THE GAUSSIAN AND WISHART ENSEMBLES: EIGENVALUE DENSITIES

1. WEYL'S INTEGRATION FORMULA

1.1. **Orthogonal and Unitary Invariance.** For the classical random matrix ensembles — the Gaussian Orthogonal, Unitary, and Symplectic Ensembles, the real and complex Wishart Ensembles, and the Circular Ensembles — the joint probability densities of the matrix entries, relative to Lebesgue measures, are functions only of the eigenvalues. This makes it possible to express the joint densities of the eigenvalues in tractable forms, by integrating out the extraneous variables. The key fact is the following theorem, a version of the *Weyl Integration Formula*, whose proof will be given in sec. **??** below.

Theorem 1. *If $X = (X_{i,j})$ is a real, symmetric random matrix with density $g(\lambda_1, \lambda_2, \dots, \lambda_N)$ relative to Lebesgue measure $\prod_i dX_{i,i} \prod_{i<j} dX_{i,j}$, where λ_i are the eigenvalues of X and g is a symmetric function of its arguments, then the joint density of the eigenvalues relative to Lebesgue measure $\prod_i d\lambda_i$ is*

$$g(\lambda_1, \lambda_2, \dots, \lambda_N) \prod_{i<j} |\lambda_i - \lambda_j| / C_N. \tag{1}$$

Similarly, if $Z = (Z_{i,j} = X_{i,j} + \sqrt{-1} Y_{i,j})$ is a complex, Hermitian random matrix[1] *with density $g(\lambda_1, \lambda_2, \dots, \lambda_N)$ relative to Lebesgue measure $\prod_i dX_{i,i} \prod_{i<j} dX_{i,j} dY_{i,j}$, where λ_i are the eigenvalues of Z and g is a symmetric function of its arguments, then the joint density of the eigenvalues relative to Lebesgue measure $\prod_i d\lambda_i$ is*

$$g(\lambda_1, \lambda_2, \dots, \lambda_N) \prod_{i<j} (\lambda_i - \lambda_j)^2 / \tilde{C}_N. \tag{2}$$

The normalizing constants C_N and $\tilde{C}_N$ in formulas (1)–(2) depend on the function g; for the Gaussian orthogonal and unitary ensembles they will be explicitly evaluated below. It is assumed in Theorem 1 that the eigenvalues λ_i are listed in increasing order $\lambda_1 \le \lambda_2 \le \dots \le \lambda_N$, and the probability densities (1)–(2) are densities relative to Lebesgue measure on the so-called *Weyl chamber* $\{\lambda_1 \le \lambda_2 \le \dots \le \lambda_N\}$. It is of course easy to deduce the joint distribution of the eigenvalues listed in *random* order: It is just (1) (or (2)) multiplied by $1/N!$.

The following lemma explains the origin of the terms "Gaussian *Orthogonal* Ensemble" and "Gaussian *Unitary* Ensemble".

Proposition 2. *Let $X = (X_{i,j})$ be a random, real symmetric $N \times N$ matrix whose entries $(X_{i,j})_{j \ge i}$ have a joint density $g(\lambda_1, \lambda_2, \dots, \lambda_N)$ with respect to Lebesgue measure $\prod_i dX_{i,i} \prod_{j>i} dX_{i,j}$, where λ_i are the eigenvalues of X and g is a symmetric function. Then for every orthogonal transformation U the random matrix UXU^T has the same distribution as X.*

[1] In the following I will revert to the custom of writing $i = \sqrt{-1}$, despite the fact that the symbol i is also generically used as the first index of the matrix entries. Observe that if Z is Hermitian then the diagonal entries $Z_{i,i} = X_{i,i}$ must have imaginary parts 0.

Similarly, if $Z = (Z_{i,j} = X_{i,j} + \sqrt{-1}Y_{i,j})$ is a complex, Hermitian random matrix whose above-diagonal entries $Z_{i,i}$ and $X_{i,j}, Y_{i,j}$ have joint density $g(\lambda_1, \lambda_2, \dots, \lambda_N)$ relative to the Lebesgue measure $\prod_i dX_{i,i} \prod_{i<j} dX_{i,j} dY_{i,j}$, then for every unitary transformation U the random matrix UZU^* has the same distribution as Z.

Proof. Consider the orthogonal case. Since the eigenvalues of UXU^T are the same as those of X, the assertion will follow if we can prove that Lebesgue measure $\prod_i dX_{i,i} \prod_{j>i} dX_{i,j}$ is invariant under conjugation by U. In particular, we will show that if

$$\tilde{X} = UXU^T$$

then

$$\prod_i d\tilde{X}_{i,i} \prod_{j>i} d\tilde{X}_{i,j} = \prod_i dX_{i,i} \prod_{j>i} dX_{i,j}.$$

