Received: 29 June 2022 | Accepted: 3 March 2023

DOI: 10.1002/rsa.21178

RESEARCH ARTICLE

WILEY

Threshold for detecting high dimensional geometry in anisotropic random geometric graphs

Matthew Brennan | **Guy Bresler** | **Brice Huang**

Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Correspondence
Brice Huang, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA.
Email: bmhuang@mit.edu

Funding information
NSF CAREER award, Grant/Award Number: CCF-1940205; NSF Graduate Research Fellowship; Siebel Scholarship; NSF Award, Grant/Award Number: DMS-2022448.

Abstract

In the anisotropic random geometric graph model, vertices correspond to points drawn from a high-dimensional Gaussian distribution and two vertices are connected if their distance is smaller than a specified threshold. We study when it is possible to hypothesis test between such a graph and an Erdős-Rényi graph with the same edge probability. If n is the number of vertices and α is the vector of eigenvalues, Eldan and Mikulincer, Geo. Aspects Func. Analysis: Israel seminar, 2017 shows that detection is possible when $n^3 \gg (\|\alpha\|_2/\|\alpha\|_3)^6$ and impossible when $n^3 \ll (\|\alpha\|_2/\|\alpha\|_4)^4$. We show detection is impossible when $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$, closing this gap and affirmatively resolving the conjecture of Eldan and Mikulincer, Geo. Aspects Func. Analysis: Israel seminar, 2017.

KEYWORDS

Erdős-Rényi graph, high dimensional geometry, hypothesis testing, random geometric graph, Wishart matrix

1 | INTRODUCTION

Extracting information from large graphs is a fundamental statistical task. Because many natural networks have underlying metric structure—for example, nearby proteins in a biological network are more likely to share function, and users with similar interests in a social network are more likely to interact—a central inference problem is to infer latent geometric structure in an observed graph. Moreover, with the proliferation of large data sets in the modern world, statistical inference is inherently

high dimensional, see, for example, the survey [10]. This motivates the study of inferring latent high dimensional geometry in a graph.

In this paper, we consider the hypothesis testing problem that determines if such inference is information-theoretically possible. This continues a line of work originated by Bubeck, Ding, Eldan, and Rácz [3] and continued by Eldan and Mikulincer [8]. Our main contribution is a tight characterization of when detection is possible in the anisotropic setting introduced in [8].

Formally, given a graph G on $[n]$, we wish to test between two hypotheses. The null hypothesis is that G is a sample from the Erdős-Rényi graph $G(n,p)$, where each edge is present with independent probability p. The alternative hypothesis is that G is a sample from a random geometric graph (RGG), which we define precisely below. In such graphs, each vertex corresponds to a random point in some metric space and an edge exists between two vertices if their distance is smaller than a given threshold.

A natural RGG is the isotropic model: each vertex $i \in [n]$ corresponds to an independent latent vector X_i sampled from the Haar measure on the sphere $\mathbb{S}^{d-1}$ or an isotropic d-dimensional Gaussian,[1] and edge (i,j) is present if $\langle X_i, X_j\rangle \geq t_{p,d}$, where $t_{p,d}$ is chosen so that each edge is present with probability p. Let $G(n,p,d)$ denote the isotropic RGG with spherical latent data; we fix $p \in (0,1)$ and allow d to vary with n. The following seminal result of Bubeck, Ding, Eldan, and Rácz characterizes, for fixed $p \in (0,1)$, when it is possible to test between $G(n,p)$ and $G(n,p,d)$. Let TV denote total variation distance.

Theorem 1. [3] Let $p \in (0,1)$ be fixed.

(a) If $n^3 \ll d$, then $\mathsf{TV}(G(n,p), G(n,p,d)) \to 0$.
(b) If $n^3 \gg d$, then $\mathsf{TV}(G(n,p), G(n,p,d)) \to 1$.

Each coordinate of the latent vector X_i represents an attribute of vertex i. The isotropic model assumes that each attribute has the same influence on the network structure. In real networks, some attributes are more important than others: for example, in a social network geographic location has a stronger influence on connectivity than preference of ice cream flavor. This motivates the following anisotropic generalization of the RGG, introduced in [8], in which attributes may have different weights.

Definition 1.1 (Anisotropic random geometric graph)**.** For $n \in \mathbb{N}$, $p \in (0,1)$, $\alpha \in \mathbb{R}^d_{\geq 0}$, let $G(n,p,\alpha)$ be the following random graph model. Generate $X_1, \ldots, X_n$ i.i.d. from $\mathcal{N}(0, D_\alpha)$ where D_α is the diagonal matrix with diagonal α. Let $t_{p,\alpha} \in \mathbb{R}$ be the unique number satisfying $\mathbb{P}(\langle X_1, X_2\rangle \geq t_{p,\alpha}) = p$. Then, $G \sim G(n,p,\alpha)$ is a graph on $[n] = \{1, \ldots, n\}$ where (i,j) is an edge if and only if $\langle X_i, X_j\rangle \geq t_{p,\alpha}$.

By rotational invariance of the model, the assumption that X_i has diagonal covariance is without loss of generality. Thus, all our results apply to the case of latent Gaussian vectors with arbitrary covariance.

Throughout, we fix $p \in (0,1)$ and allow d, α to vary with n. The central question we study is, under what limiting behaviors of (n,d,α) can one statistically distinguish $G(n,p,\alpha)$ from $G(n,p)$? This question was first studied in [8], in which the following upper and lower bounds on detection were derived.

Theorem 2. [[8], Theorem 2] Let $p \in (0,1)$ be fixed. Then,

(a) If $n^3 \ll (\|\alpha\|_2/\|\alpha\|_4)^4$, then $\mathsf{TV}(G(n,p), G(n,p,\alpha)) \to 0$.
(b) If $n^3 \gg (\|\alpha\|_2/\|\alpha\|_3)^6$, then $\mathsf{TV}(G(n,p), G(n,p,\alpha)) \to 1$.

[1]These models have the same threshold for the detection task we consider.

