MOMENTS OF RANDOM ORTHOGONAL MATRICES
18.338 FINAL REPORT

GREGORY MINTON

Abstract. In this report we study the moments of random orthogonal matrices. In particular, we give the results of an implementation of one method for computing these moments using Jack polynomials. We found limited success with this method: it appears to require significant tweaking to work in general, if indeed it is possible at all. We then discuss in more detail a particular type of polynomial in the entries of a random orthogonal matrix, a so-called "north pole" term. We review the main paper on this topic and then give a couple of new results. This section is by no means complete; we believe there to be significant low hanging fruit left in the topic, and are merely pointing at a few directions for future work.

This paper presents a summary of results from a somewhat dichotomous project. For the first part of this project, we implemented a particular method for computing moments of random orthogonal matrices. We will review this method in Section 1 and then discuss the results in Section 2. We will leave out some of the background for this discussion as much of the material was presented in our midterm report.

For the second part of the project, we looked at a particular type of moment; more specifically, we looked at the random function corresponding to the 1,1 entry of some power of a random orthogonal matrix. We shall refer to such entries as "north pole" terms. The moments of north pole terms are of course moments, in the sense used in the first part of this report, of random orthogonal matrices. We shall review the known results in Section 3 and then give a couple of new results in Section 4. We believe that many more results concerning north pole moments should be easily accessible, and look forward to pursing them once time permits.

1. Computing Moments Using Jack Polynomials: Background

1.1. **Definitions.** In this section we speedily review the method we will use to compute moments, and we fix some notation. For our computational results, we will exclusively work with real (orthogonal) matrices; but of course one could consider working over different real algebras as well.

(For example, moments of degree up to six are computed for arbitrary α in [10]. The problem of conjugation is the only issue keeping us from immediately generalizing our code, at least to $\mathbb{C}$: the case of unitary matrices over $\mathbb{R}$ is special in that each element equals its conjugate. For higher-dimensional Clifford algebras, one might need to worry about noncommutativity as well. But that is a topic for another report.)

We reserve the variable m for the dimensionality of our matrices, and will use n as a variable for indexing powers. We denote by $\mathcal{O}$ the group of $m \times m$ orthogonal matrices over $\mathbb{R}$. This is a compact topological group (in fact, a compact Lie group), and so is endowed with a Haar measure, the unique (up to scaling) finite regular Borel measure which is invariant under both left and right multiplication.

Using this measure, $\mathcal{O}$ becomes a probability space. (We shall always intend the Haar measure when we think of $\mathcal{O}$ as a probability space.) Suppose we sample a matrix

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \\ x_{2,1} & x_{2,2} & \\ & & \ddots \end{pmatrix}$$

from $\mathcal{O}$. (We will use X in the sequel to refer to a generic matrix sampled from $\mathcal{O}$.) Then any polynomial in the entries $x_{i,j}$ defines a function $\mathcal{O}$, and so we may speak of its expected value $\mathbb{E}[\cdot]$. We will say that the expected value of a monomial is a *moment* of $\mathcal{O}$; the expected value of any polynomial may of course be computed if one knows all of the moments. We define the degree of a moment in the obvious way, as the degree of the underlying polynomial in the variables $x_{i,j}$.

1.2. **Other Methods.** We will use one particular method, based on Jack polynomials, for computing moments. This is a method which computes moments "all at once" by finding a system of equations in the moments. Before discussing this method, though, we should note that it is not the only method available. (In fact, as we shall see, it is perhaps not even a complete method at all!) The standard method for computing moments relies on the Weingarten formula; see for example [1] for a discussion of this. To quickly summarize, the Weingarten formula reduces the computation of moments to an object known as the Weingarten function W_{kn}. The Weingarten function is a combinatorially defined object; under one definition, it the pseudo-inverse of the Gram matrix, which is written in terms of the number of loops obtained by superimposing two pairings of the set $\{1, 2, \dots, 2k\}$ (where $2k$ is the degree of the moment one is interested in). There are asymptotic formulas for W_{kn}, and its efficient computation is apparently a topic of ongoing study. (In fact, one recent result is a formula for W_{kn} which uses, amongst other ingredients, the zonal functions!)

Another method is given in [6]. Like the Weingarten formula, this is a method for computing a given moment directly (i.e. it is not an "all at once" method). It relies on reducing a desired moment to simpler moments, each of whose entries lie in a single row/column of X. These moments can be computed directly, as a row/column of X is just a random vector on the sphere. This method is likely similar in nature to the method using Householder reflections currently under consideration by A. Edelman. Further study of this paper seems warranted.

1.3. **Symmetric Polynomials.** To describe the method, we need to introduce Jack polynomials; to do this, we first need to introduce symmetric polynomials. Recall that a symmetric polynomial $\sigma(x_1, \dots, x_m)$ is a polynomial which is invariant under permutation of the variables x_i. For example, $x_1 + x_2 + x_3$ and $x_1 x_2 x_3$ are symmetric polynomials in x_1, x_2, x_3, but are not symmetric polynomials if the variable set is x_1, x_2, x_3, x_4.

The set of symmetric polynomials with a given degree d in a given number of variables m is a finite dimensional vector space, and its dimension is given by the number of partitions of d into at most m parts. (We shall take the definition and notation of partitions as understood.) One basis for this space is provided by the monomial symmetric polynomials m_λ, where $\lambda \vdash d$ (and λ has at most m parts). If $\lambda = (\lambda_1, \dots, \lambda_m)$, then m_λ is the sum of all monomials related to $x_1^{\lambda_1} \cdots x_m^{\lambda_m}$ by a permutation. For example,

$$\begin{aligned} m_{(3)}(x, y, z) &= x^3 + y^3 + z^3 \\ m_{(2,1)}(x, y, z) &= x^2 y + x^2 z + y^2 x + y^2 z + z^2 x + z^2 y \\ m_{(1,1,1)}(x, y, z) &= xyz. \end{aligned}$$

The Jack polynomials will provide another basis for this space.

One particularly simple degree-d symmetric polynomial is the d^{th} power sum

$$p_d(x_1, \dots, x_m) = x_1^d + \cdots + x_m^d.$$

The set $\{p_d\}$ over all d generates the ring of symmetric polynomials as an algebra over $\mathbb{R}$. (In other words, every symmetric polynomial may be written as a sum of products of power sums.) For example, we have

$$x_1x_2 + x_1x_3 + x_2x_3 = \frac{1}{2}(x_1 + x_2 + x_3)^2 - \frac{1}{2}(x_1^2 + x_2^2 + x_3^2).$$

Before moving on, we should extend the definition of a symmetric polynomial $\sigma(x_1, \ldots, x_m)$. Given a polynomial f of degree m, f has m roots (in some extension field, possibly with repetition). We can thus evaluate σ on those roots; there is no preferred ordering of the roots, but σ does not require an ordering for its input variables. This defines a way to evaluate σ on a polynomial; overloading notation, we will write this as $\sigma(f)$. To actually compute $\sigma(f)$, we could first write σ in terms of power sums. The power sums of the roots of a polynomial are related to the coefficients of the polynomial by the Newton-Girard identities. Thus we can write $\sigma(f)$ as a polynomial in the coefficients of f; solving for the roots is not required.

Similarly, given an $m \times m$ matrix A, A has m eigenvalues (in some extension field, possibly with repetition, using algebraic multiplicities). We can thus evaluate σ on these eigenvalues; further overloading notation, we will write this as $\sigma(M)$. Note that this is the same as evaluating σ on the characteristic polynomial of A. Hence $\sigma(M)$ may be computed in terms of the entries of M, without solving for the eigenvalues. Note that evaluating power sums on matrices is particularly easy: for any d, $p_d(M) = \mathrm{Tr}(M^d)$. Thus we can compute $\sigma(M)$ for any σ in terms of just traces of powers of M.

1.4. **Jack Polynomials.** The Jack polynomials, as already mentioned, are just another basis for the space of symmetric polynomials (in m variables, of degree d). Actually, there is a one-parameter family of Jack polynomials, controlled by the parameter α. The case $\alpha = 2$ corresponds to the real case that we are interested in.

The Jack polynomials are indexed by the partitions of degree d (with at most m parts). For such a partition κ, we denote by J_κ^α the corresponding Jack polynomial.

There are a variety of definitions for the Jack polynomials: depending on one's point of view, they might arise as an orthogonal set of polynomials under a certain inner product, or they might be thought of as eigenfunctions of a Laplace-Beltrami operator, or they might come about in any number of other ways.

For computing the Jack polynomials, we shall use the second definition above. Following [2] (which in turn follows [12]), the relevant Laplace-Beltrami operator is

$$D^* = \sum_{i=1}^{m} x_i^2 \frac{\mathrm{d}^2}{\mathrm{d}x_i^2} + \frac{2}{\alpha} \sum_{i \neq j} \frac{x_i^2}{x_i - x_j} \frac{\mathrm{d}}{\mathrm{d}x_i}.$$

(Note that, while it may not be manifest in this formula, D^* is indeed a differential operator on the space of symmetric polynomials. It is not an operator on the space of all polynomials.)

We expand J_κ^α in the m_λ basis as

$$J_\kappa^\alpha(x_1, \ldots, x_m) = \sum_\lambda c_{\kappa,\lambda}^\alpha m_\lambda(x_1, \ldots, x_m)$$

for some coefficients $c_{\kappa,\lambda}^\alpha$. To determine J_κ^α, we thus just need to determine the $c_{\kappa,\lambda}^\alpha$. We can do this using the eigenfunction condition; the following result is obtained in [2].

Theorem 1. *Let $\kappa = (k_1, \ldots, k_m)$. Define*

$$\rho_\kappa^\alpha = \sum_{i=1}^{m} k_i(k_i - 1 - \frac{2}{\alpha}(i-1)).$$

Then for any $\lambda = (l_1, \ldots, l_m)$,

$$c_{\kappa,\lambda}^\alpha = \frac{2/\alpha}{\rho_\kappa^\alpha - \rho_\lambda^\alpha} \sum_{\lambda \prec \mu \preceq \kappa} ((l_i + t) - (l_i - t))c_{\kappa,\mu}^\alpha,$$

where μ is defined by $\mu = (l_1, \ldots, l_i + t, \ldots, l_j - t, \ldots, l_m)$, and i, j, t vary over all possibilities with $i < j$ and $t \leq l_j$.

A few words about the interpretation of Theorem 1 is in order. First, the relation $\prec$ on partitions is the dominance relation, defined as follows.

Definition 2. *Let $\lambda = (l_1, \ldots, l_m)$ and $\kappa = (k_1, \ldots, k_m)$ be two partitions; we say that $\lambda \preceq \kappa$ if, for all $j < m$, $\sum_{i=1}^{j} l_i \leq \sum_{i=1}^{j} k_i$, and $\sum_{i=1}^{m} l_i = \sum_{i=1}^{m} k_i$. (Thus λ and κ are partitions of the same integer.) If $\lambda \preceq \kappa$ and $\lambda \neq \kappa$, we say that $\lambda \prec \kappa$.*

The partition μ could also use some more description. The summation in Theorem 1 is not taken over all μ between λ and κ with the appropriate form: it is a *weighted* sum over all such partitions, where we weight by the number of i, j, t pairs that produce it. Also, for a generic choice of i, j, t, $(l_1, \ldots, l_i + t, \ldots, l_j - t, \ldots, l_m)$ is probably no longer in weakly decreasing order. We define μ by sorting this list (and in particular, we test $\lambda \prec \mu \preceq \kappa$ using the sorted list).