By the change of variable formula for multiple integrals it suffices to show that the determinant of the $\binom{N+1}{2} \times \binom{N+1}{2}$ linear transformation

$$L = \begin{pmatrix} \frac{\partial \tilde{X}_{1,1}}{X_{1,1}} & \frac{\partial \tilde{X}_{2,2}}{X_{1,1}} & \cdots \\ \frac{\partial \tilde{X}_{1,1}}{X_{2,2}} & \frac{\partial \tilde{X}_{2,2}}{X_{2,2}} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}$$

is ± 1. For this we use the fact that $Tr\tilde{X}^2 = TrX^2$, equivalently,

$$\sum_i \tilde{X}_{i,i}^2 + 2\sum_{i<j} \tilde{X}_{i,j}^2 = \sum_i X^2 + 2\sum_{i<j} X_{i,j}^2. \tag{3}$$

Now let's change perspective and view $\tilde{X}$ and X as (column) vectors of length $\binom{N+1}{2}$, that is,

$$X = (X_{11}, \dots, X_{NN}; X_{12}, X_{13}, \dots, X_{N-1,N}) \quad \text{and}$$
$$\tilde{X} = (\tilde{X}_{11}, \dots, \tilde{X}_{NN}; \tilde{X}_{12}, \tilde{X}_{13}, \dots, \tilde{X_{N-1,N}}).$$

(This is of a serious abuse of notation, but it beats inventing something better.) Then equation (3) can be rewritten as

$$\langle LX, QLX \rangle = \langle X, QX \rangle$$

where Q is the $\binom{N+1}{2} \times \binom{N+1}{2}$ diagonal matrix with its first N diagonal entries 1 and the remaining $\binom{N}{2}$ diagonal entries 2. The following lemma now implies that $\det L = \pm 1$. □

Lemma 3. *Let L be an $m \times m$ real matrix such for some $m \times m$ diagonal matrix $D = diag(d_i)$ with all $d_i > 0$,*

$$\langle Lv, DLv \rangle = \langle v, Dv \rangle \quad \textit{for all } v \in \mathbb{R}^m.$$

Then $\det L = \pm 1$.

Proof. Exercise. □

1.2. **The Gaussian Ensembles.** Theorem 1 yields as an almost immediate consequence the joint densities of eigenvalues for both the *Gaussian Orthogonal Ensemble* (GOE) and the *Gaussian Unitary Ensemble* (GUE). Recall that in the GOE the diagonal entries are i.i.d. Normal-$(0,1)$, and the above-diagonal entries $X_{i,j}$ are i.i.d. Normal-$(0,1/2)$. Consequently, the joint density of the matrix entries relative to Lebesgue measure $\prod_i dX_{i,i} \prod_{i<j} dX_{i,j}$ is proportional to

$$\exp\left\{-\sum_{i=1}^N X_{i,i}^2/2 - \sum_{i=1}^N \sum_{j=i+1}^N X_{i,j}^2\right\} = \exp\{-Tr(XX^T)/2\} = \prod_{i=1}^n e^{-\lambda_i^2/2}. \tag{4}$$

In the GUE the diagonal entries $Z_{i,i}$ are again i.i.d. Normal-$(0,1)$, but the off-diagonal entries are *complex* normals[2] with mean zero and variance 1. Hence, the joint density of the real and imaginary parts of the matrix entries is

$$\exp\left\{-\sum_{i=1}^{N} Z_{i,i}^2/2-\sum_{i=1}^{N}\sum_{j=i+1}^{N}|Z_{i,j}|^2\right\}=\exp\{-Tr(ZZ^*)/2\}=\prod_{i=1}^{n}e^{-\lambda_i^2/2}. \tag{5}$$

Thus, by Theorem 1, the joint densities of the eigenvalues λ_i for random matrices X and Z chosen from the GOE and GUE, respectively, are

$$p_N^{GOE}(\lambda_1,\lambda_2,..,\lambda_N)=C_n|\Delta_N(\lambda)|\exp\{-\sum_{i=1}^{N}\lambda_i^2\} \quad \text{and} \tag{6}$$

$$p^{GUE}(\lambda_1,\lambda_2,..,\lambda_N)=C_N'\Delta_N(\lambda)^2\exp\{-\sum_{i=1}^{N}\lambda_i^2\}, \tag{7}$$

where

$$\Delta_N(\lambda)=\Delta_N(\lambda_1,\lambda_2,\dots,\lambda_N)=\prod_{i=1}^{N}\prod_{j=i+1}^{N}(\lambda_i-\lambda_j). \tag{8}$$

1.3. **The Orthogonal and Unitary Ensembles.** The space $U(N)$ of unitary $N\times N$ matrices is a compact group, so it has a Haar measure with finite total mass, which we can normalize to be 1. A random matrix U on U_N distributed according to normalized Haar measure is said to come from (or to be distributed according to) the *unitary ensemble*. If $U\sim UE$, then the *symmetric* random matrix U^TU is said to come from the *orthogonal ensemble*. Recall that any unitary matrix has an orthonormal basis of eigenvectors, and that the eigenvalues $e^{i\theta_j}$ are complex numbers of absolute value 1. There is no natural ordering of the unit circle, so we will assume that the eigenvalues are listed in random order.

