y\|_2$ for $r \leq 2$, and for $r \geq 2$ the fact that $\|y\|_r \leq \|y\|_2$, alongside with $\|G\|_{\text{op}} \leq \|G\|_{\mathbb{F}}$. By Theorem 3.3 we reach:

$$\mathbb{P}\left[\max_{\|x\|_2=1} \|Gx\|_r \geq \mathbb{E} \max_{\|x\|_2=1} \|Gx\|_r + \delta n^{\max(1/r, 1/2)}\right] \leq \exp\left\{-\frac{n\delta^2}{2}\right\}.$$

The proof is complete if we can show that $\mathbb{E} \max_{\|x\|_2=1} \|Gx\|_r = \mathcal{O}(n^{\max(1/r-1/2, 0)} \cdot (\sqrt{n} + \sqrt{p}))$. This follows by the same inequality as above:

$$\mathbb{E} \max_{\|x\|_2=1} \|Gx\|_r \leq n^{\max(\frac{1}{r}-\frac{1}{2}, 0)} \mathbb{E} \max_{\|x\|_2=1} \|Gx\|_2.$$

The bound $\mathbb{E} \max_{\|x\|_2=1} \|Gx\|_2 = \mathcal{O}(\sqrt{n} + \sqrt{p})$ is well-known, see e.g. [34]. \square

D.2 Proof of Theorem 4.5

As we have seen, it suffices to prove Lemmas 4.6 and 4.7. We introduce some additional notations.

- For any $\mu \in [n]$, we denote

$$\langle \cdot \rangle_{\mu} := \frac{\int P_0(dS) e^{-\sum_{\nu(\neq\mu)} \phi[\text{Tr}(U_{\nu}(t)S)]}(\cdot)}{\int P_0(dS) e^{-\sum_{\nu(\neq\mu)} \phi[\text{Tr}(U_{\nu}(t)S)]}}. \quad (65)$$

We do not write explicitly the t dependency of this average as it will be clear from context.

- For any μ , we denote $\mathbb{E}_{(\mu)}$ the expectation conditioned on $\{G_{\mu}, W_{\mu}\}$, i.e. the expectation over $\{G_{\nu}, W_{\nu}\}_{\nu(\neq\mu)}$. Note that this notation is different from [21], for which $\mathbb{E}_{(\mu)}$ was the expectation with respect to (G_{μ}, W_{μ}) . On the other hand, we denote without parenthesis the expectation with respect to these variables: e.g. $\mathbb{E}_{(1)}$ is conditioned on $\{G_1, W_1\}$, but \mathbb{E}_{G_1, W_1} is the expectation w.r.t. G_1, W_1 .

D.2.1 Proof of Lemma 4.6

We fix $t \in (0, \pi/2)$, and we start from eq. (58), which we can rewrite using eq. (65) as:

$$\mathbb{E} \frac{\partial \psi[F_d(U(t))]}{\partial t} = -\frac{n}{d^2} \mathbb{E} \left[\psi'[F_d(U(t))] \frac{\left\langle e^{-\phi(\text{Tr}[U_1(t)S])} \left(\text{Tr}[S\tilde{U}_1(t)] \phi'(\text{Tr}[U_1(t)S]) \right) \right\rangle_1}{\left\langle e^{-\phi(\text{Tr}[U_1(t)S])} \right\rangle_1} \right].$$

Since $\|\psi'\|_\infty < \infty$ and $n/d^2 \leq \alpha_2$, using the triangular inequality the proof of Lemma 4.6 is complete if one can show the following bound, which will also be useful afterwards:

Lemma D.2

There exists a universal constant $C > 0$ such that:

$$\sup_{d \geq 1} \sup_{t \in (0, \pi/2)} \sup_{\{W_\mu, G_\mu\}_{\mu=2}^n} \mathbb{E}_{W_1, G_1} \frac{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \left| \text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right| \right\rangle_1}{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1} \leq C.$$

Proof of Lemma D.2 – Note that

$$\begin{aligned} \mathbb{E}_{W_1, G_1} \frac{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \left| \text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right| \right\rangle_1}{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1} &\leq \|\phi'\|_\infty \mathbb{E}_{W_1, G_1} \left[\frac{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \left| \text{Tr}[S\tilde{U}_1(t)] \right| \right\rangle_1}{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1} \right], \\ &\leq e^{\|\phi\|_\infty} \|\phi'\|_\infty \mathbb{E}_{W_1, G_1} \left[\left\langle \left| \text{Tr}[S\tilde{U}_1(t)] \right| \right\rangle_1 \right], \end{aligned} \quad (66)$$

