THE MIDDLE-SCALE ASYMPTOTICS OF WISHART MATRICES

By Didier Chételat and Martin T. Wells

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 $F$-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 $W_p(n, I_p/n)$ distribution is the law of the symmetric matrix $X^tX/n$, 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 Marchenko 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

*Research was partially supported by NSF-DMS 1611893 and NIH grant U19 AI111143.

MSC 2010 subject classifications: Primary 60B20, 60B10; secondary 60E10
“middle-scale” regimes. Hence, one might ask: what is the asymptotic behavior of a Wishart matrix $W_p(n, I_p/n)$ in the middle-scale regimes?

To gain some intuition, it is instructive to look at the eigenvalues of a Wishart matrix $W_p(n, I_p/n)$. 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 $\text{GOE}(p)$, the distribution on $p \times p$ real symmetric matrices whose diagonal elements follow a $N(0, 2)$ distribution, while the off-diagonal elements are $N(0, 1)$, all independent. In the classical regime where $p$ is fixed, the central limit theorem implies that

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

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

$$d_{\text{TV}}(F_1, F_2) = d_{\text{TV}}(f_1, f_2) = \int |f_1(x) - f_2(x)| dx,$$

while their Kullback-Leibler divergence is given by

$$d_{\text{KL}}(f_1 \parallel f_2) = \int \log[f_1(x)/f_2(x)] f_1(x) dx.$$ With different approaches, Jiang and Li [2015] and Bubeck et al. [2016] independently established that

$$d_{\text{TV}}\left(\sqrt{n}[W_p(n, I_p/n) - I_p], \text{GOE}(p)\right) \to 0 \quad (1.1)$$

whenever $p^{3/2}/n \to 0$. Thus, when $p^{3/2}/n \to 0$, 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 $p^3/n \not\to 0$, results of Bubeck and Ganguly [2016] and Rácz and Richey [2018] show that

$$d_{TV}(\sqrt{n}[W_p(n, I_p/n) - I_p], \text{GOE}(p)) \not\to 0.$$  \hfill (1.2)

Thus a phase transition occurs when $p$ is of order $\sqrt[3]{n}$. In fact, Bubeck and Ganguly [2016] show that the statistic $\text{tr}(X/\sqrt{p})^3$ 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 $\Theta(p^3/n)$ 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 $p^3/n \not\to 0$.

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 $p^3/n \not\to 0$, 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

$$d_{KL}(\text{GOE}(p)^* \parallel \sqrt{n}[W_p(n, I_p/n) - I_p]^*)$$
\[ \begin{align*}
&= \sum_{k=1}^{K} a_k \frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} b_{k, \text{even}} \sqrt{\frac{p^{k+2}}{n^k}} + O\left(\frac{p^{K+3}}{n^{K+1}}\right) \\
as p, n \to \infty.
\end{align*} \]

But since \( F \)-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 \( p^{K+3}/n^{K+1} \to 0 \) for every \( K \in \mathbb{N} \), a change that is then reflected in its \( F \)-conjugate. Thus, we are led to conclude that \( p^3/n \to 0 \) cannot be the only phase 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 \( \lim_{n \to \infty} \log p/\log n \in \left[ \frac{K}{K+2}, \frac{K+1}{K+3} \right] \). A diagram is provided at 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 \( f_K(X) \)

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

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

$$d_{TV}\left(\sqrt{n}[W_p(n, I_p/n) - I_p], F_K\right) \to 0$$

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

In particular, since $E[\exp\{i\text{tr}(XZ)/\sqrt{8}\}]^2 = \exp\{-\text{tr} Z^2/2\}$, $f_0$ 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 et al. [2016], Bubeck and Ganguly [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 $\log p/\log n \in \left[\frac{23}{23+2}, \frac{24+1}{23+3}\right]$, an estimator exploiting the $23$th-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 $24$th-degree middle-scale asymptotics should also be reasonable. In general, a large family of nonequivalent asymptotics are 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 \( \lim_{n \to \infty} \frac{\log p}{\log n} \leq 1 \), those such that
\( \lim_{n \to \infty} \frac{\log p}{\log n} = 1 \) 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 Appendix A.

2. Notation and definitions. The transpose of a matrix is denoted \( \mathsf{T} \),
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
\( X \), \( \mathsf{tr} X^k \) means the trace of \( X \). We will write \( \mathsf{tr}^k X \) when we mean the \( k \)-th
power of the trace of \( X \). The Kronecker delta is the symbol \( \delta_{kl} = 1 \) \( \{ k = l \} \).

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

The space of symmetric matrices \( \mathcal{S}_p(\mathbb{R}) \) can be assimilated to \( \mathbb{R}^{p(p+1)/2} \) by
mapping a symmetric matrix to its upper triangle. By integration over \( \mathcal{S}_p(\mathbb{R}) \),
we mean integration with respect to the pullback Lebesgue measure under
this isomorphism, that is, \( \int_{\mathcal{S}_p(\mathbb{R})} f(X) \, dX = \int_{\mathbb{R}^{p(p+1)/2}} f(X) \prod_{i \leq j} dX_{ij} \). 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 Appendix A.

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 \( \mathcal{S}_p(\mathbb{R}) \) with density \( f \), and let \( \psi \in L^2(\mathcal{S}_p(\mathbb{R})) \). Then the*
L²-distance between \( \mathcal{F}\{f^{1/2}\} \) and \( \psi \) satisfies

\[
d_{L^2}(\mathcal{F}\{f^{1/2}\}, \psi) \leq \left[ \left\| \psi \right\|_{L^2}^2 - 1 \right] + E \left[ \Re \log \frac{\mathcal{F}\{f^{1/2}\}^2(T)}{\psi^2(T)} \right] + 2 \left\| \psi \right\|_{L^2} E \left[ \left| \Re \log \frac{\mathcal{F}\{f^{1/2}\}^2(T)}{\psi^2(T)} \right| \right]^{1/2}
\]

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 non-integer degrees of freedom yields the following.

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

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

where

\[
C_t = \frac{2^{p(n+2p)}}{\pi \frac{p(p+1)}{2} \frac{n}{p+1}} \frac{\Gamma_p^2 \left( \frac{n+p+1}{4} \right)}{\Gamma_p \left( \frac{n}{2} \right)}.
\] (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

\[
f_{NW^*}(T) = \frac{2^{2(n+2p)}}{\pi \frac{p(p+1)}{2} \frac{n}{p+1}} \frac{\Gamma_p^2 \left( \frac{n+p+1}{4} \right)}{\Gamma_p \left( \frac{n}{2} \right)} \left| I_p + \frac{16T^2}{n} \right|^{-\frac{n+p+1}{4}}.
\] (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 \( \Omega \), denoted \( \text{Sym-}t_{\nu}(\Omega) \), if it has density

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

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

as $n, p \to \infty$, where the symbol $C_{\text{GOE}/4}$ 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 $\kappa = (\kappa_1, \ldots, \kappa_q)$ in decreasing order $\kappa_1 \geq \cdots \geq \kappa_q > 0$, define its associated power sum polynomial to be

$$r_\kappa(Z) = \prod_{i=1}^{q} \text{tr} Z^{\kappa_i}.$$  \hspace{1cm} (3.3)

The norm of the partition $\kappa$ 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 $\emptyset = ()$ with length $q(\emptyset) = 0$, norm $|\emptyset| = 0$ and power sum polynomial $r_\emptyset(Z) = 1$. 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 $T \sim \text{Sym-}t_{n/2}(I_p/8)$. 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

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

for a $Y^{-1} \sim W_p^{-1}(n, I_p/n)$ and some polynomials $b_\kappa$ in $n, m, p$, indexed by integer partitions $\kappa$, whose degrees satisfy $\deg b_\kappa \leq 2k + 1 - q(\kappa)$. The sums are taken over all partitions of the integers $\kappa$ 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 $\kappa$, there exist coefficients $c_{\kappa,\lambda}$ (which depend solely on $\kappa$ and $\lambda$) such that

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

for $C_\lambda$ 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 p.237], the expected zonal polynomials for $Y^{-1} \sim W_p^{-1}(n, I_p/n)$ are

$$E[C_\lambda(Y^{-1})] = \frac{n^{|\lambda|}}{2^{|\lambda|} \prod_{i=1} q(\lambda) m^{-i+1}} C_\lambda(I_p)$$

$$= \frac{2^{|\lambda|} |\lambda|! \prod_{i<j} (2\lambda_i - 2\lambda_j - i + j) \prod_{i=1} q(\lambda) m^{-i+1}}{\prod_{i=0} q(\lambda) (2\lambda_i + q(\lambda) - i)!} n^{\lambda_1} \prod_{i=0} p + (1 - i + 2l) m - (1 - i + 2l) \quad (3.5)$$

for $\lambda \neq \emptyset$, and $E[C_\emptyset(Y^{-1})] = 1$. From this, we can exactly compute $E[r_n(Y^{-1})]$, and thus $E[tr T^{2k}]$, 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 $T \sim Sym_{n/2}(I_p/8)$. The $2k^{th}$ moment of $T$ satisfies the asymptotics $E[p^{-1} tr(T/\sqrt{p})^{2k}] = O(1)$ 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 $T \sim Sym_{n/2}(I_p/8)$. The $2k+1^{th}$ moment of $T$ satisfies the asymptotics $E[tr(T/\sqrt{p})^{2k+1}] = O(1)$ 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 $S_p(\mathbb{R})$, the Fourier transform with kernel $\exp\{-i \text{tr}(XT)\}$, normalized to be an $L^2$-isometry, satisfies $\mathcal{F}\{f\}(T) = 2^{-\frac{n}{2}} \pi^{-\frac{p(p+1)}{4}} \int_{S_p(\mathbb{R})} e^{-i \text{tr}(XT)} f(X) dX$. Thus, the $\mathcal{F}$-conjugate density of the Gaussian orthogonal ensemble is given by