When $\alpha = 1^d$, this recovers the sharp $d \asymp n^3$ detection threshold in the isotropic model. However, for general α there is a polynomially sized gap between the upper and lower bounds of Theorem 2. For example, if $\alpha_i = i^{-1/3}$, $(\|\alpha\|_2/\|\alpha\|_4)^4 \asymp d^{2/3}$ while $(\|\alpha\|_2/\|\alpha\|_3)^6 \asymp d$. [[8], Conjecture 1] conjectures that the hypothesis of part (1) can be weakened to $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$, that is, the detection threshold is $n^3 \asymp (\|\alpha\|_2/\|\alpha\|_3)^6$. The main result of this paper is to affirmatively resolve this conjecture.

Theorem 3 (Main result)**.** If $p \in (0,1)$ is fixed and $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$, then $\mathsf{TV}(G(n,p), G(n,p,\alpha)) \to 1$.

In light of Theorem 1, this result can be interpreted as meaning that for the task of detecting geometry, the effective dimension of the anisotropic RGG is $(\|\alpha\|_2/\|\alpha\|_3)^6$.

One motivation for [[8], Conjecture 1] is that Theorem 2 (b) is witnessed by the *signed triangles statistic*

$$\theta(G) = \sum_{i<j<k} (G_{i,j} - p)(G_{i,k} - p)(G_{j,k} - p),$$

which is also an optimal statistic witnessing Theorem 1 (b). So, Theorem 3 confirms the optimality of the signed triangles statistic in the anisotropic setting.

1.1 | Central limit theorem for anisotropic Wishart matrices

Closely related to the anisotropic RGG is the following matrix of inner products generating $G(n,p,\alpha)$. A sample of $G(n,p,\alpha)$ can be obtained by thresholding each entry of this matrix at $t_{p,\alpha}$.

Definition 1.2 (Diagonal-removed anisotropic Wishart matrix)**.** Let $W \sim W(n,\alpha)$ be the random $n \times n$ matrix generated as follows. Sample $X \in \mathbb{R}^{d\times n}$ with i.i.d. columns from $\mathcal{N}(0, D_\alpha)$, and set $W = \|\alpha\|_2^{-1}\left(X^\top X - \mathrm{diag}(X^\top X)\right)$.

For fixed n, if $d \to \infty$ and $\|\alpha\|_\infty/\|\alpha\|_2 \to 0$, by the multidimensional CLT $W(n,\alpha)$ converges in total variation to the following matrix of Gaussians.

Definition 1.3. Let $M \sim M(n)$ be a symmetric random $n \times n$ matrix with zero diagonal and i.i.d. standard Gaussians above the diagonal.

If we now allow d, α to vary with n, a natural question is, for which (n,d,α) can one test between $W(n,\alpha)$ and $M(n)$? This can be regarded as the random matrix analog of the question of detecting geometry in random graphs. Eldan and Mikulincer obtain the following detection lower bound.

Theorem 4. [[8], Theorem 4] If $n^3 \ll (\|\alpha\|_2/\|\alpha\|_4)^4$, then $\mathsf{KL}(W(n,\alpha), M(n)) \to 0$.

Of course, by Pinsker's inequality this also implies $\mathsf{TV}(W(n,\alpha), M(n)) \to 0$. Furthermore, the statistic $\theta(M) = \mathsf{tr}(M^3)$ distinguishes $W(n,\alpha)$ and $M(n)$ to total variation distance $1 - o(1)$ when $n^3 \gg (\|\alpha\|_2/\|\alpha\|_3)^6$, which can be verified by computing the mean and variance of this statistic under the two hypotheses.

Similarly to above, these upper and lower bounds match for $\alpha = 1^d$, but in general there is a polynomially sized gap between them. We prove the following result, which identifies the sharp threshold for this detection task by improving the lower bound in Theorem 4. This result can be regarded as a tight CLT for anisotropic Wishart matrices.

Theorem 5. If $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$, then $\mathsf{TV}(W(n,\alpha), M(n)) \to 0$.

1.2 | Techniques and discussion

Theorem 3 follows from Theorem 5 by the thresholding idea introduced in [3]. Note that $G(n,p,\alpha)$ and $G(n,p)$ are entry-wise thresholdings of $W(n,\alpha)$ and $M(n)$. Thus $\mathsf{TV}(G(n,p,\alpha),G(n,p))$ is upper bounded by $\mathsf{TV}(W(n,\alpha),M(n))$ plus a small error term from the difference of the thresholds.

Our main technical contributions are in the proof of Theorem 5. We divide the entries of α into large coordinates α^+ and small coordinates α^-, each accounting for a constant fraction of its L^2 mass. We note (Lemma 2.4) that $(\|\alpha^-\|_2/\|\alpha^-\|_4)^4$ is of the same order as $(\|\alpha\|_2/\|\alpha\|_3)^6$, so Theorem 4 is sufficient to show that $W(n,\alpha^-)$ converges in total variation to $M(n)$.

It remains to control the contributions of the large coordinates α^+. We consider a procedure (Lemma 2.3) where we add the coordinates of α^+ to α^- one by one. Note that the effect of this operation on $W\sim W(n,\alpha)$ is to add an independent rank-one spike and scale down by a constant. By a data processing argument, the increase in $\mathsf{TV}(W(n,\alpha),M(n))$ from one step of this procedure is bounded by $\mathsf{TV}(M(n,u),M(n))$, where $M(n,u)$ is a linear combination of $M(n)$ and an independent rank-one Gaussian spike (see Definition 2.1).

This last quantity is bounded (Lemma 2.2) using the Ingster-Suslina χ^2 method, as $M(n,u)$ is a mixture of shifted Gaussian matrices parametrized by the spike. This is done after conditioning on a high probability event under which the χ^2 divergence's tails are integrable. The resulting χ^2 divergence is an expectation over two independent copies of the Gaussian spike, which is bounded by hypercontractivity estimates.