For example, taking $m = 3$, $\kappa = (3)$, and $\alpha = 2$, Theorem 1 gives us the following equations.

$$\begin{aligned} c^2_{(3),(2,1)} &= \frac{3}{5} c^2_{(3),(3)} \\ c^2_{(3),(1,1,1)} &= \frac{2}{3} c^2_{(3),(2,1)} \end{aligned}$$

This system is underdetermined by one; we need to specify one more equation to be able to solve for all the coefficients. This remaining equation is determined by normalization. (Recall that we are thinking of the Jack polynomials as eigenfunctions of an operator: eigenfunctions are always defined up to scale.) There are three typical normalizations for Jack polynomials, the "C", "J", and "P" normalizations. (See Table 5 in [2].) We will not care about normalization, so we will just set $c^\alpha_{\kappa,\kappa} = 1$ for all κ for a simple normalization. (We shall write J^α_κ for the resulting polynomial, but strictly speaking it is a multiple of J^α_κ.)

In our example, we thus get

$$c^2_{(3),(3)} = 1, \qquad c^2_{(3),(2,1)} = 3/5, \quad \text{and} \quad c^2_{(3),(1,1,1)} = 2/5,$$

and so

$$\begin{aligned} J^2_{(3)}(x_1, x_2, x_3) &= m_{(3)} + \frac{3}{5} m_{(2,1)} + \frac{2}{5} m_{(1,1,1)} \\ &= (x_1^3 + x_2^3 + x_3^3) + \frac{3}{5}(x_1^2 x_2 + x_1^2 x_3 + x_2^3 x_1 x_2^3 x_3 + x_3^2 x_1 + x_3^2 x_2) + \frac{2}{5}(x_1 x_2 x_3). \end{aligned}$$

It is proven in [2] that the equations we get from Theorem 1 are linearly independent, so after choosing normalization we always obtain a full system of equations which we can solve for the coefficients $c^\alpha_{\kappa,\lambda}$.

1.5. **Using Jack Polynomials.** To see how the Jack polynomials are useful for computing moments, it is useful to consider an alternate formulation of them. Consider the following action of $\mathrm{GL}(m)$ on the real symmetric positive definite matrices:

$$L \in \mathrm{GL}(m), X \text{ symmetric, } m \times m \implies X \mapsto X' = LXL^{\mathrm{T}}.$$

We could instead think of $\mathrm{GL}(m)$ as acting on the space of homogeneous polynomials (in m variables, of some degree d) by $\phi(X) \mapsto \phi'(X) = \phi(L^{-1}X(L^{-1})^{\mathrm{T}})$. Let us denote this space by V.

This is a well defined group action; it natural to examine its irreducible subspaces. It turns out that V decomposes (over the algebraic closure, i.e. over $\mathbb{C}$) into irreducible subspaces indexed by the partitions of d into at most m parts. It further turns out that each such subspace contains a one-dimensional subspace of polynomials invariant under the action of the orthogonal group $\mathcal{O}$. These one-dimensional subspaces are precisely the span of the zonal polynomials, which are the $\alpha = 2$ special case of the Jack polynomials. (See [8] for more details on this.)

With this in mind, it is natural to think that the Jack polynomials should play a special role in understanding the behavior of the orthogonal group. Indeed they do, as exemplified by the following theorem, paraphrased from [7].

Theorem 3. *Let X denote a random matrix sampled from $\mathcal{O}$. Let A and B be any $m \times m$ symmetric matrices. Let $J^2_\kappa(x_1, \ldots, x_m)$ be a Jack polynomial (i.e. a zonal polynomial). Then*

$$\mathbb{E}[J^2_\kappa(AXBX^T)] = \frac{J^2_\kappa(A)J^2_\kappa(B)}{J_\kappa(I)},$$

where I is the $m \times m$ identity matrix.

(One could generalize Theorem 3 to different α, i.e. to different real algebras. This just requires replacing the transpose with the appropriate dual, for example the conjugate-transpose in the complex case.)

We will use Theorem 3 to compute moments of random orthogonal matrices. (Note that, as promised, Theorem 3 is obviously independent of the choice of normalization for J^α_κ.)

Our method for using Theorem 3 is rather straightforward. For any fixed symmetric matrices A and B, and any choice of Jack polynomial J^2_κ (of degree n, say), Theorem 3 gives us an equation satisfied by the degree $2n$ moments. Instead of probing with specific matrices A and B, it makes more sense to work with matrices A and B of indeterminates.

We demonstrate this by example. Take $m = 2$ and consider the Jack polynomial $J^2_{(1)}(x, y) = x + y$. To compute $J^2_{(1)}(x, y)$ applied to a matrix, we just take the trace. If we take

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix},$$

then Theorem 3 yields

$$\mathbb{E}\left[\mathrm{Tr}\left[\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{pmatrix}\begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix}\begin{pmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \end{pmatrix}\right]\right]$$

$$= \mathrm{Tr}\left[\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\right]\mathrm{Tr}\left[\begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix}\right] / \,\mathrm{Tr}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right].$$

This simplifies to the equation

$$ac\mathbb{E}[x_{1,1}^2] + ad\mathbb{E}[x_{1,2}^2] + bc\mathbb{E}[x_{2,1}^2] + bd\mathbb{E}[x_{2,2}^2] = ac/2 + ad/2 + bc/2 + bd/2$$

which must hold for all a, b, c, d. Hence we in fact have a polynomial identity in the variables a, b, c, d; equating coefficients, we find

$$\mathbb{E}[x_{1,1}^2] = 1/2 \quad , \quad \mathbb{E}[x_{1,2}^2] = 1/2 \quad , \quad \mathbb{E}[x_{2,1}^2] = 1/2 \quad , \quad \mathbb{E}[x_{2,2}^2] = 1/2.$$

2. Computing Moments Using Jack Polynomials: Results

2.1. **The Program.** We implemented a computer program which (1) computes the Jack polynomials, (2) determines polynomial equations of the form described above, (3) equates coefficients to get equations in the moments, and (4) solves these equations to get values for the moments. This application was written in `C++`, and does not rely on any external libraries. Internal arithmetic is all exact, working with rational numbers. As we shall describe below, there were a variety of tweaks made to the basic method in order to get better results, and the program reflects all these.

It contains subroutines for computing Jack polynomials for any α, but the $\alpha = 2$ (real) case is the only one relevant for computing moments as the program does not understand the distinction between elements and their conjugates.

A significant amount of the time of this project went into software development; while this is not the right venue for an in-depth discussion of the details, the implementation of this code was a nontrivial task.

2.2. **No Reductions.** The method as described gives one polynomial equation for each Jack polynomial, which then yields a system of equations in the moments by equating coefficients. It is clear that these equations separate monomials by degree, and that they only treat moments of even degree. If one wants moments of degree $2n$, then one should look at the Jack polynomials of degree n.

For example, take $m = 3$ (we will use $m = 3$ almost exclusively below) and $n = 3$, so that we are looking for moments of degree 6. As there are 9 variables $x_{i,j}$, there are

$$\left(\!\!\left(\begin{array}{c} 9 \\ 6 \end{array}\right)\!\!\right) = \left(\begin{array}{c} 15 \\ 6 \end{array}\right) = 3003$$

moments of degree 6 to consider.

There are three Jack polynomials in play: $J_{(3)}^2$, $J_{(2,1)}^2$, and $J_{(1,1,1)}^2$. Each one gives some number of equations in the moments; the number of linearly independent equations obtained is the rank of the system. We have determined every moment if and only if our system has full rank (3003, in this case).

It turns out that, in every case we ran, the equations obtained by different Jack polynomials are linearly independent, so that the total rank is the sum of the ranks obtained from each polynomial separately. (We expect that this is true in general, and suspect that it may be easy to prove.)

In the example above, we used (general) diagonal matrices for A and B. One could instead use a (general) symmetric matrix, which would have more free variables and thus might yield more equations. Indeed it does. We tried computing the moment equations in three different ways: taking both A and B to be diagonal, taking A to be diagonal and taking B to be symmetric, and taking A and B to both be symmetric. The following table lists the ranks of the system obtained from each Jack polynomial in each case.

	A,B diagonal	A diagonal, B symmetric	A,B symmetric
$J^2_{(3)}$	100	280	784
$J^2_{(2,1)}$	49	189	729
$J^2_{(1,1,1)}$	1	1	1

Note that we do indeed get more equations from using general symmetric matrices. But even in this case, the total rank of the system is $784 + 729 + 1 = 1514$, which is definitely not the full rank 3003 we want.

2.3. **Knowledge of Zeros.** Of the 3003 moments we are looking for, many of them will be zero. This is because of the following theorem.

Theorem 4. *Suppose $p(X)$ is a monomial in the entries of X such that some row (or column) appears an odd number of times. (For example, $p(X) = x_{1,1}x_{1,2}$ has the first and second columns both appearing once.) Then $\mathbb{E}[p(X)] = 0$.*

The proof of Theorem 4 is quite easy: using invariance of Haar measure under left (respectively, right) multiplication, the Haar measure must be invariant under negation of a given row (respectively, column). Hence any monomial which is negated by such an operation must have zero expectation.

Note that Theorem 4 allows us to immediately discount moments of odd degree, so we need not be concerned by the fact that Theorem 3 can only tell us about moments of even degree. It also drastically reduces the number of moments we need to consider: in the $m = n = 3$ case we are considering, the number of moments decreases from 3003 to 252.

The following table lists the ranks of the systems obtained when we set all of the other $3003 - 252$ moments to zero.

	A,B diagonal	A diagonal, B symmetric	A,B symmetric
$J^2_{(3)}$	100	100	100
$J^2_{(2,1)}$	49	63	81
$J^2_{(1,1,1)}$	1	1	1

Even after reducing the dimensionality of our problem greatly, we still have not solved it completely: even when we use general symmetric matrices for both A and B, the total rank is $100 + 81 + 1 = 182$, which is less than the 252 we are after.

2.4. **Knowledge of Permutations.** We can further reduce our problem by noting that the rows and columns of a random symmetric matrix are symmetric; so, for example, $\mathbb{E}[x^2_{1,1}]$ should equal $\mathbb{E}[x^2_{i,j}]$ for any i, j. To state this formally, we have the following.

Theorem 5. *Suppose $p(X)$ and $q(X)$ are monomials in the entries of X which are related by some permutation of rows and/or columns. Then $\mathbb{E}[p(X)] = \mathbb{E}[q(X)]$.*

Like Theorem 4, Theorem 5 follows easily from the left and right invariance of Haar measure.

Using this, we can reduce the number of free moments we need to find in our $m = n = 3$ case from 252 down to just 15. The following table lists the ranks of the systems of equations after this reduction.

	A,B diagonal	A diagonal, B symmetric	A,B symmetric
$J^2_{(3)}$	9	9	9
$J^2_{(2,1)}$	4	4	4
$J^2_{(1,1,1)}$	1	1	1

With the knowledge of zeros and permutation invariance added, it is no longer beneficial to use general symmetric matrices instead of diagonal matrices in this case. However, we still have not determined all of the moments: the rank is $9 + 4 + 1 = 14$, which is short by one from the full rank of 15.