$$f_{\text{GOE}^*}(T) = \left| \mathcal{F}\left\{ \exp\left\{ -\frac{1}{8} \sum_{i,j=1}^p X_{ij}^2 \right\} \right\} \right|^2(T)$$

$$= \frac{1}{2^{p(p+7)/4} \pi^{3p(p+1)/4}} \int_{S_p(\mathbb{R})} e^{-i \text{tr}(XT)} \exp\left\{ -\frac{1}{8} \sum_{i,j=1}^p X_{ij}^2 \right\} dX^2$$

$$= \frac{2^{p(3p+1)/4}}{\pi^{p(p+1)/4}} \exp\left\{ -4 \sum_{i,j=1}^p T_{ij}^2 \right\},$$

where we used the Fourier transform of a standard normal law [Stein and Weiss, 2016, Theorem 1.13]. But this is the density of the $\text{GOE}(p)/4$ distribution, so we conclude that the $\mathcal{F}$-conjugate of the Gaussian orthogonal ensemble must be $\text{GOE}(p)^* = \text{GOE}(p)/4$. Then, by definition of the Kullback-Leibler divergence,

$$d_{\text{KL}}\left( \text{GOE}(p)^* \parallel \sqrt{n}[W_p(n, I_p/n) - I_p]^* \right)$$

$$= \mathbb{E}\left[ \log C_{\text{GOE}/4} - 4 \text{tr} T^2 - \log C_t + \frac{n+p+1}{4} \log \frac{16T^2}{n} \right]$$

(4.1)

for a $T \sim \text{GOE}(p)/4$, where $C_{\text{GOE}/4}$ and $C_t$ are the normalization constants of the $\text{GOE}(p)/4$ and the $\text{Sym-}t_{n/2}(I_p/8)$ distributions, respectively.

A change of variables $X = 4T$ and Lemma 1 yields that for some constants $c_k^{(1)}, c_k^{(2)}$ and $X \sim \text{GOE}(p)$, the above equals

$$= \mathbb{E}\left[ -\frac{1}{4} \text{tr} X^2 + \frac{n+p+1}{4} \log \frac{16X^2}{n} \right].$$
Consider the expectation. For any integer $L$ and any real $x$, we have the inequality $\left| -\frac{1}{2} \log(1 + x^2) - \sum_{l=1}^{L}\left(-1\right)^l x^{2l}/2l \right| \leq x^{2L+2}/(2L + 2)$. Thus we have the bound

$$\left| E \left[ -\frac{n+p+1}{4} \log \left| I_p + \frac{X^2}{n} \right| \right] - E \left[ \frac{n+p+1}{2} \sum_{k=1}^{K+1} (-1)^k \frac{\text{tr} X^{2k}}{2k n^k} \right] \right| \leq \frac{n+p+1}{2} E \left[ \frac{\text{tr} X^{2(K+2)}}{2(K+2) n^{K+2}} \right] = O\left( \frac{p^{K+3}}{n^{K+1}} \right)$$

as $n, p \to \infty$, since $E[\text{tr} X^{2(K+2)}] = O(p^{K+3})$ as $p \to \infty$ by Anderson et al. [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 $E[\text{tr} X^{2k}] = \sum_{k,l=0}^{k+1} c_{k,l}^{(3)} p^l$ for constants $c_{k,l}^{(3)}$. Thus we really have

$$E \left[ \frac{n+p+1}{4} \log \left| I_p + \frac{X^2}{n} \right| \right] = \frac{n+p+1}{n} \left[ \frac{\text{tr} X^2}{4} + \sum_{k=2}^{K+1} \sum_{l=0}^{k+1} c_{k,l}^{(4)} \frac{p^l}{n^{k-1}} \right] + O\left( \frac{p^{K+3}}{n^{K+1}} \right)$$

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

$$\text{d}_{\text{KL}}\left( \text{GOE}(p)^* \parallel \sqrt{n}[W_p(n, I_p/n) - I_p]^* \right)$$

$$= \sum_{k=1}^{K} a_k \frac{p^{k+2}}{n^k} + \sum_{k=1}^{K} c_{k}^{(7)} \frac{p^{k+1}}{n^k} + O\left( \frac{p^{K+3}}{n^{K+1}} \right)$$

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

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

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

(4.3)
where $C_t$ is the normalization constant of the Sym-$t_{n/2}(I_p/8)$ distribution and $K_1 = K + 1 + 1 [K \text{ odd}], K_2 = K + 1 [K > 0, \text{ even}].$ Let $f_{NW}, f_t$ and $f_{GOE/A}$ denote the densities of the normalized Wishart, Sym-$t_{n/2}(I_p/8)$ and GOE($p$)/4 distributions, respectively.

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

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

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 $L \in \mathbb{N}$, $-\frac{1}{2} \log(1 + x^2) - \sum_{i=1}^{L} (-1)^{i} x^{2i}/(2i) = \frac{1}{2} \int_{0}^{x} [t/(1 + t)]^{L+1} dt > 0$. Thus $|\psi_K^2(T)| \leq f_t(T)$ for all $K > 0$, which is integrable whenever $n \geq p - 2$. In particular,

$$\lim_{n \to \infty} \|\psi_K\|_{L^2} \leq f_t \|_{L^1} = 1 \quad \text{for all } K > 0 \quad (4.5)$$

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

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

$$\left| \frac{1}{2} \log(1 + x^2) - \sum_{i=1}^{L} (-1)^{i} x^{2i}/(2i) \right| \leq \frac{x^{2L+2}}{2L + 2}, \quad (4.6)$$

and

$$\left| \frac{\text{atan}(x)}{\text{atan}(1)} - \sum_{i=1}^{L} (-1)^{i-1} x^{2i-1}/(2i - 1) \right| \leq \frac{x^{2L+1}}{2L + 1}, \quad (4.7)$$

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 Log $F\{f_{NW}^{1/2}\}/\psi_K^2$ satisfies

$$\mathbb{E} \left[ |\Re \text{Log} \frac{\mathcal{F}\{f_{NW}^{1/2}\}}{\psi_K^2} | \right] \leq \frac{n}{2} \mathbb{E} \left[ \left| \frac{1}{2} \log \frac{16T^2}{n} - \sum_{k=1}^{K_1} \frac{(-1)^k}{2k} \text{tr} \left( \frac{4T}{\sqrt{n}} \right)^{2k} \right| \right] + \frac{p + 1}{2} \mathbb{E} \left[ \left| \frac{1}{2} \log \frac{16T^2}{n} - \sum_{k=1}^{K_2} \frac{(-1)^k}{2k} \text{tr} \left( \frac{4T}{\sqrt{n}} \right)^{2k} \right| \right]$$

$$\leq \frac{n}{4} \mathbb{E} \left[ \text{tr} \left( \frac{4T}{\sqrt{n}} \right)^{2K_1+2} \right] + \frac{p + 1}{4} \mathbb{E} \left[ \text{tr} \left( \frac{4T}{\sqrt{n}} \right)^{2K_2+2} \right]$$

$$= O \left( \frac{p^{K+3} [K \text{ odd}]}{n^{K+1} [K > 0, \text{ even}]} \right) + O \left( \frac{p^{K+3} [K > 0, \text{ even}]}{n^{K+1}} \right) = O \left( \frac{p^{K+3}}{n^{K+1}} \right) \quad (4.8)$$

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

For the imaginary part, notice that the projection $P_{(-\pi,\pi]} x = x - 2\pi \left[ \frac{x}{2\pi} - \frac{1}{2} \right]$ satisfies $\Im \text{Log} z = P_{(-\pi,\pi]} \Im \log z$ for all branches of log $z$, as well as the
inequality $|P_{(-\pi,\pi)}x| \leq |x|$. Using this mapping, as well as Proposition 3, Equation (4.7) and Corollary 1, we find for $T \sim \text{Sym}_{t/2}(I_p/8)$ that

$$
\mathbb{E}\left[3 \log \frac{F(f^{1/2}_{\omega})}{\psi^2_K}\right] = \mathbb{E}\left[P_{(-\pi,\pi)}\left[-\frac{n+p+1}{2} \text{tr} \left(\frac{4T}{\sqrt{n}}\right) + 2\sqrt{n} \text{tr} T + \frac{n}{2} \sum_{k=1}^{K^{-1}} (-1)^k \frac{4T}{2k+1}^{2k+1} + \frac{p+1}{2} \sum_{k=1}^{K^{-1}} (-1)^k \frac{4T}{2k+1}^{2k+1}\right]\right]
$$

$$
\leq \frac{n}{2} \mathbb{E}\left[\text{tr} \left(\frac{4T}{\sqrt{n}}\right) - \frac{2\sqrt{n} \text{tr} T}{n/2} - \sum_{k=1}^{K^{-1}} (-1)^{k-1} \frac{\text{tr} (4T/\sqrt{n})^{2k-1}}{2k-1}\right] + \frac{p+1}{2} \mathbb{E}\left[\text{tr} (4T/\sqrt{n})^{2K_{k+1}}\right]
$$

$$
= O\left(\sqrt{\frac{p^{2K+3+2l[K \text{ odd}]}_{n2K+1+2l[K \text{ odd}]}}{n2K+1+2l[K \text{ odd}]}}\right) + O\left(\sqrt{\frac{p^{2K+3+2l[K \geq 0, \text{even}]}_{n2K+1+2l[K \geq 0, \text{even}]}}{n2K+1+2l[K \geq 0, \text{even}]}}\right) = O\left(\sqrt{\frac{p^{K+3}}{nK+1}}\right) \quad (4.9)
$$

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

By Equation (3.2), the $\text{Sym}_{t/2}(I_p/8)$ distribution is the $F$-conjugate of the normalized Wishart distribution. Therefore, Proposition 2 with Equations (4.8)–(4.9) and (4.4)–(4.5) implies that $\lim_{n \to \infty} d_{L^2}^2(F(f^{1/2}_{\omega}), \psi_K) = 0$ when $p^{K+3}/nK+1 \to 0$ as $p, n \to \infty$.