Our iterative method bears some resemblance to the iterative argument used in [4] to bound the KL divergence between a diagonal-removed (isotropic) Wishart matrix $W=X^\top X-\mathrm{diag}(X^\top X)$ and a symmetric matrix of independent Gaussians, which also constructs the data matrix X one vector at a time. In [4], X is constructed one *column* at a time, and each new column adds a new *row and column* to W. In contrast, we construct X one *row* at a time, and each new row adds a new *rank-one component* to W. The methods used to control the contribution of each step to the total variation distance (or in [4], the KL divergence) are consequently different.

1.3 | Related work

There is a long history of work on low-dimensional random geometric graphs, see for example, [16]. The study of high-dimensional random geometric graphs began in [7], which showed that the isotropic model $G(n,p,d)$ converges in total variation to $G(n,p)$ as $d\to\infty$ for n fixed, and moreover that their clique numbers converge if $d\gg\log^3 n$. [3] showed Theorem 1, that the threshold for convergence of $G(n,p,d)$ and $G(n,p)$ with p fixed is $d\asymp n^3$. They conjectured that if $p=o(1)$, convergence occurs at smaller d; in particular, for $p=c/n$ they conjectured the threshold $d\asymp\log^3 n$. [2] proved convergence occurs when $d=\tilde{\omega}(n^3p,n^{7/2}p^2)$, meaning the threshold does decrease with p. Recently [11] proved that for $p=c/n$ ($c\geq 1$), convergence occurs when $d\gtrsim\log^{36}n$, resolving the conjecture of [3] up to polylog factors. In a different direction, [12] obtain detection upper and lower bounds for soft random geometric graphs, wherein the inner products $\langle X_i,X_j\rangle$ determine the probability of edge (i,j) being present.

There is also a growing literature on CLTs for random matrices. [5] proved a general multidimensional CLT using Stein's method. [9] and [3] concurrently showed that $W(n,d)\triangleq W(n,1^d)$ converges in total variation to $M(n)$ if $d\gg n^3$. [4] generalized this result to arbitrary log-concave entry distributions, showing that the random matrix $W=d^{-1/2}(X^\top X-\mathrm{diag}(X^\top X))$, where $X\in\mathbb{R}^{d\times n}$ has i.i.d. entries from a log-concave measure, converges to $M(n)$ if $d/\log^2 d\gg n^3$. [17] refined the result of [3,9] by

computing the limiting value of $\mathsf{TV}(W(n,d), M(n))$ if $n, d \to \infty$ with $d/n^3 \to c$. [6] showed a countable sequence of phase transitions for the Wishart ensemble $W(n,d)$: for each $k \in \mathbb{N}$, if $n^{k+3} \gg d^{k+1}$, they show that $W(n,d)$ converges to an explicit density f_k. CLTs have been shown for Wishart tensors [13] and Wishart matrices with arbitrary deleted entries [1]. Finally, [14] considers Wishart matrices $W = \sqrt{d}(d^{-1}X^\top X - I_n)$ where the columns of X are drawn i.i.d. from $\mathcal{N}(0,\Sigma)$ for $\Sigma \in \mathbb{R}^{d\times d}$ of the form

$$\Sigma_{i,j} = s(i-j),$$

where $s : \mathbb{Z} \to \mathbb{R}$ is a covariance function with $s(0) = 1$. They show that W converges in Wasserstein distance to a Gaussian matrix if $n^3 \ll d$ and $s \in \ell^{4/3}(\mathbb{Z})$, and under various conditions if s is the correlation function of a fractional Brownian noise.

1.4 | Notation and preliminaries

We adopt standard asymptotic notations: $f \gg g$ means that $f/g \to \infty$ and $f \gtrsim g$ means that $f \geq cg$ for an absolute constant c. Throughout, $c, C > 0$ denote universal constants that may change from line to line.

We use TV, KL, and χ^2 to denote total variation, Kullback-Leibler divergence, and chi-square divergence. That is, for measures ν, μ with ν absolutely continuous with respect to μ,

$$\mathsf{TV}(\nu,\mu) = \frac{1}{2}\mathbb{E}_{\xi\sim\mu}\left|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\xi) - 1\right|, \quad \mathsf{KL}(\nu,\mu) = \mathbb{E}_{\xi\sim\mu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\xi)\log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\xi), \quad \chi^2(\nu,\mu) = \mathbb{E}_{\xi\sim\mu}\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\xi) - 1\right)^2.$$

We recall that TV satisfies the triangle inequality and the data processing inequality $\mathsf{TV}(\mathcal{K}(\nu), \mathcal{K}(\mu)) \leq \mathsf{TV}(\nu,\mu)$ for any Markov kernel $\mathcal{K}$. We also recall the Cauchy-Schwarz inequality $4\mathsf{TV}(\nu,\mu)^2 \leq \chi^2(\nu,\mu)$.

2 | PROOF OF MAIN RESULTS

For $g \in \mathbb{R}^n$, let $\Delta(g) = (gg^\top - \mathrm{diag}(gg^\top))$. We introduce the following random matrix, consisting of a linear combination of a sample from $M(n)$ and a rank-one Gaussian spike (with diagonal removed).

Definition 2.1. For $u \in [0,1]$, let $M \sim M(n,u)$ be generated by

$$M = u\Delta(g) + \sqrt{1-u^2}M', \tag{1}$$

where $g \sim \mathcal{N}(0, I_n)$ and $M' \sim M(n)$ are independent.

We defer the proof of the following lemma to Section 3. Using this lemma, we prove Theorems 3 and 5.

Lemma 2.2. We have that $\mathsf{TV}(M(n,u), M(n)) \lesssim u^3 n^{3/2}$.

2.1 | Detection lower bound for anisotropic Wishart matrices

We first prove Theorem 5. Assume without loss of generality that $\alpha_1 \geq \cdots \geq \alpha_d \geq 0$, and that $\|\alpha\|_2 = 1$. Define $\alpha^+ = (\alpha_1, \ldots, \alpha_r)$ and $\alpha^- = (\alpha_{r+1}, \ldots, \alpha_d)$, for the smallest r such that $\|\alpha^+\|_2^2 \geq \frac{1}{3}$.

