ALEA, Lat. Am. J. Probab. Math. Stat. **21**, 1835–1852 (2024)
DOI: 10.30757/ALEA.v21-69

Fitting an ellipsoid to a quadratic number of random points

Afonso S. Bandeira, Antoine Maillard, Shahar Mendelson and Elliot Paquette

ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
E-mail address: bandeira@math.ethz.ch
URL: https://people.math.ethz.ch/~abandeira

ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
E-mail address: antoine.maillard@math.ethz.ch
URL: https://anmaillard.github.io

The Australian National University
E-mail address: shahar.mendelson@gmail.com
URL: https://sites.google.com/view/shaharmendelson

McGill University, Montréal, Canada
E-mail address: elliot.paquette@mcgill.ca
URL: https://elliotpaquette.github.io

Abstract. We consider the problem (P) of fitting n standard Gaussian random vectors in $\mathbb{R}^d$ to the boundary of a centered ellipsoid, as $n, d \to \infty$. This problem is conjectured to have a sharp feasibility transition: for any $\varepsilon > 0$, if $n \leq (1-\varepsilon)d^2/4$ then (P) has a solution with high probability, while (P) has no solutions with high probability if $n \geq (1+\varepsilon)d^2/4$. So far, only a trivial bound $n \geq d^2/2$ is known on the negative side, while the best results on the positive side assume $n \leq d^2/\mathrm{plog}(d)$. In this work, we improve over previous approaches using a key result of Bartl and Mendelson (2022) on the concentration of Gram matrices of random vectors under mild assumptions on their tail behavior. This allows us to give a simple proof that (P) is feasible with high probability when $n \leq d^2/C$, for a (possibly large) constant $C > 0$.

1. Introduction

We study the following question: given n vectors in $\mathbb{R}^d$ independently sampled from the standard Gaussian measure, when does there exist an ellipsoid centered at 0 whose boundary goes through all the vectors? This question was raised by Saunderson (2011); Saunderson et al. (2012, 2013), and has received significant attention recently (Potechin et al., 2023; Kane and Diakonikolas, 2023; Hsieh et al., 2023; Tulsiani and Wu, 2023). We will discuss the motivations behind this problem and review some recent literature in Section 1.1. In the original series of work of Saunderson

Received by the editors DATE; accepted DATE.
2010 *Mathematics Subject Classification.* 60D05, 52A22, 90C22, 60B20.
Key words and phrases. Ellipsoid fitting, High-dimensional probability, Semidefinite programming, Constraint satisfaction problems.

et al. (Saunderson, 2011; Saunderson et al., 2012, 2013), it was conjectured based on numerical experiments that the ellipsoid fitting property undergoes a phase transition in the limit $d \to \infty$ for $n \sim d^2/4$. Notably, the threshold $d^2/4$ corresponds to the statistical dimension of the cone of positive semidefinite matrices (Gordon, 1988; Amelunxen et al., 2014) (see Potechin et al. (2023) for a discussion).

Conjecture 1.1 (The ellipsoid fitting conjecture)**.** *Let $n, d \geq 1$, and $x_1, \cdots, x_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d/d)$. Let $p(n, d)$ be defined as the probability of existence of a fitting ellipsoid centered in 0:*

$$p(n,d) \coloneqq \mathbb{P}\Big[\exists \Sigma \in \mathcal{S}_d \,:\, \Sigma \succeq 0 \quad \textit{and} \quad x_i^\top \Sigma x_i = 1 \quad (\forall i \in [n])\Big].$$

For any $\varepsilon > 0$, the following holds:

$$\begin{cases} \limsup\limits_{d\to\infty} \dfrac{n}{d^2} \leq \dfrac{1-\varepsilon}{4} & \Rightarrow \lim\limits_{d\to\infty} p(n,d) = 1, \\ \liminf\limits_{d\to\infty} \dfrac{n}{d^2} \geq \dfrac{1+\varepsilon}{4} & \Rightarrow \lim\limits_{d\to\infty} p(n,d) = 0. \end{cases}$$

Our main result gives a positive answer to the existence statement of Conjecture 1.1, up to a constant factor in n/d^2. We present its proof in Section 2.

Theorem 1.2 (Ellipsoid fitting up to a constant)**.** *Let $n, d \geq 1$, and $x_1, \cdots, x_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d/d)$. Given any $\beta \geq 1$, there exist a (small) constant $\alpha = \alpha(\beta) > 0$ and a (large) constant $C = C(\beta) > 0$ such that for $n \leq \alpha d^2$:*

$$\mathbb{P}[\exists \Sigma \in \mathcal{S}_d \,:\, \Sigma \succeq 0 \quad \textit{and} \quad x_i^\top \Sigma x_i = 1 \quad (\forall i \in [n])] \geq 1 - C n^{-\beta}.$$

From polynomial to exponential probability bounds – While we show a polynomial lower bound on the probability, as we will notice during the detailing of the proof, we believe that such a lower bound can be improved to an exponential lower bound of the type $1 - 2\exp(-Cd)$, for $n \leq \alpha d^2$ and a universal constant $\alpha > 0$. We highlight the principles of this improvement in the proof, and detail how it would require a slightly deeper dive into the arguments of the proof of the main result of Bartl and Mendelson (2022). Since the main conjecture of ellipsoid fitting only concerns the limit of the probability and not its scaling, we leave this improvement for future work, and will sometimes use probability estimates that are not the sharpest possible, but are sufficient for our goal.

1.1. *Motivation and related literature.* We give here a brief overview of the motivations to consider the ellipsoid fitting problem, as well as previous results on this conjecture.

Despite the fact that Conjecture 1.1 remains open, the ellipsoid fitting property is a natural question in random geometry. Notably, if the vectors $x_1, \cdots, x_n$ satisfy this property, then there is no vector x_i lying in the interior of the convex hull of the other vectors $(\pm x_j)_{j \neq i}$. Moreover, this problem has several connections with machine learning and theoretical computer science, which motivated its introduction. Examples of these connections include the decomposition of a data matrix into a sum of diagonal and low-rank components (Saunderson, 2011; Saunderson et al., 2012, 2013), overcomplete independent component analysis (Podosinnikova et al., 2019), or the discrepancy of random matrices (Saunderson et al., 2012; Potechin et al., 2023). Relations to these various problems are discussed more extensively in the introduction of Potechin et al. (2023), to which we refer the interested reader for more details.

The negative side of the conjecture – A dimension counting argument shows that ellipsoid fitting is generically not possible if $n > d(d+1)/2$, implying that the negative part of Conjecture 1.1 is non-trivial only in the range $d^2/4 \lesssim n \lesssim d^2/2$. Despite the simplicity of this argument, $d^2/2$ is still the best-known bound on the negative side of Conjecture 1.1.

Early results – In the original works that introduced the ellipsoid fitting conjecture (Saunderson, 2011; Saunderson et al., 2013), it was proven that ellipsoid fitting is feasible with high probability if $n \lesssim \mathcal{O}(d^{6/5-\varepsilon})$ (for any $\varepsilon > 0$). This bound was improved to $n \lesssim \mathcal{O}(d^{3/2-\varepsilon})$ in Ghosh et al. (2020), where the result was obtained as a corollary of the proof of a Sum-of-Squares lower bound for the Sherrington-Kirkpatrick Hamiltonian of statistical physics[1], using a pseudo-calibration construction.

Comparison with recent work – Our proof is based on an "identity perturbation" construction, an idea which was described in Potechin et al. (2023), and used in Kane and Diakonikolas (2023) to prove that $p(n, d) \to 1$ under the assumption that $n = \mathcal{O}(d^2/\mathrm{plog}(d))$. On the other hand, Potechin et al. (2023) uses a least-square construction to prove that ellipsoid fitting is possible with high probability under the similar condition $n = \mathcal{O}(d^2/\mathrm{plog}(d))$[2].

Our proof follows in part the one of Kane and Diakonikolas (2023), improving a crucial operator norm bound thanks to results of Bartl and Mendelson (2022). As mentioned in Kane and Diakonikolas (2023), using a suboptimal bound on this operator norm was the main limitation that prevented the authors to prove the existence of a fitting ellipsoid for $n \le d^2/C$. We emphasize that numerical studies (Potechin et al., 2023) suggested that the identity perturbation construction is successful only in the range $n \lesssim d^2/10$, so in order to resolve Conjecture 1.1 (or even just the existence part) it appears a new idea is needed[3].

Parallel work – As we were finalizing the current manuscript, another proof that ellipsoid fitting is possible at a quadratic number of points was proposed (Hsieh et al., 2023). Like our approach, the proof in Hsieh et al. (2023) is based on the identity perturbation construction, but the proof techniques appear to us to be quite different: Hsieh et al. (2023) relies on the theory of graph matrices, and as such strengthens similar arguments presented in Potechin et al. (2023) (while our proof can instead be viewed as a strengthening of the arguments in Kane and Diakonikolas (2023)). Finally, shortly after the present work appeared online, a third proof was proposed in the independent work of Tulsiani and Wu (2023).

More specifically, our approach relies on obtaining a crucial bound on the operator norm of a kernel Gram matrix by mapping it to the Gram matrix of flattened rank-one matrices, and using the results of Bartl and Mendelson (2022). This latter work showed the concentration of the Gram matrix of i.i.d. vectors $X_1, \cdots, X_n$ under the assumption that the first moments of the projections $\langle X, u\rangle$ satisfy (uniformly in u) a ψ_α-like tail bound for some $\alpha \in (0, 2]$.

