A smooth transition from Wishart to GOE

Miklós Z. Rácz [*]   Jacob Richey [†]

November 18, 2016

Abstract

It is well known that an $n \times n$ Wishart matrix with d degrees of freedom is close to the appropriately centered and scaled Gaussian Orthogonal Ensemble (GOE) if d is large enough. Recent work of Bubeck, Ding, Eldan, and Racz, and independently Jiang and Li, shows that the transition happens when $d = \Theta(n^3)$. Here we consider this critical window and explicitly compute the total variation distance between the Wishart and GOE matrices when $d/n^3 \to c \in (0, \infty)$. This shows, in particular, that the phase transition from Wishart to GOE is smooth.

1   Introduction

The Wishart distribution is a fundamental object appearing in many domains, such as statistics, geometry, quantum physics, and wireless communications, among others. In statistics it arises as the distribution of the sample covariance matrix of a sample from a multivariate normal distribution. In geometry it is known as the Gram matrix of inner products of n points in $\mathbb{R}^d$, and it is also the starting point for canonical models of random geometric graphs [5, 3, 8].

It is well known that an $n \times n$ Wishart matrix with d degrees of freedom is close to the appropriately centered and scaled Gaussian Orthogonal Ensemble (GOE) if d is large enough (see, e.g., [5]). Recent work [3, 9] shows that the transition happens when $d = \Theta\left(n^3\right)$ and in this paper we study this critical window. In Theorem 1.2 below we explicitly compute the total variation distance between the Wishart and GOE matrices when $d/n^3 \to c \in (0, \infty)$, showing, in particular, that the phase transition from Wishart to GOE is smooth.

1.1   Main result

Let X be an $n \times d$ matrix where the entries are i.i.d. standard normal random variables, and let $W \equiv W(n, d) = XX^T$ be the corresponding $n \times n$ Wishart matrix with d degrees of freedom.[1] Let $M(n)$ be an $n \times n$ matrix drawn from the Gaussian Orthogonal Ensemble, i.e., a symmetric $n \times n$ random matrix where the diagonal entries are i.i.d. normal random variables with mean zero and variance 2, and the entries above the diagonal are i.i.d. standard normal random variables, with the entries on and above the diagonal all independent. In order to match the first moment and the scale of the Wishart matrix, we center and scale $M(n)$ appropriately: let $M(n, d) := \sqrt{d}M(n) + dI_n$, where I_n is the $n \times n$ identity matrix.

---

[*]Microsoft Research; `miracz@microsoft.com`.

[†]University of Washington; `jfrichey@uw.edu`.

[1]In statistics the number of samples is usually denoted by n and the number of parameters is usually denoted by p, resulting in a $p \times p$ Wishart matrix with n degrees of freedom. Here our notation is taken with the geometric perspective in mind, following [5, 3, 4].

If d is large enough compared to n, then the Wishart matrix becomes approximately like the GOE. Recent work of Bubeck, Ding, Eldan, and Racz [3], and independently Jiang and Li [9], shows that the transition happens when $d = \Theta\left(n^3\right)$. Specifically, they proved the following theorem, where we write TV for total variation distance.

Theorem 1.1. *Define the random matrix ensembles $W(n,d)$ and $M(n,d)$ as above.*

(a) (Bubeck, Ding, Eldan, and Racz [3]) If $d/n^3 \to 0$ then

$$\mathrm{TV}\left(W(n,d), M(n,d)\right) \to 1.$$

(b) (Bubeck, Ding, Eldan, and Racz [3]; Jiang and Li [9]) If $d/n^3 \to \infty$ then

$$\mathrm{TV}\left(W(n,d), M(n,d)\right) \to 0.$$

Our focus is on the critical window and our main result is the explicit computation of the limiting total variation distance between $W(n,d)$ and $M(n,d)$ when $d/n^3 \to c \in (0, \infty)$.