We may assume

$$\frac{1}{3} \le \|\alpha^+\|_2^2 \le \frac{2}{3}. \tag{2}$$

Indeed, if $\|\alpha^+\|_2^2 > \frac{2}{3}$ we in fact have $r = 1$ because the α_i are decreasing. Thus $\alpha_1^2 > \frac{2}{3}$, which implies $\|\alpha\|_3 = \Theta(1)$. The condition $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$ is then vacuous so there is nothing to prove. Henceforth we assume (2). For $t = 0, 1, \dots, r$, define

$$\alpha^t = (\alpha_1, \dots, \alpha_t, \alpha_{r+1}, \dots, \alpha_d).$$

These interpolate between α^- and α in the sense that $\alpha^0 = \alpha^-$, $\alpha^r = \alpha$.

Lemma 2.3. For each $t = 1, \dots, r$,

$$\mathsf{TV}(W(n, \alpha^t), M(n)) \le \mathsf{TV}(W(n, \alpha^{t-1}), M(n)) + C\alpha_t^3 n^{3/2}.$$

Proof. Let $u_t = \alpha_t/\|\alpha^t\|_2$. By the triangle inequality,

$$\mathsf{TV}(W(n, \alpha^t), M(n)) \le \mathsf{TV}(W(n, \alpha^t), M(n, u_t)) + \mathsf{TV}(M(n, u_t), M(n)). \tag{3}$$

For $M \in \mathbb{R}^{n\times n}$, define the Markov kernel

$$\mathcal{K}(M) = u_t \Delta(g) + \sqrt{1 - u_t^2} M$$

where $g \sim \mathcal{N}(0, I_n)$. Note that $W \sim W(n, \alpha^t)$, $M \sim M(n, u_t)$ can be generated by $W = \mathcal{K}(W')$, $M = \mathcal{K}(M')$ for $W' \sim W(n, \alpha^{t-1})$, $M' \sim M(n)$. By data processing,

$$\mathsf{TV}(W(n, \alpha^t), M(n, u_t)) \le \mathsf{TV}(W(n, \alpha^{t-1}), M(n)).$$

The remaining term in (3) can be bounded by Lemma 2.2:

$$\mathsf{TV}(M(n, u_t), M(n)) \lesssim u_t^3 n^{3/2} = \frac{\alpha_t^3}{\|\alpha^t\|_2^3} n^{3/2} \lesssim \alpha_t^3 n^{3/2}$$

where the final inequality uses that $\|\alpha^t\|_2 \ge \|\alpha^-\|_2 \gtrsim 1$, by (2). ∎

Lemma 2.4. We have that $(\|\alpha^-\|_2/\|\alpha^-\|_4)^4 \gtrsim \|\alpha\|_3^{-6}$.

Proof. Note that

$$\|\alpha^+\|_2^2 \|\alpha^-\|_4^4 = \sum_{i=1}^{r}\sum_{j=r+1}^{d} \alpha_i^2 \alpha_j^4 \le \sum_{i=1}^{r}\sum_{j=r+1}^{d} \alpha_i^3 \alpha_j^3 \le \|\alpha\|_3^6,$$

where the first inequality uses that $\alpha_i > \alpha_j$ because α is decreasing. This rearranges as

$$(\|\alpha^-\|_2/\|\alpha^-\|_4)^4 \ge \|\alpha^+\|_2^2 \|\alpha^-\|_2^4 \|\alpha\|_3^{-6} \gtrsim \|\alpha\|_3^{-6},$$

where the final inequality uses that $\|\alpha^+\|_2, \|\alpha^-\|_2 \gtrsim 1$, by (2). ∎

Proof of Theorem 5. By applying Lemma 2.3 repeatedly, we get

$$\begin{aligned}\mathsf{TV}(W(n,\alpha),M(n)) &\le \mathsf{TV}(W(n,\alpha^-),M(n)) + C\sum_{t=1}^{r}\alpha_t^3 n^{3/2}\\ &\le \mathsf{TV}(W(n,\alpha^-),M(n)) + C\|\alpha\|_3^3 n^{3/2}. \end{aligned} \tag{4}$$

The hypothesis $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6 = \|\alpha\|_3^{-6}$ implies the second term of (4) is $o(1)$. By Lemma 2.4 we further have $n^3 \ll (\|\alpha^-\|_2/\|\alpha^-\|_4)^4$. Therefore Theorem 4 and Pinsker's inequality imply that the first term of (4) is $o(1)$. This concludes the proof. ∎

Remark 2.5. We conjecture that Theorem 5 remains true if the diagonal is not removed, that is, if $W(n,\alpha)$ and $M(n)$ are replaced by the law $W^*(n,\alpha)$ of $W = \|\alpha\|_2^{-1}(X^\top X - \|\alpha\|_1 I_n)$, where X is as in Definition 1.2, and the law $M^*(n)$ of a GOE matrix. With minor modifications, the proof of Lemma 2.2 in the next section generalizes if the diagonal is not removed, that is, if $\Delta(g)$ and $M(n)$ are replaced by $gg^\top - I_n$ and $M^*(n)$. So, if Theorem 4 holds without diagonal removal, the above proof can be easily modified to conclude Theorem 5 without diagonal removal. The difficulty is that the entropy chain rule argument used to prove Theorem 4 requires the diagonal to be removed.

Remark 2.6. The convergence in Theorem 5 is with respect to total variation distance, whereas that of Theorem 4 is with respect to KL divergence. We expect that Theorem 5 remains true when total variation distance is replaced by KL divergence, though our methods do not show this because our iterative argument relies on the TV triangle inequality (3).

2.2 | Detection lower bound for anisotropic RGGs

The proof of Theorem 3 is identical to that of Theorem 2 (a) (Theorem 2(b) in [8]), using Theorem 5 in place of Theorem 4.