We finish the proof by relating the $\psi_K$ to the $f_K$ of Definition 2. First note that for $K = 0$, $|\psi_0| = C_1C^{-1}_{\text{GOE}/4} f^{1/2}_{\text{GOE}(p)/4}$ is proportional to the density of a GOE$(p)/\sqrt{8}$ density, which is always integrable. In addition, $|\psi_K| \leq f^{1/2}_l$ for all $K > 0$, and $f^{1/2}_l$ is proportional to the density of a $\text{Sym}_{m/4}(\frac{n}{2m} I_p)$ distribution with $m = n - p - 1$, itself integrable whenever $n \geq 3p - 3$. Thus, in all cases $\psi_K$ is integrable whenever $n \geq 3p - 3$, and recall that we had already shown $\psi_K$ is square-integrable whenever $n \geq p - 2$. We can therefore use the Fourier inversion theorem to conclude that $|F^{-1}(\psi_K)|^2 \propto f_K$, and in particular that $f_K$ is integrable, whenever $n \geq 3p - 3 \geq p - 2$.

Define $g_K = F^{-1}(\psi_K)$ so that $f_K = \|g_K\|_{L^2}^2/\|g_K\|_{L^2}^2$. By the Plancherel theorem, $d_{L^2}(f^{1/2}_{\omega}, g_K) = d_{L^2}(F(f^{1/2}_{\omega}), \psi_K) \to 0$ and so in particular we must have $\lim_{n \to \infty} \|g_K\|_{L^2}^2 = \lim_{n \to \infty} \|\psi_K\|_{L^1} = 1$. From the triangle inequality, the reverse triangle inequality and the limits $d_{L^2}(F(f^{1/2}_{\omega}), \psi_K) \to 0$ and
\[ \|g_K\|_{L^2}^2 \to 1, \] we can conclude that
\[
\lim_{n \to \infty} d_{L^2}(f_{NW}^{1/2}, f_K^{1/2}) \leq \lim_{n \to \infty} d_{L^2}(f_{NW}^{1/2}, |g_K|^{1/2}) + \lim_{n \to \infty} d_{L^2}(|g_K|^{1/2}, f_K^{1/2}) \\
\leq \lim_{n \to \infty} d_{L^2}(f_{NW}^{1/2}, g_K) + \lim_{n \to \infty} \left|1 - \|g_K\|_{L^2}^2\right| \|g_K\|_{L^2} = 0
\]
when \(p^{K+3}/n^{K+1} \to 0\) as \(p, n \to \infty\). But by the Cauchy-Schwarz inequality,
\[
d_{TV}(f_{NW}, f_K) = \|(f_{NW}^{1/2} + f_K^{1/2})(f_{NW}^{1/2} - f_K^{1/2})\|_{L^1} \\
\leq \|f_{NW}^{1/2} + f_K^{1/2}\|_{L^2}\|f_{NW}^{1/2} - f_K^{1/2}\|_{L^2} \leq 2d_{L^2}(f_{NW}^{1/2}, f_K^{1/2}),
\]
so we conclude that the total variation distance \(d_{TV}(f_{NW}, f_K)\) must tend to zero when \(p, n \to \infty\) with \(p^{K+3}/n^{K+1} \to 0\). This concludes the proof. \( \square \)

5. Conclusion. In this paper we show that the behavior of a Wishart matrix when \(p, n \to 0\) with \(p/n \to 0\), which we call the middle-scale asymptotics, varies according to which interval \(\left(\frac{K}{K+2}, \frac{K+1}{K+3}\right)\) 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 \(F\)-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 \(F\)-conjugate of the normalized Wishart distribution to the \(F\)-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 A: PROOFS OF THE \(T\)-DISTRIBUTION RESULTS

In Section 3, results regarding \(F\)-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 \(\exp\{a \text{ tr } Z\}|Z|^b\) for a symmetric matrix \(Z\). Recall that
\( \nabla Z \) stands for the symmetric differentiation operator \( (\nabla Z)_{kl} = \frac{1+\delta_{kl}}{2} \partial Z_{kl} \), where \( \delta \) is the Kronecker delta, and that we write \( m = n - p - 1 \).

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

\[
\nabla_{Z_{i_1 j_1}} \cdots \nabla_{Z_{i_{2l} j_{2l}}} \nabla_{Z_{i_{2l+1} j_{2l+1}}} \exp \left\{ -\frac{n}{4} \text{tr} \ Z \right\} |Z|^\frac{m}{4} = \sum_{s=0}^{l} \sum_{J \in \{1, \ldots, p\}^{2l}} a_{J,s}(n,m) \prod_{t=s+1}^{l} (I_p)_{j_{2t}j_{2t-1}} \prod_{t=1}^{s} Z_{j_{2t}j_{2t-1}}^{-1} \exp \left\{ -\frac{n}{4} \text{tr} \ Z \right\} |Z|^\frac{m}{4}.
\]

**Proof.** To simplify notation, let

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

and let \( M_l = \{ M_{J,s} | J \in \{1, \ldots, p\}^{2l}, s \leq l \} \) 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

\[
\nabla_{Z_{i_1 j_1}} \cdots \nabla_{Z_{i_{2l} j_{2l}}} \nabla_{Z_{i_{2l+1} j_{2l+1}}} \exp \left\{ -\frac{n}{4} \text{tr} \ Z \right\} |Z|^\frac{m}{4} \in \langle M_l \rangle. \tag{A.1}
\]

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

\[
\nabla_{Z_{i_1 j_1}} \cdots \nabla_{Z_{i_{2l} j_{2l}}} Z_{ab}^{-1} = -\frac{1}{2} \left[ Z_{a_1}^{-1} Z_{i_1}^{-1} Z_{i_2 b}^{-1} + Z_{b_1}^{-1} Z_{i_1}^{-1} Z_{i_2 a}^{-1} \right]
\]

and

\[
\nabla_{Z_{i_1 j_1}} \exp \left\{ -\frac{n}{4} \text{tr} \ Z \right\} |Z|^\frac{m}{4} = \left[ \frac{m}{4} Z_{i_1 j_1}^{-1} - \frac{n}{4} (I_p)_{i_2 j_2} Z_{i_2 j_2}^{-1} \right] \exp \left\{ -\frac{n}{4} \text{tr} \ Z \right\} |Z|^\frac{m}{4},
\]

we conclude that

\[
\nabla_{Z_{i_1 j_1}} \cdots \nabla_{Z_{i_{2l} j_{2l}}} M_{J,s}(Z) = -\frac{1}{2} \sum_{r=1}^{s} M_{j_2 r, j_1^{2l-1}, s+1} - \frac{1}{2} \sum_{r=1}^{s} M_{j_2 r, j_1^{2l-1}, s+1} + \frac{m}{4} M_{j_2 s+1, i_2^{2l-1}, s+1} - \frac{n}{4} M_{j_2 s+1, i_2^{2l-1}, s} \in \langle M_l \rangle.
\]

Thus, by linearity, \( \nabla_{Z_{i_1 j_1}} Z_{ab} \) maps \( \langle M_{l-1} \rangle \) to \( \langle M_l \rangle \). But by definition of \( \langle M_0 \rangle \) we have \exp\{−n/4 tr Z\}Z^{m/4} ∈ ⟨M₀⟩, so by induction Equation (A.1) must then hold, as desired. \( \square \)
Lemma 4. For any \( k \in \mathbb{N} \) and any \( Z \in S_p(\mathbb{R}) \),
\[
\text{tr} \left( \nabla X^k \ e^{-\frac{Z}{2}} tr Z \frac{\partial}{\partial Z} \right) = e^{-\frac{Z}{2}} tr Z \sum_{|\kappa| \leq 2k} b_\kappa(n, m, p) r_\kappa(Z^{-1})
\]
for some polynomials \( b_\kappa(n, m, p) \) with \( \deg b_\kappa \leq 2k + 1 - q(\kappa) \) and \( r_\kappa \) as in Equation (3.3). The sums at the right hand sides are taken over all integer partitions \( \kappa \) of norm at most \( 2k \), including the empty partition.