1.2. *The dual semidefinite program.* Note that ellipsoid fitting belongs to the class of random semidefinite programs, and as such admits a dual formulation. As we find the dual problem to have a particularly interesting formulation we include a short expository snippet to highlight this dual SDP, and the consequences of Theorem 1.2 for it. Namely, it implies the following corollary.

Corollary 1.3 (Dual problem). *Let $n, d \ge 1$, and $x_1, \cdots, x_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d/d)$. Given any $\beta \ge 1$, there exist a (small) constant $\alpha = \alpha(\beta) > 0$ and a (large) constant $C = C(\beta) > 0$ such that for $n \le \alpha d^2$:*

$$\mathbb{P}\left[\exists z \in \mathbb{R}^n \, : \, \sum_{i=1}^n z_i = 0 \text{ and } \lambda_{\max}\left(\sum_{i=1}^n z_i x_i x_i^\top\right) < 0\right] \le C n^{-\beta}.$$

[1]In the revised version of Potechin et al. (2023), as well as in Hsieh et al. (2023), it was noticed that the results of Ghosh et al. (2020) actually hold for $n \lesssim \mathcal{O}(d^2/\mathrm{plog}(d))$.

[2]We note that Potechin et al. (2023) was recently updated to present an alternative proof through the identity perturbation construction, again under the assumption $n = \mathcal{O}(d^2/\mathrm{plog}(d))$.

[3]Numerical simulations of Potechin et al. (2023) suggest the least-squares approach suffers from the same shortcomings.

Corollary 1.3 rewrites ellipsoid fitting as a problem of "balancing" rank-one matrices: we show that for $n \leq \alpha d^2$ it is impossible to find a centered balancing of $(x_i x_i^\top)$ such that the resulting matrix is negative definite (nor positive definite as one can always consider $-z$). We note however that duality doesn't play any explicit role in the proof of Theorem 1.2.

Proof of Corollary 1.3 –: Ellipsoid fitting is a semidefinite program, which we can write in the canonical form

$$\min_{\substack{\Sigma \succeq 0 \\ \mathrm{Tr}[\Sigma x_i x_i^\top]=1}} \quad \mathrm{Tr}[AS] \in \{0, +\infty\},$$

with $A = 0$. By weak duality and Theorem 1.2, with probability at least $1 - Cn^{-\beta}$ for $n \leq \alpha(\beta) d^2$, its dual semidefinite program satisfies:

$$\max_{\substack{y \in \mathbb{R}^n \\ \sum_{i=1}^n y_i x_i x_i^\top \preceq 0}} \quad \sum_{i=1}^n y_i = 0.$$

We now condition on this event. Thus for all $y \in \mathbb{R}^n$, if $\sum y_i > 0$ then $\lambda_{\max}(\sum_{i=1}^n y_i x_i x_i^\top) > 0$. Let $z \in \mathbb{R}^n$ such that $\sum_{i=1}^n z_i = 0$. To prove Corollary 1.3, it suffices to show that $\lambda_{\max}(\sum_{i=1}^n z_i x_i x_i^\top) \geq 0$. Let $M(z) := \sum_{i=1}^n z_i x_i x_i^\top$. Let $\varepsilon > 0$, and $y_i(\varepsilon) := z_i + \varepsilon$. Since $\sum y_i > 0$, there exists $u_\varepsilon \in \mathcal{S}^{d-1}$ (the Euclidean unit sphere in $\mathbb{R}^d$) such that $u_\varepsilon^\top M(z) u_\varepsilon + \varepsilon \sum_{i=1}^n \langle u_\varepsilon, x_i \rangle^2 > 0$. Extracting a converging sub-sequence as $\varepsilon \to 0$ by compactness, there exists $u \in \mathcal{S}^{d-1}$ with $u^\top M(z) u \geq 0$. □

2. Proof of Theorem 1.2

Notation – Positive universal constants are generically denoted as c_k or C_k, and may vary from line to line. We will clarify possible dependencies of such constants on relevant parameters when necessary. $\mathcal{S}_d$ denotes the set of $d \times d$ real symmetric matrices, I_d is the identity matrix, and $\mathbf{1}_d$ is the all-ones vector. $\mathcal{S}^{d-1}$ is the Euclidean unit sphere in $\mathbb{R}^d$.

Remark – Since the ellipsoid fitting has a clear monotonicity property with respect to n, we assume without loss of generality in what follows that $n = \omega(d^{2-\varepsilon})$ for any fixed $\varepsilon > 0$. The polynomial exponent on the probability estimates, of the form $n^{-\beta}$, can be taken to be arbitrarily large, but it will be considered fixed throughout, with $\beta \geq 1$, and as it will be clear below constants generally depend on β.

2.1. *Identity perturbation ansatz.* In the identity perturbation ansatz (Potechin et al., 2023; Kane and Diakonikolas, 2023), we look for a fitting ellipsoid $\Sigma \in \mathcal{S}_d$ in the form:

$$\Sigma = \mathrm{I}_d + \sum_{i=1}^n q_i x_i x_i^\top, \tag{2.1}$$

for some $q \in \mathbb{R}^n$. Having $\Sigma \succeq 0$ is thus equivalent to:

$$\sum_{i=1}^n q_i x_i x_i^\top \succeq -\mathrm{I}_d. \tag{2.2}$$

We denote $x_i = \sqrt{d_i}\omega_i$, with $\omega_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[\mathcal{S}^{d-1}]$, and $d_i := \|x_i\|_2^2$, and we let $D := \mathrm{Diag}(\{d_i\}_{i=1}^n)$ and $\Theta \in \mathbb{R}^{n \times n}$ with $\Theta_{ij} := \langle \omega_i, \omega_j \rangle^2$. Note that d_i are i.i.d. variables, independent of the directions ω_i. Plugging the ansatz of eq. (2.1) into the ellipsoid fitting equations $x_i^\top \Sigma x_i = 1$ yields:

$$\mathbf{1}_n = D\mathbf{1}_n + D\Theta D q.$$

Assuming that D and Θ are invertible, this equation is solved by:

$$q = D^{-1}\Theta^{-1}(D^{-1}\mathbf{1}_n - \mathbf{1}_n).$$

Plugging it back into eq. (2.2), we see that the identity perturbation ansatz gives a semidefinite positive solution to the ellipsoid fitting problem if Θ, D are invertible, and

$$\min_{a\in\mathcal{S}^{d-1}} \sum_{i=1}^{n} \Big[\Theta^{-1}(D^{-1}\mathbf{1}_n - \mathbf{1}_n)\Big]_i \langle a, \omega_i\rangle^2 \geq -1. \tag{2.3}$$

2.2. *Concentration of a kernel Gram matrix.* We use the following critical lemma on the concentration of the matrix Θ appearing in eq. (2.3). For $\omega_1, \cdots, \omega_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[\mathcal{S}^{d-1}]$, we call $\Theta_{ij} = \langle \omega_i, \omega_j\rangle^2 = \langle \omega_i\omega_i^\top, \omega_j\omega_j^\top\rangle$ a "kernel Gram matrix" since it corresponds to the Gram matrix of $\{\omega_i\}_{i=1}^n$ under the kernel $K(x, y) = \langle x, y\rangle^2$. A key technical difficulty is that while Θ can also be seen as the Gram matrix of $\{\omega_i\omega_i^\top\}_{i=1}^n$, the random matrices $\omega_i\omega_i^\top$ are not centered, and have tails which are heavier than sub-Gaussian, preventing us from applying classical results on the concentration of Gram matrices of sub-Gaussian random vectors, see e.g. Liaw et al. (2017).

Lemma 2.1 (Concentration of a kernel Gram matrix)**.** *Let $n, d \geq 1$, and $\omega_1, \cdots, \omega_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[\mathcal{S}^{d-1}]$. Let $\Theta_{ij} \coloneqq \langle \omega_i, \omega_j\rangle^2$. For any $\beta \geq 1$, there are constants such that, with probability greater than $1 - n^{-\beta} - 2\exp(-c_0 n)$, the following occurs:*

$$\|\Theta - \mathbb{E}\Theta\|_{\mathrm{op}} \leq \frac{C_1}{d} + C_2(\beta)\Big(\sqrt{\frac{n}{d^2}} + \frac{n}{d^2}\Big) \tag{2.4}$$

Notice that $\mathbb{E}\Theta = (1 - 1/d)\mathrm{I}_n + (1/d)\mathbf{1}_n\mathbf{1}_n^\top$. This lemma is a consequence of the analysis of Bartl and Mendelson (2022), and is proven in Section 3.1.

Remark: improving the probability upper bound – A careful analysis of the proof arguments of Bartl and Mendelson (2022) reveals that in the present case in which the matrix to control is a Gram matrix of sub-exponential vectors (which will be the case here as detailed in the proof), the probability estimate could likely be improved significantly to yield a probability lower bound of $1 - 2\exp(-cn)$. We leave for future work to carry out this improvement, and keep a formulation that follows directly from the results of Bartl and Mendelson (2022).