Proof of Theorem 3. Define the threshold functions $H_{p,\alpha}, K_p : \mathbb{R} \to \{0,1\}$ by

$$H_{p,\alpha}(x) = \mathbb{1}\{x \ge t_{p,\alpha}\}, \qquad K_p(x) = \mathbb{1}\{x \ge \Phi^{-1}(p)\},$$

where $t_{p,\alpha}$ is defined in Definition 1.1 and $\Phi(t) = \mathbb{P}_{Z\sim\mathcal{N}(0,1)}(Z \ge t)$ is the complement of the cdf of the standard Gaussian. Then, $G(n,p,\alpha)$ and $G(n,p)$ can be generated as the following entry-wise thresholdings of $W(n,\alpha)$ and $M(n)$:

$$G(n,p,\alpha) = H_{p,\alpha}(W(n,\alpha)), \qquad G(n,p) = K_p(M(n)).$$

Using the TV triangle inequality and data processing inequality,

$$\begin{aligned}\mathsf{TV}(G(n,p,\alpha),G(n,p)) &\le \mathsf{TV}(H_{p,\alpha}(W(n,\alpha)),H_{p,\alpha}(M(n))) + \mathsf{TV}(H_{p,\alpha}(M(n)),K_p(M(n)))\\ &\le \mathsf{TV}(W(n,\alpha),M(n)) + \mathsf{TV}(H_{p,\alpha}(M(n)),K_p(M(n))). \end{aligned} \tag{5}$$

Since $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$, Theorem 5 implies that the first term of (5) is $o(1)$. The second term of (5) is $o(1)$ by [[8], Lemma 16]. Indeed, the proof of this lemma proceeds identically if the hypothesis $n^3 \ll (\|\alpha\|_2/\|\alpha\|_4)^4$ is weakened to $n^3 \ll (\|\alpha\|_2/\|\alpha\|_3)^6$. ∎

3 | PROOF OF TV BOUND FOR SPIKED GAUSSIAN MATRIX

In this section, we will prove Lemma 2.2. Let $M(n,u,g)$ be the random matrix M generated by (1) for $g \in \mathbb{R}^n$ fixed and $M' \sim M(n)$. Thus $M(n,u)$ is a mixture of the random matrices $M(n,u,g)$ over latent randomness $g \sim \mathcal{N}(0,I_n)$.

Further, for (always measurable and high probability) $S \subseteq \mathbb{R}^n$, let μ_S be the law of $g \sim \mathcal{N}(0,I_n)$ conditioned on $g \in S$. Let $M(n,u,S)$ be the law of M generated by (1) for $g \sim \mu_S$ and $M' \sim M(n)$. This can be regarded as $M(n,u)$ conditioned on $g \in S$, and as a mixture of the $M(n,u,g)$ over $g \sim \mu_S$.

We begin with the following series of estimates. Let $S \subseteq \mathbb{R}^n$ be a set we will specify later.

$$\begin{aligned}\mathsf{TV}(M(n,u),M(n)) &\le \mathsf{TV}(M(n,u),M(n,u,S)) + \mathsf{TV}(M(n,u,S),M(n)) \qquad (6)\\ &\le \mathbb{P}_{g\sim\mathcal{N}(0,I_n)}(g \in S^c) + \mathsf{TV}(M(n,u,S),M(n)).\\ 4\mathsf{TV}(M(n,u,S),M(n))^2 &\le \chi^2(M(n,u,S),M(n)) \qquad (7)\\ &= -1 + \mathbb{E}_{A\sim M(n)}\left(\mathbb{E}_{g\sim\mu_S}\frac{\mathrm{d}M(n,u,g)}{\mathrm{d}M(n)}(A)\right)^2\\ &= -1 + \mathbb{E}_{g,h\sim\mu_S}\mathbb{E}_{A\sim M(n)}\frac{\mathrm{d}M(n,u,g)}{\mathrm{d}M(n)}(A)\frac{\mathrm{d}M(n,u,h)}{\mathrm{d}M(n)}(A).\end{aligned}$$

The two estimates leading to (6) are by the TV triangle inequality and the data processing inequality. The estimate leading to (7) is by Cauchy-Schwarz.

These estimates are the starting point of the so-called truncated Ingster-Suslina χ^2 method. It is necessary to condition on an appropriate S in (6) so that the tails of the χ^2 divergence (7) are integrable. The following lemma evaluates the inner expectation in (7).

Lemma 3.1. For $g,h \in \mathbb{R}^n$,

$$\begin{aligned}&\mathbb{E}_{A\sim M(n)}\frac{\mathrm{d}M(n,u,g)}{\mathrm{d}M(n)}(A)\frac{\mathrm{d}M(n,u,h)}{\mathrm{d}M(n)}(A)\\ &= (1-u^4)^{-n(n-1)/4}\exp\left(\frac{u^2}{1-u^4}\sum_{1\le i<j\le n} g_ig_jh_ih_j - \frac{u^4}{1-u^4}\cdot\frac{1}{2}\sum_{1\le i<j\le n}(g_i^2g_j^2+h_i^2h_j^2)\right).\end{aligned}$$

Proof. The densities of $M(n)$ and $M(n,u,g)$ (on the subspace of symmetric matrices with zero diagonal) are

$$\begin{aligned}\frac{\mathrm{d}M(n)}{\mathrm{dLeb}}(A) &= \prod_{1\le i<j\le n}(2\pi)^{-1/2}\exp\left(-\frac{1}{2}A_{i,j}^2\right),\\ \frac{\mathrm{d}M(n,u,g)}{\mathrm{dLeb}}(A) &= \prod_{1\le i<j\le n}(2\pi\cdot(1-u^2))^{-1/2}\exp\left(-\frac{1}{2(1-u^2)}(A_{i,j}-ug_ig_j)^2\right).\end{aligned}$$