in which we used the positivity and boundedness of ϕ in the last inequality. To control the last term in eq. (66) we write:

$$\begin{aligned} \mathbb{E}_{G_1, W_1} \left[\left\langle \left| \text{Tr}[S\tilde{U}_1(t)] \right| \right\rangle_1 \right] &\stackrel{(a)}{=} \left\langle \mathbb{E}_{G_1, W_1} \left| \text{Tr}[S\tilde{U}_1(t)] \right| \right\rangle_1, \\ &\stackrel{(b)}{\leq} \left\langle \left\{ \mathbb{E}_{G_1, W_1} \left(\text{Tr}[S\tilde{U}_1(t)]^2 \right) \right\}^{1/2} \right\rangle_1, \\ &\stackrel{(c)}{\leq} \sqrt{2} \left\langle \left\{ \frac{1}{d} \text{Tr}[S^2] \right\}^{1/2} \right\rangle_1, \\ &\stackrel{(d)}{\leq} \sqrt{2} C_0, \end{aligned} \quad (67)$$

where (a) uses that $\langle \cdot \rangle_1$ is independent of $\{W_1, G_1\}$, (b) is from the Cauchy-Schwarz inequality, in (c) we use the hypothesis on the first two moments of ρ matching the ones of $\text{GOE}(d)$, and in (d) that $\text{supp}(P_0) \subseteq B_2(C\sqrt{d})$. \square

D.2.2 Proof of Lemma 4.7

We fix $t \in (0, \pi/2)$ for the rest of the proof, and we write U, \tilde{U} for $U(t), \tilde{U}(t)$. We follow the ideas of Appendix A.3 of [21], and start again from eq. (58):

$$\begin{aligned} \left| \mathbb{E} \frac{\partial \psi[F_d(U(t))]}{\partial t} \right| &\leq \alpha_2 (I_1 + I_2), \\ \begin{cases} I_1 &= \left| \mathbb{E} \left[\left\{ \psi'[F_d(U)] - \psi'[F_d(U^{(1)})] \right\} \frac{\int P_0(dS) e^{-\sum_\nu \phi(\text{Tr}[U_\nu S])} \left(\text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\int P_0(dS) e^{-\sum_\nu \phi(\text{Tr}[U_\nu S])}} \right] \right|, \\ I_2 &= \left| \mathbb{E} \left[\psi'[F_d(U^{(1)})] \frac{\int P_0(dS) e^{-\sum_\nu \phi(\text{Tr}[U_\nu S])} \left(\text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\int P_0(dS) e^{-\sum_\nu \phi(\text{Tr}[U_\nu S])}} \right] \right|, \end{cases} \end{aligned} \quad (69)$$

with $U^{(\mu)}$ obtained from U by setting $U_\mu = 0$. We show successively $I_1 \rightarrow 0$ and $I_2 \rightarrow 0$. Since ψ' is assumed to be Lipschitz, we have:

$$\begin{aligned} |\psi'[F_d(U)] - \psi'[F_d(U^{(1)})]| &\leq \frac{\|\psi'\|_{\text{Lip}}}{d^2} \left| \log \frac{\int P_0(dS) e^{-\sum_{\nu=1}^n \phi(\text{Tr}[U_\nu S])}}{\int P_0(dS) e^{-\sum_{\nu=2}^n \phi(\text{Tr}[U_\nu S])}} \right| \\ &\leq \frac{\|\psi'\|_{\text{Lip}}}{d^2} \left| \log \left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1 \right| \\ &\leq -\frac{\|\psi'\|_{\text{Lip}}}{d^2} \log \left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1 \\ &\leq \frac{\|\psi'\|_{\text{Lip}} \|\phi\|_\infty}{d^2}. \end{aligned}$$

Therefore,

$$I_1 \leq \mathcal{O}_d\left(\frac{1}{d^2}\right) \times \mathbb{E} \left| \frac{\int P_0(dS) e^{-\sum_\nu \phi(\text{Tr}[U_\nu S])} \left(\text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\int P_0(dS) e^{-\sum_\nu \phi(\text{Tr}[U_\nu S])}} \right|. \quad (70)$$

The second term in eq. (70) is bounded by Lemma D.2, uniformly in t . Therefore, we reach that $I_1 \rightarrow 0$ as $d \rightarrow \infty$.

