The Annals of Statistics
2019, Vol. 47, No. 5, 2639–2670
https://doi.org/10.1214/18-AOS1760

THE MIDDLE-SCALE ASYMPTOTICS OF WISHART MATRICES

BY DIDIER CHÉTELAT AND MARTIN T. WELLS[1]

École Polytechnique de Montréal and Cornell University

We study the behavior of a real p-dimensional Wishart random matrix with n degrees of freedom when $n, p \to \infty$ but $p/n \to 0$. We establish the existence of phase transitions when p grows at the order $n^{(K+1)/(K+3)}$ for every $K \in \mathbb{N}$, and derive expressions for approximating densities between every two phase transitions. To do this, we make use of a novel tool we call the $\mathcal{F}$-conjugate of an absolutely continuous distribution, which is obtained from the Fourier transform of the square root of its density. In the case of the normalized Wishart distribution, this represents an extension of the t-distribution to the space of real symmetric matrices.

1. Introduction. The real Wishart $\mathrm{W}_p(n, I_p/n)$ distribution is the law of the symmetric matrix $X^t X/n$, where X is a $n \times p$ matrix of independent standard normal random variables. Such matrices have important applications in statistical modeling, such as for covariance matrix estimation, financial portfolio optimization, as a prior in Bayesian models, as a model of multiple-input multiple-output channels in telecommunication systems, in quantum computing and as a real-valued analogue of random geometric graphs. Of particular interest are the asymptotics of such random matrices when the parameter n tends to infinity. Historically, these were first studied holding the dimension parameter p fixed, dating back to the classic work of Wishart (1928) and Bartlett (1933).

Starting from the work of Marčenko and Pastur (1967), it became common to study Wishart asymptotics in the setting where p also grows to infinity as the same rate as n, that is such that $p, n \to \infty$ with $p/n \to c \in (0, 1)$. This line of work has proven important in the past two decades, as the rise of large-scale data collection methodologies lead to an explosion of problems where the dimension of the Wishart matrix is large compared to the sample size. Yet, this body of work leaves open the question as to what happens to Wishart matrices when $n, p \to \infty$ with $p/n \to 0$. As these asymptotics lie between the classical regime where p is fixed as $n \to \infty$ and the high-dimensional regime where $p/n \to c \in (0, 1)$, we might refer to them as "middle-scale" regimes. Hence, one might ask: what is the asymptotic behavior of a Wishart matrix $\mathrm{W}_p(n, I_p/n)$ in the middle-scale regimes?

Received July 2017; revised August 2018.
[1]Supported in part by NSF-DMS 1611893 and NIH grant U19 AI111143.
MSC2010 subject classifications. Primary 60B20, 60B10; secondary 60E10.
Key words and phrases. Covariance estimation, Gaussian orthogonal ensemble, high-dimensional asymptotics, phase transitions, random matric theory.

To gain some intuition, it is instructive to look at the eigenvalues of a $\mathrm{W}_p(n, I_p/n)$ Wishart matrix. In the classical regime where p is fixed as $n \to \infty$, they must all almost surely tend to 1 by the strong law of large numbers. In contrast, in the high-dimensional regime where both $n, p \to \infty$ with $p/n \to c \in (0, 1)$, they distribute themselves in the shape of a Marchenko–Pastur law with parameter c. But if we let $c \to 0$, the Marchenko–Pastur law converges weakly to a Dirac measure with mass at 1, which suggests that in the middle-scale regimes, the eigenvalues should converge almost surely to unity as in the classical case.

This kind of argument motivates a binary view of Wishart asymptotics. Since the behavior of a Wishart matrix in the middle-scale regimes appears the same as in the classical regime, there seems to be only two types of asymptotics: low-dimensional where $p/n \to 0$ and high-dimensional where $p/n \to c \in (0, 1)$. This view has concrete repercussions in applications. For example, in covariance estimation, the binary view provides a useful rule of thumb: small p's call for classical covariance estimators, while large p's call for high-dimensional covariance estimators. Similar reasoning can be applied in other problems where Wishart matrices are applied.

Recent results establish that this binary view is incorrect. Consider the Gaussian orthogonal ensemble $\mathrm{GOE}(p)$, the distribution on $p \times p$ real symmetric matrices whose diagonal elements follow a $\mathrm{N}(0, 2)$ distribution, while the off-diagonal elements are $\mathrm{N}(0, 1)$, all independent. In the classical regime where p is fixed, the central limit theorem implies that

$$\sqrt{n}\big[W_p(n, I_p/n) - I_p\big] \Rightarrow \mathrm{GOE}(p),$$

as $n \to \infty$, where the arrow stands for weak convergence. In fact, something better is known: recent work has extended this result to the case where p tends to infinity. Recall that for two absolutely continuous distributions F_1 and F_2 with densities f_1 and f_2, their total variation distance is given by $\mathrm{d}_{\mathrm{TV}}(F_1, F_2) = \mathrm{d}_{\mathrm{TV}}(f_1, f_2) = \int |f_1(x) - f_2(x)|\,dx$, while their Kullback–Leibler divergence is given by $d_{\mathrm{KL}}(f_1 \| f_2) = \int \log[f_1(x)/f_2(x)] f_1(x)\,dx$. With different approaches, Jiang and Li (2015) and Bubeck et al. (2016) independently established that

$$\mathrm{d}_{\mathrm{TV}}\big(\sqrt{n}\big[W_p(n, I_p/n) - I_p\big], \mathrm{GOE}(p)\big) \to 0 \tag{1.1}$$

whenever $p^3/n \to 0$. Thus, when $p^3/n \to 0$, the same asymptotics hold as in the p fixed case, and we might regard these regimes as belonging to the classical setting.

It turns out that the converse is true. When $p^3/n \nrightarrow 0$, results of Bubeck and Ganguly (2018) and Rácz and Richey (2018) show that

$$\mathrm{d}_{\mathrm{TV}}\big(\sqrt{n}\big[W_p(n, I_p/n) - I_p\big], \mathrm{GOE}(p)\big) \nrightarrow 0. \tag{1.2}$$

Thus a phase transition occurs when p is of order $\sqrt[3]{n}$. In fact, Bubeck and Ganguly (2018) show that the statistic $\mathrm{tr}(X/\sqrt{p})^3$ asymptotically distinguishes the distributions as $n, p \to \infty$: they remark that, in Landau notation (Knuth (1976)), when X

is a normalized Wishart matrix it has mean and variance of order $\Theta(\sqrt{p^3/n})$ and $\Theta(1 + p^2/n^2)$, respectively, while for X a Gaussian orthogonal matrix it has zero mean and variance of order $\Theta(1)$. This can be shown to imply equation (1.2) when $p^3/n \nrightarrow 0$.

This result raises the following question: if the normalized Wishart distribution is not approximated by a Gaussian orthogonal ensemble in the gap between p growing like $\sqrt[3]{n}$ and growing like n, how does it behave? The goal of this article is to shed some light on this question.

Our first insight comes from studying a closely related quantity. For a given Euclidean space, let $\mathcal{F}$ denote the Fourier transform, normalized to be an isometry over square-integrable functions. Consider the following.

DEFINITION 1. The $\mathcal{F}$-conjugate of an absolutely continuous distribution F with density f on an Euclidean space is the distribution F^* with density $|\mathcal{F}\{f^{1/2}\}|^2$.

This is always well defined, since $f^{1/2}$ is square integrable, so the positive function $|\mathcal{F}\{f^{1/2}\}|^2$ is well defined and must integrate to unity by the Plancherel theorem. For the normalized Wishart, this distribution turns out to belong to a novel family that represents a generalization of the t-distribution to the real symmetric matrices, which we consequently call the symmetric t-distribution. This transformed distribution is simpler to manipulate than the normalized Wishart, which allows us to compute quantities that appear intractable for the Wishart with current tools. In particular, we were able to compute the exact asymptotics of the Kullback–Leibler divergence between this distribution and the $\mathcal{F}$-conjugate of the Gaussian orthogonal ensemble. It turns out that this discrepancy measure not only diverges when $p^3/n \nrightarrow 0$, but in fact changes asymptotics whenever p is of order $n^{(K+1)/(K+3)}$, for every $K \in \mathbb{N}$. More precisely, we prove the following.

PROPOSITION 1. *For any $K \in \mathbb{N}$, there exists constants a_k, b_k such that we have an asymptotic expansion*

$$\begin{aligned} &\mathrm{d_{KL}}\big(\mathrm{GOE}(p)^* \| \sqrt{n}\big[W_p(n, I_p/n) - I_p\big]^*\big) \\ &\quad = \sum_{k=1}^{K} a_k \frac{p^{k+2}}{n^k} + \sum_{\substack{k=1 \\ k \text{ even}}}^{K} b_k \sqrt{\frac{p^{k+2}}{n^k}} + O\left(\frac{p^{K+3}}{n^{K+1}}\right) \end{aligned}$$

as $p, n \to \infty$.

But since $\mathcal{F}$-conjugates are a function of their parent distribution, this implies that the behavior of the normalized Wishart distribution itself must vary in some way as $p^{K+3}/n^{K+1} \nrightarrow 0$ for every $K \in \mathbb{N}$, a change that is then reflected in its $\mathcal{F}$-conjugate. Thus, we are led to conclude that $p^3/n \nrightarrow 0$ cannot be the only phase

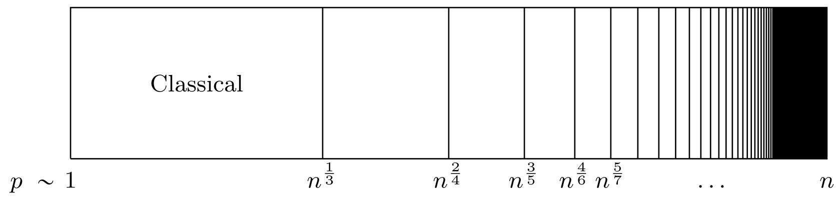

FIG. 1. *Abstract diagram of Wishart asymptotics. Each vertical bar represents a phase transition. This contrasts with the binary view, where no phase transitions occur between* p *held constant and* p *growing like* n.

transition experienced by the normalized Wishart distribution: it must be the first of an infinite sequence of phase transitions spanning the whole middle scale regimes, and every regime such that $\lim_{n\to\infty} \log p/\log n < 1$ must then sit between two phase transitions. We refer to this grouping as the degree of the regime: a middle-scale regime has degree K when $\lim_{n\to\infty} \log p/\log n \in [\frac{K}{K+2}, \frac{K+1}{K+3})$. A diagram is provided in Figure 1.

The middle-scale regimes of degree 0 correspond to the classical setting where a GOE approximation holds for the normalized Wishart distribution in Equation (1.1). In contrast, the middle-scale regimes of higher degree correspond to previously unknown behavior in equation (1.2). This raises the question whether we can find approximations of the normalized Wishart density for such middle-scale regimes. What are the extensions of the Gaussian orthogonal ensemble when $K > 0$? We propose the following candidates.

DEFINITION 2. For $n \geq 3p - 3$ and any $K \in \mathbb{N}$, we define F_K as the distribution on the space of real symmetric matrices with density $f_K(X)$,

$$\propto \left| \mathrm{E}\left[\exp\left\{ \frac{i\,\mathrm{tr}(XZ)}{\sqrt{8}} - \frac{n}{4}\sum_{k=3}^{2K_1} \frac{i^k}{k} \mathrm{tr}\left(\frac{\sqrt{2}Z}{\sqrt{n}}\right)^k + \frac{p+1}{4}\sum_{k=1}^{2K_2} \frac{i^k}{k} \mathrm{tr}\left(\frac{\sqrt{2}Z}{\sqrt{n}}\right)^k \right\}\right]\right|^2$$

for $Z \sim \mathrm{GOE}(p)$ and $K_1 = K + 1 + \mathbb{1}[K \text{ odd}]$, $K_2 = K + \mathbb{1}[K > 0, \text{ even}]$.

This definition provides us with the required approximation; namely, we can prove the following.

THEOREM 1. *For any* $K \in \mathbb{N}$, *the distribution* F_K *is well defined whenever* $n \geq 3p - 3$. *Moreover, the total variation distance between the normalized Wishart distribution* $\sqrt{n}[W_p(n, I_p/n) - I_p]$ *and* F_K *satisfies*

$$\mathrm{d}_{\mathrm{TV}}\big(\sqrt{n}\big[W_p(n, I_p/n) - I_p\big], F_K\big) \to 0$$

as $n \to \infty$ *with* $p^{K+3}/n^{K+1} \to 0$.

In particular, since $\mathrm{E}[\exp\{i \operatorname{tr}(XZ)/\sqrt{8}\}]^2 = \exp\{-\operatorname{tr} X^2/4\}$, f_0 is the Gaussian orthogonal ensemble density, so Theorem 1 also provides as a special case an independent proof of the classical GOE approximation when $p^3/n \to 0$. The results of this paper can therefore be regarded as a generalization of the Wishart asymptotics results of Jiang and Li (2015), Bubeck and Ganguly (2018), Bubeck et al. (2016) and Rácz and Richey (2018).

These results undermine the binary view upon which the practical usage of high-dimensional Wishart methodology relies. Consider, for example, a covariance estimation problem with $p = 250$ and $n = 400$. If we are only aware of the classical and high-dimensional asymptotics of a Wishart matrix, we might argue that since p is rather large compared to n, this is a situation where a covariance estimator based on high-dimensional asymptotics should be used for correct inference. But the middle-scale asymptotics of the Wishart distribution makes this reasoning untenable. For example, since $\log p/\log n \in [\frac{23}{23+2}, \frac{23+1}{23+3})$, an estimator exploiting the 23th-degree middle-scale asymptotics might make more sense, since $250^{23+2}/400^{23} \approx 1.26$ is neither big nor small. Or since $250^{24+2}/400^{24} \approx 0.79$ is also neither big nor small, the 24th-degree middle-scale asymptotics should also be reasonable. In general, a large family of nonequivalent asymptotics is plausible, and it becomes unclear which is the "proper" one for inference. At minimum, this suggests the current methodology centered around a pair of classical and high-dimensional covariance estimators might be suboptimal.

We might imagine using the approximations F_K to develop better covariance estimators, which, for example, could allow the construction of portfolios with the same expected return but decreased risk. How exactly to build and choose such estimators is at this point unclear, but this article provides a step in this direction by providing the F_K's.

We must mention that there exists regimes such that $p/n \to 0$ and yet $p \notin O(n^{(K+1)/(K+3)})$ for all $K \in \mathbb{N}$, that is, such that $\lim_{n\to\infty} \log p/\log n = 1$. An example is when p grows at the order $n^{1-1/\sqrt{\log n}}$. The results of our paper characterize almost all middle-scale regimes in the sense that among those regimes satisfying $\lim_{n\to\infty} \log p/\log n \le 1$, those such that $\lim_{n\to\infty} \log p/\log n = 1$ represent a negligible set, but we know little regarding those. A different approach might be needed to make some progress on understanding their behavior.