Thus

$$\begin{aligned}&\mathbb{E}_{A\sim M(n)}\frac{\mathrm{d}M(n,u,g)}{\mathrm{d}M(n)}(A)\frac{\mathrm{d}M(n,u,h)}{\mathrm{d}M(n)}(A)\\ &= (1-u^2)^{-n(n-1)/2}\prod_{1\le i<j\le n}\mathbb{E}_{A_{i,j}\sim\mathcal{N}(0,1)}\exp\left(-\frac{(A_{i,j}-ug_ig_j)^2}{2(1-u^2)}-\frac{(A_{i,j}-uh_ih_j)^2}{2(1-u^2)}+A_{i,j}^2\right).\end{aligned}$$

By a straightforward calculation the inner expectation equals

$$\sqrt{\frac{1+u^2}{1-u^2}}\exp\left(\frac{u^2}{1-u^4}g_ig_jh_ih_j-\frac{u^4}{2(1-u^4)}\left(g_i^2g_j^2+h_i^2h_j^2\right)\right)$$

from which the result follows. Note that the prefactors of $(1-u^2)^{-n(n-1)/2}$ and $\sqrt{\frac{1+u^2}{1-u^2}}$ (multiplied $\frac{n(n-1)}{2}$ times) from the previous two displays combine to produce the prefactor $(1-u^4)^{-n(n-1)/4}$. ∎

Before proceeding further, we record two consequences of Boolean and Gaussian hypercontractivity. The proof of the following lemma is standard, see [[15], Chapters 9 and 11].

Lemma 3.2. Let $f:\{-1,1\}^n\to\mathbb{R}$ be a polynomial of degree $d\geq 2$ and $\nu=\mathrm{unif}(\{-1,1\}^n)$. Further, let $\sigma^2=\mathbb{E}_{x\sim\nu}f(x)^2$.

(a) For any $k\geq 2$, $\mathbb{E}_{x\sim\nu}f(x)^k\leq d^{dk/2}\sigma^k$.

(b) There exist constants C_d,c_d such that $\mathbb{P}_{x\sim\nu}[|f(x)|\geq t\sigma]\leq C_d\exp(-c_dt^{2/d})$.

The same statements hold if $f:\mathbb{R}^n\to\mathbb{R}$ and $\nu=\mathcal{N}(0,I_n)$.

For $a\geq 1$, let

$$S(a)=\left\{g\in\mathbb{R}^n:\|g\|_2^2\leq(1+a)n,\|g\|_4^4\leq 3(1+a)n\right\}.$$

We will prove Lemma 2.2 by taking $S=S(a)$ in the calculation (7) for appropriate a. We first estimate the probability of $S(a)$.

Lemma 3.3. For $g\sim\mathcal{N}(0,I_n)$, $\mathbb{P}(g\in S(a)^c)\leq C\exp(-ca^{1/2}n^{1/4})$.

Proof. Let $f_2=\|g\|_2^2-n$ and $f_4=\|g\|_4^4-3n$. Note that $\mathbb{E}f_2^2=2n$ and $\mathbb{E}f_4^2=96n$. By Lemma 3.2 (b)

$$\mathbb{P}(\|g\|_2^2>(1+a)n)=\mathbb{P}(f_2>an)\leq C_2\exp\left(-c_2an/\sqrt{2n}\right)\leq C\exp\left(-can^{1/2}\right).$$

Similarly

$$\mathbb{P}(\|g\|_4^4>3(1+a)n)=\mathbb{P}(f_4>3an)\leq C_4\exp\left(-c_4(3an/\sqrt{96n})^{1/2}\right)\leq C\exp\left(-ca^{1/2}n^{1/4}\right).$$

∎

Let g,h be independent samples from $\mu_{S(a)}$ for a to be determined, and define the random variables

$$X=\sum_{1\leq i<j\leq n}g_ig_jh_ih_j,\qquad Y=\frac{1}{2}\sum_{1\leq i<j\leq n}\left(g_i^2g_j^2+h_i^2h_j^2\right).$$

The following two lemmas bound, respectively, the low and high moments of X and Y.

Lemma 3.4. The following estimates hold for all $a\geq 1$.

$$\mathbb{E}_{g,h\sim\mu_{S(a)}}X=0,$$
$$\mathbb{E}_{g,h\sim\mu_{S(a)}}XY=0,$$

$$\left|\mathbb{E}_{g,h\sim\mu_{S(a)}}X^2 - \frac{n(n-1)}{2}\right| \lesssim \mathbb{P}(S(a)^c)^{1/2}n^2,$$
$$\left|\mathbb{E}_{g,h\sim\mu_{S(a)}}Y - \frac{n(n-1)}{2}\right| \lesssim \mathbb{P}(S(a)^c)^{1/2}n^2,$$
$$\left|\mathbb{E}_{g,h\sim\mu_{S(a)}}X^3 - n(n-1)(n-2)\right| \lesssim \mathbb{P}(S(a)^c)^{1/2}n^3.$$

Proof. The first two claims follow by the symmetry of $S(a)$ under the map $(g_1, \dots, g_n) \mapsto (x_1 g_1, \dots, x_n g_n)$ for any $x \in \{-1, 1\}^n$. In the rest of this proof, let $\mathbb{E}$ denote expectation with respect to $g, h \sim \mathcal{N}(0, I_n)$. By straightforward calculation,

$$\mathbb{E}X^2 = \frac{n(n-1)}{2},$$
$$\mathbb{E}Y = \frac{n(n-1)}{2},$$
$$\mathbb{E}X^3 = n(n-1)(n-2).$$

We estimate the discrepancy caused by changing the measure from $\mu_{S(a)}$ to $\mathcal{N}(0, I_n)$ using the following generic bound. For any (g, h)-measurable ξ,

$$\mathbb{E}_{g,h\sim\mu_{S(a)}}\xi = \mathbb{P}(S(a))^{-2}\mathbb{E}\mathbb{1}\{g, h \in S(a)\}\xi = \mathbb{P}(S(a))^{-2}\left[\mathbb{E}\xi - \mathbb{E}\mathbb{1}\{(g, h \in S(a))^c\}\xi\right].$$