We get the following corollary:

Corollary 2.2 (Concentration of the inverse)**.** *Let $n, d \geq 1$, and $\omega_1, \cdots, \omega_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[\mathcal{S}^{d-1}]$. Let $\Theta_{ij} \coloneqq \langle \omega_i, \omega_j\rangle^2$. For any $\beta \geq 1$, there exists $\alpha = \alpha(\beta) > 0$ and constants such that if $n \leq \alpha d^2$ and $d \geq d_0(\beta)$, then with probability at least $1 - n^{-\beta} - 2\exp(-c_0 n)$:*

$$\Big\|\Theta^{-1} - \Big(\mathrm{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\Big)\Big\|_{\mathrm{op}} \leq \frac{C_1}{d} + C_2(\beta)\sqrt{\frac{n}{d^2}} + \frac{d}{n}. \tag{2.5}$$

In particular, assuming $n = \omega(d)$, for all $\beta \geq 1$ there is $\alpha = \alpha(\beta) > 0$ such that if $n \leq \alpha d^2$:

$$\mathbb{P}[\|\Theta^{-1}\|_{\mathrm{op}} \leq 2] \geq 1 - 2n^{-\beta}. \tag{2.6}$$

Proof of Corollary 2.2 –: Note that $\|\mathbb{E}\Theta - [\mathrm{I}_n + (1/d)\mathbf{1}_n\mathbf{1}_n^\top]\|_{\mathrm{op}} = (1/d)$, so that eq. (2.4) also holds replacing $\mathbb{E}\Theta$ by $\mathrm{I}_n + (1/d)\mathbf{1}_n\mathbf{1}_n^\top$. We use the following elementary lemma, proven in Section 3.4.

Lemma 2.3. *Let $A, B \in \mathcal{S}_n$ two symmetric matrices, such that $B \succ 0$, and for some $\varepsilon < \lambda_{\min}(B)$ we have $\|A - B\|_{\mathrm{op}} \leq \varepsilon$. Then*

$$\|A^{-1} - B^{-1}\|_{\mathrm{op}} \leq \varepsilon \frac{\|B^{-1}\|_{\mathrm{op}}^2}{1 - \varepsilon\|B^{-1}\|_{\mathrm{op}}}.$$

Applying Lemma 2.3 to $B = \mathrm{I}_n + (1/d)\mathbf{1}_n\mathbf{1}_n^\top$, such that $\lambda_{\min}(B) = 1$, and $B^{-1} = \mathrm{I}_n - (d+n)^{-1}\mathbf{1}_n\mathbf{1}_n^\top$, gives, with probability at least $1 - n^{-\beta} - 2\exp(-c_0 n)$:

$$\Big\|\Theta^{-1} - \Big(\mathrm{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\Big)\Big\|_{\mathrm{op}} \le \Big\|\Theta^{-1} - \Big(\mathrm{I}_n - \frac{1}{n+d}\mathbf{1}_n\mathbf{1}_n^\top\Big)\Big\|_{\mathrm{op}} + \frac{d}{n},$$
$$\le \frac{\frac{C_1}{d} + C_2(\beta)\Big(\sqrt{\frac{n}{d^2}} + \frac{n}{d^2}\Big)}{1 - \frac{C_1}{d} - C_2(\beta)\Big(\sqrt{\frac{n}{d^2}} + \frac{n}{d^2}\Big)} + \frac{d}{n},$$
$$\le \frac{C_1'}{d} + C_2'\sqrt{\frac{n}{d^2}} + \frac{d}{n},$$

for large enough d and small enough n/d^2 (depending only on β). □

2.3. *Reducing to a net.* We show some useful estimates in Section 3.5, summarized in the following lemma.

Lemma 2.4 (Some high-probability events)**.** *Let $\omega_1, \cdots, \omega_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[\mathcal{S}^{d-1}]$, and $\Theta_{ij} \coloneqq \langle \omega_i, \omega_j\rangle^2$. Denote $U(a)_i \coloneqq \langle \omega_i, a\rangle^2$ for $a \in \mathcal{S}^{d-1}$. We let $(a_j)_{j=1}^N$ be a $(1/2)$-net of $\mathcal{S}^{d-1}$. Let $\beta \ge 1$. There exists $\alpha = \alpha(\beta) > 0$ such that if $n \le \alpha d^2$, then we have:*

- *(i) $\mathbb{P}[E_1] \ge 1 - 2\exp(-C_1 d)$, with $E_1 \coloneqq \{\max_{j\in[N]} \|U(a_j)\|_2 \le C_2\}$ (for a sufficiently large C_2).*
- *(ii) $\mathbb{P}[E_2] \ge 1 - 2n^{-\beta}$, with $E_2 \coloneqq \{\|\Theta^{-1}\|_{\mathrm{op}} \le 2\}$.*

In the following, we fix $(a_j)_{j=1}^N$ a $(1/2)$-net of $\mathcal{S}^{d-1}$, such that $N \le 5^d$ (Vershynin, 2018). Let $\tilde{q} \coloneqq D^{-1}\mathbf{1}_n - \mathbf{1}_n$. For any matrix $M \in \mathbb{R}^{d\times d}$, we have (Vershynin, 2018):

$$\max_{a\in\mathcal{S}^{d-1}} a^\top M a \le 2 \max_{a\in\mathcal{N}} a^\top M a.$$

Therefore:

$$\mathbb{P}\Big[\min_{a\in\mathcal{S}^{d-1}} \sum_{i=1}^n (\Theta^{-1}\tilde{q})_i \langle a, \omega_i\rangle^2 \le -1\Big] \le \mathbb{P}\left[\max_{j\in[N]} \left|\sum_{i=1}^n (\Theta^{-1}\tilde{q})_i \langle a_j, \omega_i\rangle^2\right| \ge \frac{1}{2}\right]. \tag{2.7}$$

Defining $g_\Theta(a) \coloneqq \sum_{i=1}^n \tilde{q}_i [\Theta^{-1}U(a)]_i$, our goal reduced to show that $\max_{j\in[N]} |g_\Theta(a_j)| \le 1/2$ with probability at least $1 - Cn^{-\beta}$, for n/d^2 small enough. First, we show that we can truncate and center the variables $\tilde{q}_i$:

Lemma 2.5 (Truncating and centering $\tilde{q}$)**.** *Let $A_i \coloneqq \{|\tilde{q}_i| \le 1\}$ and $A \coloneqq \cap_{i=1}^n A_i$. We denote $r_i \coloneqq \tilde{q}_i | A$, and $y_i \coloneqq r_i - \mathbb{E}r_i$. Then $\{y_i\}_{i=1}^n$ are i.i.d. centered $K/\sqrt{d}$-sub-Gaussian random variables, for some universal $K > 0$. Moreover, for any $\beta \ge 1$ there exists $\alpha = \alpha(\beta) > 0$ such that if $n \le \alpha d^2$, then:*

$$\mathbb{P}\left[\max_{j\in[N]} \left|\sum_{i=1}^n (\Theta^{-1}\tilde{q})_i \langle a_j, \omega_i\rangle^2\right| \ge \frac{1}{2}\right] \le \mathbb{P}\left[\max_{j\in[N]} \left|\sum_{i=1}^n (\Theta^{-1}y)_i \langle a_j, \omega_i\rangle^2\right| \ge \frac{1}{4}\right] + Cn^{-\beta}.$$

This lemma is proven in Section 3.6.

2.4. *Controlling points on the net.* In what follows, we replace the variables $\tilde{q}_i$ by y_i by using Lemma 2.5 (assuming $n \le \alpha d^2$ for $\alpha = \alpha(\beta)$ small enough). We define, for $a \in \mathcal{S}^{d-1}$:

$$f_\Theta(a) \coloneqq \sum_{i=1}^n y_i [\Theta^{-1}U(a)]_i = \sum_{i=1}^n [\Theta^{-1}y]_i U(a)_i, \tag{2.8}$$

with $U(a) \coloneqq (\langle \omega_i, a\rangle^2)_{i=1}^n$. We prove in Section 3.7 the following elementary lemma:

Lemma 2.6. *Let* $\{y_i\}_{i=1}^n$ *be i.i.d. centered sub-Gaussian random variables, with* $\|y_1\|_{\psi_2} \leq K/\sqrt{d}$, *and* $M \in \mathcal{S}_n$. *Then:*

$$\mathbb{P}[\|My\|_\infty \geq C\|M\|_{\mathrm{op}} d^{-3/8}] \leq 2n \exp\{-d^{1/4}\}.$$

We let

$$E_3 \coloneqq \{\|\Theta^{-1}y\|_\infty \leq C\|\Theta^{-1}\|_{\mathrm{op}} d^{-3/8}\},$$

and $E \coloneqq \cap_{k=1}^3 E_k$. We have from Lemmas 2.4 and 2.6 that (recall that y is independent of Θ) there is $\alpha = \alpha(\beta) > 0$ such that for $n \leq \alpha d^2$:

$$\mathbb{P}[E] \geq 1 - Cn^{-\beta}. \tag{2.9}$$

Let us fix $a \in \mathcal{S}^{d-1}$. For $\eta \in (0,1)$ we define $S(\eta) \coloneqq \{i \in [n] : |\langle \omega_i, a\rangle| > \eta\}$. Since $\omega_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}[\mathcal{S}^{d-1}]$, $|\langle \omega_i, a\rangle|$ are i.i.d. sub-Gaussian random variables, with sub-Gaussian norm $C/\sqrt{d}$ (Vershynin, 2018). $|S(\eta)|$ is thus a binomial random variable, with parameters n and $p \leq 2\exp\{-Cd\eta^2\}$. By Theorem 1 of Klenke and Mattner (2010), $|S(\eta)|$ is stochastically dominated by a Poisson random variable with parameter $-n\log(1-p)$. Assuming that $d\eta^2 \to \infty$, we have for d large enough[4],

$$-n\log(1-p) \leq 2np \leq 4n\exp\{-Cd\eta^2\}.$$

Letting $\lambda \coloneqq 4n\exp\{-Cd\eta^2\}$ and $X \sim \mathrm{Pois}(\lambda)$, $|S(\eta)|$ is thus stochastically dominated by X. We reach that for all $x > \lambda$ (see e.g. Theorem 5.4 of Mitzenmacher and Upfal (2017) for the second inequality):

$$\mathbb{P}[|S(\eta)| \geq x] \leq \mathbb{P}[X \geq x] \leq \left(\frac{e\lambda}{x}\right)^x e^{-\lambda}. \tag{2.10}$$