And further, before one begins to think that the generality of symmetric matrices is not needed anymore, the fact that diagonals give as much information as symmetric polynomials breaks down in the next highest degree, $n = 4$ (so moments of degree 8). In this case, there are 43 moments to find after applying our reductions. There are four Jack polynomials we have at our disposal; the ranks of the corresponding systems are given below.

	A,B diagonal	A,B symmetric
$J^2_{(4)}$	16	16
$J^2_{(3,1)}$	9	16
$J^2_{(2,2)}$	4	4
$J^2_{(2,1,1)}$	1	1

So we see that in this case, we are even farther from full rank: also, the generality of symmetric matrices is again useful.

2.5. **Lower-Degree Monomials.** All of the moments for degree up to 3 were obtained in [10] by means of the following technique: one can leverage lower-degree monomials to get more equations on higher-degree monomials.

This works as follows. Supppose we want to compute moments with $m = n = 3$, i.e. moments of degree 6, as above. Suppose we happen to already know several moments of degree 4, for example

$$\mathbb{E}[x_{1,1}^4] = 1/5.$$

As X is an orthogonal matrix, there are several quadratic equations in its entries which are always satisfied: these are the orthogonality relations,

$$\sum_{k=1}^{m} x_{k,i}x_{k,j} = \delta_{i,j} \quad \text{for all } 1 \leq i, j \leq m.$$

Thus, for example, we have the equation $x_{1,1}^2 + x_{2,1}^2 + x_{3,1}^2 = 1$. Using this, we see

$$1/5 = \mathbb{E}[x_{1,1}^4] = \mathbb{E}[x_{1,1}^4(x_{1,1}^2 + x_{2,1}^2 + x_{3,1}^2)] = \mathbb{E}[x_{1,1}^6] + \mathbb{E}[x_{1,1}^4 x_{2,1}^2] + \mathbb{E}[x_{1,1}^4 x_{3,1}^2].$$

This is an equation in the moments of degree 6.

In general, if we know all of the moments of degree $2(n-1)$, and we use the orthogonality relations for each i, j, we get a large family of equations satisfied by the moments of degree $2n$.

Starting with these equations, we are able to obtain all 15 moments of degree 6, using only diagonal matrices A and B. This reproduces the corresponding results of [10].

If one then looks at moments of degree 8 and tries using diagonal matrices A and B, then one obtains in total a system of rank 40, which does not determine all 43 moments. However, if one instead uses general symmetric matrices, one does obtain all 43 moments. (It is unfortunate that diagonal matrices do not seem to provide all the information one needs, as they do afford a substantial improvement in speed.)

If one then steps up to moments of degree 10, i.e. to $n = 5$, then there are 113 free moments to determine. Using all of the reductions above, general symmetric matrices, and all of the moments for $n = 4$, one obtains a system of rank 109 in these moments. Hence, we are still stuck short of computing all of the moments.

2.6. **Transpose Symmetry.** We considered one additional reduction in the number of moments: the Haar measure also enjoys invariance under the transpose operation, so the expectations of two monomials which are related by a transpose must be equal.

Applying this symmetry, the numbers of free moments to determine in the cases of $n = 1, 2, 3, 4, 5$ are as follows.

$$n = 1: \ 1 \qquad n = 2: \ 4 \qquad n = 3: \ 11 \qquad n = 4: \ 29 \qquad n = 5: \ 71.$$

However, using all of our reductions as well as the lower-degree moments, our code is only able to determine a system of rank 70 for the $n = 5$ case. Thus we still do not have all the information needed to compute all moments in all cases.

2.7. **Saving Time.** For anyone wishing to experiment with these ideas, the following may be useful.

Claim 6. *Every equation obtained by using any Jack polynomial of degree n in Theorem 3 can be obtained at once by just using the symmetric polynomial $p_n = x_1^n + \cdots + x_m^n$, which corresponds to* $\mathrm{Tr}(X^n)$.

The interpretation of Claim 6 is as follows. As the Jack polynomials provide a basis for the space of symmetric polynomials (of degree n, in m variables) we can certainly write p_n in this basis. Suppose $p_n = \sum a_\kappa J_\kappa^2$. Then, taking the corresponding linear combination of the equations given to us by Theorem 3, we see that

$$\mathbb{E}[\mathrm{Tr}[(AXBX^{\mathrm{T}})^n]] = \mathbb{E}[p_n(AXBX^{\mathrm{T}})] = \sum_\kappa a_\kappa \frac{J_\kappa^2(A) J_\kappa^2(B)}{J_\kappa(I)}.$$

To compute the right side, we need only to deal with powers of A and B, which are polynomials in relatively few variables. Computing symmetric polynomials of $AXBX^{\mathrm{T}}$ is where most of the time is spent in the program, as this is a complicated matrix containing all of the variables. Thus it is a (large) performance win if one can only compute the trace of a single power, instead of needing all of the Jack polynomials.

We suspect that, just like the independence of the equations provided by different Jack polynomials, Claim 6 is probably easy to prove. (In fact, it probably has the same proof.)

When preparing this document, we realized that perhaps an improvement on Claim 6 might be possible; perhaps the following is true.

Claim 7. *Every equation obtained by using any Jack polynomial of degree n in Theorem 3 can be obtained at once by just using the symmetric polynomial $(x_1 + \cdots + x_m)^n$, which corresponds to* $\mathrm{Tr}(X)^n$.

This is believable as, using one normalization for the Jack polynomials (the "C" normalization), the sum of the Jack polynomials of degree n is $(\mathrm{Tr}\, X)^n$. If true, Claim 7 would provide a nice performance boost over Claim 6, as it would further reduce the amount of work we would have to do with the matrix $AXBX^{\mathrm{T}}$.

2.8. **Remarks on the Results.** The full results for all moments for $n = 1, 2, 3, 4$ are presented in Figure 8. (We present only a set of independent moments after using permutation and transpose invariance and removing the moments which are zero by Theorem 4.)

$n = 1:$

$$\mathbb{E}[x_{3,3}^2] = 1/3$$

$n = 2:$

$$\begin{aligned}
\mathbb{E}[x_{3,3}^4] &= 1/5 \\
\mathbb{E}[x_{3,2}^2 x_{3,3}^2] &= 1/15 \\
\mathbb{E}[x_{2,3}^2 x_{3,2}^2] &= 2/15 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,2} x_{3,3}] &= -1/30
\end{aligned}$$

$n = 3:$

$$\begin{aligned}
\mathbb{E}[x_{3,3}^6] &= 1/7 \\
\mathbb{E}[x_{3,2}^2 x_{3,3}^4] &= 1/35 \\
\mathbb{E}[x_{3,1}^2 x_{3,2}^2 x_{3,3}^2] &= 1/105 \\
\mathbb{E}[x_{2,3}^2 x_{3,2}^2 x_{3,3}^2] &= 2/105 \\
\mathbb{E}[x_{2,3}^2 x_{3,2}^4] &= 3/35 \\
\mathbb{E}[x_{2,3}^2 x_{3,1}^2 x_{3,2}^2] &= 1/35 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,2} x_{3,3}^3] &= -1/70 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,1}^2 x_{3,2} x_{3,3}] &= -1/210 \\
\mathbb{E}[x_{1,3}^2 x_{2,2}^2 x_{3,1}^2] &= 8/105 \\
\mathbb{E}[x_{1,3}^2 x_{2,1} x_{2,2} x_{3,1} x_{3,2}] &= -1/42 \\
\mathbb{E}[x_{1,2} x_{1,3} x_{2,1} x_{2,3} x_{3,1} x_{3,2}] &= 1/105
\end{aligned}$$

$n = 4:$

$$\begin{aligned}
\mathbb{E}[x_{3,3}^8] &= 1/9 \\
\mathbb{E}[x_{3,2}^2 x_{3,3}^6] &= 1/63 \\
\mathbb{E}[x_{3,2}^4 x_{3,3}^4] &= 1/105 \\
\mathbb{E}[x_{3,1}^2 x_{3,2}^2 x_{3,3}^4] &= 1/315 \\
\mathbb{E}[x_{2,3}^2 x_{3,2}^2 x_{3,3}^4] &= 2/315 \\
\mathbb{E}[x_{2,3}^2 x_{3,2}^4 x_{3,3}^2] &= 1/105 \\
\mathbb{E}[x_{2,3}^2 x_{3,2}^6] &= 4/63 \\
\mathbb{E}[x_{2,3}^2 x_{3,1}^2 x_{3,2}^2 x_{3,3}^2] &= 1/315 \\
\mathbb{E}[x_{2,3}^2 x_{3,1}^2 x_{3,2}^4] &= 4/315 \\
\mathbb{E}[x_{2,3}^4 x_{3,2}^4] &= 2/35 \\
\mathbb{E}[x_{2,3}^4 x_{3,1}^2 x_{3,2}^2] &= 2/105 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,2} x_{3,3}^5] &= -1/126 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,2}^3 x_{3,3}^3] &= -1/210 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,1}^2 x_{3,2} x_{3,3}^3] &= -1/630 \\
\mathbb{E}[x_{2,2} x_{2,3} x_{3,1}^4 x_{3,2} x_{3,3}] &= -1/630 \\
\mathbb{E}[x_{2,2} x_{2,3}^3 x_{3,2}^3 x_{3,3}] &= -1/140 \\
\mathbb{E}[x_{2,2} x_{2,3}^3 x_{3,1}^2 x_{3,2} x_{3,3}] &= -1/420 \\
\mathbb{E}[x_{2,2}^2 x_{2,3}^2 x_{3,2}^2 x_{3,3}^2] &= 1/210 \\
\mathbb{E}[x_{2,2}^2 x_{2,3}^2 x_{3,1}^2 x_{3,3}^2] &= 1/210 \\
\mathbb{E}[x_{2,1} x_{2,2} x_{2,3}^2 x_{3,1} x_{3,2} x_{3,3}^2] &= 0 \\
\mathbb{E}[x_{1,3}^2 x_{2,3}^2 x_{3,1}^2 x_{3,2}^2] &= 2/315 \\
\mathbb{E}[x_{1,3}^2 x_{2,2} x_{2,3} x_{3,1}^2 x_{3,2} x_{3,3}] &= -1/1260 \\
\mathbb{E}[x_{1,3}^2 x_{2,2}^2 x_{3,1}^2 x_{3,3}^2] &= 1/90 \\
\mathbb{E}[x_{1,3}^2 x_{2,2}^2 x_{3,1}^4] &= 17/315 \\
\mathbb{E}[x_{1,3}^2 x_{2,1} x_{2,2} x_{3,1} x_{3,2} x_{3,3}^2] &= -1/315 \\
\mathbb{E}[x_{1,3}^2 x_{2,1} x_{2,2} x_{3,1} x_{3,2}^3] &= -13/1260 \\
\mathbb{E}[x_{1,3}^4 x_{2,1} x_{2,2} x_{3,1} x_{3,2}] &= -11/630 \\
\mathbb{E}[x_{1,2} x_{1,3} x_{2,1} x_{2,3} x_{3,1} x_{3,2} x_{3,3}^2] &= 1/630 \\
\mathbb{E}[x_{1,2} x_{1,3} x_{2,1} x_{2,3} x_{3,1} x_{3,2}^3] &= 1/252
\end{aligned}$$

FIGURE 8. The moments for $m = 3$, $n = 1, 2, 3, 4$.

Perhaps more useful than this list of moments is the Javascript program we wrote which allows the user to input a polynomial in the entries of a random orthogonal 3×3 matrix (of degree at most 8) and have the program compute its expectation.