The rest of this article is organized as follows. We define notation used throughout the article in Section 2. We study the $\mathcal{F}$-conjugate of the normalized Wishart distribution, which can be regarded as a real symmetric matrix valued t-distribution, in Section 3. We then prove in Section 4 the main results of this paper, Proposition 1 and Theorem 1, and conclude in Section 5. Finally, we prove the claims of Section 3 in the Appendix.

2. Notation and definitions. The transpose of a matrix is denoted ^{t}, and the identity matrix of dimension p is I_p. As is standard, we take the trace operator to have lower priority than the power operator: thus for a matrix X, $\operatorname{tr} X^k$ means the trace of X^k. We will write $\operatorname{tr}^k X$ when we mean the kth power of the trace of X. The Kronecker delta is the symbol $\delta_{kl} = \mathbb{1}[k = l]$.

The space of all real-valued symmetric matrices is denoted $\mathbb{S}_p(\mathbb{R}) = \{X \in \mathbb{M}_p(\mathbb{R}) | X = X^t\}$. For a symmetric matrix X, we define the symmetric differentiation operator $\tilde{\nabla}_X$ by $(\tilde{\nabla}_X)_{kl} = \frac{1+\delta_{kl}}{2} \partial/\partial X_{kl}$. This operator has the property that $\tilde{\nabla}_X \operatorname{tr}(XY) = Y$ for any two symmetric matrices X, Y.

The space of symmetric matrices $\mathbb{S}_p(\mathbb{R})$ can be assimilated to $\mathbb{R}^{p(p+1)/2}$ by mapping a symmetric matrix to its upper triangle. By integration over $\mathbb{S}_p(\mathbb{R})$, we mean integration with respect to the pullback Lebesgue measure under this isomorphism, that is, $\int_{\mathbb{S}_p(\mathbb{R})} f(X)\, dX = \int_{\mathbb{R}^{p(p+1)/2}} f(X) \prod_{i \le j}^p dX_{ij}$. We will write the L^1 and L^2 norms with respect to this measure by $\|\cdot\|_{L^1}$ and $\|\cdot\|_{L^2}$, respectively.

While studying the Wishart distribution, the expression $n - p - 1$ appears so often that it makes sense to give it its own symbol so as to shorten the notation. We will therefore write $m = n - p - 1$ throughout the paper.

3. $\mathcal{F}$-conjugates and the symmetric t-distribution. In this section, we state results relating to the notion of $\mathcal{F}$-conjugate of a distribution, defined in the Introduction, that will be used in Section 4 to show the main results of this article. So as to not distract the reader from this goal, we relegate all proofs of this section to the Appendix.

Our first result is a variant of the Kullback–Leibler inequality in terms of $\mathcal{F}$-conjugates, which forms a key tool in our proof of Theorem 1.

PROPOSITION 2 (Kullback–Leibler inequality for $\mathcal{F}$-conjugates). *Let F be a distribution on $\mathbb{S}_p(\mathbb{R})$ with density f, and let $\psi \in L^2(\mathbb{S}_p(\mathbb{R}))$. Then the L^2-distance between $\mathcal{F}\{f^{1/2}\}$ and ψ satisfy*

$$\mathrm{d}_{L^2}^2(\mathcal{F}\{f^{1/2}\}, \psi) \le [\|\psi\|_{L^2}^2 - 1] + E\left[\Re \operatorname{Log} \frac{\mathcal{F}\{f^{1/2}\}^2(T)}{\psi^2(T)}\right] + 2\|\psi\|_{L^2} E\left[\left|\Im \operatorname{Log} \frac{\mathcal{F}\{f^{1/2}\}^2(T)}{\psi^2(T)}\right|\right]^{1/2}$$

for $T \sim F^*$, where Log *stands for the principal branch of the complex logarithm and F^* the $\mathcal{F}$-conjugate of F.*

We will also need a closed-form expression for the density of the $\mathcal{F}$-conjugate of a normalized Wishart. A reference to the characteristic function of a Wishart distribution with noninteger degrees of freedom yields the following.

PROPOSITION 3. *Let $n \ge p - 2$ and f_{NW} be the density of the normalized Wishart distribution $\sqrt{n}[W_p(n, I_p/n) - I_p]$. Then*

$$\mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}^2(T) = C_t \exp\{2i\sqrt{n} \operatorname{tr} T\} \left| I_p + i \frac{4T}{\sqrt{n}} \right|^{-\frac{n+p+1}{2}},$$

where

$$C_t = \frac{2^{\frac{p(n+2p)}{2}}}{\pi^{\frac{p(p+1)}{2}} n^{\frac{p(p+1)}{4}}} \frac{\Gamma_p^2(\frac{n+p+1}{4})}{\Gamma_p(\frac{n}{2})}. \tag{3.1}$$

In particular, by taking the modulus of the above we conclude that when $n \geq p-2$, the $\mathcal{F}$-conjugate of a normalized Wishart distribution must have a density on $\mathbb{S}_p(\mathbb{R})$ given by

$$f_{\mathrm{NW}^*}(T) = \frac{2^{\frac{p(n+2p)}{2}}}{\pi^{\frac{p(p+1)}{2}} n^{\frac{p(p+1)}{4}}} \frac{\Gamma_p^2(\frac{n+p+1}{4})}{\Gamma_p(\frac{n}{2})} \left| I_p + \frac{16T^2}{n} \right|^{-\frac{n+p+1}{4}}. \tag{3.2}$$

When $p = 1$, this is the $t_{n/2}/\sqrt{8}$ distribution, so it would be natural to interpret this distribution as the parametrization of some generalization of the t-distribution to real-valued symmetric matrices. We suggest the following definition.

DEFINITION 3 (Symmetric matrix variate t-distribution). We say a real symmetric $p \times p$ matrix T has the symmetric matrix variate t-distribution with $\nu \geq p/2 - 1$ degrees of freedom and $p \times p$ positive-definite scale matrix Ω, denoted Sym-$t_\nu(\Omega)$, if it has density

$$f_{T_n(\Omega)}(T) = \frac{2^{p(\nu-1)}\Gamma_p^2(\frac{\nu+(p+1)/2}{2})}{\pi^{\frac{p(p+1)}{2}} \nu^{\frac{p(p+1)}{4}} \Gamma_p(\nu)} |\Omega|^{-\frac{p+1}{4}} \left| I_p + \frac{T\Omega^{-1}T}{\nu} \right|^{-\frac{\nu+(p+1)/2}{2}}.$$

With this definition, the $\mathcal{F}$-conjugate of the normalized Wishart distribution is the Sym-$t_{n/2}(I_p/8)$ distribution. The fact that this is indeed a density follows from Hua (1963), Theorem 2.1.1, or by applying equation (3.2) with $n = \nu/2$ for any degrees of freedom ν. The rest of this section is focused on studying the asymptotic behavior of the Sym-$t_{n/2}(I_p/8)$ distribution.

We start by studying the asymptotic behavior of its normalization constant C_t, defined in equation (3.1).

LEMMA 1. *For every $K \in \mathbb{N}$, there exist constants a_k, b_k such that*

$$\log C_t = \log C_{\mathrm{GOE}/4} + \sum_{k=1}^{K} a_k \frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} b_k \frac{p^{k+1}}{n^k} + O\left(\frac{p^{K+3}}{n^{K+1}}\right)$$

as $n, p \to \infty$, where the symbol $C_{\mathrm{GOE}/4}$ stands for the normalization constant of GOE$(p)/4$ *distribution.*

Next, we study the empirical moments of the Sym-$t_{n/2}(I_p/8)$ distribution. For any integer partition $\kappa = (\kappa_1, \dots, \kappa_q)$ in decreasing order $\kappa_1 \geq \cdots \geq \kappa_q > 0$, define its associated power sum polynomial to be

$$r_\kappa(Z) = \prod_{i=1}^{q} \operatorname{tr} Z^{\kappa_i}. \tag{3.3}$$

The norm of the partition κ is $|\kappa| = \kappa_1 + \cdots + \kappa_q > 0$, which should not be confused with its length $q(\kappa) = q$ (number of elements). By convention, we will assume there also exists an empty partition $\varnothing = (\cdot)$ with length $q(\varnothing) = 0$, norm $|\varnothing| = 0$ and power sum polynomial $r_\varnothing(Z) = 1$. Recall the useful shorthand $m = n - p - 1$. The following lemma expresses the moments of the Sym-$t_{n/2}(I_p/8)$ distribution in terms of the moments of an inverse Wishart matrix.

LEMMA 2. *Let $T \sim$ Sym-$t_{n/2}(I_p/8)$. Then for any $k \in \mathbb{N}$, whenever n is large enough so that $n \geq p + 8k + 6$, the $2k$-th moment of T can be written*

$$\mathrm{E}[\operatorname{tr} T^{2k}] = \frac{(-1)^k}{n^k} \sum_{|\kappa| \leq 2k} b_\kappa(n, m, p)\, \mathrm{E}[r_\kappa(Y^{-1})]$$

for a $Y^{-1} \sim W_p^{-1}(n, I_p/n)$ and some polynomials b_κ in n, m, p, indexed by integer partitions κ, whose degrees satisfy $\deg b_\kappa \leq 2k + 1 - q(\kappa)$. The sums are taken over all partitions of the integers κ satisfying $|\kappa| \leq 2k$, including the empty partition.

Our next step is to compute expected power sum polynomials of an inverse Wishart, and there are two approaches in the literature. Letac and Massam (2004) found an expression in terms of a different basis, the zonal polynomials, whose expectations have a simple closed form. From this, they provided an algorithm for computing expected power sum polynomials to arbitrary order. Matsumoto (2012) found expressions of coordinate-wise moments in terms of modified Weingarten orthogonal functions, from which expectations of power sum polynomials can be computed. We follow the approach of Letac and Massam (2004) in our asymptotic analysis.

For any integer partition κ, there exist coefficients $c_{\kappa,\lambda}$ (which depend solely on κ and λ) such that

$$r_\kappa(Y^{-1}) = \sum_{|\lambda| = |\kappa|} c_{\kappa,\lambda} C_\lambda(Y^{-1}), \tag{3.4}$$

for C_λ the so-called zonal polynomials. For an overview of the topic with a focus on random matrix theory, see Muirhead (1982), Chapter 7. From Muirhead (1982), Theorem 7.2.13 and equation (18) on page 237, the expected zonal polynomials for $Y^{-1} \sim \mathrm{W}_p^{-1}(n, I_p/n)$ are

$$\begin{aligned}
&\mathrm{E}[C_\lambda(Y^{-1})] \\
&\quad = \frac{n^{|\lambda|}}{2^{|\lambda|} \prod_{i=1}^{q(\lambda)} \frac{m-i+1}{2}} C_\lambda(I_p) \qquad (3.5)\\
&\quad = \frac{2^{|\lambda|} |\lambda|! \prod_{i<j}^{q(\lambda)} (2\lambda_i - 2\lambda_j - i + j)}{\prod_{i=1}^{q(\lambda)} (2\lambda_i + q(\lambda) - i)!} n^{|\lambda|} \prod_{i=1}^{q(\lambda)} \prod_{l=0}^{\lambda_i - 1} \frac{p + (1 - i + 2l)}{m - (1 - i + 2l)}
\end{aligned}$$

for $\lambda \neq \varnothing$, and $\mathrm{E}[C_{\varnothing}(Y^{-1})] = 1$. From this, we can exactly compute $\mathrm{E}[r_\kappa(Y^{-1})]$, and thus $\mathrm{E}[\operatorname{tr} T^{2k}]$, as a function of p and n (or m). The idea is to express the moments as polynomials of p and p/m, and apply asymptotics from the two regimes where random matrix theory is well understood: the classical regime where p is held fixed as $n \to \infty$, and the linear, high-dimensional regime where p grows linearly with n. We obtain the following.

THEOREM 2. *Let $k \in \mathbb{N}$ and $T \sim \text{Sym-}t_{n/2}(I_p/8)$. The $2k$-th moment of T satisfies the asymptotics* $\mathrm{E}[p^{-1} \operatorname{tr}(T/\sqrt{p})^{2k}] = O(1)$ *as $p, n \to \infty$ with $p/n \to 0$.*

Although this result only characterizes the even moments, it can be leveraged using an integration by parts argument to also characterize the odd moments.

COROLLARY 1. *Let $k \in \mathbb{N}$ and $T \sim \text{Sym-}t_{n/2}(I_p/8)$. The $(2k+1)$-th moment of T satisfies the asymptotics* $\mathrm{E}[\operatorname{tr}(T/\sqrt{p})^{2k+1}] = O(1)$ *as $p, n \to \infty$ with $p/n \to 0$.*

Together, these two results bound the rate of growth of all moments of the symmetric t-distribution and will be key to our proof of Theorem 1.

4. The middle-scale asymptotics of the Wishart distribution. In the previous section, we summarized supporting results concerning $\mathcal{F}$-conjugates and the t-distribution. In this section, we now use these results to prove the two claims of the Introduction, which are the main goal of this article. The first result, Proposition 1, establishes the existence of middle-scale phase transitions for the normalized Wishart.