Theorem 4. *For the orthogonal ensemble, the joint density of the eigenvalues $e^{i\theta_j}$ relative to normalized Lebesgue measure $\prod_{j=1}^{N}d\theta_j/(2\pi)^N$ is proportional to*

$$|\Delta_N(e^{i\theta_1},e^i\theta_2,\dots,e^{i\theta_N})|=\prod_{j<k}|e^{i\theta_j}-e^{i\theta_k}|. \tag{9}$$

For the unitary ensemble, the joint density is

$$|\Delta_N(e^{i\theta_1},e^i\theta_2,\dots,e^{i\theta_N})|^2=\prod_{j<k}|e^{i\theta_j}-e^{i\theta_k}|^2. \tag{10}$$

This can be proved by a calculation similar to that undertaken for the GOE in section 2 below.

1.4. **The Van der Monde Determinant.** The factor $\Delta_N(\lambda):=\prod_{i=1}^{N}\prod_{j=i+1}^{N}(\lambda_i-\lambda_j)$ that occurs in the joint eigenvalue densities (1) and (2) is a commonly occurring symmetric function. The key to dealing with it is the following algebraic fact.

[2]Recall that a complex normal random variable with mean zero and variance σ^2 is a complex-valued random variable $Z=X+iY$ whose real and imaginary parts X and Y are independent (real-valued) normals with mean zero and variance $\sigma^2/2$.

Proposition 5. *The product* $\Delta_N(\lambda)$ *is the* van der Monde determinant*:*

$$\Delta_n(x_1, x_2, \dots, x_N) = \prod_{1\le i<j\le n}(x_j - x_i) = \det\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & & & \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{pmatrix} \tag{11}$$

Proof. There are various proofs of this formula. The usual proof is algebraic, using the fact that both sides vanish when $x_i = x_j$ for some pair of distinct indices $i \neq j$. Here is a more combinatorial proof that does not rely on the unique factorization theorem for polynomials: (A) Check that both sides are homogeneous polynomials of degree $\binom{n}{2}$ (*homogeneous* means that all terms have the same degree). (B) Check that both sides are *antisymmetric*, that is, if for some pair $i \neq j$ the variables x_i and x_j are switched, then both sides change sign. For this it suffices to check nearest-neighbor pairs $i, i+1$, because nearest-neighbor transpositions generate the group of permutations. (C) Use antisymmetry to deduce that there are no terms (on either side) of the form

$$C\prod_{i=1}^{n} x^{r_i}$$

with $r_i = r_j$ for some pair $i \neq j$. (D) Conclude that it suffices to check that the coefficients of $x_2x_3^2x_4^3\cdots x_n^{n-1}$ are the same on both sides. □

1.5. **Wishart Density.** Recall that the (real) Wishart distribution $W(\Sigma, n)$ is the distribution of the random symmetric matrix $S = XX^T$, where X is a $p\times n$ data matrix whose columns are independent, identically distributed $N(0,\Sigma)$. Similarly, the *complex* Wishart distribution $W_{\mathbb{C}}(I, n)$ is the distribution of the random *Hermitian* matrix $H = ZZ^*$, where Z is a $p\times n$ data matrix whose columns are independent, identically distributed *complex* $N_{\mathbb{C}}(0, I)$ (that is, the individual entries Z_{ij} are i.i.d. complex $N_{\mathbb{C}}(0,1)$).

Theorem 6. *Let* Σ *be a positive definite, symmetric* $p\times p$ *matrix and let* $n \ge p$. *If* $S \sim W(\Sigma, n)$ *then the joint density of the entries* $S_{i,i}, S_{i,j}$ *with* $j > i$ *relative to Lebesgue measure is*

$$f_{\Sigma,n}(S) = |\det(S)|^{(n-p-1)/2}\exp\{-\mathrm{Tr}\,(\Sigma^{-1}S)/2\}/C_{\Sigma,n} \tag{12}$$

for an appropriate normalizing constant $C_{\Sigma,n}$*:*

$$C_{\Sigma,n} = \sqrt{2}^{np}\sqrt{\pi}^{\binom{p}{2}}\det(\Sigma)^{n/2}\prod_{i=1}^{p}\Gamma((n+1-i)/2) \tag{13}$$

Proof for the special case $\Sigma = I$. We will assume it already known that the function $f_{I,n}$ defined by (12) is in fact a probability density (that is, that its integral against Lebesgue measure is finite). The strategy will be to show that the Laplace-Fourier transform of the density (12) is equal to the Laplace-Fourier transform of the random matrix S. Since S is symmetric, the $p+\binom{p}{2}$ entries below the main diagonal are redundant, and hence should be excluded. Thus, the Laplace-Fourier transform of S is defined to be the function

$$\varphi(\Theta) := E\exp\left\{\sum_i \Theta_{ii}S_{ii} + 2\sum_i\sum_{j>i}\Theta_{ij}S_{ij}\right\}$$

of the $p+\binom{p}{2}$ variables Θ_{ij}, where $j \geq i$. It is natural to think of this as being a function of the real symmetric matrix Θ with entries Θ_{ij} for $j \geq i$. With this convention,

$$\varphi(\Theta) = E\exp\{\text{Tr }(\Theta S)\} = E\exp\{\text{Tr }(X^T\Theta X)\}.$$