We now tackle I_2 . Note that since $U^{(1)}$ is independent of U_1 , we can rewrite it as:

$$\begin{aligned} I_2 &= \left| \mathbb{E}_{(1)} \left[\psi'[F_d(U^{(1)})] \mathbb{E}_{G_1, W_1} \left[\frac{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) \right\rangle_1}{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1} \right] \right] \right| \\ &\leq \|\psi'\|_\infty \mathbb{E}_{(1)} \left| \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1} \right] \right\rangle_1 \right|. \quad (71) \end{aligned}$$

We focus on bounding the right-hand side of eq. (71). We will show the following lemma:

Lemma D.3

Uniformly over all $t \in [0, \pi/2]$ and $\{W_\mu, G_\mu\}_{\mu=2}^n$, and under the hypotheses of Theorem 4.5:

$$\lim_{d \rightarrow \infty} \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S\tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\left\langle e^{-\phi(\text{Tr}[U_1 S])} \right\rangle_1} \right] \right\rangle_1 = 0. \quad (72)$$

One directly concludes that $I_2 \rightarrow 0$ from using the dominated convergence theorem in eq. (71) (the pointwise limit is given by Lemma D.3 and the domination hypothesis by Lemma D.2). This ends the proof of Lemma 4.7.

We thus focus on the proof of Lemma D.3. Following [21], the sketch of the proof is the following:

- (i) Show that the denominator appearing in eq. (72) can be moved to the numerator by using the expansion of $1/x$ in power series around 1. This transforms the quantity to control to a sum of terms of the type $\langle \mathbb{E}_{G_1, W_1} [f(S_1, \dots, S_k)] \rangle_1$, with S_1, \dots, S_k independent samples under $\langle \cdot \rangle_1$.
- (ii) Extend the one-dimensional CLT of Definition 4.1 to k -dimensional projections of G and W (with $k = \mathcal{O}_d(1)$), and to square-integrable locally-Lipschitz functions. This allows to apply it to the terms appearing in (i), and write (uniformly in S_1, \dots, S_k) that $\mathbb{E}_{G_1, W_1} [f(S_1, \dots, S_k)] \simeq \mathbb{E}_{G_1, \tilde{G}_1} [f(S_1, \dots, S_k)]$, with \tilde{G}_1 an independent GOE(d) matrix.
- (iii) For the case of Gaussian matrices, as explained above we have \tilde{U}_1 independent of U_1 . Using the form of the function f that appears in eq. (72) this implies that $\mathbb{E}_{G_1, \tilde{G}_1} [f(S_1, \dots, S_k)] = 0$ and concludes the proof.

Let us perform this strategy in detail. The following lemma is proven in Section D.3.

Lemma D.4 (Polynomial approximation to the fraction –)

For all $\delta > 0$, there exists a real polynomial Q (depending only on δ) such that for all $d \geq 1$, all $t \in (0, \pi/2)$ and all $\{W_\mu, G_\mu\}_{\mu=2}^n$:

$$\begin{aligned} & \left| \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1} \right] \right\rangle_1 \right| \\ & \leq \left| \mathbb{E}_{G_1, W_1} \left\{ \langle e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) \rangle_1 Q \left(\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1 \right) \right\} \right| + \delta. \end{aligned}$$

We fix $\delta > 0$, and denote $Q(X) = \sum_{k=0}^K a_k X^k$ the polynomial of Lemma D.4. Therefore, uniformly in d, t , and $\{W_\mu, G_\mu\}$:

$$\begin{aligned} & \left| \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1} \right] \right\rangle_1 \right| \\ & \leq \sum_{k=0}^K |a_k| \left| \left\langle \mathbb{E}_{G_1, W_1} \left\{ e^{-\sum_{a=0}^k \phi(\text{Tr}[U_1 S_a])} \left(\text{Tr}[S_0 \tilde{U}_1] \phi'(\text{Tr}[U_1 S_0]) \right) \right\} \right\rangle_1 \right| + \delta, \end{aligned} \quad (73)$$

with $\{S_a\}_{a=0}^k$ i.i.d. samples from $\langle \cdot \rangle_1$. We then extend the one-dimensional CLT of Definition 4.1 to finite-dimensional projections, similarly to Lemmas 29 and 30 of [21]. The proof of this lemma is deferred to Section D.3.