We get from eq. (2.10) that

$$\mathbb{P}[|S(\eta)| \geq d^{1/4}] \leq \exp\left\{d^{1/4}\log(4ne) - Cd^{5/4}\eta^2 - \frac{d^{1/4}\log d}{4}\right\} \leq \exp\left\{d^{1/4}\log n - Cd^{5/4}\eta^2\right\}. \tag{2.11}$$

We decompose $f_\Theta(a)$ in two parts, which we control separately:

$$f_\Theta(a) = \underbrace{\sum_{i\in S(\eta)} [\Theta^{-1}y]_i U(a)_i}_{=:f_1(\eta,a)} + \underbrace{\sum_{i\notin S(\eta)} [\Theta^{-1}y]_i U(a)_i}_{=:f_2(\eta,a)}. \tag{2.12}$$

First, we have that under the event E of eq. (2.9), and by the Cauchy-Schwarz inequality:

$$|f_1(\eta,a)| \leq Cd^{-3/8}\sum_{i\in S(\eta)} \langle \omega_i, a\rangle^2 \leq Cd^{-3/8}|S(\eta)|.$$

Let us pick $\eta = d^{-1/8}t$, for some $t \geq 1$ (so that $d\eta^2 \to \infty$). Using eq. (2.11) in the previous inequality, as well as the law of total probability (and $\mathbb{P}[E] \geq 1/2$), we reach:

$$\mathbb{P}[|f_1(d^{-1/8}t, a)| \geq C_1 d^{-1/8}|E] \leq 2\exp\{d^{1/4}\log n - C_2 dt^2\}. \tag{2.13}$$

We now control $f_2(\eta, a)$. For a random variable $X(\{y_i, \omega_i\})$, we denote $\|X\|_{\psi_2,y}$ the sub-Gaussian norm of the random variable with respect to the randomness of $\{y_i\}$ only (i.e. conditioned on the value of $\{\omega_i\}$). Since y_i are independent of $\{\omega_i\}$ (and thus of the choice of the set $S(\eta)$ and of Θ), we get by Hoeffding's inequality (recall that y_i are i.i.d. $K/\sqrt{d}$-sub-Gaussian), that for all $\{\omega_i\}$:

$$\|f_2(\eta,a)\|_{\psi_2,y}^2 \leq \frac{C}{d}\|\Theta^{-1}\widetilde{U}(a)\|_2^2.$$

[4] Since $\log(1-x) \geq -2x$ for $0 \leq x \leq 1/2$.

Here we denoted $\widetilde{U}(a)_i := \langle \omega_i, a\rangle^2 \mathbb{1}\{|\langle \omega_i, a\rangle| \leq \eta\}$. Therefore:

$$\|f_2(\eta, a)\|_{\psi_2,y}^2 \leq \frac{C\|\Theta^{-1}\|_{\mathrm{op}}^2}{d} \sum_{i \notin S(\eta)} \langle \omega_i, a\rangle^4. \tag{2.14}$$

We can then prove (see Section 3.8):

Lemma 2.7. *For all $q \in [1/2, 1]$, there is a constant $C = C(q) > 0$ such that for all $v \geq 0$, and all $\eta \in (0,1)$:*

$$\mathbb{P}\Bigg[\sum_{i \notin S(\eta)} \langle \omega_i, a\rangle^4 \geq \frac{n}{d^2}(3+v)\Bigg] \leq 2\exp\Bigg\{-C\min\Bigg(\frac{nd^{2/q}\eta^{4/q}}{d^4\eta^8}v^2, n^q d^{1-2q}\eta^{2-4q}v^q\Bigg)\Bigg\}.$$

2.5. *Ending the proof.* We detail now how the combination of eq. (2.13) and Lemma 2.7 allows to complete the proof. By Lemma 2.5, our task reduced to show that for a 1/2-net $(a_j)_{j=1}^N$ of $\mathcal{S}^{d-1}$, we have with probability at least $1 - Cn^{-\beta}$, and assuming $n \leq \alpha d^2$ for $\alpha = \alpha(\beta)$ small enough:

$$\max_{j \in [N]} |f_\Theta(a_j)| \leq 1/4. \tag{2.15}$$

Recall the decomposition of eq. (2.12). We fix $\eta = d^{-1/8}t$, for $t \geq 1$ large enough (not depending on n, d) such that eq. (2.13) gives, for n, d large enough:

$$\mathbb{P}[|f_1(d^{-1/8}t, a)| \geq Cd^{-1/8}|E] \leq 10^{-d}. \tag{2.16}$$

By Lemma 2.7 and eq. (2.14) we have, choosing $v = 1$ and $q = 3/5$[5], that for all $x > 0$:

$$\begin{aligned}
\mathbb{P}\Bigg[|f_2(d^{-1/8}t, a)| \geq x\|\Theta^{-1}\|_{\mathrm{op}}\sqrt{\frac{n}{d^2}}\Bigg] &\leq \mathbb{E}_\omega\Bigg[\exp\Bigg(-\frac{Cnx^2}{d\sum_{i \notin S(\eta)}\langle \omega_i, a\rangle^4}\Bigg)\Bigg], \\
&\overset{(a)}{\leq} 2\exp\Bigg\{-C_1\min\Bigg(\frac{n}{t^{4/3}\sqrt{d}}, n^{3/5}d^{-3/20}t^{-2/5}\Bigg)\Bigg\} + \exp(-C_2dx^2), \\
&\overset{(b)}{\leq} 2\exp\Big\{-C_1n^{3/5}d^{-3/20}t^{-2/5}\Big\} + \exp(-C_2dx^2), \\
&\overset{(c)}{\leq} 10^{-d} + \exp(-C_2dx^2),
\end{aligned}$$

where we used Lemma 2.7 in (a) with $v = 1$ and $q = 3/5$ (and bounding $e^{-z} \leq 1$), in (b) the fact that $n/\sqrt{d} = \omega(n^{3/5}d^{-3/20})$ since $n = \omega(d)$, and finally in (c) we used that $n = \omega(d^{23/12})$, so that we can bound the first term by 10^{-d} for n, d large enough. We fix $x > 0$ large enough (not depending on n, d) such that the second term also satisfies $\exp(-C_2dx^2) \leq 10^{-d}$. All in all, we get:

$$\mathbb{P}\Bigg[|f_2(d^{-1/8}t, a)| \geq C\|\Theta^{-1}\|_{\mathrm{op}}\sqrt{\frac{n}{d^2}}\Bigg] \leq 2 \times 10^{-d}.$$

And thus:

$$\mathbb{P}\Bigg[|f_2(d^{-1/8}t, a)| \geq C\sqrt{\frac{n}{d^2}}\Bigg|E\Bigg] \leq \frac{\mathbb{P}\Big[|f_2(d^{-1/8}t, a)| \geq C\|\Theta^{-1}\|_{\mathrm{op}}\sqrt{\frac{n}{d^2}}\Big]}{\mathbb{P}[E]} \leq 3 \times 10^{-d} \tag{2.17}$$

Notice that the event E of eq. (2.9) is independent of the net. Thus, we have for all $u > 0$:

$$\mathbb{P}\Big[\max_{j \in [N]} |f_\Theta(a_j)| \geq u\Big] \leq Cn^{-\beta} + \mathbb{P}\Big[\max_{j \in [N]} |f_\Theta(a_j)| \geq u\Big|E\Big]. \tag{2.18}$$

[5] This is an arbitrary choice, the only requirement needed is actually that $q \in (1/2, 3/4)$.

Combining eqs. (2.16) and (2.17) with the union bound (recall $N \le 5^d$) we get:

$$\mathbb{P}\left[\max_{j\in[N]} |f_\Theta(a_j)| \ge C_1\sqrt{\frac{n}{d^2}} + C_2 d^{-1/8} \middle| E\right] \le 4\cdot 5^d \cdot 10^{-d} \le 4\cdot 2^{-d}. \tag{2.19}$$

By combining eqs. (2.18) and eq. (2.19), taking d large enough, and n/d^2 small enough, this ends the proof of eq. (2.15), and thus of Theorem 1.2.

3. Auxiliary proofs

3.1. *Proof of Lemma 2.1.* We use the matrix flattening function, for $M \in \mathcal{S}_d$:

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

It is an isometry: $\langle \mathrm{vec}(M), \mathrm{vec}(N)\rangle = \mathrm{Tr}[MN]$. Note that Θ is the Gram matrix of the i.i.d. vectors $X_i := \mathrm{vec}(x_i x_i^\top) \in \mathbb{R}^p$, with $p := d(d+1)/2$.