Since the matrix Θ is symmetric, the Spectral Theorem implies that it can be diagonalized. Thus, there is an orthogonal matrix U and a diagonal matrix D whose diagonal entries d_{ii} are the eigenvalues of Θ such that $\Theta = U^T DU$. Now recall that the columns of X are independent, identically distributed $p-$variate normal $N(0, I)$, and that multiplying a $N(0, I)$ random vector by an orthogonal matrix yields another $N(0, I)$ random vector. Hence, the random matrix UX has the same distribution as X, and so

$$\varphi(\Theta) = \varphi(D) = E\exp\{\text{Tr }(X^T DX)\}.$$

Finally, since D is diagonal,

$$\begin{aligned} E\exp\{\text{Tr }(X^T DX)\} &= E\exp\left\{\sum_{i=1}^{p}\sum_{j=1}^{n} d_{ii}X_{ij}\right\} = \prod_{i=1}^{p}(1-2d_{ii})^{-n/2} \\ &= \det(I-2D)^{-n/2} = \det(I-2\Theta)^{-n/2}, \end{aligned} \tag{14}$$

provided each $|d_{ii}| < 1/2$.

Next, we calculate the corresponding Laplace-Fourier transform of the density (12) (with $\Sigma = I$) relative to the Lebesgue measure $dS := \prod_i dS_{ii}\prod_{i<j} dS_{ij}$:

$$\psi(\Theta) := \int |\det(S)|^{(n-p-1)/2}\exp\{-\text{Tr } S/2\}\exp\{\text{Tr }(\Theta S)\}\, dS/C_{I,n}$$

As above, $\Theta = U^T DU$. By Proposition 2, the Lebesgue measure dS in invariant by conjugation with U; also, $\det(USU^T) = \det(S)$ and $\text{Tr }(USU^T) = \text{Tr } S$. Hence,

$$\psi(\Theta) = \psi(D) = \int |\det(S)|^{(n-p-1)/2}\exp\{-\text{Tr }((I-2D)S)/2\}\, dS/C_{I,n}$$

The matrix $A = (I-2D)$ is diagonal with diagonal entries $\alpha_{ii} = 1-2d_{ii}$. Assume that $d_{ii} < 1/2$, so that the entries $\alpha_{ii} > 0$. Now make the linear change of variables

$$\tilde{S}_{ij} = \sqrt{\alpha_{ii}\alpha_{jj}}S_{ij}.$$

Note that this change of variables is accomplished by first multiplying the ith row by $\sqrt{\alpha_{ii}}$, and then multiplying the jth column by $\sqrt{a_{jj}}$. Consequently,

$$\det(\tilde{S}) = \prod_{i=1}^{p}\alpha_{ii}\det(S) = \det(I-2D)\det(S).$$

Also, since the ith diagonal entry is multiplied by α_{ii},

$$\text{Tr }(I-2D)S = \text{Tr }(\tilde{S}).$$

Finally, consider the effect of the change of variables on the Lebesgue measure $dS = \prod_i dS_{ii}\prod_{i<j} dS_{ij}$: Each index $1 \leq k \leq p$ is listed once on the diagonal, $(p-k)$ times as a row index S_{kj} where $j > k$, and $k-1$ time as a column index S_{ik} with $i < k$. Each appearance as a diagonal entry gives a factor α_{kk}, and every other appearance gives a $\sqrt{\alpha_{kk}}$. Thus,

$$d\tilde{S} = \left(\prod_{i=1}^{p}\alpha_{ii}\right)^{(p+1)/2} d\tilde{S} = (\det(I-2D))^{(p+1)/2}dS.$$

It now follows that

$$\begin{aligned}\psi(\Theta) = \psi(D) &= \det(I-2D)^{-n/2} \int |\det(\tilde{S})|^{(n-p-1)/2} \exp\{-\mathrm{Tr}\, \tilde{S}/2\}\, d\tilde{S}/C_{I,n} \\ &= \det(I-2D)^{-n/2} \\ &= \det(I-2\Theta)^{-n/2} \\ &= \varphi(\Theta).\end{aligned} \tag{15}$$

□

Corollary 7. *If $S \sim W(I,n)$ where I is the $p \times p$ identity matrix, then the joint density of the eigenvalues λ_i of S is*

$$C_n^{-1} \prod_i \lambda_i^{(n-p-1)/2} e^{-\lambda_i/2} \prod_{i<j} |\lambda_i - \lambda_j| \tag{16}$$

for an appropriate normalizing constant C_n. Similarly, if $H \sim W_{\mathbb{C}(I,n)}$ then the joint density of the eigenvalues λ_i of H is

$$\tilde{C}_n^{-1} \prod_i \lambda_i^{(n-p)} e^{-\lambda_i} \prod_{i<j} |\lambda_i - \lambda_j|^2. \tag{17}$$

2. Proof of Weyl's Formula: GOE

For the sake of clarity we will consider only the special cases of the Gaussian orthogonal and unitary ensembles. (The general case isn't really any different, so the proof can be translated without much difficulty to a proof of the theorem in its full generality.) The orthogonal case is slightly easier than the unitary case, so we will focus on it.