Theorem 1.2. *Define the random matrix ensembles $W(n,d)$ and $M(n,d)$ as above and let $d = d(n)$ be such that $d/n^3 \to c \in (0, \infty)$. Then*

$$\lim_{n \to \infty} \mathrm{TV}\left(W(n,d), M(n,d)\right) = \mathrm{Erf}\left(\frac{1}{4\sqrt{3}\sqrt{c}}\right), \tag{1.1}$$

where recall that the error function is defined as

$$\mathrm{Erf}\left(x\right) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

From this result we can immediately read off that, as $c \to 0$ and $c \to \infty$, the total variation distance goes to 1 and 0, respectively, recovering the previous results described in Theorem 1.1. Since $\mathrm{Erf}\left(x\right) = \frac{2}{\sqrt{\pi}} x \left(1 + o\left(1\right)\right)$ as $x \to 0$, the limiting total variation distance decays as

$$\mathrm{Erf}\left(\frac{1}{4\sqrt{3}\sqrt{c}}\right) \sim \frac{1}{2\sqrt{3\pi}\sqrt{c}}$$

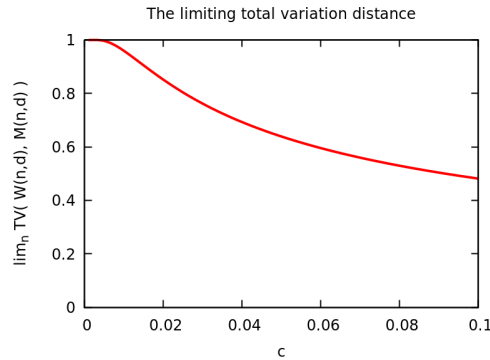as $c \to \infty$. The behavior of the limit when c is small is plotted in Figure 1.

Figure 1: The limiting total variation distance as a function of c, when c is close to 0.

From the proof we shall see that the limit in (1.1) is the expected value of an explicit function of a two-dimensional Gaussian, which comes from the central limit theorem for the first and third moments of the empirical spectral distribution of a GOE matrix.

1.2 Further related work and open problems

Several recent works have explored extensions of Theorem 1.1, and Theorem 1.2 raises further questions.

Robustness. Bubeck and Ganguly [4] showed that the critical dimension is *universal* in the following sense: Theorem 1.1 holds (up to logarithmic factors) if the entries of X are i.i.d. from a sufficiently smooth distribution. What can be said about the transition in the critical regime? Are there other distributions for which the limiting total variation distance can be computed explicitly? If not, can one prove similar qualitative behavior?

Anisotropy. Eldan and Mikulincer [8] studied the effect of anisotropy on the power of detecting geometry in random geometric graphs. This is directly related to studying Wishart matrices where each row of X is a multivariate normal with a diagonal covariance matrix. The authors introduce new notions of dimensionality and prove a theorem similar to Theorem 1.1 with appropriate upper and lower bounds on the "effective critical dimension". While the primary open problem is to close the gap between these bounds, one may also ask about the nature of the transition at the effective critical dimension: can anisotropy cause qualitatively different behavior?

Other regimes. Theorems 1.1 and 1.2 state that as $d/n^3 \to \infty$, *all* statistics of the Wishart $W(n, d)$ and the GOE $M(n, d)$ have asymptotically the same distribution, but this is not the case if d/n^3 remains bounded. In the random matrix literature there has been lots of work showing that *particular* statistics of these ensembles have asymptotically the same distribution even when $d \ll n^3$. For instance, when $d = \Theta(n)$, then the limiting empirical spectral distribution of the Wishart is the Marchenko-Pastur law, which shows the difference between the Wishart and GOE, but the largest eigenvalue of the Wishart already behaves like that of the GOE [10, 6, 7]. This naturally raises the question of whether there are other regimes of d and n where there are interesting phase transitions.

2 Proof of Theorem 1.2

The main reason that allows for an explicit computation of the limiting total variation distance in Theorem 1.2 is that both $W(n, d)$ and $M(n, d)$ have explicit densities. The proof of Theorem 1.2 is similar to the case of $d/n^3 \to \infty$ presented in [3] and proceeds by a Taylor expansion of the ratio of the densities of the two random matrix ensembles. The difference compared to the case of $d/n^3 \to \infty$ is that here the Taylor expansion has to be done to one degree higher. As we shall see, taking the limit of the total variation distance as $d/n^3 \to c \in (0, \infty)$ then requires using the central limit theorem for the moments of the empirical spectral distribution of a GOE matrix.