Lemma D.5 (Extension of the CLT to finite-dimensional projections –)

Let $R \in \mathbb{N}^*$, and $\tilde{G} \sim \text{GOE}(d)$, independent of everything else. Let $\varphi : \mathbb{R}^{2R} \rightarrow \mathbb{R}$ a locally Lipschitz function such that for both $X \in \{W, \tilde{G}\}$:

$$\sup_{d \geq 1} \sup_{\{W_\mu, G_\mu\}_{\mu=2}^n} \sup_{t \in (0, \pi/2)} \left\langle \mathbb{E}_{X, G} \left[\varphi \left(\left\{ \text{Tr}(X S_a) \right\}_{a=1}^R, \left\{ \text{Tr}(G S_a) \right\}_{a=1}^R \right)^2 \right] \right\rangle_1 < \infty. \quad (74)$$

It is understood there that $\{S_a\} \stackrel{\text{i.i.d.}}{\sim} \langle \cdot \rangle_1$. Then

$$\lim_{d \rightarrow \infty} \sup_{\{W_\mu, G_\mu\}_{\mu=2}^n} \sup_{t \in (0, \pi/2)} \left\langle \left| \mathbb{E}_{W, G} \varphi(\{\text{Tr}(W S_a)\}, \{\text{Tr}(G S_a)\}) - \mathbb{E}_{\tilde{G}, G} \varphi(\{\text{Tr}(\tilde{G} S_a)\}, \{\text{Tr}(G S_a)\}) \right| \right\rangle_1 = 0.$$

We wish to apply Lemma D.5 to eq. (73), i.e. to

$$\varphi(\{\text{Tr}(W_1 S_a)\}, \{\text{Tr}(G_1 S_a)\}) := e^{-\sum_{a=0}^k \phi(\text{Tr}[U_1 S_a])} \left(\text{Tr}[S_0 \tilde{U}_1] \phi'(\text{Tr}[U_1 S_0]) \right).$$

φ is locally Lipschitz by our hypotheses on ϕ . Moreover, note that for $X \in \{W, \tilde{G}\}$, and $\{S^a\} \in \text{supp}(P_0)$ (using that W has the same two first moments as \tilde{G}):

$$\mathbb{E}_{X, G} \left[\varphi \left(\left\{ \text{Tr}(X S_a) \right\}, \left\{ \text{Tr}(G S_a) \right\} \right)^2 \right] \leq 2 \|\phi'\|_\infty^2 \text{Tr}[S_0^2]/d \leq 2C_0^2 \|\phi'\|_\infty^2 < \infty.$$

This allows to apply Lemma D.5 in eq. (73), and to reach that, uniformly in $\{W_\mu, G_\mu\}_{\mu=2}^n$ and $t \in (0, \pi/2)$ we have:

$$\limsup_{d \rightarrow \infty} \left| \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1} \right] \right\rangle_1 \right|$$

$$\begin{aligned}
&\leq \delta + \sum_{k=0}^K |a_k| \limsup_{d \rightarrow \infty} \left\langle \left| \mathbb{E}_{G_1, W_1} \varphi(\{\text{Tr}(W_1 S_a)\}, \{\text{Tr}(G_1 S_a)\}) \right| \right\rangle_1, \\
&\leq \delta + \sum_{k=0}^K |a_k| \limsup_{d \rightarrow \infty} \left\langle \left| \mathbb{E}_{G_1, \tilde{G}_1} \varphi(\{\text{Tr}(\tilde{G}_1 S_a)\}, \{\text{Tr}(G_1 S_a)\}) \right| \right\rangle_1 \\
&+ \sum_{k=0}^K |a_k| \limsup_{d \rightarrow \infty} \left\langle \left| \mathbb{E}_{G_1, W_1} \varphi(\{\text{Tr}(W_1 S_a)\}, \{\text{Tr}(G_1 S_a)\}) - \mathbb{E}_{G_1, \tilde{G}_1} \varphi(\{\text{Tr}(\tilde{G}_1 S_a)\}, \{\text{Tr}(G_1 S_a)\}) \right| \right\rangle_1, \\
&\stackrel{(a)}{\leq} \delta + \sum_{k=0}^K |a_k| \limsup_{d \rightarrow \infty} \left| \left\langle \mathbb{E}_{G_1, \tilde{G}_1} \left\{ e^{-\sum_{a=0}^k \phi(\text{Tr}(V_1 S_a))} \left(\text{Tr}[S_0 \tilde{V}_1] \phi'(\text{Tr}[V_1 S_0]) \right) \right\} \right\rangle_1 \right|,
\end{aligned}$$

where we used Lemma D.5 in (a), and with $V_1 = \cos(t)\tilde{G}_1 + \sin(t)G_1$, and $\tilde{V}_1 = -\sin(t)\tilde{G}_1 + \cos(t)G_1$. Since G_1, \tilde{G}_1 are gaussians, so are V_1 and \tilde{V}_1 , and one verifies easily that they are independent since their covariance is zero. Therefore, we have:

$$\begin{aligned}
&\mathbb{E}_{G_1, \tilde{G}_1} \left\{ e^{-\sum_{a=0}^k \phi(\text{Tr}(V_1 S_a))} \left(\text{Tr}[S_0 \tilde{V}_1] \phi'(\text{Tr}[V_1 S_0]) \right) \right\} \\
&= \mathbb{E}_{V_1} \left\{ e^{-\sum_{a=0}^k \phi(\text{Tr}(V_1 S_a))} \underbrace{\left(\mathbb{E}_{\tilde{V}_1} \text{Tr}[S_0 \tilde{V}_1] \right)}_{=0} \phi'(\text{Tr}[V_1 S_0]) \right\} = 0.
\end{aligned}$$