2.1. **Preliminaries.** The Spectral Theorem implies that for every real, symmetric, $N \times N$ matrix $X = (X_{i,j})$ there exists an orthogonal matrix U and a real diagonal matrix $\Lambda = \mathrm{Diag}(\lambda_i)_{i \le N}$ such that

$$X = U \Lambda U^T. \tag{18}$$

This representation is not unique, even if we insist that the eigenvalues λ_i are listed in increasing order: for each representation (18) and each choice of signs $s_i = \pm 1$, if $S = \mathrm{Diag}(s_i)$ is the diagonal matrix with diagonal entries s_i then

$$X = US\Lambda S^T U^T$$

is another spectral representation of X. However, if the eigenvalues λ_i of X are distinct, then there are exactly 2^N representations (18) such that the eigenvalues are listed in increasing order along the diagonal of Λ, one for each choice of signs. If there are repeated eigenvalues then there are infinitely many representations (because there are infinitely many choices of orthonormal basis for each eigenspace of dimension > 1). Fortunately, this won't matter, because the following lemma implies that the event of repeated eigenvalues has probability zero.

Lemma 8. *The set of $\binom{N+1}{2}$ − tuples $(X_{i,j})_{j \ge i}$ such that the symmetric matrix $X = (X_{i,j})$ has repeated eigenvalues has Lebesgue measure zero.*

Proof. There are several ways to do this. The shortest is based on a basic theorem of polynomial algebra, according to which a polynomial $p(\lambda)$ of one variable has a multiple root if and only if its *discriminant*[3] vanishes. One way (of several) to represent the discriminant of a polynomial

[3]See the Wikipedia entry on discriminants for the relevant facts.

$p(x)=\sum_{j=0}^m a_j x^j$ is as the determinant of a matrix whose nonzero entries are linear functions of the coefficients a_j. This implies that the set of $(m+1)-$ vectors $(a_j)_{0\le j\le m}$ for which the polynomial $p(x)=\sum_{j=0}^m a_j x^j$ has multiple roots is the zero set of a polynomial in the $m+1$ variables. The lemma now follows from the following lemma. □

Lemma 9. *If $Q(x_1,x_2,\dots,x_k)$ is a polynomial in $k\ge 1$ variables that is not identically zero then the set of points in $\mathbb{R}^k$ at which Q vanishes has measure zero.*

Proof. Exercise. Hint: Use Fubini, together with the fact that a polynomial in *one* variable has only finitely many roots. □

The "spectral" coordinate system that we will use in the proof of Weyl's formula consists of the N eigenvalues λ_i and a special (local) parameterization of the space $O(N)$ of orthogonal matrices. One's first thought might be to use the matrix entries themselves to parametrize the orthogonal matrices; however, since the columns are constrained to be orthonormal, this becomes messy. So instead we will use the *matrix exponential map*, which we explain next. Recall that for an $N\times N$ matrix Q, the matrix exponential of Q is defined by

$$\exp(Q)=\sum_{k=0}^{\infty} Q^k/k!$$

It is routine to verify that $\exp(sQ)\exp(tQ)=\exp((s+t)Q)$ for any real numbers s,t and any square matrix Q, and it follows that $\exp(-Q)$ is the inverse matrix to $\exp(Q)$. You should check (by looking closely at the power series representation) that if

$$Q=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$$

then for any $t\in\mathbb{R}$,

$$\exp\{tQ\}=\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}.$$

Thus, the matrix exponential maps $\mathbb{R}$ onto the group $SO(2)$ of two-dimensional rotations with determinant 1. This mapping is not injective, but it is *locally* bijective.

The next result generalizes this result to higher dimensions. A square matrix Q is said to be *skew-symmetric* if $Q_{i,j}=-Q_{j,i}$ for each pair of indices i,j. Note that if Q is skew-symmetric then its diagonal entries are all 0. Recall that $SO(N)$ is the space of orthogonal matrix $U\in O(N)$ with determinant $+1$ (recall that every orthogonal matrix has determinant ± 1).

Proposition 10. *Let $Skew_N$ be the space of all skew-symmetric $N\times N$ real matrices. Then*

$$\exp: Skew_N \longrightarrow SO(N)$$

is a surjective, locally one-to-one mapping. Thus, $Skew_N$ is the Lie algebra *of the Lie group $SO(N)$.*

Proof. First, let's check that the exponential of any skew-symmetric matrix is an orthogonal matrix with determinant one. If Q is skew-symmetric, then for any vector $v\in\mathbb{R}^N$,

$$\langle Qv,v\rangle=-\langle v,Qv\rangle. \tag{19}$$

Consequently,

$$\begin{aligned}\langle \exp\{Q\}v,\exp\{Q\}v\rangle &= \langle \exp\{Q\}v,\exp\{-Q\}^T v\rangle\\ &= \langle \exp\{-Q\}\exp\{Q\}v,v\rangle\\ &= \langle v,v\rangle\end{aligned}$$

and so $\exp\{Q\}$ is an isometry. It follows that its determinant is ± 1. To see that it is $+1$, observe that $\det(\exp(tQ))$ varies continuously with t, and at $t = 0$ is $+1$. This proves that exp maps Skew_N into $SO(N)$.