Step 1: Writing out the total variation distance. Let $\mathcal{P} \subset \mathbb{R}^{\frac{n(n+1)}{2}}$ denote the cone of positive semidefinite matrices. It is well known (see, e.g., [11]) that when $d \geq n$, $W(n, d)$ has the following density with respect to the Lebesgue measure on $\mathcal{P}$:

$$f_{n,d}(A) := \frac{(\det(A))^{\frac{1}{2}(d-n-1)} \exp\left(-\frac{1}{2}\mathrm{Tr}(A)\right)}{2^{\frac{1}{2}dn} \pi^{\frac{1}{4}n(n-1)} \prod_{i=1}^{n} \Gamma\left(\frac{1}{2}(d+1-i)\right)},$$

where $\mathrm{Tr}(A)$ denotes the trace of the matrix A. The density of a GOE random matrix with respect to the Lebesgue measure on $\mathbb{R}^{\frac{n(n+1)}{2}}$ is $A \mapsto (2\pi)^{-\frac{1}{4}n(n+1)} 2^{-\frac{n}{2}} \exp\left(-\frac{1}{4}\mathrm{Tr}(A^2)\right)$ and so the density

of $M(n, d)$ with respect to the Lebesgue measure on $\mathbb{R}^{\frac{n(n+1)}{2}}$ is

$$g_{n,d}(A) := \frac{\exp\left(-\frac{1}{4d}\mathrm{Tr}\left((A - dI_n)^2\right)\right)}{(2\pi d)^{\frac{1}{4}n(n+1)}\, 2^{\frac{n}{2}}}.$$

Denote the measure given by this density by $\mu_{n,d}$, let λ denote the Lebesgue measure on $\mathbb{R}^{\frac{n(n+1)}{2}}$ and write $A \succeq 0$ if A is positive semidefinite. We can then write

$$\mathrm{TV}\left(W(n, d), M(n, d)\right) = \int_{\mathbb{R}^{\frac{n(n+1)}{2}}} \left(g_{n,d}(A) - f_{n,d}(A)\mathbf{1}_{\{A \succeq 0\}}\right)_+ d\lambda(A)$$

$$= \int_{\mathbb{R}^{\frac{n(n+1)}{2}}} \left(1 - \frac{f_{n,d}(A)\mathbf{1}_{\{A \succeq 0\}}}{g_{n,d}(A)}\right)_+ d\mu_{n,d}(A), \tag{2.1}$$

where $x_+ := \max\{x, 0\}$. Let $\mathcal{Q}$ denote the set of symmetric matrices for which all of the eigenvalues are in the interval $\left[d - 3\sqrt{dn}, d + 3\sqrt{dn}\right]$. Since $d/n^3 \to c > 0$, we have that $\mathcal{Q} \subset \mathcal{P}$ for all n large enough. It is known (see, e.g., [1]) that, with probability $1 - o(1)$, all the eigenvalues of $M(n)$ are in the interval $[-3\sqrt{n}, 3\sqrt{n}]$, which implies that $M(n, d) \in \mathcal{Q}$. Since the integrand in (2.1) is bounded, we can then write

$$\mathrm{TV}\left(W(n, d), M(n, d)\right) = \int_{\mathcal{Q}} \left(1 - \frac{f_{n,d}(A)}{g_{n,d}(A)}\right)_+ d\mu_{n,d}(A) + o(1) \tag{2.2}$$

and so we may restrict our attention to $\mathcal{Q}$.