Centering – Note that $\|X_i\|_2 = \|x_i\|_2^2 = 1$. Moreover, we have[6] $\mathbb{E}[X_i] = \mathrm{I}_d/d$, and if $Y_i := X_i - \mathbb{E}[X_i]$, then $\langle Y_i, Y_j\rangle = \langle X_i, X_j\rangle - 1/d$. Therefore, we can write

$$\Theta = H + \frac{1}{d}\mathbf{1}_n\mathbf{1}_n^\top,$$

with $H_{ij} := \langle Y_i, Y_j\rangle$ the Gram matrix of the $(Y_i)_{i=1}^n$. We also sometimes denote $H = Y^\top Y$, with Y the matrix whose columns are given by $Y_1, \cdots, Y_n$. Note that $\mathbb{E}[\Theta] = (1-1/d)\mathrm{I}_n + (1/d)\mathbf{1}_n\mathbf{1}_n^\top$. Thus, to prove Lemma 2.1 it suffices to show that with the required probability bound:

$$\|H - \mathrm{I}_n\|_{\mathrm{op}} \le \frac{C_1}{d} + C_2(\beta)\left(\sqrt{\frac{n}{d^2}} + \frac{n}{d^2}\right). \tag{3.1}$$

Projecting – Note that $\langle Y_i, \mathrm{vec}(\mathrm{I}_d)\rangle = 0$, so that $Y_i \in \{\mathrm{vec}(\mathrm{I}_d)\}^\perp$. We denote P the orthogonal projector onto $\{\mathrm{vec}(\mathrm{I}_d)\}^\perp$, i.e.

$$P := \mathrm{I}_p - \frac{1}{d}\mathrm{vec}(\mathrm{I}_d)\,\mathrm{vec}(\mathrm{I}_d)^\top. \tag{3.2}$$

We remark that $(PY_i)_{i=1}^n$ are still i.i.d., centered, and we have $\langle PY_i, PY_j\rangle = \langle Y_i, Y_j\rangle$.

Rescaling – Note that $\mathbb{E}[Y_i] = 0$, and without loss of generality (up to using the vectors $Y_i' := \varepsilon_i Y_i$ with $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{\pm 1\})$, for which the Gram matrix H' satisfies $H' = \mathrm{Diag}(\varepsilon)H\mathrm{Diag}(\varepsilon)$ and has thus the same eigenvalues as H) we can assume the Y_i to be symmetric.

Let us compute the covariance of Y. For $a \le b$ and $c \le d$, we have

$$\begin{aligned}\mathbb{E}[Y_{ab}Y_{cd}] &= [(2-\delta_{ab})(2-\delta_{cd})]^{1/2}\left[\mathbb{E}(x_a x_b x_c x_d) - \frac{\delta_{ab}\delta_{cd}}{d^2}\right],\\ &\overset{\text{(a)}}{=} \frac{[(2-\delta_{ab})(2-\delta_{cd})]^{1/2}}{d^2}\left[\frac{d}{d+2}(\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{abcd}) - \delta_{ab}\delta_{cd}\right],\\ &= \frac{1}{d^2}\left[\frac{d}{d+2}\left(\delta_{abcd} + [(2-\delta_{ab})(2-\delta_{cd})]^{1/2}\delta_{ac}\delta_{bd}\right) - \frac{2}{d+2}\delta_{ab}\delta_{cd}\right],\\ &= \frac{2}{d^2}\left[\frac{d}{d+2}\delta_{ac}\delta_{bd} - \frac{1}{d+2}\delta_{ab}\delta_{cd}\right].\end{aligned} \tag{3.3}$$

[6] We identify the matrices and their flattened versions.

In (a) we used the marginals of uniformly sampled random vectors on $\mathcal{S}^{d-1}$, which can easily be obtained e.g. by using hyperspherical coordinates[7]. In matrix notation, eq. (3.3) can be rewritten as:

$$\mathbb{E}[YY^\top] = \frac{2}{d^2}\left[\frac{d}{d+2}\mathrm{I}_p - \frac{1}{d+2}\,\mathrm{vec}(\mathrm{I}_d)\,\mathrm{vec}(\mathrm{I}_d)^\top\right],$$
$$= \frac{2}{d(d+2)}P.$$

Therefore, if we denote $V_i \coloneqq PY_i \in \mathbb{R}^{p-1}$ the coordinates of Y_i in $\{\mathrm{vec}(\mathrm{I}_d)\}^\perp$, we have that $\langle V_i, V_j\rangle = \langle Y_i, Y_j\rangle$, and

$$\mathbb{E}[VV^\top] = \frac{2}{d(d+2)}\mathrm{I}_{p-1}.$$

Denote

$$\Sigma \coloneqq (p-1)\mathbb{E}[VV^\top] = \frac{2(p-1)}{d(d+2)}\mathrm{I}_{p-1} = \left(1 - \frac{1}{d}\right)\mathrm{I}_{p-1}. \tag{3.4}$$

In particular $\|\Sigma - \mathrm{I}_{p-1}\|_{\mathrm{op}} \leq (1/d)$. Letting $Z \coloneqq \Sigma^{-1/2}V$, the vector Z satisfies $\mathbb{E}[ZZ^\top] = (p-1)^{-1}\mathrm{I}_{p-1}$, and the Gram matrix H_Z of $Z_1, \cdots, Z_n$ satisfies $H - H_Z = Z^\top(\Sigma - \mathrm{I}_{p-1})Z$, and thus for all $w \in \mathbb{R}^{p-1}$:

$$|w^\top H_Z w - w^\top H w| = |w^\top Z^\top(\Sigma - \mathrm{I}_{p-1})Zw|,$$
$$\leq (1/d)\|Zw\|_2^2,$$
$$= (1/d)w^\top H_Z w.$$

Therefore, $\|H - H_Z\|_{\mathrm{op}} \leq (1/d)\|H_Z\|_{\mathrm{op}}$. By the triangle inequality, this yields that

$$\|H - \mathrm{I}_n\|_{\mathrm{op}} \leq \frac{1}{d} + \left(1 + \frac{1}{d}\right)\|H_Z - \mathrm{I}_n\|_{\mathrm{op}}. \tag{3.5}$$

Using eq. (3.1) and eq. (3.5), it is clear that we conclude to eq. (2.4), it is enough to show that (with the required probability bound):

$$\|H_Z - \mathrm{I}_n\|_{\mathrm{op}} \leq \frac{6}{d} + C(\beta)\left(\sqrt{\frac{n}{d^2}} + \frac{n}{d^2}\right). \tag{3.6}$$

Gram matrix estimation – We will use the results of Bartl and Mendelson (2022). We need to introduce the definition of a well-behaved random vector:

Definition 3.1 (Well-behaved vector)**.** Let $q \geq 1$. A random vector $X \in \mathbb{R}^q$ is said to be well-behaved for $n \geq 1$ with constants $L, R > 0$, $\alpha \in (0,2]$, $\delta \in [0,1]$ and $\gamma \in [0,1)$ if:

- (i) X is symmetric and isotropic: $\mathbb{E}[XX^\top] = \mathrm{I}_q$.
- (ii) If one considers n i.i.d. draws $X_1, \cdots, X_n$, then with probability at least $1-\gamma$:

$$\max_{1\leq i\leq n}\left|\frac{\|X_i\|_2^2}{q} - 1\right| \leq \delta.$$

- (iii) For all $2 \leq k \leq R\log n$ and all $t \in \mathbb{R}^q$:

$$\|\langle X, t\rangle\|_{L_k} \leq Lk^{1/\alpha}\|\langle X, t\rangle\|_{L_2} = Lk^{1/\alpha}\|t\|_2.$$

[7]The two moments needed are $d^2\mathbb{E}[x_1^4] = 3d/(2+d)$ and $d^2\mathbb{E}[x_1^2x_2^2] = d/(d+2)$.

Condition (iii) corresponds to some ψ_α behavior of the projections, uniformly in t, and for some $\alpha \in (0,2]$, but only up to moments $k = \mathcal{O}(\log n)$. We can now state an immediate corollary to Theorem 1.5 of Bartl and Mendelson (2022) (precisely the particular case corresponding to T being the unit sphere):

Corollary 3.2 (Bartl and Mendelson (2022))**.** *Let $n, q \geq 1$. Let $\beta \geq 1$. Assume that the random vector $A \in \mathbb{R}^q$ is well-behaved with respect to n according to Definition 3.1, with constants $L, R = R(\beta), \alpha, \gamma, \delta$. Let $M \in \mathbb{R}^{q\times n}$ be a matrix with i.i.d. columns $A_1, \cdots A_n$. Then, with probability at least $1 - \gamma - 2\exp(-c_0 n) - n^{-\beta}$, we have*

$$\left\| \frac{1}{q} M^\top M - \mathrm{I}_n \right\|_{\mathrm{op}} \leq 2\delta + c(L,\alpha,\beta)\left(\sqrt{\frac{n}{q}} + \frac{n}{q}\right).$$

Corollary 3.2 is an application of Theorem 1.5 of Bartl and Mendelson (2022), for the simplest case in which $T = \mathcal{S}^{n-1}$, so that the Gaussian width is $\ell_\star(T) \coloneqq \mathbb{E}\|g\|_2 \simeq \sqrt{n}$ (for $g \sim \mathcal{N}(0, \mathrm{I}_n)$), $d_T \coloneqq \sup_{t\in\mathcal{S}^{n-1}} \|t\| = 1$, and $k_\star(T) \coloneqq (\ell_\star(T)/d_T)^2 \simeq n$. More precisely, we have $(1 + \mathcal{O}(n^{-1}))n \leq n^2/(n+1) \leq k_\star(T) \leq n$. Note as well that we added the factor p^{-1} in front of the Gram matrix $M^\top M$ (it is implicit in Bartl and Mendelson (2022) because the columns of M there are $A_i/\sqrt{p}$).