PROOF OF PROPOSITION 1. We saw in Section 3 that the $\mathcal{F}$-conjugate of the normalized Wishart distribution is the $\text{Sym-}t_{n/2}(I_p/8)$ distribution with density given by equation (3.2). For an integrable function f on $\mathbb{S}_p(\mathbb{R})$, the Fourier transform with kernel $\exp\{-i \operatorname{tr}(XT)\}$, normalized to be an L^2-isometry, satisfies $\mathcal{F}\{f\}(T) = 2^{-\frac{p}{2}} \pi^{-\frac{p(p+1)}{4}} \int_{\mathbb{S}_p(\mathbb{R})} e^{-i \operatorname{tr}(XT)} f(X)\, dX$. Thus, the $\mathcal{F}$-conjugate density of the Gaussian orthogonal ensemble is given by

$$\begin{aligned}
f_{\mathrm{GOE}^*}(T) &= \left| \mathcal{F}\left\{ \frac{\exp\{-\frac{1}{8}\sum_{i,j=1}^p X_{ij}^2\}}{2^{p(p+3)/8} \pi^{p(p+1)/8}} \right\} \right|^2 (T) \\
&= \frac{1}{2^{\frac{p(p+7)}{4}} \pi^{\frac{3p(p+1)}{4}}} \left| \int_{\mathbb{S}_p(\mathbb{R})} e^{-i \operatorname{tr}(XT)} \exp\left\{ -\frac{1}{8} \sum_{i,j=1}^p X_{ij}^2 \right\} dX \right|^2 \\
&= \frac{2^{p(3p+1)/4}}{\pi^{p(p+1)/4}} \exp\left\{ -4 \sum_{i,j=1}^p T_{ij}^2 \right\},
\end{aligned}$$

where we used the Fourier transform of a standard normal law (Stein and Weiss (1971), Theorem 1.13). But this is the density of the GOE(p)/4 distribution, so we conclude that the $\mathcal{F}$-conjugate of the Gaussian orthogonal ensemble must be $\mathrm{GOE}(p)^* = \mathrm{GOE}(p)/4$. Then, by definition of the Kullback–Leibler divergence,

$$\begin{aligned} &\mathrm{d_{KL}}\big(\mathrm{GOE}(p)^* \| \sqrt{n}[W_p(n, I_p/n) - I_p]^*\big) \\ &\quad = \mathrm{E}\biggl[\log C_{\mathrm{GOE}/4} - 4\operatorname{tr} T^2 - \log C_t + \frac{n+p+1}{4}\log\biggl|I_p + \frac{16T^2}{n}\biggr|\biggr] \end{aligned} \tag{4.1}$$

for a $T \sim \mathrm{GOE}(p)/4$, where $C_{\mathrm{GOE}/4}$ and C_t are the normalization constants of the GOE(p)/4 and the Sym-$t_{n/2}(I_p/8)$ distributions, respectively. A change of variables $X = 4T$ and Lemma 1 yields that for some constants $c_k^{(1)}$, $c_k^{(2)}$ and $X \sim \mathrm{GOE}(p)$, the above equals

$$\begin{aligned} &= \mathrm{E}\biggl[-\frac{1}{4}\operatorname{tr} X^2 + \frac{n+p+1}{4}\log\biggl|I_p + \frac{X^2}{n}\biggr|\biggr] \\ &\quad + \sum_{k=1}^{K} c_k^{(1)}\frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} c_k^{(2)}\frac{p^{k+1}}{n^k} + O\biggl(\frac{p^{K+3}}{n^{K+1}}\biggr). \end{aligned} \tag{4.2}$$

Consider the expectation. For any integer L and any real x, we have the inequality $|-\frac{1}{2}\log(1+x^2) - \sum_{l=1}^{L}(-1)^l x^{2l}/2l| \le x^{2L+2}/(2L+2)$. Thus we have the bound

$$\begin{aligned} &\biggl|\mathrm{E}\biggl[-\frac{n+p+1}{4}\log\biggl|I_p + \frac{X^2}{n}\biggr|\biggr] - \mathrm{E}\biggl[\frac{n+p+1}{2}\sum_{k=1}^{K+1}(-1)^k\frac{\operatorname{tr} X^{2k}}{2kn^k}\biggr]\biggr| \\ &\quad \le \frac{n+p+1}{2}\frac{\mathrm{E}[\operatorname{tr} X^{2(K+2)}]}{2(K+2)n^{K+2}} = O\biggl(\frac{p^{K+3}}{n^{K+1}}\biggr) \end{aligned}$$

as $n, p \to \infty$, since $\mathrm{E}[\operatorname{tr} X^{2(K+2)}] = O(p^{K+3})$ as $p \to \infty$ by Anderson, Guionnet and Zeitouni (2010), Lemma 2.1.6.

But in fact, induction on Theorem 2 of Ledoux (2009) shows that for any integer k, the $2k$-th moment of the Gaussian orthogonal ensemble can be written $\mathrm{E}[\operatorname{tr} X^{2k}] = \sum_{l=0}^{k+1} c_{k,l}^{(3)} p^l$ for constants $c_{k,l}^{(3)}$. Thus we really have

$$\begin{aligned} &\mathrm{E}\biggl[\frac{n+p+1}{4}\log\biggl|I_p + \frac{X^2}{n}\biggr|\biggr] \\ &\quad = \frac{n+p+1}{n}\biggl[\frac{\operatorname{tr} X^2}{4} + \sum_{k=2}^{K+1}\sum_{l=0}^{k+1} c_{k,l}^{(4)}\frac{p^l}{n^{k-1}}\biggr] + O\biggl(\frac{p^{K+3}}{n^{K+1}}\biggr) \\ &\quad = \frac{n+p+1}{n}\frac{\operatorname{tr} X^2}{4} + \sum_{k=1}^{K} c_k^{(5)}\frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} c_k^{(6)}\frac{p^{k+1}}{n^k} + O\biggl(\frac{p^{K+3}}{n^{K+1}}\biggr) \end{aligned}$$

for constants $c_{k,l}^{(4)}$, $c_k^{(5)}$, $c_k^{(6)}$. Plugging this back into equation (4.2) and using that $\mathrm{E}[\operatorname{tr} X^2] = p(2p^2+1)$ yields that

$$\begin{aligned}
&\mathrm{d_{KL}}\big(\mathrm{GOE}(p)^* \| \sqrt{n}[W_p(n, I_p/n) - I_p]^*\big) \\
&\quad = \sum_{k=1}^{K} a_k \frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} c_k^{(7)} \frac{p^{k+1}}{n^k} + O\left(\frac{p^{K+3}}{n^{K+1}}\right)
\end{aligned}$$

for constants $a_k, c_k^{(7)}$. Writing $p^{k+1}/n^k = \sqrt{p^{2k+2}/n^{2k}}$, reindexing the second sum and truncating at K concludes the proof. $\square$

We now turn to the proof of the other main result, Theorem 1, which characterizes the behavior of the normalized Wishart distribution between every two middle-scale phase transitions.

PROOF OF THEOREM 1. Define the complex-valued functions

$$\psi_K(T) = \sqrt{C_t} \exp\left\{ \frac{n}{4} \sum_{k=2}^{2K_1} \left(\frac{4i}{\sqrt{n}}\right)^k \frac{\operatorname{tr} T^k}{k} + \frac{p+1}{4} \sum_{k=1}^{2K_2} \left(\frac{4i}{\sqrt{n}}\right)^k \frac{\operatorname{tr} T^k}{k} \right\}, \tag{4.3}$$

where C_t is the normalization constant of the Sym-$t_{n/2}(I_p/8)$ distribution and $K_1 = K + 1 + \mathbb{1}[K \text{ odd}]$, $K_2 = K + \mathbb{1}[K > 0, \text{ even}]$. Let f_{NW}, f_t and $f_{\mathrm{GOE}/4}$ denote the densities of the normalized Wishart, Sym-$t_{n/2}(I_p/8)$ and $\mathrm{GOE}(p)/4$ distributions, respectively.

For $K = 0$, notice that $|\psi_0^2(T)| = C_t C_{\mathrm{GOE}/4}^{-1} f_{\mathrm{GOE}/4}(T)$ by definition. This means that ψ_0 is square-integrable and in fact, by Lemma 1, that

$$\lim_{n\to\infty} \|\psi_0\|_{L^2}^2 = \lim_{n\to\infty} \exp\{o(1)\} \|f_{\mathrm{GOE}/4}\|_{L^1} = 1 \tag{4.4}$$

when $p^3/n \to 0$ as $n \to \infty$. For $K > 0$, note that K_1, K_2 are always odd, and for any $x \in \mathbb{R}$ and odd $L \in \mathbb{N}$, $-\frac{1}{2}\log(1+x^2) - \sum_{l=1}^{L} (-1)^l x^{2l}/2l = \frac{1}{2}\int_0^{x^2} [t/(1+t)]^{L+1}\, dt > 0$. Thus $|\psi_K^2|(T) \le f_t(T)$ for all $K > 0$, which is integrable whenever $n \ge p - 2$. In particular,

$$\lim_{n\to\infty} \|\psi_K\|_{L^2}^2 \le \|f_t\|_{L^1} = 1 \qquad \text{for all } K > 0 \tag{4.5}$$

when $p/n \to 0$ as $n \to \infty$.

We now show using Proposition 2 that the L^2 distance between $\mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}$ and ψ_K tends to zero as $p^{K+3}/n^{K+1} \to 0$. For any $x \in \mathbb{R}$ and $L \in \mathbb{N}$,

$$\left| -\frac{1}{2}\log\big(1+x^2\big) - \sum_{l=1}^{L} (-1)^l \frac{x^{2l}}{2l} \right| \le \frac{x^{2L+2}}{2L+2}, \tag{4.6}$$

$$\left| \operatorname{atan}(x) - \sum_{l=1}^{L} (-1)^{l-1} \frac{x^{2l-1}}{2l-1} \right| \le \frac{x^{2L+1}}{2L+1}. \tag{4.7}$$

Let Log stand for the principal branch of the complex logarithm. By Proposition 3, equation (4.6) and Theorem 2, for $T \sim \text{Sym-}t_{n/2}(I_p/8)$, the real part of $\operatorname{Log} \mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}^2/\psi_K^2$ satisfies

$$
\begin{aligned}
&\mathrm{E}\left[\left|\Re \operatorname{Log} \frac{\mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}^2}{\psi_K^2}\right|\right] \\
&\quad \le \frac{n}{2}\mathrm{E}\left[\left|-\frac{1}{2}\log\left|I_p + \frac{16T^2}{n}\right| - \sum_{k=1}^{K_1} \frac{(-1)^k}{2k} \operatorname{tr}\left[\frac{4T}{\sqrt{n}}\right]^{2k}\right|\right] \\
&\qquad + \frac{p+1}{2}\mathrm{E}\left[\left|-\frac{1}{2}\log\left|I_p + \frac{16T^2}{n}\right| - \sum_{k=1}^{K_2} \frac{(-1)^k}{2k} \operatorname{tr}\left[\frac{4T}{\sqrt{n}}\right]^{2k}\right|\right] \qquad (4.8) \\
&\quad \le \frac{n}{4}\mathrm{E}\left[\frac{\operatorname{tr}(4T/\sqrt{n})^{2K_1+2}}{K_1+1}\right] + \frac{p+1}{4}\mathrm{E}\left[\frac{\operatorname{tr}(4T/\sqrt{n})^{2K_2+2}}{K_2+1}\right] \\
&\quad = O\left(\frac{p^{K+3+\mathbb{1}[K \text{ odd}]}}{n^{K+1+\mathbb{1}[K \text{ odd}]}}\right) + O\left(\frac{p^{K+3+\mathbb{1}[K>0, \text{ even}]}}{n^{K+1+\mathbb{1}[K>0, \text{ even}]}}\right) = O\left(\frac{p^{K+3}}{n^{K+1}}\right)
\end{aligned}
$$

as $p, n \to \infty$ with $p/n \to 0$.

For the imaginary part, notice that the projection $P_{(-\pi,\pi]}x = x - 2\pi\lceil \frac{x}{2\pi} - \frac{1}{2}\rceil$ satisfies $\Im \operatorname{Log} z = P_{(-\pi,\pi]} \Im \log z$ for all branches of $\log z$, as well as the inequality $|P_{(-\pi,\pi]}x| \le |x|$. Using this mapping, as well as Proposition 3, equation (4.7) and Corollary 1, we find for $T \sim \text{Sym-}t_{n/2}(I_p/8)$ that

$$
\begin{aligned}
&\mathrm{E}\left[\left|\Im \operatorname{Log} \frac{\mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}^2}{\psi_K^2}\right|\right] \\
&\quad = \mathrm{E}\left[\left|P_{(-\pi,\pi]}\left[-\frac{n+p+1}{2} \operatorname{tr} \operatorname{atan}\left(\frac{4T}{\sqrt{n}}\right) + 2\sqrt{n} \operatorname{tr} T\right.\right.\right. \\
&\qquad \left.\left.\left. + \frac{n}{2} \sum_{k=1}^{K_1-1} \frac{(-1)^k}{2k+1} \operatorname{tr}\left[\frac{4T}{\sqrt{n}}\right]^{2k+1} + \frac{p+1}{2} \sum_{k=1}^{K_2-1} \frac{(-1)^k}{2k+1} \operatorname{tr}\left[\frac{4T}{\sqrt{n}}\right]^{2k+1}\right]\right|\right] \\
&\quad \le \frac{n}{2}\mathrm{E}\left[\left|\operatorname{tr} \operatorname{atan}\left(\frac{4T}{\sqrt{n}}\right) - \frac{2\sqrt{n} \operatorname{tr} T}{n/2} - \sum_{k=1}^{K_1} (-1)^{k-1} \frac{\operatorname{tr}(4T/\sqrt{n})^{2k-1}}{2k-1}\right|\right] \\
&\qquad + \frac{p+1}{2}\mathrm{E}\left[\left|\operatorname{tr} \operatorname{atan}\left(\frac{4T}{\sqrt{n}}\right) - \sum_{k=1}^{K_2} (-1)^{k-1} \frac{\operatorname{tr}(4T/\sqrt{n})^{2k-1}}{2k-1}\right|\right] \qquad (4.9) \\
&\quad \le \frac{n}{2}\mathrm{E}\left[\left|\frac{\operatorname{tr}(4T/\sqrt{n})^{2K_1+1}}{2K_1+1}\right|\right] + \frac{p+1}{2}\mathrm{E}\left[\left|\frac{\operatorname{tr}(4T/\sqrt{n})^{2K_2+1}}{2K_2+1}\right|\right]
\end{aligned}
$$

$$= O\left(\sqrt{\frac{p^{2K+3+2\mathbb{1}[K \text{ odd}]}}{n^{2K+1+2\mathbb{1}[K \text{ odd}]}}}\right) + O\left(\sqrt{\frac{p^{2K+3+2\mathbb{1}[K>0, \text{ even}]}}{n^{2K+1+2\mathbb{1}[K>0, \text{ even}]}}}\right)$$
$$= O\left(\sqrt{\frac{p^{K+3}}{n^{K+1}}}\right)$$

as $p, n \to \infty$ with $p/n \to 0$.

By equation (3.2), the Sym-$t_{n/2}(I_p/8)$ distribution is the $\mathcal{F}$-conjugate of the normalized Wishart distribution. Therefore, Proposition 2 with equations (4.8)–(4.9) and (4.4)–(4.5) implies that $\lim_{n\to\infty} \mathrm{d}^2_{L^2}(\mathcal{F}\{f^{1/2}_{\mathrm{NW}}\}, \psi_K) \le 0$ when $p^{K+3}/n^{K+1} \to 0$ as $p, n \to \infty$.

We complete the proof by relating the ψ_K to the f_K of Definition 2. First, note that for $K = 0$, $|\psi_0| = C_t C^{-1}_{\mathrm{GOE}/4} f^{1/2}_{\mathrm{GOE}(p)/4}$ is proportional to the density of a $\mathrm{GOE}(p)/\sqrt{8}$ density, which is always integrable. In addition, $|\psi_K| \le f_t^{1/2}$ for all $K > 0$, and $f_t^{1/2}$ is proportional to the density of a Sym-$t_{m/4}(\frac{n}{4m} I_p)$ distribution with $m = n - p - 1$, itself integrable whenever $n \ge 3p - 3$. Thus, in all cases ψ_K is integrable whenever $n \ge 3p - 3$, and recall that we had already shown ψ_K is square-integrable whenever $n \ge p - 2$. We can therefore use the Fourier inversion theorem to conclude that $|\mathcal{F}^{-1}\{\psi_K\}|^2 \propto f_K$, and in particular that f_K is integrable, whenever $n \ge 3p - 3 \ge p - 2$.