Next, observe that the inner product relation (19) characterizes skew symmetry. To see this, apply (19) to both $v + w$ and $v - w$ and subtract to conclude that $\langle v, Qw\rangle = -\langle Qv, w\rangle$; using this for unit vectors $v = e_i$ and $w = e_j$ shows that if Q satisfies (19) for all v then it must be skew-symmetric. The nice thing about the inner product characterization is that it is coordinate-free: in particular, it shows that if Q is skew-symmetric then its matrix in any orthonormal basis is also skew-symmetric (that is, for any $U \in O(N)$ the matrix UQU^T is skew-symmetric).

Recall the spectral representation of an orthogonal matrix (see the Background Notes, Corollary 18): this implies that any orthogonal transformation $T \in SO(N)$ has, in a suitable orthonormal basis, a representation (in block form, where all but the last two blocks are 2×2)

$$\begin{pmatrix} R_{\theta_1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & R_{\theta_2} & 0 & \cdots & 0 & 0 & 0 \\ & & & \cdots & & & \\ 0 & 0 & 0 & \cdots & R_{\theta_k} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -I & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & I \end{pmatrix}$$

where the matrices R_{θ_i} are two-dimensional rotations. By the observation above, each of these can be represented as the exponential of a 2×2 skew matrix. This implies that any $T \in SO(N)$ is the exponential of an $N \times N$ skew matrix. Thus, the map exp is surjective.

Finally, to show that the exponential mapping is locally one-to-one it suffices to show that it is one-to-one on some neighborhood of the zero matrix in Skew_N (by group invariance). For this, we use the Inverse Function Theorem of advanced calculus, according to which it is enough to show that the matrix of partial derivatives at the zero matrix is invertible. Since a skew-symmetric matrix Q satisfies $Q_{ij} = -Q_{ji}$, the space Skew_N is parameterized by the above-diagonal entries $(q_{ij})_{j>i}$, of which there are $\binom{N}{2}$. Set

$$X = (x_{ij}) = \exp(Q) = \exp((q_{ij}))$$

and consider the above diagonal entries $(x_{ij})_{j>i}$ as a function of $(q_{ij})_{j>i}$. By the power series representation of exp,

$$\exp(Q) = I + Q + O(\|Q\|^2);$$

consequently, the partial derivatives at $Q = 0$ are

$$\begin{aligned} \frac{\partial x_{ij}}{\partial q_{i'j'}} &= 1 \quad \text{if } i = i' \quad \text{and } j = j', \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Thus, the Jacobian matrix is the identity matrix at $Q = 0$. □

2.2. **Gaussian Orthogonal Ensemble: Proof of Theorem 1.** A random matrix X chosen from the GOE always has a diagonalization with a complete set of real eigenvalues $\lambda_1, \dots, \lambda_N$, so the vector $\lambda = (\lambda_1, \dots, \lambda_N)$ has some distribution F_N^{GOE} on $\mathbb{R}^N$. (Recall that we have agreed to write the eigenvalues in increasing order, so their joint distribution is concentrated on the Weyl chamber of $\mathbb{R}^N$.) We don't yet know that this distribution has a density relative to Lebesgue measure, but later we will prove that it does.

We know, by Proposition 2, that the distribution of a random matrix chosen from the GOE is invariant under conjugation by any element of the orthogonal group $O(N)$, that is, if $X \sim GOE$ and $U \in O(N)$ then $UXU^T \sim GOE$. Consequently, if U is randomly chosen from the uniform distribution (i.e., Haar measure) on $O(N)$, independently of X, then $UXU^T \sim GOE$. Now X has a diagonalization $X = V\Lambda V^T$ with $V \in O(N)$, so $UXU^T = (UV)\Lambda(UV)^T$. But since U has the uniform distribution on $O(N)$, so does UV (proof: condition on V), and furthermore, UV is independent of X, and hence also of Λ. This proves

Lemma 11. *If Λ is a random diagonal matrix whose diagonal entries $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ are chosen according to the density $p(\lambda)$, and if U is uniformly distributed on $O(N)$ independently of Λ, then*

$$U\Lambda U^T \sim GOE. \tag{20}$$

The strategy for proving of Theorem 1 will be based on this fact. We will show that if $U \in O(N)$ is independent of Λ and uniformly distributed, and if $U\Lambda U^T \sim GOE$, then the joint distribution of the diagonal entries of Λ (which will be denoted by F_N^{GOE}) must have density proportional to

$$p_N^{GOE}(\lambda) \propto |\Delta_N(\lambda)| \exp\left\{-\frac{1}{2}\sum_{j=1}^N \lambda_j^2\right\}. \tag{21}$$

To do this we will compute the probability that $X = U\Lambda U^T$ is in a small (infinitesimal) neighborhood of the diagonal matrix

$$D = \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ & & & \cdots & 0 \\ 0 & 0 & 0 & \cdots & d_N \end{pmatrix}$$

in two different ways, first using the fact that $X \sim GOE$ and then using the fact that $\Lambda \sim F_N^{GOE}$ and U ~Uniform. This will yield an equality between the two distributions, which will then lead directly to (21).