An important remark – We emphasize a technical point, related to the final probability bounds we obtain in Theorem 1.2. In what follows, we will apply Corollary 3.2 with $R = \infty$, as the moment bound will be valid for all orders. In this context, the analysis of Bartl and Mendelson (2022) would naturally imply that Corollary 3.2 holds with probability at least $1-\gamma-2\exp(-c_0 n)$, and with a constant $c(L,\alpha)$ not depending on β. In turn, a more careful analysis would reveal that the probability bound of Theorem 1.2 can be made exponentially small in d. However, as proving this would require a possibly lengthy technical analysis of the arguments of Bartl and Mendelson (2022), for reasons of clarity we chose to restrict to the most direct application of Theorem 1.5 of Bartl and Mendelson (2022), which gives then a suboptimal polynomial probability upper bound.

In order to deduce eq. (3.6) from Corollary 3.2, with the dimension $q = p-1$ (recall $p = d(d+1)/2$), we need to verify that the distribution of the columns Z_i is well-behaved for some $\alpha, L, R, \delta, \gamma$. We let $A_i \coloneqq \sqrt{q} Z_i$, and we check that it satisfies Definition 3.1.

Condition (i) – Because of the random sign that we can add wlog, we have seen that the distribution of A is symmetric. Moreover, by our analysis above, $\mathbb{E}[AA^\top] = q\mathbb{E}[ZZ^\top] = \mathrm{I}_q$, so that A is isotropic.

Condition (ii) – Notice that for all i, $\|Y_i\|_2^2 = \|V_i\|_2^2 = 1 - 1/d$. Thus, with the notations from above:

$$\begin{aligned}\left|\frac{1}{q}\|A\|_2^2 - 1\right| &= \left|V^\top(\Sigma^{-1} - \mathrm{I}_q)V + \frac{1}{d}\right|,\\ &\leq \|V\|_2^2 \|\Sigma^{-1} - \mathrm{I}_q\|_{\mathrm{op}} + \frac{1}{d},\\ &\overset{(a)}{\leq} \frac{3}{d}.\end{aligned}$$

In (a) we used that

$$\|\Sigma - \mathrm{I}_q\|_{\mathrm{op}} \leq \frac{1}{d} \Rightarrow \|\Sigma^{-1} - \mathrm{I}_q\|_{\mathrm{op}} \leq \frac{\frac{1}{d}}{1 - \frac{1}{d}} \leq \frac{2}{d}.$$

Thus, A satisfies the condition (ii) with $\gamma = 0$ and $\delta = 3/d$ (since the bound is deterministic, there is no need to consider n i.i.d. samples).

Condition (iii) – We are going to see that it actually holds for all $k \geq 2$ with $\alpha = 1$, i.e. the random vector A is uniformly sub-exponential. Let $t \in \mathbb{R}^q$. Then[8]:

$$\begin{aligned} |\langle A, t\rangle - \sqrt{q}\langle V, t\rangle| &= |\sqrt{q} V^\top (\Sigma^{-1/2} - \mathrm{I}_q) t|, \\ &\leq \sqrt{q}\|V\|_2 \times \frac{2}{d} \times \|t\|_2, \\ &\overset{\text{(a)}}{\leq} C\|t\|_2, \end{aligned}$$

using in (a) that $q + 1 = d(d+1)/2$ and that $\|V\|_2^2 = \|Y\|_2^2 = \mathrm{Tr}[(xx^\top - \mathrm{I}_d/d)^2] = 1 - 1/d \leq 1$. We have then for all $k \geq 2$:

$$\begin{aligned} \|\langle A, t\rangle\|_k &\overset{\text{(a)}}{\leq} 2\Bigg[q^{k/2}\|\langle V, t\rangle\|_k^k + C^k\|t\|_2^k\Bigg]^{1/k}, \\ &\overset{\text{(b)}}{\leq} 2\Bigg[\sqrt{q}\|\langle V, t\rangle\|_k + C\|t\|_2\Bigg], \end{aligned}$$

using in (a) that $(x+y)^k \leq 2^{k-1}(x^k + y^k)$ for $x, y > 0$, and in (b) Minkowski's inequality $(x+y)^{1/k} \leq (x^{1/k} + y^{1/k})$. Therefore, it is enough to check that for all $k \geq 2$:

$$\|\langle V, t\rangle\|_k \leq \frac{L}{d} k^{1/\alpha}\|t\|_2, \tag{3.7}$$

for some $\alpha \in (0, 2]$. We will use the Hanson-Wright inequality for random vectors on the sphere:

Lemma 3.3 (Hanson-Wright)**.** *Let $d \geq 1$ and $x \sim \mathrm{Unif}(\mathcal{S}^{d-1})$. For any $M \in \mathcal{S}_d$ and any $u > 0$:*

$$\mathbb{P}\Big[\Big|dx^\top Mx - \mathrm{Tr}[M]\Big| \geq u\Big] \leq 2\exp\left\{-C\min\left(\frac{u^2}{\|M\|_F^2}, \frac{u}{\|M\|_{\mathrm{op}}}\right)\right\}.$$

Remark – We prove Lemma 3.3 as a consequence of a general Hanson-Wright inequality for random vectors satisfying a convex Lipschitz concentration property (Adamczak, 2015), easily satisfied by the Haar measure on $\mathcal{S}^{d-1}$. We give details in Section 3.2.

Recall that $V = PY \in \mathbb{R}^q$, with P the orthogonal projector onto $\mathrm{vec}(\mathrm{I}_d)^\perp$, and that $t \in \mathbb{R}^q$. If we identify t with the corresponding element of $\mathbb{R}^p$ (or the corresponding $d \times d$ symmetric matrix), then $\mathrm{Tr}[t] = 0$, and $\langle V, t\rangle = \langle Y, t\rangle = x^\top tx - \mathrm{Tr}[t]/d = x^\top tx$ for $x \sim \mathrm{Unif}(\mathcal{S}^{d-1})$. Using Lemma 3.3 with $M = t$ gives:

$$\mathbb{P}\Big[d|\langle V, t\rangle| \geq u\Big] \leq 2\exp\left\{-C\min\left(\frac{u^2}{\|t\|_2^2}, \frac{u}{\|t\|_{\mathrm{op}}}\right)\right\}.$$

[8] Again, since $\|\Sigma - \mathrm{I}_q\|_{\mathrm{op}} \leq 1/d \Rightarrow \|\Sigma^{-1/2} - \mathrm{I}_q\|_{\mathrm{op}} \leq 2/d$.

It is now classical to deduce the moments from the tails:

$$
\begin{aligned}
d^k \|\langle V, t\rangle\|_k^k &= \int_0^\infty k u^{k-1} \mathbb{P}[d|\langle V, t\rangle| \geq u] \mathrm{d}u, \\
&\leq 2k \int_0^\infty u^{k-1} \exp\left\{ - C \min\left(\frac{u^2}{\|t\|_2^2}, \frac{u}{\|t\|_{\mathrm{op}}}\right)\right\} \mathrm{d}u, \\
&\leq 2k \int_0^{\|t\|_2^2/\|t\|_{\mathrm{op}}} u^{k-1} \exp\{-Cu^2/\|t\|_2^2\} \mathrm{d}u + 2k \int_{\|t\|_2^2/\|t\|_{\mathrm{op}}}^\infty u^{k-1} \exp\{-Cu/\|t\|_{\mathrm{op}}\} \mathrm{d}u, \\
&\leq 2k \int_0^\infty u^{k-1} \exp\{-Cu^2/\|t\|_2^2\} \mathrm{d}u + 2k \int_0^\infty u^{k-1} \exp\{-Cu/\|t\|_{\mathrm{op}}\} \mathrm{d}u, \\
&\leq k C^{-k/2} \|t\|_2^k \Gamma\left[\frac{k}{2}\right] + 2k \Big(\frac{\|t\|_{\mathrm{op}}}{C}\Big)^k \Gamma(k), \\
&\leq k \|t\|_2^k \left\{ C^{-k/2} \Gamma\left[\frac{k}{2}\right] + 2 C^{-k} \Gamma(k)\right\},
\end{aligned}
$$

since $\|t\|_{\mathrm{op}} \leq \|t\|_2 = \|t\|_F$. This is simply the sum of the sub-Gaussian and sub-exponential part of the tail given by Hanson-Wright's inequality. Thus we have

$$
d\|\langle V, t\rangle\|_k \leq Lk\|t\|_2,
$$

which is exactly eq. (3.7) for $\alpha = 1$.

Applying Corollary 3.2 to $A = \sqrt{q}Z$ with $L, R = \infty, \alpha = 1, \gamma = 0, \delta = 3/d$, we reach that for all $\beta \geq 1$:

$$
\|Z^\top Z - \mathrm{I}_n\|_{\mathrm{op}} \leq \frac{6}{d} + C_1(\beta)\left(\sqrt{\frac{n}{d^2}} + \frac{n}{d^2}\right),
$$

with probability at least $1 - n^{-\beta} - 2\exp(-c_0 n)$. This implies eq. (3.6) and concludes the proof. □

3.2. *Proof of Lemma 3.3.* We use a generalization of Hanson-Wright's inequality (usually stated for i.i.d. sub-Gaussian vectors) which is due to Adamczak (2015).