Proof. We give a spectral proof. Let \( \lambda_1 \geq \cdots \geq \lambda_p \) be the eigenvalues of \( Z \) and \( OLO^t \) its corresponding spectral decomposition, with diagonal matrix \( L = \text{diag}(\lambda_1, \ldots, \lambda_p) \) and orthogonal matrix \( O \), and notice that
\[
\nabla Z_{ij} O_{hl} = \frac{1}{2} \sum_{a,b}^{p} O_{ha} O_{ai}^t \frac{\partial}{\partial \lambda_a} O_{jl} + \frac{1}{2} \sum_{a \neq l} O_{ha} O_{aj}^t \delta \lambda_a,
\]
for any \( 1 \leq i, j, h, l \leq p \). As a consequence, for any differentiable real-valued functions \( F_1(L), \ldots, F_p(L) \), we have
\[
\sum_{j=1}^{p} \nabla Z_{j} O_{F_{j}^{t}} = \frac{1}{2} \sum_{a,b}^{p} O_{ha} \frac{F_b - F_a}{\lambda_b - \lambda_a} O_{ai}^t + \sum_{a=1}^{p} O_{ha} \frac{\partial F_a}{\partial \lambda_a} O_{ai}^t.
\]
This suggests we define a new operator \( D_L \) that maps the space of diagonal matrices \( F(L) = \text{diag}(F_1(L), \ldots, F_p(L)) \) that differentiably depends on \( L \), to itself, by
\[
D_L(F)_a = \frac{1}{2} \sum_{b \neq a}^{p} \frac{F_b - F_a}{\lambda_b - \lambda_a} + \frac{\partial F_a}{\partial \lambda_a}, \quad \text{so that} \quad \sum_{j=1}^{p} \nabla Z_{j} O_{F_{j}^{t}} = O_{F_{j}^{t} I_p}.
\]
In particular, we would then have
\[
\text{tr} \left( \nabla X^k \ e^{-\frac{Z}{2}} tr Z \frac{\partial}{\partial Z} \right) = \text{tr} D_L^{2k} \left\{ e^{-\frac{Z}{2}} tr Z \frac{\partial}{\partial Z} I_p \right\}.
\]
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 \( L \),
\[
D_L(a F + b G) = a D_L(F) + b D_L(G).
\]
(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 \( D_L \),
\[
D_L \left( e^{-\frac{Z}{2}} \frac{\partial}{\partial Z} I_p \right) = -\frac{n}{4} e^{-\frac{Z}{2}} \frac{\partial}{\partial Z} I_p, \quad D_L \left\{ \frac{\partial}{\partial Z} I_p \right\} = -\frac{m}{4} |Z|^{\frac{m}{2}} I_p,
\]
\[
D_L \left( \text{tr} \left( L^{-s} \right) I_p \right) = -s L^{-(s+1)}.
\]
Now define the spaces
\[ M_l = \left\{ b(n, m, p)e^{-\frac{q}{2} \text{tr} L}|L|^{m} r_{\kappa}(L^{-1}) L^{-s} \mid b(n, m, p) \text{ is a polynomial with} \right\}
\]
\[ \text{degree at most } l - q(\kappa), \text{ and } \kappa \]
\[ \text{and } s \text{ satisfy } |\kappa| \leq l - s. \]
for \( l = 1, \ldots, 2k \), and let \( \langle M_l \rangle \) denote the linear span of \( M_l \), i.e. the space of all real linear combinations of elements of \( M_l \). Moreover, for a partition \( \kappa \), let \( \kappa \pm i \) denote \( \kappa \) 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 \( F \in M_l \),
\[ D_L\{F\} = D_L\left\{ b(n, m, p)e^{-\frac{q}{2} \text{tr} L}|L|^{m} r_{\kappa}(L^{-1}) L^{-s} \right\} \]
\[ = b(n, m, p)D_L\left\{ e^{-\frac{q}{2} \text{tr} L} I_p \right\}|L|^{m} r_{\kappa}(L^{-1}) L^{-s} \]
\[ + b(n, m, p)e^{-\frac{q}{2} \text{tr} L} D_L\left\{ |L|^{m} I_p \right\} r_{\kappa}(L^{-1}) L^{-s} \]
\[ + b(n, m, p)e^{-\frac{q}{2} \text{tr} L} |L|^{m} D_L\left\{ r_{\kappa}(L^{-1}) I_p \right\} L^{-s} \]
\[ + b(n, m, p)e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa}(L^{-1}) D_L\left\{ L^{-s} \right\} \]
\[ = \left[ -\frac{n}{4} b(n, m, p) \right] e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa}(L^{-1}) L^{-s} \]
\[ + \left[ \frac{m}{4} b(n, m, p) \right] e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa}(L^{-1}) L^{-s} \]
\[ + \sum_{i=1}^{q(\kappa)} \left[ -\kappa_i b(n, m, p) \right] e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa - \kappa_i}(L^{-1}) L^{-s} \]
\[ + \left[ -\frac{s}{2} b(n, m, p) \right] e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa}(L^{-1}) L^{-(s+1)} \]
\[ + \sum_{t=1}^{s} \left[ -\frac{1}{2} b(n, m, p) \right] e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa + (s+1-t)} L^{-t}. \]
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, \( e^{-\frac{q}{2} \text{tr} L} |L|^{m} I_p \in M_0 \), so by induction \( D_L^{2k}\left\{ e^{-\frac{q}{2} \text{tr} L} |L|^{m} I_p \right\} \in \langle M_{2k} \rangle \).
Hence, for some polynomials \( b_{\kappa, s}(n, m, p) \) of degree at most \( 2k - q(\kappa) \),
\[ \text{tr} D_L^{2k}\left\{ e^{-\frac{q}{2} \text{tr} L} |L|^{m} I_p \right\} = \sum_{|\kappa| + s \leq 2k} b_{\kappa, s}(n, m, p) e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa}(L^{-1}) \text{ tr}(L^{-s}) \]
\[ = \sum_{|\kappa'| \leq 2k} b_{\kappa'}(n, m, p) e^{-\frac{q}{2} \text{tr} L} |L|^{m} r_{\kappa'}(L^{-1}) \quad (A.3) \]
for \( \kappa' = \kappa + s \), \( b_{\kappa'} = b_{\kappa, s} \) when \( s \neq 0 \), while \( \kappa' = \kappa \), \( b_{\kappa'} = pb_{\kappa, s} \) when \( s = 0 \).
Notice that when \( s \neq 0 \), the degree of the \( b_{n'} \)'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, \( \deg b_{n'} \leq 2k - q(\kappa') + 1 \), which by the top line of Equation (A.2) shows the lemma. \( \square \)

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 \( n \gg p + 4s + 2 \), the \( s \)th inverse moment satisfies the recursive bound

\[
\left( 1 - \frac{(p + 1)s}{n} \right) \mathbb{E} \left[ \text{tr} Y^{-s} \right] \leq \mathbb{E} \left[ \text{tr} Y^{-(s-1)} \right].
\]

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

**Proof.** For any differentiable function \( f : \mathbb{R} \to \mathbb{R} \) such that \((\partial/\partial Z - Z)f(Z)\) is integrable when \( Z \sim N(0, 1) \), integration by parts yields that \( \mathbb{E}[(\partial/\partial Z - Z)f(Z)] = 0 \), a result also known as Stein’s lemma. Let \( Z \) be an \( n \times p \) matrix of independent standard normal random variables, and let \( Y = \frac{1}{n}Z^tZ \sim W_p(n, I_p/n) \). For any \( 1 \leq \alpha \leq n \) and \( 1 \leq \beta, i, j \leq p \), we have

\[
\nabla_{Y_{ij}} Y_{ij}^{-s} = -\frac{1}{n} \sum_{l=1}^{s} \left[ Y_{ij}^{-l} Y_{ij}^{-(s-l+1)} + Y_{ij}^{-l} Y_{ij}^{-(s-l+1)} \right],
\]

so for \( \delta \) the Kronecker delta,

\[
\left( \frac{\partial}{\partial \delta_{\alpha\beta}} - Z_{\alpha\beta} \right) (ZY^{-s})_{\alpha\beta} = \sum_{j=1}^{p} \delta_{j\beta} Y_{j\beta}^{-s} = \frac{1}{n} \sum_{i=1}^{p} Z_{\alpha i} \nabla_{Y_{ij}} Y_{ij}^{-s} - Z_{\alpha\beta} Y_{ij}^{-s}
\]

\[
= Y_{\beta\beta}^{-s} - \frac{1}{n} \sum_{l=1}^{s} (ZY^l)_{\alpha\beta} (ZY^{-(s-l+1)})_{\alpha\beta} - \frac{1}{n} \sum_{l=1}^{s} (ZY^{-(s-l+1)})_{\alpha\beta} - Z_{\alpha\beta} (ZY^{-s})_{\alpha\beta}. \quad (A.4)
\]