Define $\alpha_{n,d}(A) := \log\left(f_{n,d}(A)/g_{n,d}(A)\right)$. Denote the eigenvalues of an $n \times n$ matrix A by $\lambda_1(A) \le \cdots \le \lambda_n(A)$; when the matrix is obvious from the context, we omit the dependence on A. Recall that $\det(A) = \prod_{i=1}^n \lambda_i$ and $\mathrm{Tr}(A) = \sum_{i=1}^n \lambda_i$. We then have that

$$\alpha_{n,d}(A) = \frac{1}{2}\sum_{i=1}^n \left\{(d - n - 1)\log\lambda_i - \lambda_i + \frac{1}{2d}(\lambda_i - d)^2\right\}$$

$$+ \left\{\frac{n(n+3)}{4} - \frac{dn}{2}\right\}\log 2 + \frac{n}{2}\log\pi + \frac{n(n+1)}{4}\log d - \sum_{i=1}^n \log\Gamma\left(\frac{1}{2}(d + 1 - i)\right).$$

By Stirling's formula we know that $\log\Gamma(z) = \left(z - \frac{1}{2}\right)\log z - z + \frac{1}{2}\log(2\pi) + O\left(\frac{1}{z}\right)$ as $z \to \infty$, so

$$\alpha_{n,d}(A) = \frac{1}{2}\sum_{i=1}^n \left\{(d - n - 1)\log\lambda_i - \lambda_i + \frac{1}{2d}(\lambda_i - d)^2\right\}$$

$$+ \frac{n(n+1)}{4}\log d - \frac{1}{2}\sum_{i=1}^n (d - i)\log(d + 1 - i) + \frac{1}{2}\sum_{i=1}^n (d + 1 - i) + O\left(\frac{n}{d}\right).$$

Now writing $\log(d + 1 - i) = \log d + \log\left(1 - \frac{i-1}{d}\right) = \log d - \frac{i-1}{d} - \frac{(i-1)^2}{2d^2} + O\left(\frac{(i-1)^3}{d^3}\right)$ we get that

$$\alpha_{n,d}(A) = \frac{1}{2}\sum_{i=1}^n \left\{(d - n - 1)\log\lambda_i - \lambda_i + \frac{1}{2d}(\lambda_i - d)^2\right\} - \frac{n}{2}\left\{(d - n - 1)\log d - d\right\} - \frac{n^3}{12d} + o(1).$$

Defining $h(x) := \frac{1}{2}\left\{(d - n - 1)\log(x/d) - (x - d) + \frac{1}{2d}(x - d)^2\right\}$, we have that

$$\alpha_{n,d}(A) = \sum_{i=1}^n h(\lambda_i) - \frac{n^3}{12d} + o(1). \tag{2.3}$$

4

Step 2: Taylor expansion and taking the limit. The derivatives of h at d are $h(d) = 0$, $h'(d) = -\frac{n+1}{2d}$, $h''(d) = \frac{n+1}{2d^2}$, $h^{(3)}(d) = \frac{d-n-1}{d^3}$, $h^{(4)}(d) = -\frac{3(d-n-1)}{d^4}$, and also $h^{(5)}(x) = \frac{12(d-n-1)}{x^5}$. Approximating h with its fourth order Taylor polynomial around d we get that

$$h(x) = -\frac{n+1}{2d}(x-d) + \frac{n+1}{4d^2}(x-d)^2 + \frac{d-n-1}{6d^3}(x-d)^3 - \frac{d-n-1}{8d^4}(x-d)^4 + \frac{d-n-1}{10\xi^5}(x-d)^5,$$

where ξ is some real number between x and d. From 2.3 we see that to compute $\alpha_{n,d}(A)$, we need to compute the sum over the eigenvalues $\{\lambda_i\}_{i=1}^n$ of each term in the expansion.