Define $g_K = \mathcal{F}^{-1}\{\psi_K\}$ so that $f_K = |g_K|^2/\|g_K\|^2_{L^2}$. By the Plancherel theorem, $\mathrm{d}_{L^2}(f^{1/2}_{\mathrm{NW}}, g_K) = \mathrm{d}_{L^2}(\mathcal{F}\{f^{1/2}_{\mathrm{NW}}\}, \psi_K) \to 0$ and so in particular we must have $\lim_{n\to\infty} \|g_K\|^2_{L^2} = \lim_{n\to\infty} \|\psi_K\|_{L^1} = 1$. From the triangle inequality, the reverse triangle inequality and the limits $\mathrm{d}_{L^2}(\mathcal{F}\{f^{1/2}_{\mathrm{NW}}\}, \psi_K) \to 0$ and $\|g_K\|^2_{L^2} \to 1$, we can conclude that

$$\begin{aligned}\lim_{n\to\infty} \mathrm{d}_{L^2}\big(f^{1/2}_{\mathrm{NW}}, f^{1/2}_K\big) &\le \lim_{n\to\infty} \mathrm{d}_{L^2}\big(f^{1/2}_{\mathrm{NW}}, |g_K|\big) + \lim_{n\to\infty} \mathrm{d}_{L^2}\big(|g_K|, f^{1/2}_K\big) \\ &\le \lim_{n\to\infty} \mathrm{d}_{L^2}\big(f^{1/2}_{\mathrm{NW}}, g_K\big) + \lim_{n\to\infty} \big|1 - \|g_K\|^{-1}_{L^2}\big| \|g_K\|_{L^2} = 0\end{aligned}$$

when $p^{K+3}/n^{K+1} \to 0$ as $p, n \to \infty$. But by the Cauchy–Schwarz inequality,

$$\begin{aligned}\mathrm{d}_{\mathrm{TV}}(f_{\mathrm{NW}}, f_K) &= \big\|\big(f^{1/2}_{\mathrm{NW}} + f^{1/2}_K\big)\big(f^{1/2}_{\mathrm{NW}} - f^{1/2}_K\big)\big\|_{L^1} \\ &\le \|f^{1/2}_{\mathrm{NW}} + f^{1/2}_K\|_{L^2} \|f^{1/2}_{\mathrm{NW}} - f^{1/2}_K\|_{L^2} \le 2\mathrm{d}_{L^2}\big(f^{1/2}_{\mathrm{NW}}, f^{1/2}_K\big),\end{aligned}$$

so we conclude that the total variation distance $\mathrm{d}_{\mathrm{TV}}(f_{\mathrm{NW}}, f_K)$ must tend to zero when $p, n \to \infty$ with $p^{K+3}/n^{K+1} \to 0$. This concludes the proof. □

5. Conclusion. In this paper, we show that the behavior of a Wishart matrix when $p, n \to \infty$ with $p/n \to 0$, which we call the middle-scale asymptotics, varies according to which interval $[\frac{K}{K+2}, \frac{K+1}{K+3})$ the ratio $\log p/\log n$ tends to, when smaller than one. We show this by associating to any distribution a closely related distribution called its $\mathcal{F}$-conjugate.

In the case of the normalized Wishart distribution $\sqrt{n}[W_p(n, I_p/n) - I_p]$, this closely related distribution happens to be a generalization of the t-distribution to the real symmetric matrices. We show that the distance from the $\mathcal{F}$-conjugate of the normalized Wishart distribution to the $\mathcal{F}$-conjugate of the Gaussian orthogonal ensemble does not vary continuously with $c = \lim_{n\to\infty} \log p / \log n$, but rather jumps discontinuously as c increases, with discontinuities at the points $(K+3)/(K+1)$ for $K \in \mathbb{N}$. Thus the Wishart distribution itself must exhibit phase transitions at these discontinuities.

Moreover, in the same way that the Gaussian orthogonal ensemble approximates the normalized Wishart distribution in the classical regimes, we derive approximations for all the higher degrees. In particular, we rederive the Gaussian orthogonal ensemble approximation independently of the published proofs.

APPENDIX: PROOFS OF THE t-DISTRIBUTION RESULTS

In Section 3, results regarding $\mathcal{F}$-conjugates and the t-distribution were merely stated and not proven so as to not distract from the main results of this article. This appendix contains the proofs of these claims.

These proofs themselves require technical lemmas, so we start by stating and proving those. The first two concern repeated differentiation of expressions of the form $\exp\{a \operatorname{tr} Z\}|Z|^b$ for a symmetric matrix Z. Recall that $\tilde{\nabla}_Z$ stands for the symmetric differentiation operator $(\tilde{\nabla}_Z)_{kl} = \frac{1+\delta_{kl}}{2}\partial/\partial Z_{kl}$, where δ is the Kronecker delta, and that we write $m = n - p - 1$.

LEMMA 3. *For any indices* $1 \le i_1, \dots, i_{2l} \le p$ *and real symmetric matrix* Z, *there exist polynomials* $a_{J,s}(n,m)$ *in* n *and* m, *indexed by* $0 \le s \le l$ *and* $J = (j_1, \dots, j_{2l})$, *such that*

$$\begin{aligned}\tilde{\nabla}_{Zi_{2l}i_{2l-1}} \cdots \tilde{\nabla}_{Zi_4i_3}\tilde{\nabla}_{Zi_2i_1} \exp\left\{-\frac{n}{4}\operatorname{tr} Z\right\}|Z|^{\frac{m}{4}} \\ = \sum_{s=0}^{l} \sum_{\substack{J\in\\ \{1,\dots,p\}^{2l}}} a_{J,s}(n,m) \prod_{t=s+1}^{l} (I_p)_{j_{2t}j_{2t-1}} \prod_{t=1}^{s} Z^{-1}_{j_{2t}j_{2t-1}} \exp\left\{-\frac{n}{4}\operatorname{tr} Z\right\}|Z|^{\frac{m}{4}}.\end{aligned}$$

PROOF. To simplify notation, let

$$M_{J,s}(Z) = \prod_{t=s+1}^{l} (I_p)_{j_{2t}j_{2t-1}} \prod_{t=1}^{s} Z^{-1}_{j_{2t}j_{2t-1}} \exp\left\{-\frac{n}{4}\operatorname{tr} Z\right\}|Z|^{\frac{m}{4}},$$

and let $M_l = \{M_{J,s} | J \in \{1,\dots,p\}^{2l}, s \le l\}$ be the set of all such terms "on $2l$ indices". Let $\langle M_l \rangle$ denote the linear span of M_l, that is, the space of all linear combinations of elements of M_l, with real polynomials in n and m as coefficients. Then we are really claiming that

$$\tilde{\nabla}_{Zi_{2l}i_{2l-1}} \cdots \tilde{\nabla}_{Zi_4i_3} \tilde{\nabla}_{Zi_2i_1} \exp\left\{-\frac{n}{4}\operatorname{tr} Z\right\}|Z|^{\frac{m}{4}} \in \langle M_l\rangle. \tag{A.1}$$

To see this, let $J = (j_1, \dots, j_{2l-2}) \in \{1, \dots, p\}^{2l-2}$ and define the extension $J^q_{a,b} = (j_1, \dots, j_{q-1}, a, b, j_{q+1}, \dots, j_{2l-2}) \in \{1, \dots, p\}^{2l}$ to be J with indices a, b inserted (in this order) at the qth position. Then using that

$$\tilde{\nabla}_{Zi_{2l}i_{2l-1}} Z^{-1}_{ab} = -\frac{1}{2}[Z^{-1}_{ai_{2l}} Z^{-1}_{i_{2l-1}b} + Z^{-1}_{ai_{2l-1}} Z^{-1}_{i_{2l}b}]$$

and

$$\begin{aligned}\tilde{\nabla}_{Zi_{2l}i_{2l-1}} &\exp\left\{-\frac{n}{4}\operatorname{tr} Z\right\}|Z|^{\frac{m}{4}} \\ &= \left[\frac{m}{4} Z_{i_{2l}i_{2l-1}} - \frac{n}{4}(I_p)_{i_{2l}i_{2l-1}}\right]\exp\left\{-\frac{n}{4}\operatorname{tr} Z\right\}|Z|^{\frac{m}{4}},\end{aligned}$$

we conclude that

$$\begin{aligned}\tilde{\nabla}_{Zi_{2l}i_{2l-1}} &M_{J,s}(Z) \\ &= -\frac{1}{2}\sum_{r=1}^{s} M_{J^{2r}_{i_{2l}i_{2l-1}},s+1} - \frac{1}{2}\sum_{r=1}^{s} M_{J^{2r}_{i_{2l-1}i_{2l}},s+1} \\ &\quad + \frac{m}{4} M_{J^{2s+1}_{i_{2l}i_{2l-1}},s+1} - \frac{n}{4} M_{J^{2s+1}_{i_{2l}i_{2l-1}},s} \in \langle M_l\rangle.\end{aligned}$$

Thus, by linearity, $\tilde{\nabla}_{Zi_{2l}i_{2l-1}}$ maps $\langle M_{l-1}\rangle$ to $\langle M_l\rangle$. But by definition of $\langle M_0\rangle$ we have $\exp\{-\frac{n}{4}\operatorname{tr} Z\}|Z|^{m/4} \in \langle M_0\rangle$, so by induction, equation (A.1) must then hold, as desired. $\square$

LEMMA 4. *For any $k \in \mathbb{N}$ and any $Z \in \mathbb{S}_p(\mathbb{R})$,*

$$\operatorname{tr}\left(\tilde{\nabla}^k_X e^{-\frac{n}{4}\operatorname{tr} Z}|Z|^{\frac{m}{4}}\right) = e^{-\frac{n}{4}\operatorname{tr} Z}|Z|^{\frac{m}{4}} \sum_{|\kappa|\le 2k} b_\kappa(n, m, p) r_\kappa\left(Z^{-1}\right)$$

for some polynomials $b_\kappa(n, m, p)$ with $\deg b_\kappa \le 2k + 1 - q(\kappa)$ *and r_κ as in equation* (3.3). *The sums on the right-hand sides are taken over all integer partitions κ of norm at most $2k$, including the empty partition.*

PROOF. We give a spectral proof. Let $\lambda_1 \ge \cdots \ge \lambda_p$ be the eigenvalues of Z and OLO^t its corresponding spectral decomposition, with diagonal matrix $L = \operatorname{diag}(\lambda_1, \dots, \lambda_p)$ and orthogonal matrix O, and notice that

$$\tilde{\nabla}_{Zij} O_{hl} = \frac{1}{2}\sum_{a\ne l}^{p} \frac{O_{ha} O^t_{ai}}{\lambda_l - \lambda_a} O_{jl} + \frac{1}{2}\sum_{a\ne l} \frac{O_{ha} O^t_{aj}}{\lambda_l - \lambda_a} O_{il}, \qquad \tilde{\nabla}_{Zij}\lambda_h = O_{ih} O_{jh}$$

for any $1 \le i, j, h, l \le p$. As a consequence, for any differentiable real-valued functions $F_1(L), \dots, F_p(L)$, we have

$$\sum_{j=1}^{p} \tilde{\nabla}_{Zhj} \left(\sum_{a=1}^{p} O_{ja} F_a O_{ai}^t \right) = \frac{1}{2} \sum_{\substack{a,b \\ b \neq a}}^{p} O_{ha} \frac{F_b - F_a}{\lambda_b - \lambda_a} O_{ai}^t + \sum_{a=1}^{p} O_{ha} \frac{\partial F_a}{\partial \lambda_a} O_{ai}^t.$$

This suggests we define a new operator D_L that maps the space of diagonal matrices $F(L) = \operatorname{diag}(F_1(L), \dots, F_p(L))$ that differentiably depends on L, to itself, by

$$D_L\{F\}_a = \frac{1}{2} \sum_{b \neq a}^{p} \frac{F_b - F_a}{\lambda_b - \lambda_a} + \frac{\partial F_a}{\partial \lambda_a} \qquad \text{so that } \sum_{j=1}^{p} \tilde{\nabla}_{Zhj} O F O_{ji}^t = O F O_{ki}^t.$$

In particular, we would then have

$$\operatorname{tr}\left(\tilde{\nabla}_X^k e^{-\frac{n}{4} \operatorname{tr} Z} |Z|^{\frac{m}{4}}\right) = \operatorname{tr} D_L^{2k} \left\{ e^{-\frac{n}{4} \operatorname{tr} Z} |Z|^{\frac{m}{4}} I_p \right\}. \tag{A.2}$$

Let us look more closely at this operator D_L. It satisfies the following:

(i) D_L is linear, in the sense that for diagonals $F(L)$, $G(L)$ and constants a, b with respect to L,

$$D_L\{aF + bG\} = a D_L\{F\} + b D_L\{G\}.$$

(ii) D_L satisfies a restricted product rule, in the sense that for a diagonal $F(L)$ of the form $F(L) = f(L) I_p$ for some function $f(L)$, and any diagonal $G(L)$,

$$D_L\{FG\} = D_L\{F\} G + F D_L\{G\}.$$

Moreover, from the definition of D_L,

$$D_L\left\{e^{-\frac{n}{4} \operatorname{tr} L} I_p\right\} = -\frac{n}{4} e^{-\frac{n}{4} \operatorname{tr} L} I_p, \qquad D_L\left\{|L|^{\frac{m}{4}} I_p\right\} = \frac{m}{4} |L|^{\frac{m}{4}} I_p,$$

$$D_L\left\{\operatorname{tr}\left(L^{-s}\right) I_p\right\} = -s L^{-(s+1)} \quad \text{and}$$

$$D_L\left\{L^{-s}\right\} = -\frac{s}{2} L^{-(s+1)} - \frac{1}{2} \sum_{t=1}^{s} \operatorname{tr}\left(L^{-[s+1-t]}\right) L^{-t}.$$

Now define the spaces

$$M_l = \left\{ b(n, m, p) e^{-\frac{n}{4} \operatorname{tr} L} |L|^{\frac{m}{4}} r_\kappa\left(L^{-1}\right) L^{-s} \,\middle|\, \begin{array}{l} b(n, m, p) \text{ is a polynomial} \\ \text{with degree at most } l - \\ q(\kappa), \text{ and } \kappa \text{ and } s \text{ satisfies} \\ |\kappa| \le l - s. \end{array} \right\}$$

for $l = 1, \dots, 2k$, and let $\langle M_l \rangle$ denote the linear span of M_l, that is, the space of all real linear combinations of elements of M_l. Moreover, for a partition κ, let $\kappa \pm i$ denote κ with the integer i added or removed, respectively. For example,

$(3, 1, 1, 1) + 2 = (3, 2, 1, 1, 1)$ and $(3, 2, 1, 1, 1) - 1 = (3, 2, 1, 1)$. Note that $|\kappa \pm i| = |\kappa| \pm i$. Then, for any $F \in M_l$,

$$\begin{aligned}
D_L\{F\} &= D_L\{b(n, m, p)e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_\kappa(L^{-1})L^{-s}\} \\
&= b(n, m, p)D_L\{e^{-\frac{n}{4}\operatorname{tr} L} I_p\}|L|^{\frac{m}{4}} r_\kappa(L^{-1})L^{-s} \\
&\quad + b(n, m, p)e^{-\frac{n}{4}\operatorname{tr} L} D_L\{|L|^{\frac{m}{4}} I_p\} r_\kappa(L^{-1})L^{-s} \\
&\quad + b(n, m, p)e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} D_L\{r_\kappa(L^{-1}) I_p\}L^{-s} \\
&\quad + b(n, m, p)e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_\kappa(L^{-1}) D_L\{L^{-s}\} \\
&= \left[-\frac{n}{4} b(n, m, p)\right] e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_\kappa(L^{-1})L^{-s} \\
&\quad + \left[\frac{m}{4} b(n, m, p)\right] e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_\kappa(L^{-1})L^{-s} \\
&\quad + \sum_{i=1}^{q(\kappa)} [-\kappa_i b(n, m, p)] e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_{\kappa-\kappa_i}(L^{-1})L^{-s} \\
&\quad + \left[-\frac{s}{2} b(n, m, p)\right] e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_\kappa(L^{-1})L^{-(s+1)} \\
&\quad + \sum_{t=1}^{s} \left[-\frac{1}{2} b(n, m, p)\right] e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_{\kappa+(s+1-t)}(L^{-1})L^{-t}.
\end{aligned}$$

Thus $D_L\{F\} \in \langle M_{l+1}\rangle$. It follows by linearity that D_L maps $\langle M_l\rangle$ to $\langle M_{l+1}\rangle$.