Let us first show that this expression is integrable. For any matrix \( X \), we have \( |X_{ij}| \leq \|X\|_2 = \|X^tX\|_2^{1/2} \). Thus by Equation (A.4),

\[
\mathbb{E} \left[ \left( \frac{\partial}{\partial \delta_{\alpha\beta}} - Z_{\alpha\beta} \right) (ZY^{-s})_{\alpha\beta} \right] \leq \mathbb{E} \left[ Y^{-s} \right] + \sum_{l=1}^{s} \|Y^{2l+1}\|_2^{1/2} \|Y^{2s+2l-1}\|_2^{1/2}
\]
As $Y$ is positive definite, $\|Y^{-a}\|_2 \leq \text{tr} Y^{-a}$ for any $a \in \mathbb{N}$, so by the Cauchy-Schwarz inequality,
\[
\leq E[\text{tr} Y^{-s}] + \sum_{l=1}^{s} E\left[\text{tr} Y^{-2l+1}\right] \leq E\left[\text{tr} Y^{-2s+2l}\right] + \sum_{l=1}^{s} E\left[\text{tr} Y^{-2l}\right] + n E[\text{tr} Y] + n E[\text{tr} Y] + n E[\text{tr} Y^{-1}] + \frac{1}{2},
\]
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.4) and obtain
\[
0 = E\left[\frac{1}{n} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{p} \left( - \frac{\partial}{\partial Z_{\alpha\beta}} - Z_{\alpha\beta}\right) (ZY^{-s})_{\alpha\beta}\right],
\]
\[
= E\left[\text{tr} Y^{-s} - \frac{s}{n} \text{tr} Y^{-s} - \frac{1}{n} \sum_{l=1}^{s} \text{tr}(Y^{-l}) \text{tr}(Y^{-(s-l)}) - \text{tr} Y^{-(s-1)}\right].
\] (A.5)

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 et al., 1967, Theorem 43] entails $\text{tr}(Y^{-s}) \text{tr}(Y^{-(s-l)}) \leq p \text{tr}(Y^{-s})$. Employing this result in Equation (A.6), whose terms are all individually integrable as $n \geq p + 4s + 2$, then yields that
\[
\left(1 - \frac{(p+1)s}{n}\right) E\left[\text{tr} Y^{-s}\right] \leq E\left[\text{tr} Y^{-(s-1)}\right].
\] (A.7)

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.7) and dividing by $p$, we obtain
\[
\prod_{l=1}^{s} \left(1 - \frac{(p+1)s}{n}\right) \cdot \frac{1}{p} E\left[\text{tr} Y^{-s}\right] \leq \frac{1}{p} E\left[\text{tr} Y^{0}\right] = 1.
\]

Taking a limit in the above yields
\[
\prod_{l=1}^{s} (1 - \alpha l) \lim_{n \to \infty} \frac{1}{p} E\left[\text{tr} Y^{-s}\right] \leq 1.
\]
so for any $s < \alpha^{-1}$ we have
\[
\lim_{n \to \infty} \frac{1}{p} \mathbb{E} \left[ \text{tr} \, Y^{-s} \right] \leq \prod_{l=1}^{s} \frac{1}{1 - \alpha l} < \infty,
\] (A.8)
as desired. \hfill \Box

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 $\mathcal{F}\{ t^{1/2} \} = \phi$ to lighten the notation. We can write
\[
d_{L^2}(\phi, \psi) = \left[ \left\| \frac{\psi^2}{L^2} - 1 \right\| + 2 \int_{S_p(\mathbb{R})} \left[ 1 - \Re \left\{ \frac{\psi^{1/2}(T)}{\phi^{1/2}(T)} \right\} \right] \phi(T) \, dT \right]
\]
for $R = \Re \log[\phi(T)/\psi(T)]$ and $I = \Im \log[\phi(T)/\psi(T)]$. Using the inequality $-\cos(x) \leq -1 + \sqrt{2|x|}$, this can be bounded by
\[
\leq \left[ \left\| \frac{\psi^2}{L^2} - 1 \right\| + 2 \int_{S_p(\mathbb{R})} \left[ 1 - \exp \left\{ \frac{R}{2} \right\} + \exp \left\{ \frac{R}{2} \right\} \right] \phi(T) \, dT,\right.
\]
which, by the inequality $1 - \exp(x) \leq -x$ and the Cauchy-Schwarz inequality, can be further bounded by
\[
\leq \left[ \left\| \psi^2 \right\||L^2 - 1 \right\| + \int_{S_p(\mathbb{R})} R |\phi(T)| \, dT + 2 \int_{S_p(\mathbb{R})} \exp \left\{ -R \right\} |\phi(T)| \, dT \frac{1}{2} \int_{S_p(\mathbb{R})} |I| \, |\phi(T)| \, dT \frac{1}{2}.
\]
But by definition of $R$, $\exp\{-R\} = |\psi(T)|/|\phi(T)|$. Plugging this back in the above and simplifying yields the desired result. \hfill \Box

**Proof of Proposition 3.** Let $\text{NW}(n, p) = \sqrt{n}[W_p(n, I_p/n) - I_p]$ 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
\[
f_{\text{NW}(n, p)}(X) = \frac{n^{(2n+p-2)/2}}{\Gamma_p \left( \frac{n}{2} \right)} e^{-\frac{1}{2} \text{tr} \left[ I_p + \frac{X}{\sqrt{n}} \right] I_p + \frac{X}{\sqrt{n}}} \mathbb{1} \left[ I_p + \frac{X}{\sqrt{n}} > 0 \right],
\]
where $\Gamma_p$ stands for the multivariate Gamma function. But
\[
f_{\text{NW}(n, p)}^{1/2}(X) = \frac{2^{p(2n+p+1)/2} \Gamma_p \left( \frac{n+p+1}{2} \right)}{\Gamma_p \left( \frac{n}{2} \right)} \sqrt{n} \mathbb{1} \left[ W_p \left( \frac{n+p+1}{2}, I_p \right) - I_p \right] (X) \quad (A.9)
\]
which is integrable when \((n+p+1)/2 > p - 1\), i.e. \(n \geq p - 2\) [Muirhead, 1982, Theorem 3.2.1 again and comment on p.87]. For an integrable function \(f\) on \(\mathbb{S}_p(\mathbb{R})\), the Fourier transform with kernel \(\exp\{-i \text{tr}(XT)\}\), normalized to be an \(L^2\)-isometry, satisfies \(\mathcal{F}\{f\}(T) = 2^{-\frac{p}{2}} \frac{\pi^{p(p+1)/4}}{\Gamma_p \left(\frac{n}{2}\right)} e^{i\sqrt{p} \text{tr} T} \mathcal{F}\{f_{\mathbb{W}_p}(n+\frac{1}{2}, \frac{1}{2} T_p)\}(\sqrt{n}T)\). Thus, we obtain that

\[
\begin{aligned}
\mathcal{F}\{f_{\mathbb{W}_p}(n,p)\}(T) &= 2^{-\frac{p}{2}} \frac{\pi^{p(p+1)/4}}{\Gamma_p \left(\frac{n}{2}\right)} e^{i\sqrt{p} \text{tr} T} \mathcal{F}\{f_{\mathbb{W}_p}(n+\frac{1}{2}, \frac{1}{2} T_p)\}(\sqrt{n}T) \\
&= 2^{-\frac{p}{2}} \frac{\pi^{p(p+1)/4}}{\Gamma_p \left(\frac{n}{2}\right)} e^{i\sqrt{p} \text{tr} T} \int_{\mathbb{S}_p(\mathbb{R})} e^{-i \text{tr}(XT)} f_{\mathbb{W}_p}(n+\frac{1}{2}, \frac{1}{2} T_p)(X) dX \\
&= 2^{-\frac{p}{2}} \frac{\pi^{p(p+1)/4}}{\Gamma_p \left(\frac{n}{2}\right)} e^{i\sqrt{p} \text{tr} T} f_{\mathbb{W}_p}(n+\frac{1}{2}, \frac{1}{2} T_p)(X) dX,
\end{aligned}
\]

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 \(2^{-p(p+3)/4} \pi^{-p(p+1)/4} \exp\{-\text{tr} \; X^2/4\}\), we find that the normalizing constant of the GOE\((p)/4\) distribution is \(C_{\text{GOE}/4} = 2^{p(p+1)/4} \pi^{p(p+1)/4}\). Then by Stirling’s approximation as well as Muirhead [1982, Theorem 2.1.12],

\[
\log C_t = \frac{p(p+1)}{4} \log 2 - \frac{p(p+1)}{4} \log \pi - \frac{p(p+3)}{4} \\
+ \frac{1}{2} \sum_{i=1}^{p} \left( n - [2i-\frac{1}{2}p] \right) \log \left( 1 - \frac{2i-p-3}{n} \right) \\
- \frac{1}{2} \sum_{i=1}^{p} \left( n - i \right) \log \left( 1 - \frac{i-\frac{1}{2}}{n} \right) + o(1) \quad (A.10)
\]
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 \(x = 0\),

\[
- \sum_{i=1}^{p} \frac{n-i}{2} \log \left( 1 - \frac{i-\frac{1}{2}}{n} \right) = \sum_{k=1}^{K} \frac{C_k^{(1)} p^{k+2}}{n^k} + o \left( \frac{p^{K+3}}{n^{K+1}} \right)
\]