The first calculation is easy: if B_ε are any open neighborhoods of D with diameters $< \varepsilon$ (relative to the usual rectangular coordinates x_{ij}) in the space of symmetric matrices, then as $\varepsilon \to 0$,

$$P\{X \in B_\varepsilon\} \sim \exp\left\{-\sum_{j=1}^N d_j^2/2\right\} \mathrm{vol}(B_\varepsilon), \tag{22}$$

by the definition of the GOE.

The remainder of the argument will consist of estimating the probability that $X \in B_\varepsilon$ when $X = U\Lambda U^T$, with $\Lambda \sim F_N^{GOE}$ and U ~Uniform on $O(N)$. For this we will rely on the parameterization of $SO(N)$ provided by the exponential mapping. Let $X = (x_{ij})_{i,j\le N}$ be a real, symmetric $N \times N$ matrix. Since $x_{ij} = x_{ji}$, the coordinates x_{ij} with $i > j$ are redundant, so the space of symmetric matrices is parameterized by the $\binom{N+1}{2}$ variables x_{ij} with $j \ge i$. Let $D = \mathrm{Diag}(d_i)_{i\le N}$ be a real diagonal matrix. Each symmetric matrix X near D has a spectral representation (not necessarily distinct) $X = U\Lambda U^T$ with U an orthogonal matrix near the identity and Λ a real diagonal matrix with diagonal entries λ_i close to the corresponding entries d_i of D. By Proposition 10, each orthogonal matrix can be written as the exponential of a skew-symmetric matrix; hence, every symmetric matrix X near Λ has a representation

$$X = \exp\{S\}\Lambda\exp\{-S\}$$

with S skew-symmetric and close to the zero matrix. Thus, we have a mapping

$$T : ((\lambda_i)_{i\leq N}; (s_{ij})_{j>i}) \mapsto (x_{ij})_{j\geq i}$$

from *spectral coordinates* to *rectangular coordinates.*

Lemma 12. *The Jacobian matrix of the transformation T at any $(\Lambda; S)$ with $S = 0$ is given by*

$$\frac{\partial x_{ii}}{\partial \lambda_j} = \delta_{ij}; \qquad \frac{\partial x_{ii}}{\partial s_{ij}} = 0; \tag{23}$$
$$\frac{\partial x_{ij}}{\partial \lambda_j} = 0; \qquad \frac{x_{ij}}{\partial s_{i'j'}} = \delta_{ii'}\delta_{jj'}(\lambda_i - \lambda_j).$$

Consequently, the absolute value of the determinant of the Jacobian is the van der Monde determinant $|\Delta_N(\lambda)|$.

Proof. The partial derivatives with respect to the variables λ_j are easy to verify, because along $S = 0$ the transformation T is just $X = \Lambda$. To compute the partial derivative with respect to the variable s_{ij}, we use the power series of the exponential mapping. For $i < j$ let E_{ij} be the $N \times N$ matrix with entries $E_{kl} = 0$ except when $k = i$ and $l = j$ or when $k = j$ and $l = i$, where the entries are ± 1. Consider the dependence of the rectangular coordinates on ε as $\varepsilon \to 0$ in the mapping

$$X = X(\varepsilon) = (I + \varepsilon E_{ij} + O(\varepsilon^2))\Lambda(I - \varepsilon E_{ij} + O(\varepsilon^2)).$$

The only coordinates x_{kl} that change with ε are those in the 2×2 sub-matrix with rows and columns indexed by i, j, so it enough to consider the case $i = 1, j = 2$. Matrix multiplication shows that

$$\left(I + \varepsilon\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + O(\varepsilon^2)\right)\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\left(I + \varepsilon\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + O(\varepsilon^2)\right)$$
$$= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} + \varepsilon\begin{pmatrix} 0 & -\lambda_1 \\ \lambda_2 & 0 \end{pmatrix} + \varepsilon\begin{pmatrix} 0 & \lambda_2 \\ -\lambda_1 & 0 \end{pmatrix} + O(\varepsilon^2).$$

The only terms of order ε are those in the off-diagonal slots. It follows that the partial derivatives with respect to the variables s_{ij} are given by (23). □

Now we turn to the asymptotic evaluation of the probabilities $P\{X \in B_\varepsilon\}$ for $X = U^T \Lambda U$, where U and Λ are independent random matrices, U with the uniform distribution on $O(N)$ and Λ a diagonal matrix whose diagonal entries λ_i are jointly distributed according to F_N^{GOE}. Let $D = \text{diag}(d_i)$ be a diagonal matrix with distinct diagonal entries d_i, and let B_ε be the open neighborhood of D specified (in rectangular coordinates) by

$$B_\varepsilon = \{(x_{ij}) : |x_{ij} - D_{ij}| < \varepsilon\}. \tag{24}$$

If ε is sufficiently small then $X \in B_\varepsilon$ only if its eigenvalues λ_i are distinct. (Recall that the eigenvalues of a symmetric matrix X vary continuously with the entries of X, by Proposition 23 of the Background Notes.)