Now, $e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} I_p \in M_0$, so by induction $D_L^{2k}\{e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} I_p\} \in \langle M_{2k}\rangle$. Hence, for some polynomials $b_{\kappa,s}(n, m, p)$ of degree at most $2k - q(\kappa)$,

$$\begin{aligned}
&\operatorname{tr} D_L^{2k}\{e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} I_p\} \\
&\quad = \sum_{|\kappa|+s \le 2k} b_{\kappa,s}(n, m, p)e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_\kappa(L^{-1}) \operatorname{tr}(L^{-s}) \\
&\quad = \sum_{|\kappa'| \le 2k} b_{\kappa'}(n, m, p)e^{-\frac{n}{4}\operatorname{tr} L}|L|^{\frac{m}{4}} r_{\kappa'}(L^{-1})
\end{aligned}$$

for $\kappa' = \kappa + s$, $b_{\kappa'} = b_{\kappa,s}$ when $s \neq 0$, while $\kappa' = \kappa$, $b_{\kappa'} = p b_{\kappa,s}$ when $s = 0$. Notice that when $s \neq 0$, the degree of the $b_{\kappa'}$'s is at most $2k - q(\kappa) = 2k - (q(\kappa') - 1)$, while when $s = 0$ it is at most $2k - q(\kappa) + 1 = 2k - q(\kappa') + 1$. Thus in both cases, $\deg b_{\kappa'} \le 2k - q(\kappa') + 1$, which equation (A.2) shows the lemma. □

We will also need in our proofs a result about the asymptotics of inverse moments of the Wishart distribution.

LEMMA 5. *Let $Y \sim W_p(n, I_p/n)$ and s be any integer $s \geq 1$. Then as long as $n \geq p + 4s + 2$, the s-th inverse moment satisfies the recursive bound*

$$\left(1 - \frac{(p+1)s}{n}\right)\mathrm{E}\big[\operatorname{tr} Y^{-s}\big] \leq \mathrm{E}\big[\operatorname{tr} Y^{-(s-1)}\big].$$

In particular, when $n, p \to \infty$ such that $\lim_{n\to\infty} p/n = \alpha < 1$, if $s < \alpha^{-1}$ then $\mathrm{E}[\operatorname{tr} Y^{-s}] = O(p)$.

PROOF. For any differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that $(\partial/\partial Z - Z) f(Z)$ is integrable when $Z \sim \mathrm{N}(0, 1)$, integration by parts yields that $\mathrm{E}[(\partial/\partial Z - Z) f(Z)] = 0$, a result also known as Stein's lemma. Let Z be an $n \times p$ matrix of independent standard normal random variables, and let $Y = \frac{1}{n} Z^t Z \sim \mathrm{W}_p(n, I_p/n)$. For any $1 \leq \alpha \leq n$ and $1 \leq \beta, i, j \leq p$, we have $\partial/\partial Z_{\alpha\beta} = \frac{2}{n}\sum_{i=1}^{p} Z_{\alpha i} \tilde{\nabla}_{Yi\beta}$ and

$$\tilde{\nabla}_{Yi\beta} Y_{\beta j}^{-s} = -\frac{1}{2}\sum_{l=1}^{s}\big[Y_{\beta i}^{-l} Y_{\beta j}^{-(s-l+1)} + Y_{\beta\beta}^{-l} Y_{ij}^{-(s-l+1)}\big],$$

so for δ the Kronecker delta,

$$\begin{aligned}
&\left(\frac{\partial}{\partial Z_{\alpha\beta}} - Z_{\alpha\beta}\right)(ZY^{-s})_{\alpha\beta} \\
&\quad = \sum_{j=1}^{p}\left[\delta_{\beta j} Y_{\beta j}^{-s} + \frac{2}{n}\sum_{i=1}^{p} Z_{\alpha j} Z_{\alpha i} \tilde{\nabla}_{Yi\beta} Y_{\beta j}^{-s} - Z_{\alpha\beta} Z_{\alpha j} Y_{\beta j}^{-s}\right] \\
&\quad = Y_{\beta\beta}^{-s} - \frac{1}{n}\sum_{l=1}^{s}(ZY^{l})_{\alpha\beta}(ZY^{-(s-l+1)})_{\alpha\beta} \\
&\qquad - \frac{1}{n}\sum_{l=1}^{s} Y_{\beta\beta}^{-l}(ZY^{-(s-l+1)}Z^t)_{\alpha\alpha} - Z_{\alpha\beta}(ZY^{-s})_{\alpha\beta}.
\end{aligned} \tag{A.3}$$

Let us first show that this expression is integrable. For any matrix X, we have $|X_{ij}| \leq \|X\|_2 = \|X^t X\|_2^{1/2}$. Thus by equation (A.3),

$$\begin{aligned}
&\mathrm{E}\left[\left|\left(\frac{\partial}{\partial Z_{\alpha\beta}} - Z_{\alpha\beta}\right)(ZY^{-s})_{\alpha\beta}\right|\right] \\
&\quad \leq \mathrm{E}\Bigg[\|Y^{-s}\|_2 + \sum_{l=1}^{s}\|Y^{2l+1}\|_2^{\frac{1}{2}}\|Y^{2s+2l-1}\|_2^{\frac{1}{2}} \\
&\qquad + \sum_{l=1}^{s}\|Y^{-2l}\|_2^{\frac{1}{2}}\|Y^{-2s+2l}\|_2^{\frac{1}{2}} + n\|Y\|_2^{\frac{1}{2}}\|Y^{-2s+1}\|_2^{\frac{1}{2}}\Bigg].
\end{aligned}$$

As Y is positive definite, $\|Y^{\pm a}\|_2 \leq \operatorname{tr} Y^{\pm a}$ for any $a \in \mathbb{N}$, so by the Cauchy–Schwarz inequality,

$$\leq \mathrm{E}[\operatorname{tr} Y^{-s}] + \sum_{l=1}^{s} \mathrm{E}[\operatorname{tr} Y^{-2l+1}]^{\frac{1}{2}} \mathrm{E}[\operatorname{tr} Y^{-2s+2l-1}]^{\frac{1}{2}}$$

$$+ \sum_{l=1}^{s} \mathrm{E}[\operatorname{tr} Y^{-2l}]^{\frac{1}{2}} \mathrm{E}[\operatorname{tr} Y^{-2s+2l}]^{\frac{1}{2}} + n\,\mathrm{E}[\operatorname{tr} Y]^{\frac{1}{2}} \mathrm{E}[\operatorname{tr} Y^{-2s+1}]^{\frac{1}{2}},$$

which is finite for $n \geq p + 4s + 2$. Moreover, $(ZY^{-s})_{\alpha\beta}$ can be expressed using minors and determinants as a rational function of the entries of Z, so $\lim_{Z_{\alpha\beta}\to\pm\infty}(ZY^{-s})_{\alpha\beta} e^{-Z_{\alpha\beta}^2/2} = 0$. So all conditions are fulfilled to apply Stein's lemma to equation (A.3) and obtain

$$\text{(A.4)} \quad \begin{aligned} 0 &= \mathrm{E}\left[\frac{1}{n}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{p}\left(\frac{\partial}{\partial Z_{\alpha\beta}} - Z_{\alpha\beta}\right)(ZY^{-s})_{\alpha\beta}\right] \\ &= \mathrm{E}\left[\operatorname{tr} Y^{-s} - \frac{s}{n}\operatorname{tr} Y^{-s} - \frac{1}{n}\sum_{l=1}^{s}\operatorname{tr}(Y^{-l})\operatorname{tr}(Y^{-(s-l)}) - \operatorname{tr} Y^{-(s-1)}\right]. \end{aligned}$$

Now let $\lambda_1 \geq \cdots \geq \lambda_p$ be the eigenvalues of Y. For any $1 \leq l \leq s$, we have $\lambda_p^{-l} \geq \cdots \geq \lambda_1^{-l}$ and $\lambda_p^{-(s-l)} \geq \cdots \geq \lambda_1^{-(s-l)}$, so Chebyshev's sum inequality (Hardy, Littlewood and Pólya (1965), Theorem 43) entails $\operatorname{tr}(Y^{-s})\operatorname{tr}(Y^{-(s-l)}) \leq p\operatorname{tr}(Y^{-s})$. Employing this result in equation (A.4), whose terms are all individually integrable as $n \geq p + 4s + 2$, then yields that

$$\text{(A.5)} \quad \left(1 - \frac{(p+1)s}{n}\right)\mathrm{E}[\operatorname{tr} Y^{-s}] \leq \mathrm{E}[\operatorname{tr} Y^{-(s-1)}].$$

This concludes the first part of the proof.

For the second part, if we let $n \to \infty$ such that $\lim_{n\to\infty} \frac{p}{n} = \alpha < 1$, then eventually $n \geq p + 4s + 2$ and $n \geq (p+1)s$ for n large enough. So by repeatedly applying equation (A.5) and dividing by p, we obtain

$$\prod_{l=1}^{s}\left(1 - \frac{(p+1)l}{n}\right) \cdot \frac{1}{p}\mathrm{E}[\operatorname{tr} Y^{-s}] \leq \frac{1}{p}\mathrm{E}[\operatorname{tr} Y^{-0}] = 1.$$

Taking a limit in the above yields

$$\prod_{l=1}^{s}(1 - \alpha l)\lim_{n\to\infty}\frac{1}{p}\mathrm{E}[\operatorname{tr} Y^{-s}] \leq 1,$$

so for any $s < \alpha^{-1}$ we have

$$\text{(A.6)} \quad \lim_{n\to\infty}\frac{1}{p}\mathrm{E}[\operatorname{tr} Y^{-s}] \leq \prod_{l=1}^{s}\frac{1}{1-\alpha l} < \infty,$$

as desired. □

With these technical lemmas in hand, we now proceed to prove the claims of Section 3, in order of appearance.

PROOF OF PROPOSITION 2. Let $\mathcal{F}\{f^{1/2}\} = \phi$ to lighten the notation. We can write

$$d_{L^2}(\phi, \psi) = [\|\psi\|_{L^2}^2 - 1] + 2\int_{\mathbb{S}_p(\mathbb{R})}\left[1 - \Re\left\{\frac{\psi^{1/2}(T)}{\phi^{1/2}(T)}\right\}\right]|\phi|(T)\,dT$$

$$= [\|\psi\|_{L^2}^2 - 1] + 2\int_{\mathbb{S}_p(\mathbb{R})}\left[1 - \exp\left\{\frac{R}{2}\right\}\cos\left(\frac{I}{2}\right)\right]|\phi|(T)\,dT$$

for $R = \Re\,\mathrm{Log}[\phi(T)/\psi(T)]$ and $I = \Im\,\mathrm{Log}[\phi(T)/\psi(T)]$. Using the inequality $-\cos(x) \leq -1 + \sqrt{2|x|}$, this can be bounded by

$$\leq [\|\psi\|_{L^2}^2 - 1] + 2\int_{\mathbb{S}_p(\mathbb{R})}\left[1 - \exp\left\{\frac{R}{2}\right\} + \exp\left\{\frac{R}{2}\right\}|I|^{\frac{1}{2}}\right]|\phi|(T)\,dT,$$

which, by the inequality $1 - \exp(x) \leq -x$ and the Cauchy–Schwarz inequality, can be further bounded by

$$\leq [\|\psi\|_{L^2}^2 - 1] + \int_{\mathbb{S}_p(\mathbb{R})} R|\phi|(T)\,dT$$
$$+ 2\int_{\mathbb{S}_p(\mathbb{R})}\exp\{-R\}|\phi|(T)\,dT^{\frac{1}{2}}\int_{\mathbb{S}_p(\mathbb{R})}|I||\phi|(T)\,dT^{\frac{1}{2}}.$$

But by definition of R, $\exp\{-R\} = |\psi|(T)/|\phi|(T)$. Plugging this back in the above and simplifying yields the desired result. □