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

\[
\sum_{i=1}^{p} \frac{n-[2i-p-1]}{2} \log \left( 1 - \frac{2i-p-3}{n} \right)
\]
because the Sym-
try to repeatedly differentiate
distribution evaluated at zero provide its moments, up to a constant, we can
on the open set
creasing positive-definite matrices at both sides instead. In this spirit, define
to compute the derivatives at zero by taking a limit over a sequence of de-
symbols. Because the integrand is orthogonally invariant, we found it easier
makes it difficult to directly interchange the differentiation and integration
for fixed
Thus the characteristic function of the Sym-
malized Wishart
"c
for constants
Proof of Lemma 2. Let \( f_{\text{NW}} \) and \( f_t \) stand for the densities of the normal-
ized Wishart \( \sqrt{n}[W_p(n, I_p/n) - I_p] \) and the Sym-\( t_{n/2}(I_p/8) \) distributions,
respectively. Let \( R(X) = -X \) be the flip operator and \( \ast \) stand for the con-
volution. Since \( f_{\text{NW}}^{1/2} \) is proportional to a \( \sqrt{n}[W_p(\frac{n+p+1}{2}, \frac{n}{2} I_p) - I_p] \) density (see
Equation (A.9)), it is integrable. Therefore, \( f_{\text{NW}}^{1/2} \ast (f_{\text{NW}}^{1/2} \circ R) \) is well-defined
and integrable as well. But then, as \( f_{\text{NW}} \) is real-valued,

\[
2^{-\frac{p}{2}} e^{-\frac{p(p+1)}{4}} \mathcal{F}\{f_{\text{NW}}^{1/2} \ast (f_{\text{NW}}^{1/2} \circ R)\} = \mathcal{F}\{f_{\text{NW}}^{1/2}\} \mathcal{F}\{f_{\text{NW}}^{1/2} \circ R\} = \mathcal{F}\{f_{\text{NW}}^{1/2}\} \mathcal{F}\{f_{\text{NW}}^{1/2}\} = f_t.
\]

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

\[
f_{\text{NW}}^{1/2} \ast (f_{\text{NW}}^{1/2} \circ R) (X) = \int_{\mathbb{S}_p(\mathbb{R})} e^{i tr(TX)} f_t(T) dT.
\]

(A.11)

Thus the characteristic function of the Sym-\( t_{n/2}(I_p/8) \) distribution is given
by \( f_{\text{NW}}^{1/2} \ast (f_{\text{NW}}^{1/2} \circ R) \). 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 \( f_{\text{NW}} \ast (f_{\text{NW}} \circ R) \) at zero to compute \( \mathbb{E}[\text{tr} T^{2k}] \).

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 de-
creasing 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

\[
H(X) = \frac{(-1)^k}{n^k} \text{tr} \left( \nabla_X f_{\text{NW}}^{1/2} \ast (f_{\text{NW}}^{1/2} \circ R)(\sqrt{n}X) \right)
\]

for fixed \( k, p \) and \( n \). Here \( (\nabla_X)_{ij} = \frac{1+\delta_{ij}}{2} \partial_i X_{j} \) is the symmetric differen-
tiation operator, as defined in Section 2, with the property that \( \nabla_X \text{tr}(XT) = T \)
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 $1 \leq l \leq k$, indices $1 \leq i_1, \ldots, i_{2l} \leq p$ and $X \in S_p(\mathbb{R})$,

$$\left| \nabla_{X_{i_1}} \cdots \nabla_{X_{i_{2l}}} e^{i \sqrt{n} \text{tr}(TX)} \right| = n^l \left| T_{i_2i_3} \cdots T_{i_{2l}i_1} \right|. \quad (\text{A.12})$$

We first show this expression is integrable for large enough $n$ when $T \sim \text{Sym}-t_{n/2}(I_p/8)$. 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 $T$, $|T_{ij}| \leq \lambda_1^{1/2}(T^2) \leq |I_p + T^2|^{1/2}$, where $\lambda_1(T^2) \geq \cdots \geq \lambda_p(T^2) \geq 0$ are the ordered eigenvalues of the positive-definite matrix $T^2$. Thus

$$\int_{S_p(\mathbb{R})} n^l \left| T_{i_2i_3} \cdots T_{i_{2l}i_1} \right| f_l(T) \, dT \leq n^{2l} C_l \int_{S_p(\mathbb{R})} \left| I_p + \frac{16T^2}{n} \right|^{-(n-2l)l+1} dT \leq \frac{n^{2l} C_l}{4^l} \left( \frac{n}{n-2l} \right)^{\frac{p(p+1)}{4}} \int_{S_p(\mathbb{R})} \left| I_p + \frac{16T^2}{n-2l} \right| dT. \quad (\text{A.13})$$

When $n - 2l \geq p - 2$, 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 $n \geq p + 2k - 2$, the right hand side of Equation (A.12) is an integrable function for all $1 \leq l \leq k$ and $1 \leq i_1, \ldots, i_{2l} \leq p$. By Equation (A.11), and repeated differentiation under the integral sign justified by the integrability bounds given by Equations (A.12) and (A.13), we find that

$$H(X) = \int_{S_p(\mathbb{R})} \text{tr} \left( T^{2k} e^{i \sqrt{n} \text{tr}(TX)} \right) f_l(T) \, dT \quad (\text{A.14})$$

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

Now let’s relate $H$ to the definition of $f_{NW}^{1/2} \ast (f_{NW}^{1/2} \circ R)$ 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

$$f_{NW}^{1/2} \ast (f_{NW}^{1/2} \circ R)(\sqrt{n}X) = \int_{S_p(\mathbb{R})} f_{NW}^{1/2}(Z)f_{NW}^{1/2}(Z-\sqrt{n}X) \, dZ$$

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

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 \( 1[Y + X > 0, Y > 0] = 1[Y > 0] \), and thus \( H \) satisfies

\[
H(X) = \frac{(-1)^k}{n^k} \text{tr} \left( \nabla^k_X \int_{Y > 0} e^{-\frac{n}{2} \operatorname{tr}(Y + X) |Y + X|^{rac{m}{2}}} \frac{\sum_{i=1}^{n} \sum_{J_p} (I_p)_{j_1,j_2} |(I_p)_{j_1,j_2}|^2}{\frac{2^n \Gamma_p \left( \frac{n}{2} \right)}{2^n \Gamma_p \left( \frac{n}{2} \right)}} dY \right) \quad (A.15)
\]

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

\[
\left| \nabla^k X_{i_2i_{2l-1}} \cdots \nabla^k X_{i_1i_3} \nabla^k X_{i_2i_1} e^{-\frac{n}{2} \operatorname{tr}(Y + X) |Y + X|^{rac{m}{2}}} \right| \leq \sum_{s=0}^{l} \sum_{J_p} (n, m) \prod_{t=s+1}^{l} (I_p)_{j_2, j_2} |(I_p)_{j_2, j_2}| \prod_{t=1}^{s} |(Y + X)_{j_2, j_2}^{-1} | e^{-\frac{n}{2} \operatorname{tr}(Y + X) |Y + X|^{rac{m}{2}}}
\]

for some polynomials \( a_{i,s} \) that do not depend on \( X \) or \( Y \). This can be uniformly bounded for all \( 0 \leq X \leq I_p \) by

\[
\leq C(n, m, p) \sum_{s=0}^{l} \text{tr}^s(Y^{-1}) e^{-\frac{n}{2} \text{tr} Y} \left[ 1 + \text{tr} Y \right] \frac{mp}{4} \quad (A.16)
\]

for some constant \( C(n, m, p) \) that does not depend on \( X \) or \( Y \). But for any \( n, m, p \)

\[
\int_{Y > 0} C(n, m, p) \sum_{s=0}^{l} \text{tr}^s(Y^{-1}) e^{-\frac{n}{2} \text{tr} Y} \left[ 1 + \text{tr} Y \right] \frac{mp}{4} \frac{\sum_{i=1}^{n} \sum_{J_p} (I_p)_{j_1,j_2} |(I_p)_{j_1,j_2}|^2}{\frac{2^n \Gamma_p \left( \frac{n}{2} \right)}{2^n \Gamma_p \left( \frac{n}{2} \right)}} dY
\]

\[
= \frac{n^{mp}}{2^n \Gamma_p \left( \frac{n}{2} \right)} C(n, m, p) \sum_{s=0}^{l} \int_{Y > 0} \left[ 1 + \text{tr} Y \right] \frac{mp}{4} \text{tr}^s(Y^{-1}) e^{-\frac{n}{2} \text{tr} Y} |Y| \frac{n^{p+1}}{4} \frac{\sum_{i=1}^{n} \sum_{J_p} (I_p)_{j_1,j_2} |(I_p)_{j_1,j_2}|^2}{\frac{2^n \Gamma_p \left( \frac{n}{2} \right)}{2^n \Gamma_p \left( \frac{n}{2} \right)}} dY
\]

\[
= \left( \frac{n}{2} \right)^{mp} \frac{\Gamma_p \left( \frac{n+p+1}{2} \right)}{\Gamma_p \left( \frac{n}{2} \right)} E \left[ (1 + \text{tr} Y) \frac{mp}{4} \text{tr}^s(Y^{-1}) \right]
\]

for a \( Y \) with a Wishart distribution \( W_p \left( \frac{n+p+1}{2}, \frac{n}{2} I_p \right) \). The Cauchy-Schwarz inequality then entails the bound

\[
\leq \left( \frac{n}{2} \right)^{mp} \frac{\Gamma_p \left( \frac{n+p+1}{2} \right)}{\Gamma_p \left( \frac{n}{2} \right)} \left[ (1 + \text{tr} Y) \frac{mp}{4} \right]^2 \left[ \text{tr}^2 s(Y^{-1}) \right]^{\frac{1}{2}} \quad (A.17)
\]