First, we argue that the contribution from the remainder term is negligible. Recall that $A \in \mathcal{Q}$, and hence $\lambda_i \in \left[d - 3\sqrt{dn}, d + 3\sqrt{dn}\right]$ for every $i \in [n]$. If $x \in \left[d - 3\sqrt{dn}, d + 3\sqrt{dn}\right]$, then

$$\left|\frac{d-n-1}{10\xi^5}(x-d)^5\right| \leq \frac{d-n-1}{10\left(d - 3\sqrt{dn}\right)^5}\left(3\sqrt{dn}\right)^5 = O\left(\frac{1}{n^2}\right),$$

where we used that $d = \Theta(n^3)$. Summing n such terms gives a term of order $O(1/n)$, which is negligible in the limit. Turning to the four terms that matter, defining

$$S_1(n, d) := -\frac{n+1}{2d}\sum_{i=1}^n (\lambda_i - d), \qquad S_2(n, d) := \frac{n+1}{4d^2}\sum_{i=1}^n (\lambda_i - d)^2,$$

$$S_3(n, d) := \frac{d-n-1}{6d^3}\sum_{i=1}^n (\lambda_i - d)^3, \qquad S_4(n, d) := -\frac{d-n-1}{8d^4}\sum_{i=1}^n (\lambda_i - d)^4,$$

and also letting $S_0(n, d) := -n^3/(12d)$, we thus have that

$$\alpha_{n,d}(A) = S_0(n, d) + S_1(n, d) + S_2(n, d) + S_3(n, d) + S_4(n, d) + o(1). \tag{2.4}$$

For $i \in [n]$ define $\mu_i := \frac{1}{\sqrt{dn}}(\lambda_i - d)$. If $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of $M(n, d)$, then $\{\mu_i\}_{i=1}^n$ are the eigenvalues of $\frac{1}{\sqrt{n}}M(n)$. Recall that the empirical spectral distribution $\frac{1}{n}\sum_{i=1}^n \delta_{\mu_i}$ converges weakly to the semicircle distribution with density $\rho_{\text{sc}}(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{\{|x|\leq 2\}}$ (see, e.g., [1]). With this notation we can rewrite the four quantities above as follows:

$$S_1(n, d) = -\frac{\sqrt{n}(n+1)}{2\sqrt{d}} \times \sum_{i=1}^n \mu_i, \qquad S_2(n, d) = \frac{n^2(n+1)}{4d} \times \frac{1}{n}\sum_{i=1}^n \mu_i^2,$$

$$S_3(n, d) = \frac{d-n-1}{6d} \times \sqrt{\frac{n^3}{d}} \times \sum_{i=1}^n \mu_i^3, \qquad S_4(n, d) = -\frac{d-n-1}{8d} \times \frac{n^3}{d} \times \frac{1}{n}\sum_{i=1}^n \mu_i^4.$$

Let U be a random variable distributed according to the semicircle law. We know that for any fixed $k \in \mathbb{N}$, the kth moment of the empirical spectral distribution converges in probability to the kth moment of the semicircle law (see [1, Lemmas 2.1.6 and 2.1.7]), i.e.,

$$\frac{1}{n}\sum_{i=1}^n \mu_i^k \to \mathbb{E}\left[U^k\right].$$

Since $\mathbb{E}[U^2] = 1$ and $\mathbb{E}[U^4] = 2$, we have that, as $d/n^3 \to c \in (0, \infty)$, $S_2(n, d) \to 1/(4c)$ and $S_4(n, d) \to -1/(4c)$, so put together we have that $S_2(n, d) + S_4(n, d) \to 0$.

5

The central limit theorem for the moments of the empirical spectral distribution of a GOE matrix (see [1, Theorem 2.1.31 and Exercise 2.1.35] and [2]) shows that

$$\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \mu_i^3 \right) \Longrightarrow (N_1, N_3),$$

where (N_1, N_3) are jointly normal with mean zero and covariance matrix $C = \begin{pmatrix} 2 & 6 \\ 6 & 24 \end{pmatrix}$. The entries of the covariance matrix C are special cases of the more general formulas found in [1, 2], so we do not describe the computations here, but one can verify these numbers by using the identity

$$\sum_{i=1}^{n} \mu_i^k = \operatorname{Tr}\left(\left(\tfrac{1}{\sqrt{n}} M(n) \right)^k \right) = \sum_{i=1}^{n} \left(\left(\tfrac{1}{\sqrt{n}} M(n) \right)^k \right)_{i,i}$$

and computing appropriate moments of normal random variables.