Thus, we reach, for any $\delta > 0$:

$$\limsup_{d \rightarrow \infty} \sup_{t \in (0, \pi/2)} \sup_{\{W_\mu, G_\mu\}_{\mu=2}^n} \left| \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1} \right] \right\rangle_1 \right| \leq \delta.$$

Letting $\delta \rightarrow 0$ finishes the proof of Lemma D.3. \square

D.3 Additional proofs

D.3.1 Proof of Lemma D.4

We use the power expansion of $x \mapsto 1/x$ around $x = 1$, defining, for $M \geq 1$:

$$Q_M(x) := \sum_{k=0}^M (1-x)^k, \quad R_M(x) := \frac{1}{x} - Q_M(x).$$

We make use of Lemma 27 of [21]:

Lemma D.6 ([21])

For any integer $M \geq 1$ we have

- For all $x \neq 0$, $R_M(x) = (1-x)^{M+1}/x$.
- $x \mapsto R_M(x)^2$ is convex on $(0, 1]$.
- For any $s \in (0, 1)$ and $\eta > 0$, there exists $M \geq 1$ such that $\sup_{t \in [s, 1]} |R_M(t)| \leq \eta$.

Since $e^{-\|\phi\|_\infty} \leq e^{-\phi} \leq 1$, we have that for all $\eta > 0$ there exists $M_\eta \geq 1$ such that for all $M \geq M_\eta$, all matrices $\{W_\mu, G_\mu\}_{\mu=1}^n$, all $t \in (0, \pi/2)$:

$$\left| R_M \left(\langle e^{-\phi(\text{Tr}[S U_1])} \rangle_1 \right) \right| \leq \eta. \tag{75}$$

Thus, since $1/x = Q_M(x) + R_M(x)$:

$$\begin{aligned}
& \left| \left\langle \mathbb{E}_{G_1, W_1} \left[\frac{e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right)}{\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1} \right] \right\rangle_1 \right| \\
& \leq \left| \left\langle \mathbb{E}_{G_1, W_1} \left[e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) Q_M \left(\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1 \right) \right] \right\rangle_1 \right| \\
& \quad + \left| \left\langle \mathbb{E}_{G_1, W_1} \left[e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) R_M \left(\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1 \right) \right] \right\rangle_1 \right|, \\
& \stackrel{(a)}{\leq} \left| \left\langle \mathbb{E}_{G_1, W_1} \left[e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) Q_M \left(\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1 \right) \right] \right\rangle_1 \right| \\
& \quad + \eta \left\langle \mathbb{E}_{G_1, W_1} \left[e^{-\phi(\text{Tr}[U_1 S])} \left| \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) \right| \right] \right\rangle_1, \\
& \leq \left| \left\langle \mathbb{E}_{G_1, W_1} \left[e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) Q_M \left(\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1 \right) \right] \right\rangle_1 \right| + \eta \|\phi'\|_\infty \langle \mathbb{E}_{G_1, W_1} |\text{Tr}[S \tilde{U}_1]| \rangle_1, \\
& \stackrel{(b)}{\leq} \left| \left\langle \mathbb{E}_{G_1, W_1} \left[e^{-\phi(\text{Tr}[U_1 S])} \left(\text{Tr}[S \tilde{U}_1] \phi'(\text{Tr}[U_1 S]) \right) Q_M \left(\langle e^{-\phi(\text{Tr}[U_1 S])} \rangle_1 \right) \right] \right\rangle_1 \right| + C\eta \|\phi'\|_\infty,
\end{aligned}$$

using eq. (75) in (a), and the Cauchy-Schwarz inequality (cf. the proof of Lemma D.2) in (b). We emphasize that this bound is uniform in t and $\{W_\mu, G_\mu\}$. Choosing $\eta = \delta/(C\|\phi'\|_\infty)$ ends the proof of Lemma D.4.