The first expectation is always finite when \( n \geq p - 2 \). Since \( \text{tr}^2 s(Y^{-1}) \) 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.15) 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.16) and (A.17), and obtain in that case

$$H_1(X) = \frac{(-1)^k}{n^k} \int_{Y > 0} \left( \nabla_X e^{-\frac{\tr(Y+X)}{2}} \left| Y+X \right|^{\frac{n}{2}} \right) \frac{\frac{np}{2} e^{-\frac{\tr Y}{2}} \left| Y \right|^{\frac{m}{2}}}{2^{\frac{np}{2}} \Gamma_p \left( \frac{n}{2} \right) } dY \quad \text{(A.18)}$$

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

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

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.14),

$$\lim_{X \to 0} H_1(X) = \int_{0 < X < I_p} \lim_{X \to 0} \left( \nabla_X e^{i \sqrt{n} \tr(TX)} f_t(T) \right) dT = E \left[ \tr T^{2k} \right]$$

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

On the other hand, the integrand at Equation (A.18) takes a particularly simple form. Lemma 4 establishes by induction that there must be polynomials $b_\kappa$ in $n$, $m$ and $p$ with deg $b_\kappa \leq 2k + 1 - q(\kappa)$ such that

$$H(X) = \frac{(-1)^k}{n^k} \int_{Y > 0} \sum_{|\kappa| \leq 2k} b^{(1)}(n, m, p) r_\kappa([Y + X]^{-1}) e^{-\frac{\tr(Y+X)}{2}} \left| Y+X \right|^{\frac{m}{2}} \frac{n^\frac{np}{2} e^{-\frac{\tr Y}{2}} \left| Y \right|^{\frac{m}{2}}}{2^{\frac{np}{2}} \Gamma_p \left( \frac{n}{2} \right) } dY \quad \text{(A.20)}$$

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

$$r_\kappa([Y + X]^{-1}) e^{-\frac{\tr(Y+X)}{2}} \left| Y+X \right|^{\frac{m}{2}} \leq \tr|\kappa|([Y]^{-1}) e^{-\frac{\tr Y}{2}} \left[ 1 + \tr Y \right]^{\frac{m}{2}}$$

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.17). Thus for such $n$, by the dominated convergence theorem and Equation (A.20), we obtain that

$$\lim_{X \to 0} H(X) = \frac{(-1)^k}{n^k} \int_{0 < X < I_p} \sum_{|\kappa| \leq 2k} b^{(1)}(n, m, p) r_\kappa([Y + X]^{-1})$$
RFnomials of degree at most \(|\lambda|\) for some coefficients \(P\), where \(Y\) follows a \(W_p(n, I_p/n)\) distribution. Combining Equation (A.19) with Equation (A.21) and Lemma 4 concludes the proof. \(\square\)

**Proof of Theorem 2.** Recall the expected zonal polynomial of an inverse Wishart \(W_p^{-1}(n, I_p/n)\) is given by Equation (3.5). Based on the previous calculations, it is tempting to define

\[
c'_\lambda = \frac{2^{2|\lambda|}|\lambda|!}{\prod_{\lambda \in \lambda} \prod_{\lambda \in \lambda}} (2\lambda_i - 2\lambda_j - i + j), \quad R_\lambda(m) = \prod_{i=1}^{\lambda} \prod_{l=0}^{m-1} \frac{m}{m - (1 - i + 2l)},
\]

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

\[
E[r_\alpha(Y^{-1})] = \sum_{|\lambda|=|\alpha|} c_{\alpha,\lambda} c'_\lambda n^{\lambda} \left( \prod_{|\mu|=|\lambda|} R_\mu(m) \prod_{|\mu|=|\lambda|} R_\mu^{-1}(m) \right) P_\lambda(m, p)
\]

\[
= \frac{n^{\lambda}}{m^{\lambda}} R'_{|\lambda|}(m) P'_{|\lambda|}(m, p) \tag{A.23}
\]

where

\[
R'_{|\mu|} = \prod_{|\mu|=|\lambda|} R_\mu(m), \quad P'_{|\lambda|}(m, p) = m^{\lambda} \sum_{|\mu|=|\lambda|} c_{\alpha,\lambda} c'_\lambda \prod_{|\mu|=|\lambda|} R_\mu^{-1}(m) P_\lambda(m, p). \tag{A.24}
\]

But \(R_\mu^{-1}(m) = \prod_{\lambda=1}^{\lambda} \prod_{l=0}^{m-1} \left(1 - \frac{1}{m} + \frac{2l}{m}\right)\) is a polynomial in \(1/m\), while \(P_\lambda(m, p) = \prod_{\lambda=1}^{\lambda} \prod_{l=0}^{m-1} \left(\frac{p}{m} + \frac{1}{m} + \frac{2l}{m}\right)\) is a polynomial in \(p/m\) and \(1/m\), both of degree at most \(|\mu| = |\lambda| = |\kappa|\). Thus

\[
P'_{|\lambda|}(m, p) = m^{\lambda} \sum_{|\lambda|=|\alpha|} c_{\alpha,\lambda} c'_\lambda \prod_{|\mu|=|\lambda|} R_\mu^{-1}(m) P_\lambda(m, p)
\]

\[
= m^{\lambda} \sum_{i=0}^{\lambda} b_i \left(\frac{p}{m}\right)^i \frac{1}{m^j} \tag{A.25}
\]

for some coefficients \(b_i \) that don’t depend on \(m, p\) (or \(n\)). Define the polynomials \(f_j(\alpha) = \sum_{i=0}^{\lambda} b_i \alpha^i\), so that

\[
P'_{|\lambda|}(m, p) = m^{\lambda} \sum_{j=0}^{\lambda} \frac{f\left(\frac{p}{m}\right)}{m^{-j}} m^{j}. \tag{A.26}
\]
Let us show that for all \(0 \leq j < |\kappa| - q(\kappa)\), 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 \(0 \leq j_0 < |\kappa| - q(\kappa)\) be the smallest \(j\) with the property that \(f_{j_0}(\alpha_0) \neq 0\) for some \(\alpha_0 \in (0, 1/\max(|\kappa| - 2, 0))\). As \(f_{j_0}\) is a polynomial, by continuity it must be non-zero in a neighborhood of \(\alpha_0\), so we may as well assume \(\alpha_0\) is rational without loss of generality. Now look at what happens to \(E[r_\kappa(Y^{-1})]\) as \(p\) grows to infinity at the very specific linear rate \(p = \left[\frac{\alpha_0}{1+\alpha_0}(n-1)\right]\). Since \(\alpha_0\) is rational, there must be a subsequence \(n_l\) such that \(p_l\) is exactly an integer (for example, 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,

\[
\lim_{l \to \infty} \frac{1}{m_l^{\|\kappa| - q(\kappa) - j_0\}} \cdot \frac{1}{p_l^{q(\kappa)}} E[r_\kappa(Y^{-1})] \leq 0 \cdot \lim_{l \to \infty} \frac{1}{p_l} E[\text{tr} Y^{-|\kappa|}] = 0.
\]

On the other hand, by Equations (A.23) and (A.26), the definition of \(j_0\) and the fact that \(R_{|\kappa|}(m) \to 1\) as \(m \to \infty\),

\[
\lim_{l \to \infty} \frac{1}{m_l^{\|\kappa| - q(\kappa) - j_0\}} \cdot \frac{1}{p_l^{q(\kappa)}} E[r_\kappa(Y^{-1})] = \lim_{l \to \infty} \left(\frac{n_l}{m_l}\right)^{|\kappa|} R'_{|\kappa|}(m_l) \left(\frac{m_l}{p_l}\right)^{q(\kappa)} \sum_{j=j_0}^{\|\kappa|} f_j(\alpha_0) m_j^{j_0-j} = (1 + \alpha_0)^{|\kappa|} \alpha_0^{-q(\kappa)} f_{j_0}(\alpha_0).
\]

As \(\alpha_0 > 0\), \(f_{j_0}(\alpha_0)\) must therefore equal zero, a contradiction. Hence, as claimed, the polynomials \(f_j(\alpha)\) for \(0 \leq j < |\kappa| - q(\kappa)\) 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 \(b_{ij} = 0\) for \(0 \leq j < |\kappa| - q(\kappa)\).