This program is available at

`http://math.mit.edu/~gminton/momentinterface.html` and has what is (hopefully) a usable interface. For convenience, in addition to the entries `X(i,j)` of the matrix itself, one may also use the entries `Xn(i,j)` of X^n.

(As a word of warning, the input syntax is rather strict. In particular, no implied multiplication is supported: one must always use an asterisk to denote multiplication, even by a constant. Integer constants are supported, but fractional constants are not.)

The obvious question in closing this section is the following: is it possible to add enough information to the system to compute all of the moments in general? And, even if so, is it possible to *prove* that one will get every moment?

Looking for additional conditions to be added to the system, one is guided by the results in Figure 8. If we had fully characterized all of the symmetries of the problem, we might expect that the free monomials left to find the expectations for should all have different expected value: indeed, if two monomials have the same expected values, then maybe we should have been able to tell that a priori. We certainly do not live up to that goal in our results: for example, $\mathbb{E}[x_{3,2}^2 x_{3,3}^4] = \mathbb{E}[x_{2,3}^2 x_{3,1}^2 x_{3,2}^2] = 1/35$, but these monomials were not identified as having the same expectation by any of our rules. Is there a rule which shows that these two should have the same expectation?

A less ambitious goal would be to identify a priori all moments which are zero. We come very close to meeting this goal: the only moment in the table which is zero, but which we did not identify as such by Theorem 4, is $\mathbb{E}[x_{2,1}x_{2,2}x_{2,3}^2 x_{3,1}x_{3,2}x_{3,3}^2]$. Is there some reason why we should be able to know in advance that this expectation is zero?

(An observation which is somewhat related but irrelevant to our main point is the following: there are moments which are nonzero, but which are not equal to the expectation of their absolute value. This is obvious from the fact that there are monomials with negative expectation. This is a sort of nontriviality condition on moments of real orthogonal matrices, and becomes relevant when one starts thinking about generalizing to other real algebras.)

An alternative mission is to try to characterize the moments which we do know. It seem quite likely that the Jack polynomials can tell us the expectation of any symmetric poylnomial in the random orthogonal matrix X; the problem is then that most polynomials we think of are not symmetric. One might think that this method should at least give us the expectation of any "symmetric" (invariant under transpose) polynomial, guided by some intuitive feeling of the role of symmetric matrices in Theorem 3. This is false, as we saw by the fact that adding in knowledge of transpose symmetry does not specify all moments. (Every polynomial can be written as a symmetric part plus an antisymmetric part, and the expectation of the antisymmetric part is known to be zero by the transpose symmetry.) Perhaps there is some similar class of polynomials for which we can always find the expectation with this system.

Finally, we note an oddity in the moments listed in Figure 8: certain denominators appear more frequently than others. (For example, 7 divides the denominator of each moment of degree 6, and 9 divides many, but not all, of the moments of degree 8.) This is perhaps not terribly surprising on an intuitive level, but the precise description of what is going on may help give further insights into the moments.

3. The North Pole Problem

3.1. **Introduction.** For the remainder of this paper, we look at the so-called "north pole" terms of random orthogonal matrices. This study was inspired by [5], and in this section we shall mostly be repeating results from that paper. While the previous sections were computational in nature, this present study is more theoretical.

As before, we let X denote a random sample from $\mathcal{O}$, the group of $m \times m$ random orthogonal matrices endowed with the Haar measure. Fix (column) vectors $v, w \in \mathbb{R}^m$ and fix some $n \geq 1$. The product $w^{\mathrm{T}} X^n v$ is a scalar which can be expressed as a polynomial in the entries of X. Hence moments of this scalar are special cases of the general moments of random orthogonal matrices considered above. (For example, in the case $m = 3$, we can compute the first, second, third, and fourth moments of $w^{\mathrm{T}} X^2 v$ for any fixed w, v from the data in Figure 8, or more simply from the Javascript application.)

However, this is a special type of polynomial in the entries of X, so we may expect to be able to say more about its behavior. With this in mind, our goal is not to compute moments of $w^{\mathrm{T}} X^n v$, but rather to determine its distribution exactly. Following [5], we will use the notation $\mathcal{L}(\cdot)$ to refer to the distribution law of a random variable.

By scaling v and w (which, of course, just scales the distribution of $w^{\mathrm{T}} X^n v$), we may as well assume that v and w lie on the unit sphere. Let p be the first coordinate vector $p = (1, 0, \dots)$; this may be thought of as the "north pole" of the unit sphere. Fix an orthogonal matrix Q such that $Qv = p$, and recall that the Haar measure on $\mathcal{O}$ is invariant under conjugation.

Thus

$$\mathcal{L}(X) = \mathcal{L}(Q^{\mathrm{T}} X Q) \quad \text{so} \quad \mathcal{L}(w^{\mathrm{T}} X^n v) = \mathcal{L}((Qw)^{\mathrm{T}} X^n (Qv)) = \mathcal{L}((w')^{\mathrm{T}} X^n p),$$

where $w' = Qw$. Hence, by replacing w with w', we may as well take $v = p$.

Let ξ_m denote the random variable of a coordinate of a random vector sampled uniformly on the sphere. Define the probability density functions

$$f_m(x) = \frac{\Gamma(\frac{1}{2}m)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(m-1))}(1-x^2)^{(m-3)/2} \quad \text{for } |x| \leq 1. \tag{1}$$

It is well-known that $f_m(x)$ is the probability density function for ξ_m.

Now suppose we condition on $p^{\mathrm{T}} X^n p$. As X is orthogonal, so is X^n; thus $X^n p$ is a unit vector. Thus we have

$$X^n p = (p^{\mathrm{T}} X^n p) \cdot p + \sqrt{1 - (p^{\mathrm{T}} X^n p)^2} \cdot p' \tag{2}$$

for some unit vector p' orthogonal to p. Having conditioned on $p^{\mathrm{T}} X^n p$, we may now think of $w^{\mathrm{T}} X^n p$, the quantity which we are really interested in, as a function of the random variable p'.

Let q be any (nonrandom) unit vector orthogonal to p. Then, from (2), the projection $q^{\mathrm{T}} X^n p$ is given by $\sqrt{1 - (p^{\mathrm{T}} X^n p)^2} \cdot (q^{\mathrm{T}} p')$. Now if q' is any other (nonrandom) unit vector orthogonal to p, then we may find a (nonrandom) orthogonal matrix Q fixing p and taking q to q'. Conjugating X by this matrix, we see that $\mathcal{L}(q^{\mathrm{T}} X^n p) = \mathcal{L}((q')^{\mathrm{T}} X^n p)$. Conjugating by Q does not affect the scalar $p^{\mathrm{T}} X^n p$ on which we are conditioning; thus we deduce that $\mathcal{L}(q^{\mathrm{T}} p') = \mathcal{L}((q')^{\mathrm{T}} p')$. As q and q' were arbitrary, this shows that p' is an isotropic random vector on the intersection of the unit sphere with the orthogonal complement to p. Thus p' is uniformly distributed on an $(m-1)$-dimensional sphere.

But then if we write $w = \alpha p + \beta q$ where q is a unit vector orthogonal to p (and so $\alpha^2 + \beta^2 = 1$), then we have

$$w^{\mathrm{T}} X^n p = \alpha(p^{\mathrm{T}} X^n p) + \beta(q^{\mathrm{T}} X^n p),$$

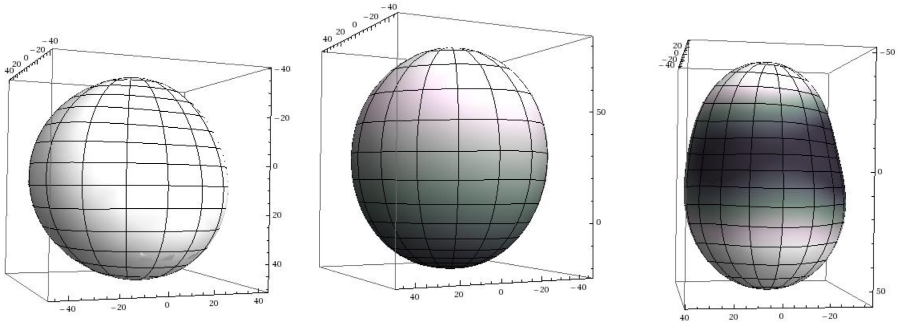
the second term is given by $\beta(q^{\mathrm{T}} p')$, and this is distributed like β times a random variable ξ_{m-1} independent of $p^{\mathrm{T}} X^n p$. As we understand $\mathcal{L}(\beta \xi_{m-1})$, to find $\mathcal{L}(w^{\mathrm{T}} X^n p)$ it suffices to find $\mathcal{L}(p^{\mathrm{T}} X^n p)$. In other words, we do not lose any generality by considering only $w = p$.

We have reduced our goal to studying the random function $p^{\mathrm{T}}X^n p$, which is just the 1,1 entry of the matrix X^n. This measure the length of the projection of $X^n p$ onto the span of p; it measures the latitude of $X^n p$, so to speak.

To emphasize the dependence on n, we will denote by U_n the random function $p^{\mathrm{T}}X^n p$. This random function appears to have first been studied in [11] (as attributed by [5]).

3.2. **Some Plots.** A priori it is possible that the random variables U_n might be entirely uninteresting. For example, when $n = 1$ we are considering $p^{\mathrm{T}}Xp$; it is well-known that Xp is uniform on the sphere and independent of p, so this is just distributed like a random variable ξ_m.

To justify our study, in this section we present some plots of samples of north pole terms. We will focus on $m = 3$ so that the vectors may be visualized. If one randomly samples X and fixes n, then $X^n p$ is a random vector on the sphere with some distribution. In Figure 9 we present "histograms" of $X^n p$; these are spherical plots where the radius function is determined by the count of random samples in the vicinity of that point. (Thus a direction where the surface is far out corresponds to a dense point in the distribution.) This may also be thought of as a plot on the sphere of the density function for $X^n p$.



$n = 1$ $\quad$ $n = 2$ $\quad$ $n = 3$

FIGURE 9. Density plots of the vector $X^n p$ for $n = 1, 2, 3$.

Figure 9 demonstrates what we argued above: the only interesting behavior is in the p-direction, i.e. on the axis of the north pole. For $n = 1$ the plot is basically spherical, demonstrating (as we know) that $X^1 p$ is uniform on the sphere. For $n = 2$, though, we start to see interesting behavior: $X^2 p$ clearly tends to be near the north pole, and is rarely near the south pole. For $n = 3$, we see that $X^3 p$ tends to prefer being near a pole, but appears to be roughly unbiased as to pole it is near. (Indeed, as X is invariant under negation, $\mathcal{L}(X^3 p) = \mathcal{L}(-X^3 p)$, so $X^3 p$ certainly should treat the north and south poles the same!)

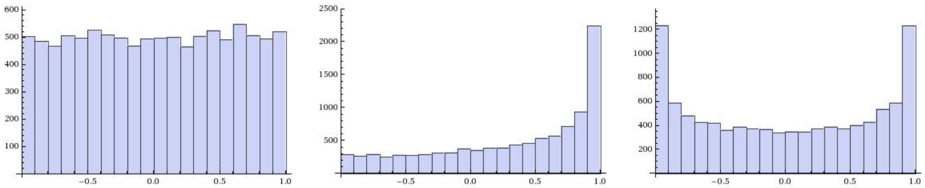
While these three dimensional plots are nice, having already found above that we can focus on $p^{\mathrm{T}}X^n p$, it makes more sense to histogram this scalar quantity on its own. We do so in Figure 10.