D.4 Proof of Lemma D.5

We first prove that the conclusion of Lemma D.5 holds uniformly over all $S_1, \dots, S_R \in A_d$ when the function φ is assumed to be continuous with compact support:

Lemma D.7

Let $R \in \mathbb{N}^*$, and $\tilde{G} \sim \text{GOE}(d)$, independent of everything else. Let $\varphi : \mathbb{R}^{2R} \rightarrow \mathbb{R}$ be Lipschitz and compactly supported. Recall that $W \sim \rho$, a measure who is assumed to satisfy a one-dimensional CLT with respect to a set $A_d \subseteq S_d$, see Definition 4.1. We assume that A_d is convex and symmetric. Then:

$$\lim_{d \rightarrow \infty} \sup_{S_1, \dots, S_R \in A_d} \left| \mathbb{E}_{W, G} \varphi \left(\{\text{Tr}(W S_a)\}, \{\text{Tr}(G S_a)\} \right) - \mathbb{E}_{\tilde{G}, G} \varphi \left(\{\text{Tr}(\tilde{G} S_a)\}, \{\text{Tr}(G S_a)\} \right) \right| = 0.$$

Proof of Lemma D.7 – Recall that we use the matrix flattening function of eq. (25). Let us denote:

$$\begin{cases} H & := \begin{pmatrix} \text{vec}(S^1) & 0 & \text{vec}(S_2) & 0 & \cdots & \text{vec}(S^R) & 0 \\ 0 & \text{vec}(S^1) & 0 & \text{vec}(S_2) & \cdots & 0 & \text{vec}(S^R) \end{pmatrix} \in \mathbb{R}^{d(d+1) \times 2R}, \\ v & := (\text{vec}(W)^\top, \text{vec}(G)^\top)^\top \in \mathbb{R}^{d(d+1)}, \\ h & := (\text{vec}(\tilde{G})^\top, \text{vec}(G)^\top)^\top \in \mathbb{R}^{d(d+1)}. \end{cases} \quad (76)$$

Using these notations, we have:

$$\begin{cases} \left(\{\text{Tr}(W S_a)\}_{a=1}^R, \{\text{Tr}(G S_a)\}_{a=1}^R \right) & = H^\top v \in \mathbb{R}^{2R}, \\ \left(\{\text{Tr}(\tilde{G} S_a)\}_{a=1}^R, \{\text{Tr}(G S_a)\}_{a=1}^R \right) & = H^\top h \in \mathbb{R}^{2R}. \end{cases} \quad (77)$$

We add a small Gaussian noise to help us deal with characteristic functions later on. Let $\delta > 0$, and $Z \sim \mathcal{N}(0, \delta^2 \mathbf{I}_{2R})$. For all S_1, \dots, S_R we have:

$$\begin{aligned}
|\mathbb{E} \varphi(H^\top v) - \mathbb{E} \varphi(H^\top h)| & \leq |\mathbb{E} \varphi(H^\top v) - \mathbb{E} \varphi(H^\top v + Z)| + |\mathbb{E} \varphi(H^\top h) - \mathbb{E} \varphi(H^\top h + Z)| \\
& \quad + |\mathbb{E} \varphi(H^\top v + Z) - \mathbb{E} \varphi(H^\top h + Z)|,
\end{aligned}$$

$$\begin{aligned}
&\leq 2\|\varphi\|_L \mathbb{E}\|Z\|_2 + |\mathbb{E}\varphi(H^\top v + Z) - \mathbb{E}\varphi(H^\top h + Z)|, \\
&\leq CR^{1/2}\|\varphi\|_L \delta + |\mathbb{E}\varphi(H^\top v + Z) - \mathbb{E}\varphi(H^\top h + Z)|.
\end{aligned} \tag{78}$$

We now control the last term of eq. (78). For $X \in \mathbb{R}^{2R}$ a random variable, we define its characteristic function as $\phi_X(u) := \mathbb{E}e^{-iu^\top X}$. We have then

$$\begin{aligned}
\frac{1}{(2\pi)^{2R}} \int_{\mathbb{R}^{2R}} \varphi(t) \int_{\mathbb{R}^{2R}} e^{it^\top u - \frac{\delta^2}{2}\|u\|^2} \phi_X(u) du dt &= \frac{1}{(2\pi\delta^2)^R} \mathbb{E}_X \int_{\mathbb{R}^{2R}} \varphi(t) e^{-\frac{\|t-X\|^2}{2\delta^2}}, \\
&= \mathbb{E}\varphi(X + Z).
\end{aligned}$$