PROOF OF PROPOSITION 3. Let $\mathrm{NW}(n, p) = \sqrt{n}[W_p(n, I_p/n) - I_p]$ denote the normalized Wishart distribution. By a change of variables from the Wishart density (Muirhead (1982), Theorem 3.2.1), the density of the normalized Wishart is

$$f_{\mathrm{NW}(n,p)}(X) = \frac{n^{\frac{p(2n-p-1)}{4}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})}e^{-\frac{n}{2}\mathrm{tr}[I_p + \frac{X}{\sqrt{n}}]}\left|I_p + \frac{X}{\sqrt{n}}\right|^{\frac{n-p-1}{2}}\mathbb{1}\left[I_p + \frac{X}{\sqrt{n}} > 0\right],$$

where Γ_p stands for the multivariate Gamma function. But

$$f_{\mathrm{NW}(n,p)}^{1/2}(X) = \frac{2^{\frac{p(2n+p+1)}{2}}\Gamma_p(\frac{n+p+1}{4})}{n^{\frac{p(3n+p+1)}{4}}\Gamma_p(\frac{n}{2})}f_{\sqrt{n}[\mathrm{W}_p(\frac{n+p+1}{2}, \frac{2}{n}I_p) - I_p]}(X), \tag{A.7}$$

which is integrable when $(n + p + 1)/2 > p - 1$, that is, $n \geq p - 2$ (Muirhead (1982), Theorem 3.2.1 again and comment on page 87). For an integrable function f on $\mathbb{S}_p(\mathbb{R})$, the Fourier transform with kernel $\exp\{-i\,\mathrm{tr}(XT)\}$, normalized to be an L^2-isometry, satisfies $\mathcal{F}\{f\}(T) = 2^{-\frac{p}{2}}\pi^{-\frac{p(p+1)}{4}} \int_{\mathbb{S}_p(\mathbb{R})} e^{-i\operatorname{tr}(XT)} f(X)\,dX$. Thus, we obtain that

$$
\begin{aligned}
&\mathcal{F}\{f_{\mathrm{NW}(n,p)}^{1/2}\}(T) \\
&\quad = \frac{2^{\frac{p(2n+p+1)}{2}}\Gamma_p(\frac{n+p+1}{4})}{n^{\frac{p(3n+p+1)}{4}}\Gamma_p(\frac{n}{2})} e^{i\sqrt{n}\operatorname{tr} T}\mathcal{F}\{f_{\mathrm{W}_p(\frac{n+p+1}{2},\frac{2}{n}I_p)}\}(\sqrt{n}T) \\
&\quad = \frac{2^{\frac{p(2n+p+1)}{2}}\Gamma_p(\frac{n+p+1}{4})}{n^{\frac{p(3n+p+1)}{4}}\Gamma_p(\frac{n}{2})} \frac{e^{i\sqrt{n}\operatorname{tr} T}}{2^{\frac{p}{2}}\pi^{\frac{p(p+1)}{4}}} \int_{\mathbb{S}_p(\mathbb{R})} e^{-i\operatorname{tr}(XT)} f_{\mathrm{W}_p(\frac{n+p+1}{2},\frac{2}{n}I_p)}(X)\,dX \\
&\quad = \frac{2^{\frac{p(n+2p)}{4}}}{\pi^{\frac{p(p+1)}{4}} n^{\frac{p(p+1)}{8}}} \frac{\Gamma_p(\frac{n+p+1}{4})}{\Gamma_p^{1/2}(\frac{n}{2})} e^{i\sqrt{n}\operatorname{tr} T}\left|I_p + i\frac{4T}{\sqrt{n}}\right|^{-\frac{n+p+1}{4}},
\end{aligned}
$$

using the characteristic function of the Wishart distribution (Muirhead (1982), Theorem 3.2.3). Squaring this result yields the desired expression. $\square$

PROOF OF LEMMA 1. By a change of variable in the GOE(p) density $2^{-p(p+3)/4}\pi^{-p(p+1)/4}\exp\{-\operatorname{tr} X^2/4\}$, we find that the normalizing constant of the GOE(p)/4 distribution is $C_{\mathrm{GOE}/4} = 2^{p(3p+1)/4}/\pi^{p(p+1)/4}$. Then by Stirling's approximation as well as Muirhead (1982), Theorem 2.1.12,

$$
\begin{aligned}
\log C_t = {} & \frac{p(3p+1)}{4}\log 2 - \frac{p(p+1)}{4}\log\pi - \frac{p(p+3)}{4} \\
& + \frac{1}{2}\sum_{i=1}^{p}\bigl(n - [2i-p-1]\bigr)\log\left(1 - \frac{2i-p-3}{n}\right) \\
& - \frac{1}{2}\sum_{i=1}^{p}(n-i)\log\left(1 - \frac{i-1}{n}\right) + o(1)
\end{aligned}
\tag{A.8}
$$

as $n \to \infty$ with $p/n \to 0$. We focus on the two sums in this expression. By Taylor's theorem applied to $-\log(1-x)$ around $x = 0$,

$$
\begin{aligned}
&-\sum_{i=1}^{p}\frac{n-i}{2}\log\left(1 - \frac{i-1}{n}\right) \\
&\quad = \sum_{k=1}^{K+1}\sum_{i=1}^{p}\frac{n-i}{2k}\left(\frac{i-1}{n}\right)^k + O\left(\frac{p^{K+3}}{n^{K+1}}\right) \\
&\quad = \frac{p(p-1)}{4} + \sum_{k=1}^{K} c_k^{(1)}\frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} c_k^{(2)}\frac{p^{k+1}}{n^k} + O\left(\frac{p^{K+3}}{n^{K+1}}\right)
\end{aligned}
$$

for some constants $c_k^{(1)}, c_k^{(2)}$, while

$$\begin{aligned}\sum_{i=1}^{p} & \frac{n-[2i-p-1]}{2}\log\left(1-\frac{2i-p-3}{n}\right) \\ &= -\sum_{k=1}^{K+1}\sum_{i=1}^{p}\frac{n-[2i-p-1]}{2k}\left(\frac{2i-p-3}{n}\right)^k + O\left(\frac{p^{K+3}}{n^{K+1}}\right) \\ &= p+\sum_{k=1}^{K} c_k^{(3)}\frac{p^{k+2}}{n^k}+\sum_{k=1}^{K} c_k^{(4)}\frac{p^{k+1}}{n^k}+O\left(\frac{p^{K+3}}{n^{K+1}}\right)\end{aligned}$$

for constants $c_k^{(3)}, c_k^{(4)}$. Filling these two expressions back in equation (A.8) yields the desired result. □

PROOF OF LEMMA 2. Let f_{NW} and f_t stand for the densities of the normalized Wishart $\sqrt{n}[\mathrm{W}_p(n, I_p/n) - I_p]$ and the Sym-$t_{n/2}(I_p/8)$ distributions, respectively. Let $R(X) = -X$ be the flip operator and $\star$ stand for the convolution. Since $f_{\mathrm{NW}}^{1/2}$ is proportional to a $\sqrt{n}[\mathrm{W}_p(\frac{n+p+1}{2}, \frac{2}{n}I_p) - I_p]$ density [see equation (A.7)], it is integrable. Therefore, $f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)$ is well defined and integrable as well. But then, as f_{NW} is real-valued,

$$\begin{aligned}2^{-\frac{p}{2}}\pi^{-\frac{p(p+1)}{4}}\mathcal{F}\{f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)\} &= \mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}\mathcal{F}\{f_{\mathrm{NW}}^{1/2} \circ R\} \\ &= \mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}\overline{\mathcal{F}\{f_{\mathrm{NW}}^{1/2}\}} = f_t\end{aligned}$$

because the Sym-$t_{n/2}(I_p/8)$ is the $\mathcal{F}$-conjugate of the normalized Wishart distribution by Equation (3.2). The Fourier inversion formula then yields that

$$f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)(X) = \int_{\mathbb{S}_p(\mathbb{R})} e^{i\,\mathrm{tr}(TX)} f_t(T)\,dT. \tag{A.9}$$

Thus the characteristic function of the Sym-$t_{n/2}(I_p/8)$ distribution is given by $f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)$. As the derivatives of the characteristic function of a distribution evaluated at zero provide its moments, up to a constant, we can try to repeatedly differentiate $f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)$ at zero to compute $\mathrm{E}[\mathrm{tr}\, T^{2k}]$.

Unfortunately, the convolution is given by an integral whose domain makes it difficult to directly interchange the differentiation and integration symbols. Because the integrand is orthogonally invariant, we found it easier to compute the derivatives at zero by taking a limit over a sequence of decreasing positive-definite matrices at both sides instead. In this spirit, define on the open set $\{0 < X < I_p\} \subset \mathbb{S}_p(\mathbb{R})$ the real-valued function

$$H(X) = \frac{(-1)^k}{n^k}\,\mathrm{tr}\big(\tilde{\nabla}_X^k f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)(\sqrt{n}X)\big)$$

for fixed k, p and n. Here, $(\tilde{\nabla}_X)_{ij} = \frac{1+\delta_{ij}}{2}\partial/\partial X_{ij}$ is the symmetric differentiation operator, as defined in Section 2, with the property that $\tilde{\nabla}_X \operatorname{tr}(XT) = T$ for any two symmetric matrices X, T. The $\sqrt{n}$ scaling in the argument of H links the convolution to an expectation with respect to an inverse Wishart distribution.

We can relate this function to the moments of the Sym-$t_{n/2}(I_p/8)$ distribution. For any $1 \le l \le k$, indices $1 \le i_1, \dots, i_{2l} \le p$ and $X \in \mathbb{S}_p(\mathbb{R})$,

$$\left|\tilde{\nabla}_{Xi_{2l}i_{2l-1}} \cdots \tilde{\nabla}_{Xi_4i_3}\tilde{\nabla}_{Xi_2i_1} e^{i\sqrt{n}\operatorname{tr}(TX)}\right| = n^l |T_{i_{2l}i_{2l-1}} \cdots T_{i_4i_3}T_{i_2i_1}|. \tag{A.10}$$

We first show this expression is integrable for large enough n when $T \sim$ Sym-$t_{n/2}(I_p/8)$. This is not obvious, as when $p = 1$, asking if this expression is integrable is the same as asking if the t-distribution with $n/2$ degrees of freedom has an l-th moment, which is true only when $l < n/2$.

For any symmetric matrix T, $|T_{ij}| \le \lambda_1^{1/2}(T^2) \le |I_p + T^2|^{1/2}$, where $\lambda_1(T^2) \ge \cdots \ge \lambda_p(T^2) \ge 0$ are the ordered eigenvalues of the positive-definite matrix T^2. Thus

$$\begin{aligned} &\int_{\mathbb{S}_p(\mathbb{R})} n^l |T_{i_{2l}i_{2l-1}} \cdots T_{i_2i_1}| f_t(T)\, dT \\ &\quad\le \frac{n^{\frac{3l}{2}}}{4^l} C_t \int_{\mathbb{S}_p(\mathbb{R})} \left| I_p + \frac{16T^2}{n}\right|^{-\frac{(n-2l)+p+1}{4}} dT \\ &\quad\le \frac{n^{\frac{3l}{2}} C_t}{4^l}\left(\frac{n}{n-2l}\right)^{\frac{p(p+1)}{4}} \int_{\mathbb{S}_p(\mathbb{R})} \left| I_p + \frac{16T^2}{n-2l}\right|^{-\frac{(n-2l)+p+1}{4}} dT. \end{aligned} \tag{A.11}$$

When $n - 2l \ge p - 2$, the last integrand is proportional to the density of a Sym-$t_{n/2-l}(I_p/8)$ distribution, so the integral is finite. Thus, when $n \ge p + 2k - 2$, the right-hand side of equation (A.10) is an integrable function for all $1 \le l \le k$ and $1 \le i_1, \dots, i_{2l} \le p$. By equation (A.9), and repeated differentiation under the integral sign justified by the integrability bounds given by equations (A.10) and (A.11), we find that

$$H(X) = \int_{\mathbb{S}_p(\mathbb{R})} \operatorname{tr} T^{2k} e^{i\sqrt{n}\operatorname{tr}(TX)} f_t(T)\, dT \tag{A.12}$$

for any $X \in \mathbb{S}_p(\mathbb{R})$ and any $n \ge p + 2k - 2$.

Now let us relate H to the definition of $f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)$ as a convolution. This is where restricting H to small positive-definite matrices becomes useful. From the definition of the Wishart density (e.g., Muirhead (1982), Theorem 3.2.1), the expression equals

$$\begin{aligned} &f_{\mathrm{NW}}^{1/2} \star (f_{\mathrm{NW}}^{1/2} \circ R)(\sqrt{n}X) \\ &\quad= \int_{\mathbb{S}_p(\mathbb{R})} f_{\mathrm{NW}}^{1/2}(Z) f_{\mathrm{NW}}^{1/2}(Z - \sqrt{n}X)\, dZ \end{aligned}$$

$$= \frac{n^{\frac{np}{2}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})} \int_{Y+X>0,Y>0} e^{-\frac{n}{4}\operatorname{tr}(Y+X)}|Y+X|^{\frac{m}{4}} e^{-\frac{n}{4}\operatorname{tr} Y}|Y|^{\frac{m}{4}}\,dY$$

using the change of variables $Y = I_p + Z/\sqrt{n} - X$ with $dZ = n^{\frac{p(p+1)}{4}}\,dY$. For $X > 0$, we have $\mathbb{1}[Y + X > 0, Y > 0] = \mathbb{1}[Y > 0]$, and thus H satisfies

$$H(X) = \frac{(-1)^k}{n^k}\operatorname{tr}\biggl(\tilde{\nabla}_X^k \int_{Y>0} e^{-\frac{n}{4}\operatorname{tr}(Y+X)}|Y+X|^{\frac{m}{4}} \frac{n^{\frac{np}{2}} e^{-\frac{n}{4}\operatorname{tr}(Y)}|Y|^{\frac{m}{4}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})}\,dY\biggr). \tag{A.13}$$

We would now like to interchange the integral and differentiation signs. From Lemma 3, for any indices $1 \le l \le k$ and $1 \le i_1, \dots, i_{2l} \le p$, and any symmetric matrices $X, Y \in \mathbb{S}_p(\mathbb{R})$, we must have some crude bound

$$\begin{aligned}
&\bigl|\tilde{\nabla}_{Xi_{2l}i_{2l-1}} \cdots \tilde{\nabla}_{Xi_4i_3}\tilde{\nabla}_{Xi_2i_1} e^{-\frac{n}{4}\operatorname{tr}(Y+X)}|Y+X|^{\frac{m}{4}}\bigr| \\
&\quad\le \sum_{s=0}^{l} \sum_{\substack{J\in\\ \{1,\dots,p\}^{2l}}} |a_{J,s}(n,m)| \prod_{t=s+1}^{l} |(I_p)_{j_{2t}j_{2t-1}}| \\
&\qquad\times \prod_{t=1}^{s} |(Y+X)^{-1}_{j_{2t}j_{2t-1}}| e^{-\frac{n}{4}\operatorname{tr}(Y+X)}|Y+X|^{\frac{m}{4}}
\end{aligned}$$

for some polynomials $a_{J,s}$ that do not depend on X or Y. This can be uniformly bounded for all $0 \le X \le I_p$ by

$$\le C(n,m,p)\sum_{s=0}^{l} \operatorname{tr}^s(Y^{-1}) e^{-\frac{n}{4}\operatorname{tr} Y}[1+\operatorname{tr} Y]^{\frac{mp}{4}} \tag{A.14}$$

for some constant $C(n,m,p)$ that does not depend on X or Y. But for any $n \ge p-2$ and $l \ge 0$,

$$\begin{aligned}
&\int_{Y>0} C(n,m,p)\sum_{s=0}^{l} \operatorname{tr}^s(Y^{-1}) e^{-\frac{n}{4}\operatorname{tr} Y}[1+\operatorname{tr} Y]^{\frac{mp}{4}} \frac{n^{\frac{np}{2}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})} e^{-\frac{n}{4}\operatorname{tr} Y}|Y|^{\frac{m}{4}}\,dY \\
&\quad= \frac{n^{\frac{np}{2}} C(n,m,p)}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})} \sum_{s=0}^{l} \int_{Y>0} [1+\operatorname{tr} Y]^{\frac{mp}{4}} \operatorname{tr}^s(Y^{-1}) e^{-\frac{n}{4}\operatorname{tr} Y}|Y|^{\frac{n+p+1}{4}-\frac{p+1}{2}}\,dY \\
&\quad= \left(\frac{n}{2}\right)^{\frac{mp}{4}} \frac{\Gamma_p(\frac{n+p+1}{4})}{\Gamma_p(\frac{n}{2})} \mathrm{E}\bigl[(1+\operatorname{tr} Y)^{\frac{mp}{4}} \operatorname{tr}^s(Y^{-1})\bigr]
\end{aligned}$$

for a Y with a Wishart distribution $\mathrm{W}_p(\frac{n+p+1}{2}, \frac{n}{2}I_p)$. The Cauchy–Schwarz inequality then entails the bound

$$\le \left(\frac{n}{2}\right)^{\frac{mp}{4}} \frac{\Gamma_p(\frac{n+p+1}{4})}{\Gamma_p(\frac{n}{2})} \mathrm{E}\bigl[(1+\operatorname{tr} Y)^{\frac{mp}{2}}\bigr]^{\frac{1}{2}} \mathrm{E}\bigl[\operatorname{tr}^{2s}(Y^{-1})\bigr]^{\frac{1}{2}}. \tag{A.15}$$