Putting everything together we see that, as $d/n^3 \to c \in (0, \infty)$, we have that

$$(S_0, S_1, S_2, S_3, S_4) \Longrightarrow \left(-\frac{1}{12c}, -\frac{1}{2\sqrt{c}} N_1, \frac{1}{4c}, \frac{1}{6\sqrt{c}} N_3, -\frac{1}{4c} \right). \tag{2.5}$$

Therefore, since the function $(s_0, s_1, s_2, s_3, s_4) \mapsto (1 - \exp(s_0 + s_1 + s_2 + s_3 + s_4))_+$ is continuous and bounded, we have by (2.2), (2.4), and (2.5) that

$$\operatorname{TV}(W(n, d), M(n, d)) \to \mathbb{E}\left[\left(1 - \exp\left(-\frac{1}{12c} - \frac{1}{2\sqrt{c}} N_1 + \frac{1}{6\sqrt{c}} N_3 \right) \right)_+ \right]. \tag{2.6}$$

Step 3: Evaluating the limit. What remains is to evaluate the expectation on the right hand side of (2.6). Let Y and Z be independent normal random variables with mean zero and variances 2 and 6, respectively. Then (N_1, N_3) and $(Y, 3Y + Z)$ have the same distribution, since both are Gaussian with the same mean vector and covariance matrix. Notice that

$$-\frac{1}{2\sqrt{c}} Y + \frac{1}{6\sqrt{c}} (3Y + Z) = \frac{1}{6\sqrt{c}} Z$$

and hence the right hand side of (2.6) is equal to

$$\mathbb{E}\left[\left(1 - \exp\left(-\frac{1}{12c} + \frac{1}{6\sqrt{c}} Z \right) \right)_+ \right].$$

Since $1 - \exp(-1/(12c) + z/(6\sqrt{c})) \geq 0$ if and only if $z \leq 1/(2\sqrt{c})$, we have that

$$\mathbb{E}\left[\left(1 - \exp\left(-\frac{1}{12c} + \frac{1}{6\sqrt{c}} Z \right) \right)_+ \right] = \int_{-\infty}^{\frac{1}{2\sqrt{c}}} \left(1 - e^{-\frac{1}{12c} + \frac{1}{6\sqrt{c}} z} \right) \cdot \frac{1}{\sqrt{12\pi}} e^{-\frac{z^2}{12}} dz$$

$$= \frac{1}{\sqrt{12\pi}} \int_{-\infty}^{\frac{1}{2\sqrt{c}}} e^{-\frac{z^2}{12}} dz - \frac{1}{\sqrt{12\pi}} \int_{-\infty}^{\frac{1}{2\sqrt{c}}} e^{-\frac{(z - 1/\sqrt{c})^2}{12}} dz$$

$$= \mathbb{P}\left(Z < \frac{1}{2\sqrt{c}} \right) - \mathbb{P}\left(Z < -\frac{1}{2\sqrt{c}} \right) = \operatorname{Erf}\left(\frac{1}{4\sqrt{3}\sqrt{c}} \right),$$

which concludes the proof.

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2010.

[2] G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probability Theory and Related Fields*, 134(2):283–338, 2006.

[3] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

[4] S. Bubeck and S. Ganguly. Entropic CLT and phase transition in high-dimensional Wishart matrices. Preprint available at `http://arxiv.org/abs/1509.03258`, 2015.

[5] L. Devroye, A. György, G. Lugosi, and F. Udina. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16:2481–2508, 2011.

[6] N. El Karoui. On the largest eigenvalue of Wishart matrices with identity covariance when n, p and $p/n \to \infty$. *arXiv preprint math/0309355*, 2003.

[7] N. El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, pages 663–714, 2007.

[8] R. Eldan and D. Mikulincer. Information and dimensionality of anisotropic random geometric graphs. Preprint available at `https://arxiv.org/abs/1609.02490`, 2016.

[9] T. Jiang and D. Li. Approximation of Rectangular Beta-Laguerre Ensembles and Large Deviations. *Journal of Theoretical Probability*, 28:804–847, 2015.

[10] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[11] J. Wishart. The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, 20A(1/2):32–52, 1928.