Coming back to eq. (78) we get:

$$\begin{aligned}
|\mathbb{E}\varphi(H^\top v + Z) - \mathbb{E}\varphi(H^\top h + Z)| &\leq \frac{1}{(2\pi)^{2R}} \int_{\mathbb{R}^{2R}} |\varphi(t)| \left| \int_{\mathbb{R}^{2R}} e^{it^\top u - \frac{\delta^2}{2}\|u\|^2} [\phi_{H^\top v}(u) - \phi_{H^\top h}(u)] du \right| dt, \\
&\leq \frac{\|\varphi\|_{L_1}}{(2\pi)^{2R}} \int_{\mathbb{R}^{2R}} e^{-\frac{\delta^2}{2}\|u\|^2} |\phi_{H^\top v}(u) - \phi_{H^\top h}(u)| du.
\end{aligned} \tag{79}$$

Here $\|\varphi\|_{L_1} := \int |\varphi(t)| dt$. We will show that for any $u \in \mathbb{R}^{2R}$:

$$\lim_{d \rightarrow \infty} \sup_{S_1, \dots, S_R \in A_d} |\phi_{H^\top v}(u) - \phi_{H^\top h}(u)| = 0. \tag{80}$$

Combining eq. (80) with the dominated convergence theorem applied in eq. (79), we get:

$$\lim_{d \rightarrow \infty} \sup_{S_1, \dots, S_R \in A_d} |\mathbb{E}\varphi(H^\top v + Z) - \mathbb{E}\varphi(H^\top h + Z)| = 0.$$

Plugging this back into eq. (78), we get that for any $\delta > 0$:

$$\limsup_{d \rightarrow \infty} \sup_{S_1, \dots, S_R \in A_d} |\mathbb{E}\varphi(H^\top v) - \mathbb{E}\varphi(H^\top h)| \leq CR^{1/2}\|\varphi\|_L \delta.$$

Letting $\delta \rightarrow 0$ ends the proof of Lemma D.7. There remains to prove eq. (80). Let $u = (u^{(1)}, u^{(2)}) \in \mathbb{R}^{2R}$, with $u^{(1)}, u^{(2)} \in \mathbb{R}^R$, and let us fix $S_1, \dots, S_R \in A_d$. We have

$$\begin{aligned}
&|\phi_{H^\top v}(u) - \phi_{H^\top h}(u)| \\
&= \left| \mathbb{E} e^{-i \sum_{a=1}^r u_a^{(1)} \text{Tr}[W S_a] - i \sum_{a=1}^r u_a^{(2)} \text{Tr}[G S_a]} - \mathbb{E} e^{-i \sum_{a=1}^r u_a^{(1)} \text{Tr}[\tilde{G} S_a] - i \sum_{a=1}^r u_a^{(2)} \text{Tr}[G S_a]} \right|, \\
&\stackrel{(a)}{=} \left| \mathbb{E} \exp \left\{ -i \text{Tr} \left[W \sum_{a=1}^r u_a^{(1)} S_a \right] \right\} - \mathbb{E} \exp \left\{ -i \text{Tr} \left[\tilde{G} \sum_{a=1}^r u_a^{(1)} S_a \right] \right\} \right|,
\end{aligned} \tag{81}$$

using in (a) that G is independent of W, \tilde{G} . We can assume that $u^{(1)} \neq 0$, otherwise the result of eq. (80) is clear. Since A_d is symmetric and convex we have

$$\hat{S} := \frac{1}{\|u^{(1)}\|_1} \sum_{a=1}^r u_a^{(1)} S_a \in A_d.$$

Therefore, letting $\varphi_u(x) := e^{-i\|u\|_1 x}$, we have by eq. (81):

$$|\phi_{H^\top v}(u) - \phi_{H^\top h}(u)| = |\mathbb{E}\varphi_u(\text{Tr}[\hat{S}W]) - \mathbb{E}\varphi_u(\text{Tr}[\hat{S}\tilde{G}])| \tag{82}$$

Moreover,

$$|\varphi_u(x) - \varphi_u(y)| = \|u\|_1 \left| \int_0^{y-x} e^{i\|u\|_1 t} dt \right| \leq \|u\|_1 |y - x|,$$

so that $\|\varphi_u\|_L \leq \|u\|_1$. We can then therefore apply the one-dimensional CLT of Definition 4.1 in eq. (82), and we get that

$$\sup_{S_1, \dots, S_R \in A_d} |\phi_{H^\top v}(u) - \phi_{H^\top h}(u)| \leq \sup_{S \in A_d} |\mathbb{E} \varphi_u(\text{Tr}[\hat{S}W]) - \mathbb{E} \varphi_u(\text{Tr}[\hat{S}\tilde{G}])| \xrightarrow{d \rightarrow \infty} 0.$$

This ends the proof of eq. (80). \square

We then deduce the full statement of Lemma D.5 by a truncation argument.