The first expectation is always finite when $n \geq p - 2$. Since $\operatorname{tr}^{2s}(Y^{-1})$ can be written as a sum of zonal polynomials indexed by partitions of the integer $2s$, the results of Muirhead (1982), Theorem 7.2.13, imply that the second expectation is finite whenever $(n + p + 1)/4 > 2s + (p - 1)/2$, or $n \geq p + 8s - 2$. Thus, in equation (A.13) with $l \leq k$, whenever $n \geq p + 8k - 2$ we are justified in repeatedly differentiating under the integral sign by the integrability bounds given by equations (A.14) and (A.15), and obtain

$$H_1(X) = \frac{(-1)^k}{n^k} \int_{Y>0} \operatorname{tr}(\tilde{\nabla}_X^k e^{-\frac{n}{4}\operatorname{tr}(Y+X)} |Y + X|^{\frac{m}{4}}) \frac{n^{\frac{np}{2}} e^{-\frac{n}{4}\operatorname{tr} Y} |Y|^{\frac{m}{4}}}{2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})} \, dY. \tag{A.16}$$

Let us now look at how $H(X)$ behaves as $X \to 0$. On one hand, for any symmetric matrix T we have $|\operatorname{tr} T^k| \leq \sqrt{p \operatorname{tr} T^{2k}} \leq \sqrt{p}|I_p + T^2|^{k/2}$, so we must have the bound

$$\left|\operatorname{tr} T^{2k} e^{i\sqrt{n}\operatorname{tr}(TX)} f_t(T)\right| \leq \frac{n^k C_t}{16^k} \left| I_p + \frac{16T^2}{n} \right|^{-\frac{(n-4k)+p+1}{4}}$$

holding uniformly in X. But the right-hand side is proportional to the Sym-$t_{(n-4k)/2}(I_p/8)$ density, so is integrable whenever $(n - 4k)/2 \geq p/2 - 1$, or $n \geq p + 4k - 2$. Thus, by the dominated convergence theorem and equation (A.12),

$$\lim_{\substack{X \to 0 \\ 0 < X < I_p}} H(X) = \int_{\mathbb{S}_p(\mathbb{R})} \operatorname{tr} T^{2k} \lim_{\substack{X \to 0 \\ 0 < X < I_p}} e^{i\sqrt{n}\operatorname{tr}(TX)} f_t(T) \, dT = \mathrm{E}[\operatorname{tr} T^{2k}] \tag{A.17}$$

for a $T \sim$ Sym-$t_{n/2}(I/8)$.

On the other hand, the integrand at equation (A.16) takes a particularly simple form. Lemma 4 establishes by induction that there must be polynomials b_κ in n, m and p with $\deg b_\kappa \leq 2k + 1 - q(\kappa)$ such that

$$\begin{aligned} H(X) = {} & \frac{(-1)^k}{n^k} \int_{Y>0} \sum_{|\kappa| \leq 2k} b_\kappa^{(1)}(n, m, p) r_\kappa([Y + X]^{-1}) e^{-\frac{n}{4}\operatorname{tr}(Y+X)} |Y + X|^{\frac{m}{4}} \\ & \times \frac{n^{\frac{np}{2}} e^{-\frac{n}{4}\operatorname{tr} Y} |Y|^{\frac{m}{4}}}{2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})} \, dY \end{aligned} \tag{A.18}$$

for any $0 < X < I_p$ and $n \geq p + 8k - 2$. The sum is taken over all partitions of the integers κ satisfying $|\kappa| \leq 2k$, including the empty partition. But for any κ, the bound

$$r_\kappa([Y + X]^{-1}) e^{-\frac{n}{4}\operatorname{tr}(Y+X)} |Y + X|^{\frac{m}{4}} \leq \operatorname{tr}^{|\kappa|}(Y^{-1}) e^{-\frac{n}{4}\operatorname{tr} Y} [1 + \operatorname{tr} Y]^{\frac{mp}{4}}$$

holds uniformly in $0 \leq X \leq I_p$. Thus for $|\kappa| \leq 2k$, the right-hand side is integrable for $n \geq p + 8k + 6$, by the same argument as for equation (A.15). Thus for such n, by the dominated convergence theorem and equation (A.18), we obtain that

$$\lim_{\substack{X\to 0\\ 0<X<I_p}} H(X) = \frac{(-1)^k}{n^k}\int_{Y>0} \lim_{\substack{X\to 0\\ 0<X<I_p}} \sum_{|\kappa|\le 2k} b_\kappa^{(1)}(n,m,p) r_\kappa\big([Y+X]^{-1}\big) \tag{A.19}$$
$$\times e^{-\frac{n}{4}\operatorname{tr}(Y+X)}|Y+X|^{\frac{m}{4}} \frac{n^{\frac{np}{2}} e^{-\frac{n}{4}\operatorname{tr} Y}|Y|^{\frac{m}{4}}}{2^{\frac{np}{2}}\Gamma_p(\frac{n}{2})}\,dY,$$

where Y follows a $\mathrm{W}_p(n, I_p/n)$ distribution. Combining equation (A.17) with equation (A.19) and Lemma 4 concludes the proof. □

PROOF OF THEOREM 2. Recall the expected zonal polynomial of an inverse Wishart $\mathrm{W}_p^{-1}(n, I_p/n)$ is given by equation (3.5). Based on the previous calculations, it is tempting to define

$$c'_\lambda = \frac{2^{|\lambda|}|\lambda|!\prod_{i<j}^{q(\lambda)}(2\lambda_i - 2\lambda_j - i + j)}{\prod_{i=1}^{q(\lambda)}(2\lambda_i + q(\lambda) - i)!},$$
$$R_\lambda(m) = \prod_{i=1}^{q(\lambda)}\prod_{l=0}^{\lambda_i - 1}\frac{m}{m-(1-i+2l)}, \tag{A.20}$$
$$P_\lambda(m,p) = \prod_{i=1}^{q(\lambda)}\prod_{l=0}^{\lambda_i - 1}\left(\frac{p}{m} + \frac{1-i+2l}{m}\right)$$

so that $\mathrm{E}[C_\lambda(Y^{-1})] = c'_\lambda n^{|\lambda|} R_\lambda(m) P_\lambda(m,p)$. With these expressions, the expected power sum polynomials can be written

$$\mathrm{E}\big[r_\kappa(Y^{-1})\big] = \sum_{|\lambda|=|\kappa|}^{p} c_{\kappa,\lambda} c'_\lambda n^{|\kappa|}\Bigg(\prod_{|\mu|=|\kappa|} R_\mu(m) \prod_{\substack{|\mu|=|\kappa|\\ \mu\ne\lambda}} R_\mu^{-1}(m)\Bigg) P_\lambda(m,p) \tag{A.21}$$
$$= \frac{n^{|\kappa|}}{m^{|\kappa|}} R'_{|\kappa|}(m) P'_{|\kappa|}(m,p),$$

where

$$R'_{|\mu|} = \prod_{|\mu|=|\kappa|} R_\mu(m),$$
$$P'_{|\lambda|}(m,p) = m^{|\kappa|}\sum_{|\lambda|=|\kappa|} c_{\kappa,\lambda} c'_\lambda \prod_{\substack{|\mu|=|\kappa|\\ \mu\ne\lambda}} R_\mu^{-1}(m) P_\lambda(m,p). \tag{A.22}$$

But $R_\mu^{-1}(m) = \prod_{i=1}^{q(\lambda)}\prod_{l=0}^{\lambda_i-1}(1 - \frac{1-i+2l}{m})$ is a polynomial in $1/m$, while $P_\lambda(m,p) = \prod_{i=1}^{q(\lambda)}\prod_{i=1}^{\lambda_i-1}(\frac{p}{m} + \frac{1-i+2l}{m})$ is a polynomial in p/m and $1/m$, both of degree at most

$|\mu| = |\lambda| = |\kappa|$. Thus,

$$\begin{aligned} P'_{|\kappa|}(m,p) &\equiv m^{|\kappa|} \sum_{|\lambda|=|\kappa|} c_{\kappa,\lambda} c'_\lambda \prod_{\substack{|\mu|=|\kappa| \\ \mu\neq\lambda}} R_\mu^{-1}(m) P_\lambda(m,p) \\ &= m^{|\kappa|} \sum_{i=0}^{|\kappa|} \sum_{j=0}^{|\kappa|} b_{ij} \left(\frac{p}{m}\right)^i \frac{1}{m^j} \end{aligned} \tag{A.23}$$

for some coefficients b_{ij} that do not depend on m, p (or n). Define the polynomials $f_j(\alpha) = \sum_{i=0}^{|\kappa|} b_{ij}\alpha^i$, so that

$$P'_{|\kappa|}(m,p) = m^{|\kappa|} \sum_{j=0}^{|\kappa|} f\left(\frac{p}{m}\right) m^{-j}. \tag{A.24}$$

Let us show that for all $0 \le j < |\kappa| - q(\kappa)$, the polynomial f_j must be identically zero over the interval $\alpha \in (0, 1/\max(|\kappa|-2, 0))$. Indeed, say this was not the case, and let $0 \le j_0 < |\kappa| - q(\kappa)$ be the smallest j with the property that $f_{j_0}(\alpha_0) \neq 0$ for some $\alpha_0 \in (0, \frac{1}{\max(|\kappa|-2,0)})$. As f_{j_0} is a polynomial, by continuity it must be nonzero in a neighborhood of α_0, so we may as well assume α_0 is rational without loss of generality. Now look at what happens to $\mathrm{E}[r_\kappa(Y^{-1})]$ as p grows to infinity at the very specific linear rate $p = \lfloor \frac{\alpha_0}{1+\alpha_0}(n-1) \rfloor$. Since α_0 is rational, there must be a subsequence n_l such that p_l is exactly an integer (e.g., if $\alpha_0 = a/b$ with a, b integers, we can take $n_l = (a+b)l + 1$). Then for $p_l = \frac{\alpha_0}{1+\alpha_0}(n_l - 1)$, we have exactly $p_l = \alpha_0 m_l$.

Since $\alpha_0 < \frac{1}{\max(|\kappa|-2,0)}$, then $|\kappa| < 1 + (\frac{\alpha_0}{1+\alpha_0})^{-1}$. Thus by Hölder's inequality and Lemma 5,

$$\lim_{l\to\infty} \frac{1}{m_l^{|\kappa|-q(\kappa)-j_0}} \cdot \frac{1}{p_l^{q(\kappa)}} \mathrm{E}\bigl[r_\kappa(Y^{-1})\bigr] \le 0 \cdot \lim_{l\to\infty} \frac{1}{p_l} \mathrm{E}\bigl[\operatorname{tr} Y^{-|\kappa|}\bigr] = 0.$$

On the other hand, by equations (A.21) and (A.24), the definition of j_0 and the fact that $R_{|\kappa|}(m) \to 1$ as $m \to \infty$,

$$\begin{aligned} &\lim_{l\to\infty} \frac{1}{m_l^{|\kappa|-q(\kappa)-j_0}} \cdot \frac{1}{p_l^{q(\kappa)}} \mathrm{E}\bigl[r_\kappa(Y^{-1})\bigr] \\ &\quad= \lim_{l\to\infty} \left(\frac{n_l}{m_l}\right)^{|\kappa|} R'_{|\kappa|}(m_l) \left(\frac{m_l}{p_l}\right)^{q(\kappa)} \sum_{j=j_0}^{|\kappa|} f_j(\alpha_0) m_l^{j_0-j} \\ &\quad= (1+\alpha_0)^{|\kappa|} \alpha_0^{-q(\kappa)} f_{j_0}(\alpha_0). \end{aligned}$$

As $\alpha_0 > 0$, $f_{j_0}(\alpha_0)$ must therefore equal zero, a contradiction. Hence, as claimed, the polynomials $f_j(\alpha)$ for $0 \le j < |\kappa| - q(\kappa)$ all vanish over the interval $(0, \frac{1}{\max(|\kappa|-2,0)})$.