Thus, from Equations (A.23) and (A.25) we have

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

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 $n \geq p + 8k + 6$ that

$$E[\text{tr } T^{2k}] = \frac{m^{2k+1}}{m^k} R_{2k}(m) Q_n(m, p), \quad (A.27)$$

where

$$Q_n(m, p) = (-1)^k \sum_{|\kappa| \leq 2k} \left(1 + \frac{p}{m} + \frac{1}{m}\right)^{|\kappa|} R_{|\kappa|}'(m) b_n(m, m, p, m^{2k+1-q(|\kappa|)}) P_{\kappa}'(m, p).$$

Now, for any $a \leq b$, we can associate a partition $\mu$ of norm $|\mu| = a$ with the partition $\mu^* = (\mu_1 + b - a, \mu_2, \ldots, \mu_q(\mu))$ of norm $|\mu^*| = b$, which satisfies

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

By definition for the $R_\mu(m)$’s at Equation (A.22), this means that every factor that appears in $R_\mu^{-1}(m)$ appears in $R_{\mu^*}(m)$, so by definition of the $R_{|\mu|}(m)$’s at Equation (A.24), $R_q(m) R^{-1}(m)$ is a polynomial in $\frac{1}{m}$. Moreover, as the $b_n$ are polynomials of degrees $d(\kappa) \equiv 2k + 1 - q(\kappa)$, there exists coefficients $c_{ijl}$ such that

$$b_n(m, m, p, m^{2k+1-q(\kappa)}) = \frac{1}{m^{d(\kappa)}} \sum_{i=0}^{d(\kappa)} \sum_{j=0}^{d(\kappa)-i} \sum_{l=0}^{d(\kappa)-i-j} c_{ijl} m^i n^j p^l$$

$$= \sum_{i=0}^{d(\kappa)} \sum_{j=0}^{d(\kappa)-i} \sum_{l=0}^{d(\kappa)-i-j} c_{ijl} \frac{m^{d(\kappa)-i-j-l}}{m^{d(\kappa)-i-j-l}} \left(1 + \frac{p}{m} + \frac{1}{m}\right)^j \left(\frac{p}{m}\right)^l.$$

As $d(-i - j - l \geq 0), j, l \geq 0$, we conclude that this expression is a polynomial in $\frac{p}{m}$ and $\frac{1}{m}$. Therefore, looking back at (A.27), we conclude that the $Q_n(m, p)$’s are polynomials in $\frac{p}{m}$ and $\frac{1}{m}$. In other words, when $n \geq p + 8k + 6$ there must be polynomials $g_i(p) = \sum_{j=0}^{D} a_{ij} p^j$ and a large enough integer $D$ such that

$$E[\text{tr } T^{2k}] = \left(\frac{m^k}{n}\right) R_{2k}(m) \sum_{i=0}^{D} g_i(p) \frac{m^{i-k-1}}{m^i - k - 1}. \quad (A.28)$$

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 $E[\text{tr } T^{2k}]$ 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

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

But for $p$ fixed, $\lim_{n \to \infty} C_t = C_{\text{GOE}/4}$, the normalization constant of the GOE(p)/4 distribution, by Lemma 1. Moreover, $|I_p + 16T^2/n|^{-\frac{p}{2}} = \prod_{i=1}^{p} [1 + \ldots]$
\( \lambda_i(4T^2)/(n/4) \) 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 towards \( \exp(x) \). Therefore, for a fixed dimension \( p \) we can apply the monotone convergence theorem to obtain that
\[
\lim_{n \to \infty} E \left[ \text{tr} T^{2k} \right] \leq C_{\text{GOE}/4} \int_{S_p(\mathbb{R})} \text{tr} T^{2k} e^{-\text{tr} T^2} dT = E \left[ \text{tr} Z^{2k} \right] < \infty \quad (A.29)
\]
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 wasn’t 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.28) and the definition of \( i_0 \), the limit of \( E[\text{tr} T^{2k}] \) as \( n \to \infty \) with \( p \) fixed at \( p_0 \) satisfies
\[
\lim_{n \to \infty} \frac{E[\text{tr} T^{2k}]}{m^{k+1-i_0}} = 1^k \cdot 1 \cdot \lim_{n \to \infty} \sum_{i=i_0}^D g_i(p_0)m^{i-i_0} = g_{i_0}(p_0).
\]
But \( m = n - p - 1 \) tends to infinity as \( n \) tends to infinity, and since \( k + 1 - i_0 > 0 \), Equation (A.29) means that \( E[\text{tr} T^{2k}] / m^{k+1-i_0} \) 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.28) and the above,
\[
\lim_{n \to \infty} \frac{1}{p^{k+1}} E \left[ \text{tr} T^{2k} \right] = \lim_{n \to \infty} \left( \frac{m}{n} \right)^k R_{2k}(m) \sum_{i=k+1}^D \sum_{j=0}^k a_{ij} \frac{p^{j-(k+1)}}{m^{i-(k+1)}} + \sum_{i=k+1}^D \sum_{j=k+1}^\infty \left( \frac{p}{m} \right)^{j-(k+1)} \frac{a_{ij}}{m^{i-j}}
\]
\[
= \sum_{j=0}^k \left( \lim_{p \to \infty} \frac{a_{k+1}}{p^{k+1-j}} \right) + a_{(k+1)(k+1)} < \infty,
\]
as desired.

\[
\text{Proof of Corollary 1.} \text{ Let } f_t \text{ be the density of the Sym-}t_{n/2}(I_p/8) \text{ distribution, and define for } L \in \mathbb{N}
\]
\[
h_L(T) = \frac{n+p+1}{n} \text{tr} \left[ (I_p + \frac{16T^2}{n})^{-1}(4T)^{2L+1} \right]
\]
\[
- \left[ 2L \text{tr}(4T)^{2L-1} + \sum_{l=0}^{2L-1} \text{tr}(4T)^{2L-1-l} \right] 1[L > 0].
\]
Then
\[
- \frac{1}{2} h_L^2(T) f_1(T) = h_L(T) \left( \text{tr} \left[ (4T)^{2k} \tilde{\nabla}_4 h(T) \right] + \text{tr} \left[ \tilde{\nabla}_4 (4T)^{2k} \right] f_1(T) \right) 
\]
\[
= tr \tilde{\nabla}_4 \left[ (4T)^{2k} h_L(T) f_1(T) \right] - tr \left[ (4T)^{2k} \tilde{\nabla}_4 h_L(T) \right] f_1(T) 
\]  \tag{A.30}
But when \( n + p > 16L + 3 \), for any \( 1 \leq \alpha \leq \beta \leq p \),
\[
\int_{S_p(\mathbb{R})} \left( \tilde{\nabla}_4 \right)_{\alpha \beta} \left[ (4T)^{2L} h_L(T) f_1(T) \right]_{\beta \alpha} dT 
\]
\[
= \frac{1 + \delta_{\alpha \beta}}{8} \int_{\mathbb{R}^{2(p+1)}} \left[ ([4T]^{2L})_{\beta \alpha} h_L(T) f_1(T) \right] \prod_{i \neq j} dT_{ij} = 0 
\]
since \( T^{2L} h_L(T) f_1(T) \) is integrable for \( \frac{n+p+1}{4} > 4L + 1 \). Using the notation \( \| \cdot \|_T \) for the \( L^2 \)-norm under \( T \sim \text{Sym}^{-n/2}(I_p/8) \), we find from Equation (A.30) that
\[
\frac{1}{p^{2k+1}} \| h_L(T) \|_T^2 = 0 + \frac{2}{p^{2k+1}} \text{tr} \left[ (4T)^{2k} \tilde{\nabla}_4 h_L(T) \right] 
\]
\[
= E \left[ \frac{2(2L + 1)}{p^{2k+1}} \frac{n+p+1}{n} \frac{1}{p} \text{tr} \left[ \left( I_p + \frac{16T^2}{n} \right)^{-1} \left( \frac{4T}{\sqrt{p}} \right)^{4L+2} \right] \right] 
\]
\[
+ 2(2L - 1) \frac{n+p+1}{n} \frac{1}{p} \text{tr} \left[ \left( I_p + \frac{16T^2}{n} \right)^{-1} \left( \frac{4T}{\sqrt{p}} \right)^{4L} \right] 
\]
\[
- 4L(2L - 1) \frac{1}{p^2} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{4L-2} 
\]
\[
- \sum_{l=0}^{2L-1} \frac{1}{p} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2L-l} \frac{1}{p} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2L-l} = O(1) \tag{A.31}
\]
as \( p, n \to \infty \) with \( p/n \to 0 \), using that \( x^{2l}/(1+16x^2/n) \leq x^{2l} \) \( \forall l \in \mathbb{N} \) and Theorem 2.

To show the corollary, we proceed by induction on \( L \) and assume that
\( \| \text{tr}(4T/\sqrt{p})^{2l+1} \|_T = O(1) \) for every \( 0 \leq l < L \). If we take Equation (A.31) with \( L = k + 1 \) and multiply it by \( p/n \), we find by the triangle inequality that
\[
\| \frac{n+p+1}{n} \text{tr} \left[ \left( I_p + \frac{16T^2}{n} \right)^{-1} \left( \frac{4T}{\sqrt{p}} \right)^{2k+3} \right] - 2p + k \frac{1}{n} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k+1} \|_T 
\]
\[
\leq \frac{p}{n} \sum_{l=1}^{2k} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k+1-l} \]
\[ + \sum_{l=1}^{2k} \frac{1}{p} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^l \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k+1-l} \| T \| + O \left( \frac{p}{n} \right) = o(1) \quad (A.32) \]

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
\[
\left\| \frac{n+p+1}{n} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k+1} - \frac{2p+k}{n} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k+1} \| T \right. 
\leq \left\| \frac{n+p+1}{n} \text{tr} \left( I_p + \frac{16T^2}{n} \right)^{-1} \left( \frac{4T}{\sqrt{p}} \right)^{2k+3} \right. 
+ \sum_{l=0, \text{odd}}^{2k-1} \frac{1}{l} \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^l \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k-1-l} \| T 
+ O(1),
\]
that is, from Equation (A.32), Theorem 2 and the inductive hypotheses,
\[
\left\| \text{tr} \left( \frac{4T}{\sqrt{p}} \right)^{2k+1} \right\| T \leq \left( \frac{n+p+1}{n} - \frac{2p+k}{n} \right)^{-1} \left( o(1) + O(1) + O(1) \right) = O(1), \]
as desired. \( \square \)

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

REFERENCES


Miklós Z. Rácz and Jacob Richey. A smooth transition from Wishart to GOE. *Journal of Theoretical Probability*, pages 1–9, 2018.