End of the proof of Lemma D.5 – φ is now only assumed to be locally Lipschitz and square integrable, in the sense of eq. (74). We take the same notations as in the proof of Lemma D.7, see in particular eq. (76), so that

$$\mathbb{E}_{W, G} \varphi(\{\text{Tr}(WS_a)\}, \{\text{Tr}(GS_a)\}) - \mathbb{E}_{\tilde{G}, G} \varphi(\{\text{Tr}(\tilde{G}S_a)\}, \{\text{Tr}(GS_a)\}) = \mathbb{E} \varphi(H^\top v) - \mathbb{E} \varphi(H^\top h).$$

Let $B > 0$, and let us denote $u_B : \mathbb{R}_+ \rightarrow [0, 1]$ a \mathcal{C}^∞ function such that $u_B(x) = 1$ if $x \leq B$ and $u_B(x) = 0$ if $x \geq B + 1$. We denote $\varphi_B(z) := \varphi(z)u_B(\|z\|)$. One can then check easily that φ_B is Lipschitz (because φ is locally-Lipschitz), and compactly supported. Moreover, we have:

$$\begin{aligned} & \langle \mathbb{E} \varphi(H^\top v) - \mathbb{E} \varphi(H^\top h) \rangle_1 \\ & \leq \langle \mathbb{E} \varphi_B(H^\top v) - \mathbb{E} \varphi_B(H^\top h) \rangle_1 + \sum_{z \in \{h, v\}} \langle \mathbb{E} |\varphi(H^\top z)|(1 - u_B(\|H^\top z\|)) \rangle_1. \end{aligned} \quad (83)$$

We now control the different terms in eq. (83) successively. Notice that

$$\langle \mathbb{E} \varphi_B(H^\top v) - \mathbb{E} \varphi_B(H^\top h) \rangle_1 \leq \sup_{S_1, \dots, S_R \in A_d} |\mathbb{E} \varphi_B(H^\top v) - \mathbb{E} \varphi_B(H^\top h)|,$$

so that by Lemma D.7:

$$\lim_{d \rightarrow \infty} \sup_{\{W_\mu, G_\mu\}_{\mu=2}^n} \sup_{t \in (0, \pi/2)} \langle \mathbb{E} \varphi_B(H^\top v) - \mathbb{E} \varphi_B(H^\top h) \rangle_1 = 0. \quad (84)$$

We now tackle the remaining terms in eq. (83). Let $z \in \{h, v\}$. Using the Cauchy-Schwarz inequality twice we get:

$$\begin{aligned} \langle \mathbb{E} |\varphi(H^\top z)|(1 - u_B(\|H^\top z\|)) \rangle_1 & \leq \langle \mathbb{E} |\varphi(H^\top z)| \mathbf{1}\{\|H^\top z\|_2 \geq B\} \rangle_1, \\ & \leq \langle (\mathbb{E}_z [\varphi(H^\top z)^2])^{1/2} \mathbb{P}_z\{\|H^\top z\|_2 \geq B\}^{1/2} \rangle_1, \\ & \leq \langle (\mathbb{E}_z [\varphi(H^\top z)^2]) \rangle_1^{1/2} \cdot \langle \mathbb{P}_z\{\|H^\top z\|_2 \geq B\} \rangle_1^{1/2}. \end{aligned} \quad (85)$$

The first term in eq. (85) is bounded by the square integrability assumption, cf. eq. (74). To bound the second term, we use Markov's inequality:

$$\langle \mathbb{P}_z\{\|H^\top z\|_2 \geq B\} \rangle_1 \leq \frac{1}{B^2} \langle \mathbb{E}_z \|H^\top z\|_2^2 \rangle_1.$$

From eq. (77) and the matching of the first two moments of ρ with GOE(d), we have:

$$\mathbb{E}_z \|H^\top z\|_2^2 = \frac{4}{d} \sum_{a=1}^R \text{Tr}[S_a^2].$$

Thus, we get:

$$\langle \mathbb{P}_z\{\|H^\top z\|_2 \geq B\} \rangle_1 \leq \frac{4R}{B^2} \left\langle \frac{\text{Tr} S^2}{d} \right\rangle_1 \leq \frac{4RC_0}{B^2}.$$

All in all, we get:

$$\sup_{d \geq 1} \sup_{\{W_\mu, G_\mu\}_{\mu=2}^n} \sup_{t \in (0, \pi/2)} \langle \mathbb{E} |\varphi(H^\top z)|(1 - u_B(\|H^\top z\|)) \rangle_1 \leq \frac{C(R, \varphi)}{B}. \quad (86)$$

Combining eqs. (84) and eq. (86) into eq. (83), and taking $B \rightarrow \infty$ after $d \rightarrow \infty$, we conclude the proof of Lemma D.5. \square