Definition 3.4 (Convex concentration property)**.** Let $n \geq 1$ and X be a random vector in $\mathbb{R}^n$. We say that X has the convex concentration property with constant K if, for all $\varphi : \mathbb{R}^n \to \mathbb{R}$ convex and 1-Lipschitz, we have $\mathbb{E}|\varphi(X)| < \infty$, and for every $t > 0$:

$$
\mathbb{P}[|\varphi(X) - \mathbb{E}[\varphi(X)]| \geq t] \leq 2\exp(-t^2/K^2).
$$

Note that if $X = \sqrt{d}x$, with $x \sim \mathrm{Unif}[\mathcal{S}^{d-1}]$, then X satisfies Definition 3.4 for some absolute constant $K > 0$ (the function φ does not even need to be convex), it is one of the most classical results of concentration of measure, cf. e.g. Theorem 5.1.4 of Vershynin (2018). The main result of Adamczak (2015) is the following:

Proposition 3.5 (Hanson-Wright (Adamczak, 2015))**.** *Let $n \geq 1$ and X be a zero-mean vector in $\mathbb{R}^n$ that has the convex concentration property with constant K. Then for all symmetric $M \in \mathbb{R}^{n\times n}$ and $t > 0$:*

$$
\mathbb{P}[|X^\top M X - \mathbb{E}(X^\top M X)| \geq t] \leq 2\exp\left(- C \min\left(\frac{t^2}{2K^4\|M\|_F^2}, \frac{t}{K^2\|M\|_{\mathrm{op}}}\right)\right).
$$

Applying Proposition 3.5 to the vector X described above yields Lemma 3.3.

3.3. *Tail bounds for χ^2 random variables.* The following is a useful tail bound on χ^2_d random variables, from Laurent and Massart (2000).

Lemma 3.6 (Tail bounds for χ^2_d)**.** *Let $d \geq 1$, and $x_1, \cdots, x_d \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Let $z \coloneqq (1/d)\sum_{i=1}^d x_i^2$. Then for all $u \geq 0$:*

$$\begin{cases} \mathbb{P}\left[z - 1 \geq 2\sqrt{\frac{u}{d}} + 2\frac{u}{d}\right] & \leq \exp(-u), \\ \mathbb{P}\left[z - 1 \leq -2\sqrt{\frac{u}{d}}\right] & \leq \exp(-u). \end{cases}$$

Corollary 3.7 (Tail bounds for $\tilde{q}$)**.** *Let $d \geq 1$, and $x_1, \cdots, x_d \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Let $z \coloneqq (1/d)\sum_{i=1}^d x_i^2$, and we denote $\tilde{q} \coloneqq 1/z - 1$. Notice that $\tilde{q} \geq -1$. Then, for all $t \in (0,1)$:*

$$\mathbb{P}[|\tilde{q}| \geq t] \leq 2\exp\left(-\frac{dt^2}{16}\right).$$

Proof of Corollary 3.7 –: We start with the upper tail $\tilde{q} \geq t$. Notice that $\tilde{q} \geq t \Leftrightarrow z \leq (1+t)^{-1}$. Using Lemma 3.6 with $4u = d[t/(1+t)]^2$, we have (using that $t < 1$):

$$\mathbb{P}[\tilde{q} \geq t] \leq \exp\left\{-\frac{dt^2}{4(1+t)^2}\right\} \leq \exp\left\{-\frac{dt^2}{16}\right\}.$$

Similarly, for the lower tail, $\tilde{q} \leq -t \Leftrightarrow z \geq (1-t)^{-1}$. Using Lemma 3.6 with $2u = d[1/(1-t) - \sqrt{(1+t)/(1-t)}]$, we have (again using that $t \in (0,1)$):

$$\mathbb{P}[\tilde{q} \leq -t] \leq \exp\left\{-\frac{d}{2}\left[\frac{1}{1-t} - \sqrt{\frac{1+t}{1-t}}\right]\right\} \leq \exp\left\{-\frac{dt^2}{4}\right\}.$$

This ends the proof. □

3.4. *Proof of Lemma 2.3.*

Proof: Note that $\lambda_{\min}(A) \geq \lambda_{\min}(B) - \varepsilon$, so that $A \succ 0$ and $\|A^{-1}\|_{\mathrm{op}} \leq \|B^{-1}\|_{\mathrm{op}}/(1 - \varepsilon\|B^{-1}\|_{\mathrm{op}})$. We can use the standard estimate:

$$\|A^{-1} - B^{-1}\|_{\mathrm{op}} = \|B^{-1}(B-A)A^{-1}\|_{\mathrm{op}} \leq \|B^{-1}\|_{\mathrm{op}}\|A-B\|_{\mathrm{op}}\|A^{-1}\|_{\mathrm{op}}.$$

Using the remark above and the fact that $\|A - B\|_{\mathrm{op}} \leq \varepsilon$ completes the proof. □

3.5. *Proof of Lemma 2.4.* The probability bound for the event E_2 is the conclusion of Corollary 2.2, so we focus on the bound for E_1. To control $\|U(a)\|_2$, we make use of the following tail bound (Talagrand, 1994; Hitczenko et al., 1997; Adamczak et al., 2011).

Lemma 3.8 (Tail of sum of i.i.d. sub-Weibull random variables (Adamczak et al., 2011))**.** *Let $q \in [1/2, 1]$, and $W_1, \cdots, W_n$ be i.i.d. centered random variables satisfying $\mathbb{P}[|W_1| \geq t] \leq C_1 e^{-C_2 t^q}$. Then for all $t > 0$:*

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

Lemma 3.8 is a generalization of Bernstein's inequality for ψ_q tails, with $q \in [1/2, 1]$. This lemma is stated in Adamczak et al. (2011), see Lemma 3.7 and eq. (3.7) there, and is a classical consequence of the same result for symmetric Weibull random variables (Hitczenko et al., 1997).

We fix $a \in \mathcal{S}^{d-1}$. Note that:

$$\|U(a)\|_2^2 = \sum_{i=1}^n \langle \omega_i, a\rangle^4.$$

Since $\langle \omega_i, a\rangle \overset{\mathrm{d}}{=} (\omega_i)_1$ by rotation invariance of the Haar measure on $\mathcal{S}^{d-1}$, it is easy to check that $\mathbb{E}[\langle \omega_i, a\rangle^4] = (3/d^2) \cdot d/(2+d) \leq 3/d^2$. Moreover, we have for all $t \geq 0$ (Vershynin, 2018):

$$\mathbb{P}[\langle \omega_i, a\rangle^4 \geq t] \leq 2\exp\{-Cd\sqrt{t}\}.$$

Therefore, applying Lemma 3.8 and using the union bound (recall $N \leq 5^d$), we get:

$$\mathbb{P}\left[\sup_{j\in[N]} \|U(a_j)\|_2^2 \geq \frac{3n}{d^2} + t\right] \leq 2\exp\left\{d\log 5 - C\min\left(\frac{d^4t^2}{n}, d\sqrt{t}\right)\right\}.$$

Taking e.g. $t = (2\log 5/C)^2$, and since $d^4/n = \omega(d)$, we reach the conclusion.

3.6. *Proof of Lemma 2.5*. Note that $\tilde{q}_i = 1/d_i - 1 \overset{\mathrm{d}}{=} d/\chi_d^2 - 1$. We let $r_i \coloneqq \tilde{q}_i|A_i$. The A_i are independent, and by Corollary 3.7, $\mathbb{P}[A_i] \geq 1 - 2\exp(-d/16)$. By the law of total expectation and the union bound, we thus have:

$$\mathbb{P}\left[\max_{j\in[N]}\left|\sum_{i=1}^n (\Theta^{-1}\tilde{q})_i\langle a_j, \omega_i\rangle^2\right| \geq \frac{1}{2}\right] \leq \mathbb{P}\left[\max_{j\in[N]}\left|\sum_{i=1}^n (\Theta^{-1}r)_i\langle a_j, \omega_i\rangle^2\right| \geq \frac{1}{2}\right] + 2ne^{-d/16}. \tag{3.8}$$

Since $\tilde{q}_i \geq -1$, for all $x \in \mathbb{R}$: $\mathbb{P}[r_i \leq x] = \mathbb{P}[\tilde{q}_i \leq x \wedge 1]/\mathbb{P}[\tilde{q}_i \leq 1]$, and thus for all $x \in (0,1)$, by Corollary 3.7:

$$\begin{cases} \mathbb{P}[r_i \geq x] & \leq \mathbb{P}[\tilde{q}_i \geq x] \leq 2e^{-dx^2/16}, \\ \mathbb{P}[r_i \leq -x] & \leq \dfrac{\mathbb{P}[\tilde{q}_i \leq -x]}{1 - 2e^{-d/16}} \leq 4e^{-dx^2/16}. \end{cases}$$