But a polynomial can have an infinite number of zeros only if all its coefficients are zero, so we conclude that

$$b_{ij} = 0 \qquad \text{for} \quad 0 \le j < |\kappa| - q(\kappa).$$

Thus, from equations (A.21) and (A.23) we have

$$\mathrm{E}\big[r_\kappa(Y^{-1})\big] = \left(\frac{n}{m}\right)^{|\kappa|} m^{q(\kappa)} R'_{|\kappa|}(m) P''_\kappa(m,p),$$

where

$$P''_\kappa(m,p) = \sum_{i=0}^{|\kappa|} \sum_{j=|\kappa|-q(\kappa)}^{|\kappa|} b_{ij} \left(\frac{p}{m}\right)^i \frac{1}{m^{j-|\kappa|+q(\kappa)}}.$$

Plugging the above in Lemma 2 yields for $n \ge p + 8k + 6$ that

$$\mathrm{E}\big[\operatorname{tr} T^{2k}\big] = \frac{m^{2k+1}}{n^k} R'_{2k}(m) Q_\kappa(m,p), \tag{A.25}$$

where

$$Q_\kappa(m,p) = (-1)^k \sum_{|\kappa| \le 2k} \left(1 + \frac{p}{m} + \frac{1}{m}\right)^{|\kappa|} \frac{R'_{|\kappa|}(m)}{R'_{2k}(m)} \frac{b_\kappa(n,m,p)}{m^{2k+1-q(\kappa)}} P''_\kappa(m,p).$$

Now, for any $a \le b$, we can associate a partition μ of norm $|\mu| = a$ with the partition $\mu^* = (\mu_1 + b - a, \mu_2, \dots, \mu_{q(\mu)})$ of norm $|\mu^*| = b$, which satisfies

$$\prod_{i=1}^{q(\mu^*)} \prod_{j=0}^{\mu_i^*-1} \left(1 - \frac{1-i+2j}{m}\right) = \prod_{i=1}^{q(\mu)} \prod_{j=0}^{\mu_i-1} \left(1 - \frac{1-i+2j}{m}\right) \prod_{\substack{j=\mu_1\\ -a-1}}^{\mu_1+b} \left(1 - \frac{2j}{m}\right).$$

By definition for the $R_\mu(m)$'s at equation (A.20), this means that every factor that appears in $R_\mu^{-1}(m)$ appears in $R_{\mu^*}^{-1}(m)$, so by definition of the $R_{|\mu|}(m)$'s at equation (A.22), $R_a(m) R^{-1}(m)$ is a polynomial in $\frac{1}{m}$. Moreover, as the b_κ are polynomials of degrees $d(\kappa) \equiv 2k + 1 - q(\kappa)$, there exists coefficients c_{ijl} such that

$$\begin{aligned} \frac{b_\kappa(n,m,p)}{m^{2k+1-q(\kappa)}} &= \frac{1}{m^{d(\kappa)}} \sum_{i=0}^{d(\kappa)} \sum_{j=0}^{d(\kappa)-i} \sum_{l=0}^{d(\kappa)-i-j} c_{ijl} m^i n^j p^l \\ &= \sum_{i=0}^{d(\kappa)} \sum_{j=0}^{d(\kappa)-i} \sum_{l=0}^{d(\kappa)-i-j} \frac{c_{ijl}}{m^{d(\kappa)-i-j-l}} \left(1 + \frac{p}{m} + \frac{1}{m}\right)^j \left(\frac{p}{m}\right)^l. \end{aligned}$$

As $d(-i-j-l \ge 0)$, $j, l \ge 0$, we conclude that this expression is a polynomial in $\frac{p}{m}$ and $\frac{1}{m}$. Therefore, looking back at (A.25), we conclude that the $Q_\kappa(m,p)$'s are polynomials in $\frac{p}{m}$ and $\frac{1}{m}$. In other words, when $n \geq p + 8k + 6$ there must be polynomials $g_i(p) = \sum_{j=0}^{i} a_{ij} p^j$ and a large enough integer D such that

$$\text{E}[\operatorname{tr} T^{2k}] = \left(\frac{m}{n}\right)^k R_{2k}(m) \sum_{i=0}^{D} \frac{g_i(p)}{m^{i-k-1}}. \tag{A.26}$$

We will now proceed to show that the g_i must vanish on $\mathbb{N}$ for $0 \leq i_0 < k + 1$. Observe first that $\text{E}[\operatorname{tr} T^{2k}]$ must have a finite limit as $n \to \infty$ with p held fixed. Indeed, since $16T^2/n$ is positive definite, $|I_p + 16T^2/n|^{-(n-p-1)/4} < |I_p + 16T^2/n|^{-n/4}$ and so we have the bound

$$\text{E}[\operatorname{tr} T^{2k}] \leq C_t \int_{\mathbb{S}_p(\mathbb{R})} \operatorname{tr} T^{2k} \left| I_p + \frac{16T^2}{n} \right|^{-\frac{n}{4}} dT.$$

But for p fixed, $\lim_{n\to\infty} C_t = C_{\text{GOE}/4}$, the normalization constant of the $\text{GOE}(p)/4$ distribution, by Lemma 1. Moreover, $|I_p + 16T^2/n|^{-\frac{n}{4}} = \prod_{i=1}^{p} [1 + \lambda_i(4T^2)/(n/4)]^{-\frac{n}{4}}$ for $\lambda_1(4T^2) \geq \cdots \geq \lambda_p(4T^2) \geq 0$ the eigenvalues of $4T^2$, and $(1 + x/n)^{-n}$ is monotone decreasing toward $\exp(x)$. Therefore, for a fixed dimension p, we can apply the monotone convergence theorem to obtain that

$$\lim_{n\to\infty} \text{E}[\operatorname{tr} T^{2k}] \leq C_{\text{GOE}/4} \int_{\mathbb{S}_p(\mathbb{R})} \operatorname{tr} T^{2k} e^{-4 \operatorname{tr} T^2} dT = \text{E}[\operatorname{tr} Z^{2k}] < \infty \tag{A.27}$$

for a $Z \sim \text{GOE}(p)/4$.

We can use this to show that g_i must vanish on $\mathbb{N}$ for $0 \leq i_0 < k + 1$ as follows. Say the first statement was not true, and let $0 \leq i_0 < k + 1$ be the smallest i such that for some $p_0 \in \mathbb{N}$, $g_{i_0}(p_0) \neq 0$. Then by equation (A.26) and the definition of i_0, the limit of $\text{E}[\operatorname{tr} T^{2k}]$ as $n \to \infty$ with p fixed at p_0 satisfies

$$\lim_{n\to\infty} \frac{\text{E}[\operatorname{tr} T^{2k}]}{m^{k+1-i_0}} = 1^k \cdot 1 \cdot \lim_{n\to\infty} \sum_{i=i_0}^{D} g_i(p_0) m^{i_0 - i} = g_{i_0}(p_0).$$

But $m = n - p - 1$ tends to infinity as n tends to infinity, and since $k + 1 - i_0 > 0$, equation (A.27) means that $\text{E}[\operatorname{tr} T^{2k}]/m^{k+1-i_0}$ must tend to zero. Thus, $g_{i_0}(p_0)$ has to equal zero, which contradicts our assumption. Thus, for every $0 \leq i < k+1$, the polynomial g_i must vanish on $\mathbb{N}$. But a polynomial can only have an infinite number of zeroes if its coefficients are all zero, so we must have $a_{ij} = 0$ for $0 \leq i < k + 1$.

Now say that $p/n \to 0$ as $n \to \infty$. Then eventually $n \geq p + 8k + 6$, so by equation (A.26) and the above,

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{p^{k+1}} \text{E}[\operatorname{tr} T^{2k}] \\ &\quad = \lim_{n\to\infty} \left(\frac{m}{n}\right)^k R_{2k}(m) \sum_{i=k+1}^{D_1} \sum_{j=0}^{i} a_{ij} \frac{p^{j-(k+1)}}{m^{i-(k+1)}} \end{aligned}$$

$$= 1^k \cdot 1 \cdot \lim_{n\to\infty}\Bigg[\sum_{i=k+1}^{D}\sum_{j=0}^{k}\frac{a_{ij}}{m^{i-(k+1)}p^{(k+1)-j}} + \sum_{i=k+1}^{D}\sum_{j=k+1}^{i}\left(\frac{p}{m}\right)^{j-(k+1)}\frac{a_{ij}}{m^{i-j}}\Bigg]$$

$$= \sum_{j=0}^{k}\frac{a_{k+1}}{(\lim_{p\to\infty} p)^{k+1-j}} + a_{(k+1)(k+1)} < \infty,$$

as desired. □

PROOF OF COROLLARY 1. Let f_t be the density of the Sym-$t_{n/2}(I_p/8)$ distribution, and define for $L \in \mathbb{N}$,

$$h_L(T) = \frac{n+p+1}{n}\operatorname{tr}\big[(I_p + 16T^2/n)^{-1}(4T)^{2L+1}\big] - \Bigg[2L\operatorname{tr}(4T)^{2L-1} + \sum_{l=0}^{2L-1}\operatorname{tr}(4T)^l\operatorname{tr}(4T)^{2L-1-l}\Bigg]\mathbb{1}[L>0].$$

Then

$$-\frac{1}{2}h_L^2(T)f_t(T)$$
$$= h_L(T)\big(\operatorname{tr}\big[(4T)^{2k}\tilde{\nabla}_{4T}f_t(T)\big] + \operatorname{tr}\big[\tilde{\nabla}_{4T}(4T)^{2k}\big]f_t(T)\big) \tag{A.28}$$
$$= \operatorname{tr}\tilde{\nabla}_{4T}\big[(4T)^{2k}h_L(T)f_t(T)\big] - \operatorname{tr}\big[(4T)^{2k}\tilde{\nabla}_{4T}h_L(T)\big]f_t(T)$$

But when $n + p > 16L + 3$, for any $1 \le \alpha \le \beta \le p$,

$$\int_{\mathbb{S}_p(\mathbb{R})}(\tilde{\nabla}_{4T})_{\alpha\beta}\big[(4T)^{2L}h_L(T)f_t(T)\big]_{\beta\alpha}\,dT$$
$$= \frac{1+\delta_{\alpha\beta}}{8}\int_{\mathbb{R}^{\frac{p(p+1)}{2}-1}}\big[\big([4T]^{2L}\big)_{\beta\alpha}h_L(T)f_t(T)\big]_{4T_{\alpha\beta}=-\infty}^{4T_{\alpha\beta}=\infty}\prod_{\substack{i\le j\\ \ne(\alpha,\beta)}}dT_{ij} = 0$$

since $T^{2L}h_L(T)f_t(T)$ is integrable for $\frac{n+p+1}{4} > 4L+1$. Using the notation $\|\cdot\|_T$ for the L^2-norm under $T \sim$ Sym-$t_{n/2}(I_p/8)$, we find from equation (A.28) that

$$\frac{1}{p^{2k+1}}\|h_L(T)\|_T^2$$
$$= 0 + \frac{2}{p^{2k+1}}\operatorname{tr}\mathrm{E}\big[(4T)^{2k}\tilde{\nabla}_{4T}h_L(T)\big]$$
$$= \mathrm{E}\Bigg[2(2L+1)\frac{n+p+1}{n}\frac{p}{n}\frac{1}{p}\operatorname{tr}\Bigg[\left(I_p + \frac{16T^2}{n}\right)^{-1}\left(\frac{4T}{\sqrt{p}}\right)^{4L+2}\Bigg] \tag{A.29}$$

$$+ 2(2L-1)\frac{n+p+1}{n}\frac{1}{p}\operatorname{tr}\left[\left(I_p + \frac{16T^2}{n}\right)^{-1}\left(\frac{4T}{\sqrt{p}}\right)^{4L}\right]$$
$$- 4L(2L-1)\frac{1}{p^2}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{4L-2}$$
$$- \sum_{l=0}^{2L-1}\frac{1}{p}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2L-1-l}\frac{1}{p}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2L-1+l}\Bigg] = O(1)$$

as $p, n \to \infty$ with $p/n \to 0$, using that $x^{2l}/(1+16x^2/n) \le x^{2l} \forall l \in \mathbb{N}$ and Theorem 2.

To show the corollary, we proceed by induction on L and assume that $\|\operatorname{tr}(4T/\sqrt{p})^{2l+1}\|_T = O(1)$ for every $0 \le l < L$. If we take equation (A.29) with $L = k+1$ and multiply it by p/n, we find by the triangle inequality that

$$\left\|\frac{n+p+1}{n}\frac{p}{n}\operatorname{tr}\left[\left(I_p+\frac{16T^2}{n}\right)^{-1}\left(\frac{4T}{\sqrt{p}}\right)^{2k+3}\right] - 2\frac{p+k}{n}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1}\right\|_T$$
$$\le \frac{p}{n}\Bigg\|\sum_{\substack{l=1\\ \text{odd}}}^{2k}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{l}\frac{1}{p}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1-l} \tag{A.30}$$
$$+ \sum_{\substack{l=1\\ \text{even}}}^{2k}\frac{1}{p}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{l}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1-l}\Bigg\|_T + O\left(\frac{p}{n}\right) = o(1)$$

by Theorem 2 and the inductive hypotheses. But since

$$\left(I_p+\frac{16T^2}{n}\right)^{-1}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1} = \left(\frac{4T}{\sqrt{p}}\right)^{2k+1} - \frac{p}{n}\left(I_p+\frac{16T^2}{n}\right)^{-1}\left(\frac{4T}{\sqrt{p}}\right)^{2k+3},$$

we can conclude that

$$\left\|\frac{n+p+1}{n}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1} - 2\frac{p+k}{n}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1}\right\|_T$$
$$\le \left\|\frac{n+p+1}{n}\frac{p}{n}\operatorname{tr}\left[\left(I_p+\frac{16T^2}{n}\right)^{-1}\left(\frac{4T}{\sqrt{p}}\right)^{2k+3}\right] - 2\frac{p+k}{n}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1}\right\|_T$$
$$+ \Bigg\|\sum_{\substack{l=0\\ \text{odd}}}^{2k-1}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{l}\frac{1}{p}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k-1-l}$$
$$+ \sum_{\substack{l=0\\ \text{even}}}^{2k-1}\frac{1}{p}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{l}\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k-1-l}\Bigg\|_T$$
$$+ O(1),$$

that is, from equation (A.30), Theorem 2 and the inductive hypotheses,

$$\left\|\operatorname{tr}\left(\frac{4T}{\sqrt{p}}\right)^{2k+1}\right\|_T \leq \left(\frac{n+p+1}{n} - 2\frac{p+k}{n}\right)^{-1}\bigl(o(1)+O(1)+O(1)\bigr) = O(1),$$

as desired. □

Acknowledgments. We thank Michael Nussbaum for helpful discussions. We also thank the anonymous referees for significantly improving the manuscript through their helpful comments.

REFERENCES

ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

BARTLETT, M. S. (1933). On the theory of statistical regression. *Proc. Roy. Soc. Edinburgh Sect. A* **53** 260–283.

BUBECK, S. and GANGULY, S. (2018). Entropic CLT and phase transition in high-dimensional Wishart matrices. *Int. Math. Res. Not. IMRN* **2** 588–606. MR3801440

BUBECK, S., DING, J., ELDAN, R. and RÁCZ, M. Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms* **49** 503–532. MR3545825

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1967). *Inequalities*, 2nd ed. Cambridge University Press.

HUA, L. K. (1963). *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*. Amer. Math. Soc., Providence, RI. MR0171936

JIANG, T. and LI, D. (2015). Approximation of rectangular beta-Laguerre ensembles and large deviations. *J. Theoret. Probab.* **28** 804–847. MR3413957

KNUTH, D. (1976). Big omicron and big omega and big theta. *SIGACT News* 18–24.

LEDOUX, M. (2009). A recursion formula for the moments of the Gaussian orthogonal ensemble. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 754–769. MR2548502

LETAC, G. and MASSAM, H. (2004). All invariant moments of the Wishart distribution. *Scand. J. Stat.* **31** 295–318. MR2066255

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **72** 507–536. MR0208649

MATSUMOTO, S. (2012). General moments of the inverse real Wishart distribution and orthogonal Weingarten functions. *J. Theoret. Probab.* **25** 798–822. MR2956213

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

RÁCZ, M. Z. and RICHEY, J. (2018). A smooth transition from Wishart to GOE. *J. Theoret. Probab.* 1–9.

STEIN, E. M. and WEISS, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. *Princeton Mathematical Series* **32**. Princeton Univ. Press, Princeton, NJ. MR0304972

WISHART, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.

DEPARTMENT OF MATHEMATICS
AND INDUSTRIAL ENGINEERING
ÉCOLE POLYTECHNIQUE DE MONTRÉAL
MONTRÉAL, QUÉBEC H3T 1J4
CANADA
E-MAIL: didier.chetelat@polymtl.ca

DEPARTMENT OF STATISTICAL SCIENCE
CORNELL UNIVERSITY
1190 COMSTOCK HALL
ITHACA, NEW YORK 14853-3801
USA
E-MAIL: mtw1@cornell.edu