3.3. **Previous Results.** Understanding the distributions in Figure 10, i.e. the random functions $p^{\mathrm{T}}X^n p$ for $n = 1, 2, 3$, was the main goal of [5]. In this section we present their results.

We have already noted the following result.

Theorem 11. *$\mathcal{L}(U_1) = \mathcal{L}(\xi_m)$, and this has density function $f_m(x)$.*

Looking at equation (1), for $m = 3$ we see that the density function for U_1 is constant. This, of course, matches what we see in Figure 10.



$n = 1$ $n = 2$ $n = 3$

FIGURE 10. Histograms of $U_n = p^{\mathrm{T}} X^n p$ for $n = 1, 2, 3$.

Consider U_2. To study this quantity, let $x_{1,1}$ denote the 1,1 entry of the random orthogonal matrix X, let r denote the rest of the first row of X (i.e. r is the row vector $(x_{1,2}, \ldots, x_{1,m})$) and let c denote the rest of the first column of X (i.e. c is the column vector $(x_{2,1}, \ldots, x_{m,1})$). Then by definition of matrix multiplication, $p^{\mathrm{T}} X^2 p$, the 1,1 entry of X^2, is just $x_{1,1}^2 + r \cdot c$, as shown.

$$X^2 = \begin{pmatrix} x_{1,1} & r \\ c & ? \end{pmatrix} \begin{pmatrix} x_{1,1} & r \\ c & ? \end{pmatrix} = \begin{pmatrix} x_{1,1}^2 + r \cdot c & ? \\ ? & ? \end{pmatrix}.$$

Now r and c are both $(m-1)$-dimensional vectors with magnitude $\sqrt{1 - x_{1,1}^2}$. Thus, if we let $\widehat{r}$ and $\widehat{c}$ denote the normalizations of r and c, respectively, then we have $p^{\mathrm{T}} X^2 p = x_{1,1}^2 + (1 - x_{1,1}^2)(\widehat{r} \cdot \widehat{c})$.

There is no reason to expect that the vectors $\widehat{r}$ and $\widehat{c}$ should have any particular relationship; one might reasonably guess that they are independent and uniformly distributed on the sphere. (Showing that they are uniformly distributed on the sphere is easy; the nontrivial part of this conjecture would be the independence.) In fact this is correct, and it is proven in [4]. Now $x_{1,1}$ behaves like a ξ_m random variable, and the dot product $\widehat{r} \cdot \widehat{c}$ of independent uniform vectors on the sphere behaves like a ξ_{m-1} random variable. Thus we obtain the following.

Theorem 12. *Let ξ_m and ξ_{m-1} be independent scalars with density functions $f_m(x)$ and $f_{m-1}(x)$, respectively. Then $\mathcal{L}(U_2) = \mathcal{L}(\xi_m^2 + (1 - \xi_m^2)\xi_{m-1})$.*

Theorem 12 explains the phenomenon witnessed in Figure 10 that U_2 tends to be positive: it is given by the sum of a term which is centered at 0 with a square, and certainly the square is biased towards positive values.

As an aside, in [3], it is posited that the proof of independence of r and c requires "a substantial amount of group invariance theory." It seems like such a simple result really should have a simpler proof; because of how simple it is to state, we believe that this problem deserves significant attention.

By extending this type of decomposition result, the following theorem is proven in [5].

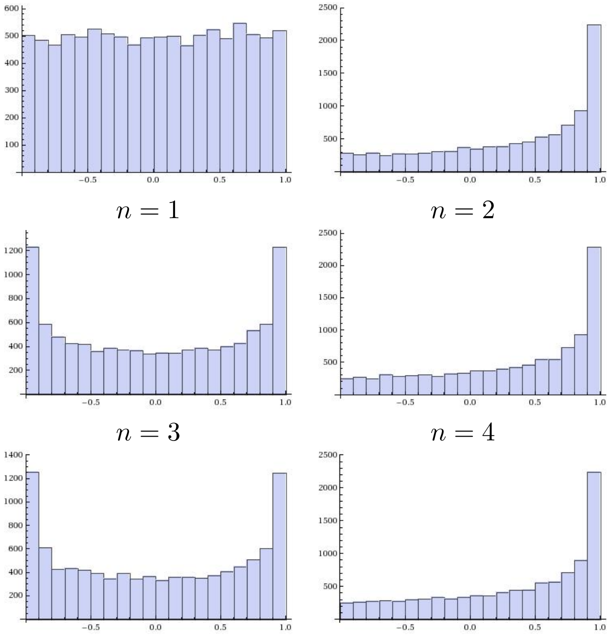
Theorem 13. *Let ξ_m, ξ_{m-1}, and ξ_{m-2} be independent scalars with density functions $f_m(x)$, $f_{m-1}(x)$, and $f_{m-2}(x)$, respectively. Then $\mathcal{L}(U_3) = \mathcal{L}(\xi_m^3 + 2\xi_m(1 - \xi_m^2)\xi_{m-1} + (1 - \xi_m^2)[-\xi_m \xi_{m-1}^2 + (1 - \xi_{m-1}^2)\xi_{m-2}])$.*

This is as far as the authors of [5] were able to take their techniques. Theorems 11, 12, and 13 suggest a clear form for $\mathcal{L}(U_n)$ for general n; it seems like it should be given as a polynomial in n independent scalars $\xi_m, \xi_{m-1}, \ldots, \xi_{m-n+1}$ which are distributed like the coordinates of a random uniform vector on the $m, (m-1), \ldots, (m-n+1)$ spheres, respectively.

This seems like an appropriate form for the answer to the north pole problem, and we believe that with the right point of view, it should not be too difficult to prove. The fact that the present approach becomes unwieldy for $n > 3$ suggests, we claim, that it is not the right approach for this problem. We have two suggestions for alternate approaches. First, one might decompose X as a product of Householder reflections; this is explicitly an expansion in terms of random uniform vectors on spheres of decreasing dimensions. This compares nicely with the form we expect for $\mathcal{L}(U_n)$. As another suggestion, one might study an eigenvalue decomposition of a random orthogonal matrix. This technique is especially powerful for U_n for large n, as we shall see in the next section.

4. New Results in the North Pole Problem

4.1. **High Powers.** In this final section of the paper, we present (basically) two new results under the heading of north pole problems. The first concerns the distribution of U_n for large n, and is motivated by looking at the histograms in Figure 14 for U_n past $n = 3$. As always, we are considering the case of $m = 3$.



$n = 1$ $n = 2$

$n = 3$ $n = 4$

$n = 5$ $n = 6$

Figure 14. Histograms of $U_n = p^{\mathrm{T}} X^n p$ for $n = 1, 2, 3, 4, 5, 6$.

Figure 14 immediately suggests a conjecture: $\mathcal{L}(U_n)$ for $n \geq 2$ depends only on the parity of n, so that Theorems 12 and 13 actually characterize $\mathcal{L}(U_n)$ for all $n \geq 2$. In fact, the tools for proving this conjecture basically already exist; we just need to put them together. (We will present the pieces, but in the spirit of applied mathematics, we will technically not give a full proof of the claim.)

4.2. **Eigenvalues of Powers of Compact Lie Groups.** The tool needed to prove this conjecture is provided by [13]. This paper presents a lovely result from the author's dissertion, which is stated most cleanly in the case of unitary matrices over $\mathbb{C}$. Because it is cleanest, we shall state it first in this case, even though our immediate interest is in the real orthogonal group. Let $\mathcal{U}$ denote the $m \times m$ unitary group; this is a compact Lie group, just like $\mathcal{O}$, so also has a Haar measure. We think of $\mathcal{U}$ as a probability space by using this Haar measure.

Theorem 15. *Fix some $n \geq m$ and sample U from $\mathcal{U}$. Then the eigenvalues of U^n are independent and uniformly distributed on S^1, the unit circle.*

The proof, or at least the basic concept, is also quite nice, so we shall spend a moment to discuss it. The joint density for the eigenvalues $\lambda_1, \ldots, \lambda_m$ of a random $U \in \mathcal{U}$ is given by

$$\frac{1}{m!} \prod_{1 \leq j < k \leq m} |\lambda_j - \lambda_k|^2 , \tag{3}$$

where of course the λ_j are confined to S^1. Alternately, one could write this as

$$\frac{1}{(2\pi)^m m!} \prod_{1 \leq j < k \leq m} \left| e^{i\theta_j} - e^{i\theta_k} \right|^2 ,$$

where the θ_j are confined to $[0, 2\pi)$. (This density is well-known; in fact, it is basically present already in our class notes, in the form of the joint density for the eigenvalues of a matrix in the GUE. To justify our claim that it is well-known, see appearances of this in [13], [11], or [9].)

Now the density (3) is a Laurent polynomial, and each eigenvalue in each term has degree between $-(m-1)$ and $(m-1)$. As the eigenvalues of U are $\lambda_1, \ldots, \lambda_m$, of course the eigenvalues of U^n are $\lambda_1^n, \ldots, \lambda_m^n$.

The ubiquitous method of moments tells us the following. Given random variables $\{\lambda_j\}$ whose joint density function is a Laurent polynomial, the joint density function for $\{\lambda_j^n\}$ is given by dividing every exponent in the original Laurent polynomal by n and dropping any monomials with fractional exponents. (Thanks to a touch of Fourier analysis, λ_j^n "does not see" any effects involving terms λ_j^k unless $n \mid k$.) If $n \geq m$, then the only term in our Laurent polynomial (3) which has all exponents divisible by m is the constant term, thanks to our bounds on the exponent. Thus the random functions $\lambda_1^n, \ldots, \lambda_m^n$ have constant joint density function, so they are independent and uniform on the circle, just as Theorem 15 asserts.

Theorem 15 is not a special property of the unitary group $\mathcal{U}$; the result of [13] generalizes it to any compact Lie group. To paraphrase, the generalized result says that for sufficiently large n, say $n > d$ for some d, the distribution of the "eigenvalues" (technically, the eigenvalues of the image under a continuous unitary representation) of an element U of a compact Lie group L depends only on which connected component of L the element U falls into. It is asserted in [13] that the "independence threshold" d for the orthogonal group $\mathcal{O}$ is $m - 2$.

The orthogonal group $\mathcal{O}$, which would usually be denoted by $\mathrm{O}(m)$, has two connected components: $\mathrm{SO}(m)$, the group of orthogonal matrices with determinant 1, and its complement $\mathrm{O}(m) - \mathrm{SO}(m)$, the coset of orthogonal matrices with determinant -1. This, in an intuitive sense, explains the parity issue in U_n; the parity of n is picking up on the distinction between the special orthogonal group and its complement.