Thus

$$\begin{aligned}\left|\mathbb{E}_{g,h\sim\mu_{S(a)}}\xi - \mathbb{E}\xi\right| &\le \left(\mathbb{P}(S(a))^{-2} - 1\right)|\mathbb{E}\xi| + \mathbb{P}(S(a))^{-2}|\mathbb{E}\mathbb{1}\{(g, h \in S(a))^c\}\xi| \\ &\le \frac{2\mathbb{P}(S(a)^c) + \sqrt{2\mathbb{P}(S(a)^c)}}{\mathbb{P}(S(a))^2}\sqrt{\mathbb{E}(\xi^2)} \lesssim \sqrt{\mathbb{P}(S(a)^c)\mathbb{E}(\xi^2)}.\end{aligned}$$

For $\xi = X^2$, by Lemma 3.2 (a)

$$\mathbb{E}\xi^2 = \mathbb{E}X^4 \le 4^8(\mathbb{E}X^2)^2 \lesssim n^4,$$

which proves the third conclusion. For $\xi = X^3$, we similarly have $\mathbb{E}\xi^2 \lesssim n^6$, proving the fifth conclusion. For $\xi = Y$, a straightforward calculation shows $\mathbb{E}\xi^2 \lesssim n^4$, proving the fourth conclusion. ∎

Lemma 3.5. For $a \ge 1$ and integer $i, j \ge 0$,

$$\mathbb{E}_{g,h\sim\mu_{S(a)}}|X^i Y^j| \le (6i)^i(2an)^{i+2j}.$$

Proof. By Cauchy-Schwarz,

$$\mathbb{E}_{g,h\sim\mu_{S(a)}}|X^i Y^j| \le \left(\mathbb{E}_{g,h\sim\mu_{S(a)}}X^{2i}\right)^{1/2}\left(\mathbb{E}_{g,h\sim\mu_{S(a)}}Y^{2j}\right)^{1/2}. \tag{8}$$

For all $g \in S(a)$,

$$\sum_{1\le k<\ell\le n} g_k^2 g_\ell^2 \le \|g\|_2^4 \le (1+a)^2 n^2 \le (2an)^2$$

and similarly for h. Thus if $g, h \in S(a)$, then $|Y| \le (2an)^2$, which implies

$$\left(\mathbb{E}_{g,h\sim\mu_{S(a)}} Y^{2j}\right)^{1/2} \le (2an)^{2j}. \tag{9}$$

If $i = 0$, this implies the result. Otherwise assume $i \ge 1$. By symmetry of the set $S(a)$, the distribution of X under $g, h \sim \mu_{S(a)}$ is the same as that of

$$\widetilde{X} = \sum_{1\le k<\ell\le n} x_k x_\ell g_k g_\ell h_k h_\ell$$

where $x \sim \mathrm{unif}(\{-1,1\}^n)$ and $g, h \sim \mu_{S(a)}$ are independent. By Lemma 3.2 (a), conditioned on g, h,

$$\mathop{\mathbb{E}}_{x} \widetilde{X}^{2i} \le (2i)^{2i}\left(\mathop{\mathbb{E}}_{x}\widetilde{X}^2\right)^i = (2i)^{2i}\left(\sum_{1\le k<\ell\le n} g_k^2 g_\ell^2 h_k^2 h_\ell^2\right)^i.$$

For $g, h \in S(a)$,

$$\sum_{1\le k<\ell\le n} g_k^2 g_\ell^2 h_k^2 h_\ell^2 \le \frac{1}{2}\sum_{1\le k<\ell\le n}\left(g_k^4 g_\ell^4 + h_k^4 h_\ell^4\right) \le \frac{1}{2}\left(\|g\|_4^8 + \|h\|_4^8\right) \le 3^2(1+a)^2 n^2 \le 3^2(2an)^2.$$

So,

$$\mathop{\mathbb{E}}_{g,h\sim\mu_{S(a)}} X^{2i} = \mathop{\mathbb{E}}_{g,h\sim\mu_{S(a)}} \mathbb{E}_x \widetilde{X}^{2i} \le (6i\cdot 2an)^{2i}.$$

Recalling (8) and (9) this implies the result. ∎

Proof of Lemma 2.2. We may assume $u^2 n \le 10^{-4}$ because otherwise the lemma is trivial. Take $S = S(a)$ for $a = (u^2 n)^{-1/4}$. By Lemma 3.3, $\mathbb{P}(S^c) \le C\exp(-cu^{-1/4}n^{1/8})$. Equation (7) and Lemma 3.1 imply

$$\begin{aligned} 1 + 4\mathsf{TV}(M(n,u,S), M(n))^2 &\le (1-u^4)^{-n(n-1)/4}\mathbb{E}_{g,h\sim\mu_S}\exp\left(\frac{u^2}{1-u^4}X - \frac{u^4}{1-u^4}Y\right) \\ &= (1-u^4)^{-n(n-1)/4}\left(\sum_{i,j\ge 0}\frac{(-1)^j}{i!j!}\cdot\frac{u^{2i+4j}}{(1-u^4)^{i+j}}\mathbb{E}_{g,h\sim\mu_S} X^i Y^j\right). \end{aligned}$$

Let $T = \{(i,j) \in \mathbb{Z}_{\ge 0}^2 : i + 2j < 4\}$ and $T^c = \mathbb{Z}_{\ge 0}^2 \setminus T$. By the estimates in Lemma 3.4,

$$\begin{aligned} &\sum_{(i,j)\in T}\frac{1}{i!j!}\frac{u^{2i+4j}}{(1-u^4)^{i+j}}\mathbb{E}_{g,h\sim\mu_S} X^i Y^j \\ &\le 1 - \frac{u^4}{1-u^4}\cdot\frac{n(n-1)}{2} + \frac{u^4}{(1-u^4)^2}\cdot\frac{n(n-1)}{4} + \frac{u^6}{(1-u^4)^3}\cdot\frac{n(n-1)(n-2)}{6} + Cn^3 e^{-cu^{-1/4}n^{1/8}} \\ &\le 1 - u^4\cdot\frac{n(n-1)}{4} + Cu^6 n^3. \end{aligned}$$