Moreover, $\mathbb{P}[|r_i| > 1] = 0$. r_i are thus i.i.d. sub-Gaussian random variables, with sub-Gaussian norm smaller than $K/\sqrt{d}$. Moreover, by the law of total expectation:

$$\mathbb{E}[\tilde{q}_i] = \mathbb{E}[r_i]\mathbb{P}(A_i) + \mathbb{E}[\tilde{q}_i \mathbb{1}\{|\tilde{q}_i| \geq 1\}],$$

so that since $\mathbb{P}[A_i] \geq 1 - 2e^{-d/16}$, and using Cauchy-Schwarz:

$$\begin{aligned} |\mathbb{E}[\tilde{q}_i] - \mathbb{E}[r_i]| &\leq |\mathbb{E}[r_i]| \cdot 2e^{-d/16} + \sqrt{2}\mathbb{E}[\tilde{q}_i^2]^{1/2}e^{-d/32}, \\ &\overset{\text{(a)}}{\leq} 2e^{-d/16} + Ce^{-d/32}/\sqrt{d}, \end{aligned}$$

using in (a) that $|r_i| \leq 1$ and that $\mathbb{E}[\tilde{q}_i^2]^{1/2} \leq C/\sqrt{d}$. Since $\mathbb{E}\tilde{q}_i = 2/(d-2)$, we get

$$|\mathbb{E}r_i| \leq \frac{3}{d}.$$

Recall that $y_i = r_i - \mathbb{E}r_i$. Therefore we have, for all $a \in \mathcal{S}^{d-1}$:

$$\begin{aligned} \left|\sum_{i=1}^n [\Theta^{-1}(y-r)]_i\langle \omega_i, a\rangle^2\right| &\leq \|\mathbb{E}r\|_2\|\Theta^{-1}\|_{\mathrm{op}}\|U(a)\|_2, \\ &\leq \frac{3\sqrt{n}}{d}\|\Theta^{-1}\|_{\mathrm{op}}\|U(a)\|_2. \end{aligned}$$

Using Lemma 2.4, it is clear that if $n \leq \alpha d^2$ for $\alpha = \alpha(\beta) > 0$ small enough, we have

$$\mathbb{P}\left[\max_{j\in[N]}\left|\sum_{i=1}^n (\Theta^{-1}r)_i\langle a_j, \omega_i\rangle^2\right| \geq \frac{1}{2}\right] \leq \mathbb{P}\left[\max_{j\in[N]}\left|\sum_{i=1}^n (\Theta^{-1}y)_i\langle a_j, \omega_i\rangle^2\right| \geq \frac{1}{4}\right] + Cn^{-\beta}. \tag{3.9}$$

Combining eqs. (3.8) and eq. (3.9) gives the sought result. Finally, $(y_i)_{i=1}^n$ are i.i.d. centered sub-Gaussian random variables with sub-Gaussian norm $K/\sqrt{d}$. $\square$

3.7. *Proof of Lemma 2.6.* Let $M \in \mathcal{S}_n$, and denote $z \coloneqq My$. By Hoeffding's inequality, for all $i \in [n]$:

$$\mathbb{P}[|z_i| \geq t] \leq 2\exp\left\{-\frac{Cdt^2}{\|M_i\|_2^2}\right\} \leq 2\exp\left\{-\frac{Cdt^2}{\|M\|_{\mathrm{op}}^2}\right\},$$

with $(M_i)_{i=1}^n$ the rows of M, since $\|M\|_{\mathrm{op}} \geq \max_{i\in[n]} \|M_i\|_2$. Thus by the union bound:

$$\mathbb{P}[\|z\|_\infty \geq t] \leq 2n\exp\left\{-\frac{Cdt^2}{\|M\|_{\mathrm{op}}^2}\right\}.$$

Letting $t = C\|M\|_{\mathrm{op}}d^{-3/8}$ ends the proof. $\square$

3.8. *Proof of Lemma 2.7.* Let $q \in [1/2, 1]$. Recall that

$$\sum_{i\notin S(\eta)} \langle\omega_i, a\rangle^4 = \sum_{i=1}^n \langle\omega_i, a\rangle^4 \mathbb{1}\{|\langle\omega_i, a\rangle| \leq \eta\}$$

We let $z_i \coloneqq \langle\omega_i, a\rangle^4 \mathbb{1}\{|\langle\omega_i, a\rangle| \leq \eta\}$. They are i.i.d. random variables, with $\mathbb{E}[z_i] \leq \mathbb{E}[\langle\omega_i, a\rangle^4] \leq 3/d^2$, and for all $t \geq 0$:

$$\begin{aligned}\mathbb{P}[z_i \geq t] &\leq \min\left[2e^{-Cd\sqrt{t}}, \mathbb{1}\{t^{1/4} \leq \eta\}\right], \\ &\leq 2\exp\left\{-Cd\eta^{2-4q}t^q\right\}.\end{aligned}$$

Consequently $z_i' = z_i d^{1/q}\eta^{2/q-4}$ satisfy $\mathbb{P}[z_i' \geq t] \leq 2\exp\{-Ct^q\}$. We use again Lemma 3.8 to get:

$$\mathbb{P}\left[\sum_{i=1}^n z_i \geq n\mathbb{E}[z_i] + nd^{-1/q}\eta^{-2/q+4}t\right] \leq 2\exp\{-C_q\min(nt^2, (nt)^q)\}.$$

This last inequality can be rewritten as, for all $v \geq 0$:

$$\mathbb{P}\left[\sum_{i=1}^n z_i \geq \frac{n}{d^2}(3+v)\right] \leq 2\exp\left\{-C_q\min\left(nd^{-4+2/q}\eta^{4/q-8}v^2, n^q d^{1-2q}\eta^{2-4q}v^q\right)\right\}. \quad \square$$

Acknowledgements

The authors are grateful to Tim Kunisky, from whom they learned about this problem, and to Joel Tropp for insightful discussions. We would also like to thank the anonymous referees for useful suggestions.

References

Adamczak, R. A note on the Hanson-Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.*, **20**, no. 72, 13 (2015). MR3407216.

Adamczak, R., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constr. Approx.*, **34** (1), 61–88 (2011). MR2796091.

Amelunxen, D., Lotz, M., McCoy, M. B., and Tropp, J. A. Living on the edge: phase transitions in convex programs with random data. *Inf. Inference*, **3** (3), 224–294 (2014). MR3311453.

Bartl, D. and Mendelson, S. Random embeddings with an almost Gaussian distortion. *Adv. Math.*, **400**, Paper No. 108261, 31 (2022). MR4386543.

Ghosh, M., Jeronimo, F. G., Jones, C., Potechin, A., and Rajendran, G. Sum-of-squares lower bounds for Sherrington-Kirkpatrick via planted affine planes. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pp. 954–965. IEEE Computer Soc., Los Alamitos, CA (2020). MR4232101.

Gordon, Y. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric aspects of functional analysis (1986/87)*, volume 1317 of *Lecture Notes in Math.*, pp. 84–106. Springer, Berlin (1988). MR950977.

Hitczenko, P., Montgomery-Smith, S. J., and Oleszkiewicz, K. Moment inequalities for sums of certain independent symmetric random variables. *Studia Math.*, **123** (1), 15–42 (1997). MR1438303.

Hsieh, J.-T., Kothari, P. K., Potechin, A., and Xu, J. Ellipsoid fitting up to a constant. In *50th International Colloquium on Automata, Languages, and Programming*, volume 261 of *LIPIcs. Leibniz Int. Proc. Inform.*, pp. Art. No. 78, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2023). MR4612948.

Kane, D. and Diakonikolas, I. A nearly tight bound for fitting an ellipsoid to Gaussian random points. In Neu, G. and Rosasco, L., editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 3014–3028. PMLR (2023). Available at https://proceedings.mlr.press/v195/kane23a.html.

Klenke, A. and Mattner, L. Stochastic ordering of classical discrete distributions. *Adv. in Appl. Probab.*, **42** (2), 392–410 (2010). MR2675109.

Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, **28** (5), 1302–1338 (2000). MR1805785.

Liaw, C., Mehrabian, A., Plan, Y., and Vershynin, R. A simple tool for bounding the deviation of random matrices on geometric sets. In *Geometric aspects of functional analysis*, volume 2169 of *Lecture Notes in Math.*, pp. 277–299. Springer, Cham (2017). MR3645128.

Mitzenmacher, M. and Upfal, E. *Probability and computing. Randomization and probabilistic techniques in algorithms and data analysis.* Cambridge University Press, Cambridge, second edition (2017). ISBN 978-1-107-15488-9. MR3674428.

Podosinnikova, A., Perry, A., Wein, A. S., Bach, F., d'Aspremont, A., and Sontag, D. Overcomplete Independent Component Analysis via SDP. In Chaudhuri, K. and Sugiyama, M., editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 2583–2592. PMLR (2019). https://proceedings.mlr.press/v89/podosinnikova19a.html.

Potechin, A., Turner, P. M., Venkat, P., and Wein, A. S. Near-optimal fitting of ellipsoids to random points. In Neu, G. and Rosasco, L., editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 4235–4295. PMLR (2023). https://proceedings.mlr.press/v195/potechin23a.html.

Saunderson, J. *Subspace identification via convex optimization.* Ph.D. thesis, Massachusetts Institute of Technology (2011).

Saunderson, J., Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM J. Matrix Anal. Appl.*, **33** (4), 1395–1416 (2012). MR3028972.

Saunderson, J., Parrilo, P. A., and Willsky, A. S. Diagonal and low-rank decompositions and fitting ellipsoids to random points. In *52nd IEEE Conference on Decision and Control*, pp. 6031–6036 (2013). DOI: 10.1109/CDC.2013.6760842.

Talagrand, M. The supremum of some canonical processes. *Amer. J. Math.*, **116** (2), 283–325 (1994). MR1269606.

Tulsiani, M. and Wu, J. Ellipsoid fitting up to constant via empirical covariance estimation. *ArXiv Mathematics e-prints* (2023). arXiv: 2307.10941.

Vershynin, R. *High-dimensional probability. An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University

Press, Cambridge (2018). ISBN 978-1-108-41519-4. MR3837109.