To be more concrete about this, we will actually spell out the structure of the eigenvalues of an orthogonal matrix $X \in \mathcal{O}$. It depends on the parity of m as well as the coset which X falls into, i.e. on the determinant of X. Suppose first that m is odd, say $m = 2k + 1$. As the characteristic polynomial of X, a real matrix, is a real polynomial, it has some number of complex conjugate pairs of roots and then an odd number of left over real roots. The eigenvalues of an orthogonal matrix all lie on the unit circle, so the only possible real eigenvalues are ± 1. As there are an odd number of total real roots, one of $1, -1$ must have an odd multiplicity and the other must have even multiplicity. Peeling off the root d with odd multiplicity, the remaining $2k$ eigenvalues are composed of some number of complex conjugate pairs together with 1 and -1, both appearing with even multiplicity. We may think of $(1,1)$ and $(-1,-1)$ also as complex conjugate pairs, so we may say the following: the eigenvalues of X consist of d, which is either 1 or -1, and k complex conjugate pairs $\{(e^{i\theta_j}, e^{-i\theta_j})\}_{j=1}^{k}$. Note that the determinant of X is d, so this tells us whether X falls into $\mathrm{SO}(m)$ or $\mathrm{O}(m) - \mathrm{SO}(m)$.

In the case where $m = 2k$ is even, the eigenvalues in the two cosets look more different. If $X \in \mathrm{SO}(m)$, then the eigenvalues of X consist of k complex conjugate pairs $\{(e^{i\theta_j}, e^{-i\theta_j})\}_{j=1}^{k}$. If $X \in \mathrm{O}(m) - \mathrm{SO}(m)$, then the eigenvalues of X consist of $k-1$ complex conjugate pairs $\{(e^{i\theta_j}, e^{-i\theta_j})\}_{j=1}^{k-1}$ together with the two additional eigenvalues 1 and -1.

We believe that the following is the correct statement of the analogue of Theorem 15 for $\mathcal{O}$. (We call it a claim just because we have not worked out the precise details; we really believe this to be a theorem.)

Claim 16. *Let X be a random sample from $\mathcal{O}$. Condition on $\det X$, and let $\{\theta_j\}$ be the angles corresponding to the complex conjugate pairs of eigenvalues of X. (Depending on $\det X$ and on the parity of m, there may be $\lfloor m/2 \rfloor$ or $\lfloor m/2 \rfloor - 1$ such angles. Each angle θ_j is only defined up to sign; randomly choose a sign.) The set $\{e^{ni\theta_j}\}$ of points on the unit circle S^1 is uniformly distributed for any $n \geq m-1$.*

In particular, notice that Claim 16 tells us that the distribution of the complex conjugate pairs of eigenvalues of X^n is independent of n for $n \geq m-1$. The remaining eigenvalues λ of X are in $\{1, -1\}$, so the remaining eigenvalues λ^n of X^n depend only on the parity of n. In other words, we have the following: for $n \geq m-1$, the distribution of the eigenvalues of X^n (where X is sampled from $\mathcal{O}$) depends only on the parity of n.

4.3. **Eigenvectors of Random Orthogonal Matrices.** Having studied the eigenvalues of X^n for X randomly sampled from $\mathcal{O}$, we now need to understand something about the eigenvectors. For this, we first need the following result.

Theorem 17. *Let X be a unitary matrix. Then X is unitarily diagonalizable.*

Theorem 17 is the sort of nice result which one really believes should be true, but perhaps never knows exactly how to prove. Following [11], we can give a straightforward proof using the Schur decomposition. The Schur decomposition asserts that there exists a unitary matrix Q and a (lower, say) triangular matrix R such that $X = QRQ^*$, where Q^* denotes the conjugate-transpose of Q (which is also Q^{-1} as Q is unitary). But now X, Q, and Q^* are unitary, so $R = Q^*XQ$ must be unitary as well. But a triangular matrix which is unitary must be diagonal. (Indeed, look at the first row. There is only one nonzero entry, in the 1,1 position. As the row must be a unit vector, the entry in the 1,1 position must be a complex number of unit modulus. But then the first column is also a unit vector, so it follows that the rest of the first column must be zero. This argument continues.) Hence $X = QRQ^*$ is actually the eigenvalue decomposition of X.

As with Theorem 15, working with the unitary group $\mathcal{U}$ yields simpler statements than working with $\mathcal{O}$. The following result is proven in [11].

Theorem 18. *The eigenvector matrix Q of a random matrix $U \in \mathcal{U}$ is isotropic (i.e. it is itself a random matrix in $\mathcal{U}$) and is independent of the eigenvalues.*

We assert that the same is basically true in the real case, once one takes care to handle the real nature of the problem. Consider first the case when $m = 2k+1$ is odd, which is simpler to describe. Let X be a random sample from $\mathcal{O}$. Then recall from above that X has one real eigenvalue $\lambda_0 = d$ and k complex conjugate pairs of eigenvalues $\{(\lambda_{2j-1}, \lambda_{2j})\}_{j=1}^{k}$. As the distribution of (some power, at least) of the eigenvalues λ_{2j-1} is uniform on the circle, with probability one each λ_{2j-1} will truly be complex (i.e. not real) and they will be distinct. Now X is orthogonal, so is also unitary; thus by Theorem 17, X may be unitarily diagonalized. Let $v_0, \ldots, v_{2k}$ be the eigenvectors of X, written in order corresponding to the eigenvalues $\lambda_0, \ldots, \lambda_{2k}$. Then by Theorem 17 the matrix composed of these vectors is unitary; so the vectors are orthogonal and of unit length (using the Hermitian inner product). Now v_0 is a real unit vector, and as $\lambda_{2j-1} = \overline{\lambda_{2j}}$ for each j, we have $v_{2j-1} = \overline{v_{2j}}$.

For each j, write $v_{2j} = u_j + iw_j$, where u_j and w_j are real vectors. Then $v_{2j-1} = u_j - iw_j$. As $|v_{2j-1}| = 1$, we have $|u_j|^2 + |w_j|^2 = 1$. Each pair of distinct eigenvectors is orthogonal, so in particular v_{2j-1} and v_{2j} are orthogonal; hence $\overline{v_{2j-1}}^{\mathrm{T}} v_{2j} = 0$. This inner product is $|u_j|^2 - |w_j|^2 + 2i(u_j \cdot w_j)$. Thus $|u_j|^2 = |w_j|^2$ (which then implies that they equal 1/2, as we know that their sum is 1) and u_j and w_j are orthogonal.

Using orthogonality of v_{2j-1} and v_{2j} with v_0, we find that u_j, w_j, and v_0 are mutually orthogonal.

For $j \neq j'$, we get four more orthogonality conditions: we have orthgonality of v_{2j-1} and $v_{2j'-1}$, of v_{2j-1} and $v_{2j'}$, of v_{2j} and $v_{2j'-1}$, and of v_{2j} and $v_{2j'}$. Writing out what this means for the real and imaginary parts, we find that u_j, w_j, $u_{j'}$, and $w_{j'}$ are mutually orthogonal.

Putting this all together, we see that $\{v_0, u_1\sqrt{2}, \ldots, u_k\sqrt{2}, w_1\sqrt{2}, \ldots, w_k\sqrt{2}\}$ is a set of $m = 2k + 1$ mutually orthogonal unit vectors; thus they form the columns of an orthogonal matrix. We claim that the resulting matrix is distributed according to the Haar measure on $\mathcal{O}$. More importantly, we claim that this matrix is independent of the eigenvalues.

Of course, the same sort of construction works for m even. In both cases, we are making the following claim. (We are actually claiming much more, but the stronger claim will not be needed.)

Claim 19. *Sample X from $\mathcal{O}$. Condition on* $\det X$*, and let $\{\theta_j\}$ be the angles corresponding to the complex conjugate pairs of eigenvalues of X. The eigenvectors of X are independent of the angles θ_j.*

4.4. Finishing the Claim.

We are now in a position to the complete the proof of our claim that, for $n \geq 2$ (in the case of $m = 3$), $\mathcal{L}(U_n)$ depends only on the parity of n. In fact, we can show the corresponding result for all m: for $n \geq m - 1$, $\mathcal{L}(U_n)$ depends only on the parity of n.

Let X be a random sample from $\mathcal{O}$. By Theorem 17, we may find an eigendecomposition

$$X = \begin{pmatrix} & & \\ v_1 & \cdots & v_m \\ & & \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix} \begin{pmatrix} v_1^* \\ \vdots \\ v_m^* \end{pmatrix}$$

for X. (As above, we use a superscripted asterisk to denote the conjugate-transpose.) For each j, k, let $v_{j,k}$ denote the k^{th} entry of the vector v_j. Then notice that $U_n = p^{\mathrm{T}} X^n p$ is given by

$$U_n = \sum_{j=1}^{m} v_{j,1} \lambda_j^n \overline{v_{j,1}} = \sum_{j=1}^{m} |v_{j,1}|^2 \lambda_j^n.$$

Applying Claim 19, the sets $\{v_{j,1}\}$ and $\{\lambda_j^n\}$ of random functions are independent. By Claim 16, the distribution of the set $\{\lambda_j^n\}$ of random functions depends only on the parity of n, as long as $n \geq m - 1$. We deduce immediately that $\mathcal{L}(U_n)$ depends only on the parity of n, as long as $n \geq m - 1$; this is what we set out to show. Modulo Claims 16 and 19, we have proven the following.

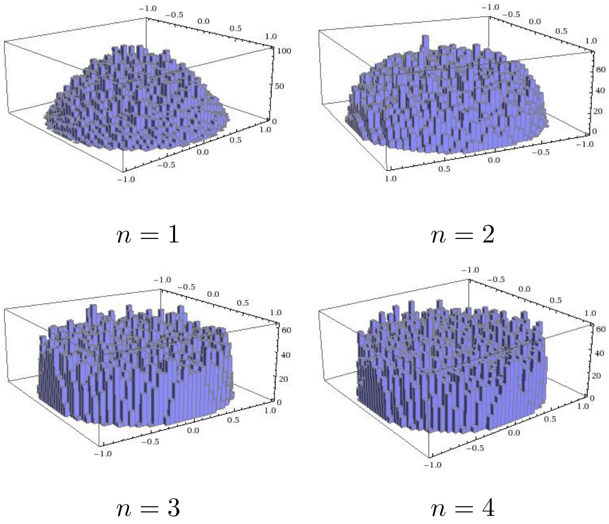
Theorem 20. *For $n \geq m - 1$, $\mathcal{L}(U_n)$ depends only on the parity of n.*

4.5. **The Complex North Pole.** The second and final new result of this paper concerns complex (unitary) matrices. In the last section we saw several times that distributions arising from random unitary matrices are simpler than distributions arising from random orthogonal matrices. Thus it seems reasonable to look for results in the complex case.

One can of course define the analogue of north pole terms for random unitary matrices: given a random unitary matrix $U \in \mathcal{U}$ and some positive integer n, we look at $p^{\mathrm{T}} U^n p$, the 1,1 entry of U^n. We denoted the corresponding quantity in the real case by U_n; to avoid conflicts from this (fairly poor) notation, we will now write $M_n = p^{\mathrm{T}} U^n p$. This is in general a complex scalar value.

Recall that in the real case, we saw a bias towards positive values for even n. We cannot have any such bias in M_n; as $\mathcal{L}(U)$ is invariant under under multiplication by any complex number of unit modulus, it is easy to see that M_n is an isotropic random variable. Thus the density function for M_n is a function of $|M_n|$ only.

The isotropy does not mean that M_n is uninteresting, though; the density functions for various n may be nontrivial. Taking $m = 3$, Figure 21 shows histograms for M_n for $n = 1, 2, 3, 4$.