In the last line we have used that $n^3 e^{-cu^{-1/4}n^{1/8}} \ll u^6 n^3$, because $e^{cu^{-1/4}n^{1/8}}$ is larger than any polynomial in u^{-1}. By Lemma 3.5,

$$\sum_{(i,j)\in T^c} \frac{1}{i!j!} \frac{u^{2i+4j}}{(1-u^4)^{i+j}} \mathbb{E}_{g,h\sim\mu_S} X^i Y^j \le \sum_{(i,j)\in T^c} \frac{1}{(i/e)^i} \frac{u^{2i+4j}}{(1-u^4)^{i+2j}} \mathbb{E}_{g,h\sim\mu_S} |X^i Y^j|$$
$$\le \sum_{(i,j)\in T^c} \frac{(6i)^i}{(i/e)^i} \left(\frac{2au^2 n}{1-u^4}\right)^{i+2j}$$
$$\le \sum_{(i,j)\in T^c} \left(\frac{12e\cdot au^2 n}{1-u^4}\right)^{i+2j}.$$

Since $au^2 n = (u^2 n)^{3/4} \le 10^{-3}$, this is a convergent double-geometric sum. As $i + 2j \ge 4$ for $(i,j) \in T^c$,

$$\sum_{(i,j)\in T^c} \frac{1}{i!j!} \frac{u^{2i+4j}}{(1-u^4)^{i+j}} \mathbb{E}_{g,h\sim\mu_S} X^i Y^j \lesssim (au^2 n)^4 = u^6 n^3.$$

Combining the above,

$$\log\left(1 + 4\mathsf{TV}(M(n,u,S), M(n))^2\right) \le -\frac{n(n-1)}{4} \log(1-u^4) + \log\left(1 - u^4 \cdot \frac{n(n-1)}{4} + Cu^6 n^3\right)$$
$$\le Cu^6 n^3.$$

Therefore

$$\mathsf{TV}(M(n,u,S), M(n)) \lesssim u^3 n^{3/2}.$$

Finally, since $\mathbb{P}(S^c) \le C\exp(-cu^{-1/4} n^{1/8}) \ll u^3 n^{3/2}$, (6) implies $\mathsf{TV}(M(n,u), M(n)) \lesssim u^3 n^{3/2}$. ∎

ACKNOWLEDGMENTS

We are grateful to Dheeraj Nagaraj and Dan Mikulincer for helpful and stimulating discussions, and to the anonymous reviewer whose comments and suggestions have improved this paper. This work was done in part while the authors were participating in the Probability, Geometry, and Computation in High Dimensions program at the Simons Institute for the Theory of Computing in Fall 2020. Guy Bresler was supported by NSF CAREER award CCF-1940205. Brice Huang was supported by a NSF Graduate Research Fellowship, a Siebel Scholarship, and NSF award DMS-2022448.

REFERENCES

1. M. Brennan, G. Bresler, and B. Huang. De Finetti-style results for Wishart matrices: Combinatorial structure and phase transitions. arXiv preprint arXiv:2103.14011. 2021.
2. M. Brennan, G. Bresler, and D. Nagaraj, *Phase transitions for detecting latent geometry in random graphs*, Probab. Theory Rel. Fields **178** (2020), no. 3, 1215–1289.
3. S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz, *Testing for high-dimensional geometry in random graphs*, Rand. Struct. Alg. **49** (2016), no. 3, 503–532.
4. S. Bubeck and S. Ganguly, *Entropic CLT and phase transition in high-dimensional Wishart matrices*, Int. Math. Res. Not. **2018** (2018), no. 2, 588–606.
5. S. Chatterjee and E. Meckes, *Multivariate normal approximation using exchange-able pairs*, Alea **4** (2008), 257–283.
6. D. Chételat and M. T. Wells, *The middle-scale asymptotics of Wishart matrices*, Ann. Stat. **47** (2019), no. 5, 2639–2670.

7. L. Devroye, A. György, G. Lugosi, and F. Udina, *High-dimensional random geometric graphs and their clique number*, Elec. J. Probab. **16** (2011), 2481–2508.
8. R. Eldan and D. Mikulincer, "*Information and dimensionality of anisotropic random geometric graphs*," Geo. Aspects Func. Analysis: Israel seminar, Cham, Switzerland: Springer, 2017, pp. 273–324.
9. T. Jiang and D. Li, *Approximation of rectangular beta-Laguerre ensembles and large deviations*, J. Theor. Probab. **28** (2015), no. 3, 804–847.
10. I. M. Johnstone. High dimensional statistical inference and random matrices. Paper presented at: Proc. 2006 ICM. 2007 307–333.
11. S. Liu, S. Mohanty, T. Schramm, and E. Yang. Testing thresholds for high-dimensional sparse random geometric graphs. Paper presented at: Proc. 54th STOC. 2022 672–677.
12. S. Liu and M. Z. Rácz. Phase transition in noisy high-dimensional random geometric graphs. arXiv preprint arXiv:2103.15249. 2021.
13. D. Mikulincer, *A CLT in Stein's distance for generalized Wishart matrices and higher-order tensors*, Int. Math. Res. Not. **2022** (2022), no. 10, 7839–7872.
14. I. Nourdin and G. Zheng, *Asymptotic behavior of large Gaussian correlated Wishart matrices*, J. Theor. Probab. **35** (2021), 2239–2268.
15. R. O'Donnell, Analysis of Boolean functions, Cambridge, UK: Cambridge University Press, 2014.
16. M. Penrose, Random geometric graphs, Vol **5**, Oxford, UK: Oxford University Press, 2003.
17. M. Z. Rácz and J. Richey, *A smooth transition from Wishart to GOE*, J. Theor. Probab. **32** (2019), no. 2, 898–906.

How to cite this article: M. Brennan, G. Bresler, and B. Huang, *Threshold for detecting high dimensional geometry in anisotropic random geometric graphs*, Random Struct. Alg. **64** (2024), 125–137. https://doi.org/10.1002/rsa.21178