When the eigenvalues λ_i of X are distinct there are 2^N different spectral representations $X = U\Sigma\Lambda\Sigma U^T$ with $\Lambda = \text{Diag}(\lambda_i)$ and λ_i in increasing order, one for each of the 2^N diagonal matrices Σ with ± 1 entries on the diagonal. If X lies in the neighborhood B_ε, for $\varepsilon > 0$ sufficiently small, then *one* of the 2^N spectral representations $U\Lambda U^T$ will be such that U is close to the identity matrix I and $\Lambda - D$ is small. Therefore, the inverse image of the neighborhood B_ε under the mapping $(\Lambda;U) \mapsto U^T\Lambda u$ will consist of 2^N non-overlapping components $K_j = K_j(\varepsilon;D)$, *all of which are congruent*. More precisely, if

$$K_I = \{(\Lambda;U)\}$$

is one of the components, then the others are

$$K_\Sigma = \{(\Lambda;U\Sigma)\}$$

where Σ ranges over the 2^N diagonal ± 1 matrices. Consequently, since the uniform distribution on $O(N)$ is invariant by "translations" $U \mapsto US$, if $\Lambda \sim F_N^{GOE}$ and U ~Uniform on $O(N)$, then

$$P((\Lambda;U) \in K_\Sigma) = P((\Lambda;U) \in K_I)$$

for all choices of S. Thus,

$$P\{U\Lambda U^T \in B_\varepsilon\} = 2^N P\{(\Lambda;U) \in K_I\}.$$

To complete the calculation we must estimate the probability $P\{(\Lambda;U) \in K_I\}$ for small $\varepsilon > 0$. There are two pertinent facts: first, the random matrices Λ and U are independent; second, the region $K_I = K_I(\varepsilon;D)$ is the inverse image of the rectangle B_ε under the transformation from spectral to rectangular coordinates. Since for small ε the region K_I is a small neighborhood of the identity, for each $(\Lambda;U) \in K_I$ there is a unique representation

$$U = \exp(S) \quad \text{where} \quad S \in \text{Skew}_N.$$

By Lemma 12, the transformation T from spectral to rectangular coordinates is approximately linear:

$$T \approx I_{N\times N} \times V_N$$

where V_N is the $\binom{N}{2} \times \binom{N}{2}$ diagonal matrix with diagonal entries $\lambda_j - \lambda_i$. Consequently, for small ε the region K_I is well-approximated (in spectral coordinates $(\Lambda;S)$) by a rectangle:

$$K_I \approx \{(\lambda_i)_{i\le N} : |\lambda_i - d_i| < \varepsilon\} \times \{(s_{ij})_{j>i} : |s_{ij}| < \varepsilon/|\lambda_j - \lambda_i|\}.$$

Hence, as $\varepsilon \to 0$,

$$P\{(\Lambda;U) \in K_I\} \sim C_N \varepsilon^{\binom{N}{2}} \prod_{i<j} |\lambda_i - \lambda_j| F_N^{GOE}\{\Lambda : \|\Lambda - D\| < \varepsilon\} \quad \Longrightarrow$$

$$P\{U\Lambda U^T \in B_\varepsilon\} \sim 2^N C_N \prod_{i<j} |\lambda_i - \lambda_j|^{-1} F_N^{GOE}\{\Lambda : \|\Lambda - D\| < \varepsilon\} \varepsilon^{\binom{N}{2}}$$

where $\|\cdot\|$ is the usual matrix norm and C_N is a constant that does not depend on D. By construction, $U^T\Lambda U$ has the same distribution as $X \sim GOE$; consequently, by equation (22) (and using the fact that $\text{vol}(B_\varepsilon) = (2\varepsilon)^{N(N+1)/2}$), we have

$$2^N C_N \prod_{i<j} |\lambda_i - \lambda_j|^{-1} F_N^{GOE}\{\Lambda : \|\Lambda - D\| < \varepsilon\} \sim \exp\{-\sum_{i=1}^N d_i^2/2\} 2^N \varepsilon^N.$$

It follows that as $\varepsilon \to 0$,

$$F_N^{GOE}\{\Lambda : \|\Lambda - D\| < \varepsilon\} \sim C_N^{-1} \exp\{-\sum_{i=1}^N d_i^2/2\} \prod_{i<j} |\lambda_i - \lambda_j| \varepsilon^N,$$

and this implies that F_N^{GOE} has density (21). □