$n = 1$

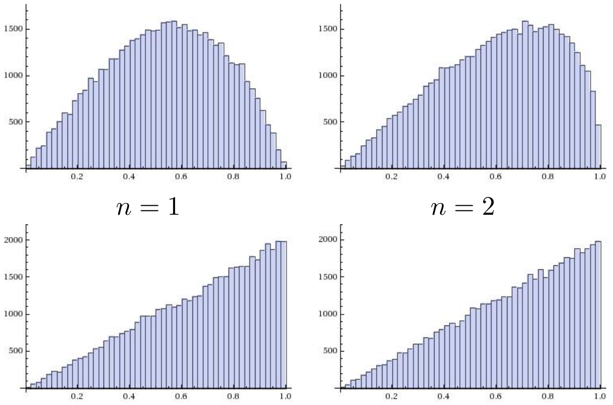
$n = 2$

$n = 3$

$n = 4$

Figure 21. Histograms of $M_n = p^{\mathrm{T}} U^n p$ for $n = 1, 2, 3, 4$.

In light of the isotropy of M_n, we might find it more useful to look at the distribution of its magnitude. Figure 22 presents histograms of $|M_n|$.



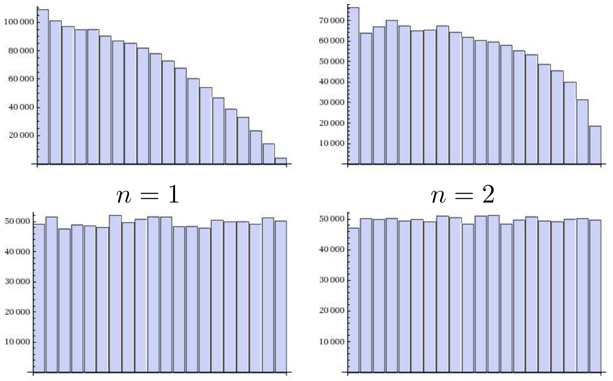
$n = 1$ $n = 2$

$n = 3$ $n = 4$

FIGURE 22. Histograms of $|M_n| = \left|p^{\mathrm{T}} U^n p\right|$ for $n = 1, 2, 3, 4$.

The histograms of Figures 21 and 22 do not seem to agree at first glance. We see, for example, in Figure 22 that the probability density of $|M_n|$ at zero appears to be zero. However, in Figure 21 the density at zero does not appear to be zero. This issue is a standard problem when comparing two-dimensional isotropic plots with plots of the magnitude. The explanation, in a loose sense, is that there are just more points with large magnitude, and so even if the two-dimensional distribution favors smaller magnitude, the distribution of the magnitude will still avoid zero. (In a more concrete sense, the Jacobian of the transformation to polar coordinates vanishes at $r = 0$.)

If what we really want is a one-dimensional histogram which agrees visually with Figure 21, then we might take a cross-section of the two-dimensional histogram. Alternately (and equivalently, given isotropy), we could just scale the histogram of the magnitude by the Jacobian of the polar transformation. (In geometric terms, we are scaling by the area of the region represented by each bin of the histogram.) Figure 23 presents histograms scaled like this.



$n = 1$ $n = 2$

$n = 3$ $n = 4$

FIGURE 23. Scaled histograms of $|M_n| = \left|p^{\mathrm{T}} U^n p\right|$ for $n = 1, 2, 3, 4$.

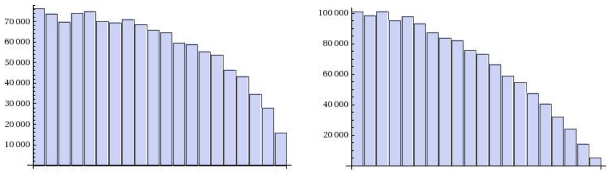
4.6. **First Observations.** The phenomenon suggested by Figure 21 is that M_1 tends to be fairly small in magnitude (note that M_1 is just a component of a random unit vector on the complex sphere), but that M_n tends to get larger in magnitude as n increases. We believe that the distributions of M_n for small n should be easy to describe, in the manner of Theorems 11, 12, and 13. However, as of now we have not looked into this.

Another observation is that the distributions for $n = 3$ and $n = 4$ appear to be the same: in light of our independence results about $\mathcal{L}(U_n)$ for large n, this is not terribly surprising.

In fact, we can immediate prove this fact, in the same way as we proved Theorem 20. Let U be a random unitary matrix, and eigendecompose U: let $v_1, \ldots, v_m$ be the eigenvectors, and let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues. Let $v_{j,k}$ refer to the k^{th} entry of v_j. Then $M_n = \sum_{j=1}^{m} |v_{j,1}|^2 \lambda_j^n$. By Theorem 18, the eigenvectors and eigenvalues are independent. By Theorem 15, the distribution of $\{\lambda_j^n\}$ for $n \geq m$ is independent of n. Thus we have the following theorem.

Theorem 24. *For $n \geq m$, $\mathcal{L}(M_n)$ is independent of n.*

4.7. **Limiting Distribution.** As $\mathcal{L}(M_n)$ stabilizes for large n, we may talk about its limiting distribution. (The limiting distribution is achieved, in particular, by $n = m$.) Looking at Figure 23, say, it appears that the limiting distribution for $m = 3$ is uniform on the unit disk. This is not true for all m, though, as Figure 25 indicates.



$m = 4$ $m = 5$

FIGURE 25. Scaled histograms of $|M_m|$ for $m = 4, 5$.

The rest of this paper is dedicated to determining the limiting distribution. We will do this by looking at exponential moment generating functions; there is no doubt a simpler approach, but this was the first approach we were able to make work. (In fact, given the nice form of the result we shall find, it is likely that there is a very simple proof. Probably one obtain the result by mimicking the derivation of the density function for a coordinate of a vector uniform on the sphere. However, we find the present proof to be a nice exercise in manipulating exponential generating functions, and rather enjoy the way it all works out.) Throughout we will be cavalier about differentiating under integral signs and commuting various limits (for example, commuting integrals with infinite sums); the revelant details do all work out, though, if one is inclined to look at them closely.

We can get the limiting distribution by looking at M_n for any $n \geq m$. We already know that M_n is isotropic, so to find its distribution it suffices to study the distribution of $\text{Re}[M_n]$. We shall show the following.

Theorem 26. *The probability density function for* $\text{Re}[M_n]$ *(where $n \geq m$) is proportional to $(1 - x^2)^{m/2-1}$.*

If we take $m = 3$ in Theorem 26, then we find that $\mathrm{Re}[M_n]$ has density function proportional to $\sqrt{1-x^2}$. This is of course the density function for the real part of a point chosen uniformly in the unit disk; thus we see that the limiting distribution in the case of $m = 3$ is indeed uniform. From Theorem 26, we can also see that the limiting distribution is not uniform for any other m. As m increases, $\mathrm{Re}[M_n]$ becomes more biased towards small values; in other words, the tail of the limiting distribution shrinks. This agrees with what we see in Figure 25.

With these observations out of the way, we shall now start the proof of Theorem 26. Recall from above that if U is a random unitary matrix with eigenvectors $v_1, \ldots, v_m$ and eigenvalues $\lambda_1, \ldots, \lambda_m$, then $M_n = \sum_{j=1}^{m} |v_{j,1}|^2 \lambda_j^n$.

Because we are interested in $n \geq m$, Theorem 15 asserts that the points λ_j^n are independent and uniform on the complex circle. Thus, if we write $\lambda_j = e^{i\theta_j}$ for $\theta_j \in [0, 2\pi)$, then the θ_j are independent and uniformly distributed.

Now from Theorem 18, the eigenvectors are isotropic and independent of the eigenvalues. The meaning of isotropy is that the matrix with columns $v_1, \ldots, v_m$ is a random (with Haar measure) unitary matrix. This means that v_1, say, is uniform on the complex sphere. But the Haar measure is invariant under the transpose operation; thus the vector $(v_{1,1}, \ldots, v_{m,1})$ is also uniform on the complex sphere.

Recall one way to generate a vector uniform on the complex sphere: one can draw independent Gaussians $G_1, \ldots, G_{2m}$ and then normalize the vector $(G_1 + iG_2, \ldots, G_{2m-1} + iG_{2m})$.

Putting this together, we see that the limiting distribution is described by

$$\mathcal{L}(M_n) = \mathcal{L}\left[\frac{(G_1^2+G_2^2)e^{i\theta_1} + \cdots + (G_{2m-1}^2+G_{2m}^2)e^{i\theta_m}}{G_1^2+G_2^2+\cdots+G_{2m-1}^2+G_{2m}^2}\right], \tag{4}$$

where the G_j are normally distributed, the θ_j are uniform on $[0, 2\pi)$, and the G_j and θ_j are all independent.

Supposing some mild summability conditions (which do work out), this distribution is specified uniquely by its moments. Now the exponential generating function for the moments is $\mathbb{E}[\exp(x \cdot \mathrm{Re}[M_n])]$. Let us denote this generating function by $\eta(x)$. Using equation (4) and our knowledge of the density functions for the uniform and normal distributions, this is given by

$$\begin{aligned}\eta(x) \;=\; & \int_{\theta_1=0}^{2\pi} \cdots \int_{\theta_m=0}^{2\pi} \int_{G_1=-\infty}^{\infty} \cdots \int_{G_{2m}=-\infty}^{\infty} \exp\left[x \cdot \frac{(G_1^2+G_2^2)\cos\theta_1 + \cdots + (G_{2m-1}^2+G_{2m}^2)\cos\theta_m}{G_1^2+G_2^2+\cdots+G_{2m-1}^2+G_{2m}^2}\right] \\ & \frac{1}{(2\pi)^{2m}} e^{-1/2(G_1^2+\cdots+G_{2m}^2)}\ \mathrm{d}G_1 \cdots \mathrm{d}G_{2m}\mathrm{d}\theta_1 \cdots \mathrm{d}\theta_m.\end{aligned}$$

For each $j = 1, 2, \ldots, m$, consider the polar change of coordinates $(G_{2j-1}, G_{2j}) \mapsto (r_j, \phi_j)$. As our integrand does not depend on ϕ_j, we may immediately integrate it out. Thus we may make the replacement $\int_{G_{2j-1}=-\infty}^{\infty} \int_{G_{2j}=-\infty}^{\infty} \mathrm{d}G_{2j-1}\mathrm{d}G_{2j} \longrightarrow \int_{r_j=0}^{\infty} 2\pi r_j\ \mathrm{d}r_j$. We may as well then make the change of variables $r_j \mapsto s_j = r_j^2$, which replaces $\int_{r_j=0}^{\infty} 2\pi r_j\ \mathrm{d}r_j$ with $\int_{s_j=0}^{\infty} \pi\ \mathrm{d}s_j$.

We now have

$$\eta(x) = \int_{\theta_1=0}^{2\pi} \cdots \int_{\theta_m=0}^{2\pi} \int_{s_1=0}^{\infty} \cdots \int_{s_m=0}^{\infty} \frac{e^{-(s_1+\cdots+s_m)/2}}{(4\pi)^m} \exp\left[x \cdot \frac{s_1\cos\theta_1 + \cdots + s_m\cos\theta_m}{s_1+\cdots+s_m}\right] \prod_{j=1}^{m} \mathrm{d}s_j \mathrm{d}\theta_j.$$

Next, for each j, we make the change of variables $(s_j, \theta_j) \mapsto (s_j, p_j)$, where $p_j = s_j \cos\theta_j$. This transformation has Jacobian $\sqrt{s_j^2 - p_j^2}$, and p_j covers the interval $[-s_j, s_j]$ exactly twice. Hence we get

$$\eta(x) = \int_{s_1=0}^{\infty} \cdots \int_{s_m=0}^{\infty} \int_{p_1=-s_1}^{s_1} \cdots \int_{p_m=-s_m}^{s_m} \frac{e^{-(s_1+\cdots+s_m)/2}}{(2\pi)^m} \exp\left[x \cdot \frac{p_1+\cdots+p_m}{s_1+\cdots+s_m}\right] \prod_{j=1}^{m} \frac{\mathrm{d}s_j \mathrm{d}p_j}{\sqrt{s_j^2 - p_j^2}}. \tag{5}$$

We now integrate out the variables p_j. Consider the function

$$h(\alpha) = \int_{-1}^{1} \frac{e^{\alpha x}}{\sqrt{1-x^2}}\ \mathrm{d}x.$$

(Taking $x = p_j/s_j$, we see that h is indeed the appropriate integral to look at.)

Differentiating under the integral sign, we find

$$\alpha^2 h''(\alpha) + \alpha h'(\alpha) - \alpha^2 h(\alpha) = \int_{-1}^{1} \frac{e^{\alpha x}}{\sqrt{1-x^2}} \cdot (\alpha^2 x^2 + \alpha x - \alpha^2)\ \mathrm{d}x.$$

This splits into the difference of two integrals,

$$\int_{-1}^{1} \frac{\alpha x e^{\alpha x}}{\sqrt{1-x^2}}\ \mathrm{d}x - \int_{-1}^{1} e^{\alpha x}\alpha^2\sqrt{1-x^2}\ \mathrm{d}x. \tag{6}$$

We can perform the second integral by parts, integrating the piece $\alpha^2 e^{\alpha x}$ and differentiating $\sqrt{1-x^2}$. This gives

$$\int_{-1}^{1} e^{\alpha x}\alpha^2\sqrt{1-x^2}\ \mathrm{d}x = \alpha e^{\alpha x}\sqrt{1-x^2}\Big|_{x=-1}^{1} - \int_{-1}^{1} \alpha e^{\alpha x}\frac{-x}{\sqrt{1-x^2}}\ \mathrm{d}x.$$

The first term vanishes and the remaining integral cancels exactly the other integral in (6). Thus we get

$$\alpha^2 h''(\alpha) + \alpha h'(\alpha) - \alpha^2 h(\alpha) = 0. \tag{7}$$

This is the differential equation defining $I_0(\alpha)$, a modified Bessel function of the first kind. Now $h(\alpha)$ is an even function as one can see by making the replacements $\alpha \mapsto -\alpha$ and $x \mapsto -x$. Thus $h(\alpha)$ must be some multiple of $I_0(\alpha)$. As $h(0) = \sin^{-1}(1) - \sin^{-1}(-1) = \pi$, we in fact have $h(\alpha) = \pi I_0(\alpha)$. We thus have the series expansion

$$h(\alpha) = \pi \cdot \sum_{n=0}^{\infty} \frac{\alpha^{2n}}{(2^n n!)^2}. \tag{8}$$

(If one does not want to think about modified Bessel functions of the first kind, then one could just derive this series directly from the differential equation (7) together with the initial conditions $h(0) = \pi$ and $h'(0) = 0$. We will not need any properties besides the series expansion.)

(Note that we have started hijacking the letter n as an indexing variable for summations. We will continue to do this occasionally; hopefully there should be no confusion with the n subscript of M_n, which just stands for any integer at least m.)

Returning to equation (5), we can now integrate out the variables p_j using the function I_0. We get

$$\eta(x) = \int_{s_1=0}^{\infty} \cdots \int_{s_m=0}^{\infty} \frac{e^{-(s_1+\cdots+s_m)/2}}{2^m} \prod_{j=1}^{m} I_0\left(\frac{s_j x}{s_1+\cdots+s_m}\right)\ \mathrm{d}s_1\cdots\mathrm{d}s_m. \tag{9}$$

Let $e_1, \ldots, e_m$ be nonnegative integers and consider the related integral

$$q(\alpha) = \int_{s_1=0}^{\infty} \cdots \int_{s_m=0}^{\infty} e^{\alpha(s_1+\cdots+s_m)} \frac{s_1^{e_1}\cdots s_m^{e_m}}{(s_1+\cdots+s_m)^{e_1+\cdots+e_m}}\ \mathrm{d}s_1\cdots\mathrm{d}s_m,$$

defined when α is negative. For brevity, write $n = e_1 + \cdots + e_m$. Taking n derivatives of $q(\alpha)$, we get

$$q^{(n)}(\alpha) = \int_{s_1=0}^{\infty} \cdots \int_{s_m=0}^{\infty} e^{\alpha(s_1+\cdots+s_m)} s_1^{e_1}\cdots s_m^{e_m}\ \mathrm{d}s_1\cdots\mathrm{d}s_m,$$

and this integral factors as a product of integrals of the form

$$\int_0^{\infty} e^{\alpha x} x^k\ \mathrm{d}x = \frac{k!}{(-\alpha)^{k+1}}.$$

Thus we get

$$q^{(n)}(\alpha) = \frac{e_1! \cdots e_m!}{(-\alpha)^{n+m}}.$$

Integrating n times, we see that

$$q(\alpha) = (\text{some polynomial of degree at most } n-1) + \frac{e_1! \cdots e_m! (m-1)!}{(-\alpha)^m (m+n-1)!}. \tag{10}$$

As the s_j are always nonnegative, we have $s_1 + \cdots + s_m \geq s_j$ for any j. Thus $s_1^{e_1} \cdots s_m^{e_m}/(s_1 + \cdots + s_m)^n$ is nonnegative and bounded above by 1. Thus $q(\alpha)$ is bounded by the n-fold integral of $e^{\alpha(s_1+\cdots+s_m)}$, which is $(-\alpha)^{-m}$. As $\alpha \to -\infty$, this vanishes: hence $q(\alpha)$ vanishes at infinity. But then the polynomial in equation (10) must be zero; we have found $q(\alpha)$ exactly.

If we return to equation (9), substitute the series expansion (8), and then apply our formula in equation (10) for $q(\alpha)$ (taking $\alpha = -1/2$), we find that

$$\eta(x) = (m-1)! \sum_{d_1,\ldots,d_m \geq 0} (x/2)^{2n} \frac{1}{(d_1!)^2 \cdots (d_m!)^2} \cdot \frac{(2d_1)! \cdots (2d_m)!}{(m+2n-1)!},$$

where we write n for $d_1 + \cdots + d_m$.

This may be rewritten as

$$\eta(x) = \sum_{n \geq 0} \frac{(m-1)!(x/2)^{2n}}{(m+2n-1)!} \left(\sum_{d_1+\cdots+d_m=n} \prod_{j=1}^{m} \binom{2d_j}{d_j} \right),$$

and the sum over products of diagonal binomial coefficients may be simplified further. For this we use the fact that

$$(1-4x)^{-1/2} = \sum_{d \geq 0} x^d \binom{2d}{d}.$$

Raising this to the m^{th} power, we get

$$(1-4x)^{-m/2} = \sum_{n \geq 0} x^n \sum_{d_1+\cdots+d_m=n} \prod_{j=1}^{m} \binom{2d}{d}.$$

Taking the coefficient of x^n, we find

$$\sum_{d_1+\cdots+d_m=n} \prod_{j=1}^{m} \binom{2d}{d} = (-4)^n \binom{-m/2}{n} = 4^n \binom{n+m/2-1}{n}.$$

Substituting this above, we get

$$\eta(x) = \sum_{n \geq 0} x^{2n} \cdot \frac{(m-1)!}{(m+2n-1)!} \cdot \binom{n+m/2-1}{n}.$$

Multiplying each term by a particular expression for one, this may be rewritten as

$$\eta(x) = \sum_{n \geq 0} \frac{x^{2n}}{(2n)!} \cdot \binom{n+m/2}{n} \cdot \binom{2n+m}{2n}^{-1},$$

and from this expression we can immediately read off the moments of the random variable $\mathrm{Re}[M_n]$.

Before moving on, we will rewrite these moments using double factorials. Consider the $(2n)^{\text{th}}$ moment, $\binom{n+m/2}{n} \cdot \binom{2n+m}{2n}^{-1}$. Let $x_{(n)}$ denote the falling factorial $x(x-1)\cdots(x-n+1)$; then $\binom{n+m/2}{n} = (n+m/2)_{(n)}/n!$.

Similarly, $\binom{2n+m}{2n} = (2n+m)_{(2n)}/(2n)!$. Now

$$\frac{(2n+m)_{(2n)}}{(n+m/2)_{(n)}} = \frac{(2n+m)(2n+m-2)\cdots(m+2)(2n+m-1)(2n+m-3)\cdots(m+1)}{(n+m/2)(n+m/2-1)\cdots(m/2+1)} = \frac{2^n(2n+m-1)!!}{(m-1)!!}.$$

We can express the ratio $(2n)!/n!$ in terms of a double factorial using the relation $(2n-1)!! = (2n)!/(2^n n!)$. Combining this with the observation above, we have that the $(2n)^{\text{th}}$ moment of $\mathrm{Re}[M_n]$ is given by

$$\frac{(n+m/2)_{(n)}}{n!} \cdot \frac{(2n)!}{(2n+m)_{(2n)}} = \frac{(2n)!}{n!} \cdot \frac{(n+m/2)_{(n)}}{(2n+m)_{(2n)}} = \frac{(2n-1)!!(m-1)!!}{(2n+m-1)!!}. \tag{11}$$

Finding the moments was not really our goal: we are after the density function. We want to show that the density function is (proportional to) $(1-x^2)^{m/2-1}$; we shall do this by showing that the moments of that distribution agree with the moments of $\mathrm{Re}[M_n]$. We shall suppose that $m \geq 2$, so that the exponent $(m/2-1)$ is nonnegative. (For the case of $m=1$, one can directly determine the distribution of $\mathrm{Re}[M_n]$, as M_n is just a random point on the unit circle. Thus $\mathrm{Re}[M_n]$ is distributed like $\cos(\theta)$, where $\theta \in [0, 2\pi)$ is uniform. The density function for such a random variable is $(1-x^2)^{-1/2}/\pi$.)

Fix some real $\alpha \geq 0$ and some integer $n \geq 0$ and define the integral $i(n,\alpha) = \int_{-1}^{1} x^{2n}(1-x^2)^{\alpha}\,\mathrm{d}x$. We can write this as $\int_{-1}^{1} x^{2n-1} \cdot x(1-x^2)^{\alpha}\,\mathrm{d}x$, and then integrate by parts; doing so, we find the recurrence relation

$$i(n,\alpha) = \frac{2n-1}{2(\alpha+1)} i(n-1, \alpha+1). \tag{12}$$

By direct integration we have $i(n,0) = 2/(2n+1)$. With this we can compute $i(n,\alpha)$ for all integers $n, \alpha \geq 0$. Specifically, we find

$$i(n,\alpha) = \frac{2\alpha \cdot 2(\alpha-1)\cdots 2}{(2n+1)(2n+3)\cdots(2n+2\alpha-1)} \cdot \frac{2}{(2n+2\alpha+1)} = \frac{2(2\alpha)!!(2n-1)!!}{(2n+2\alpha+1)!!}.$$

Taking $n=0$, we find $\int_{-1}^{1}(1-x^2)^{\alpha}\,\mathrm{d}x$, which tells us the normalization needed for a probability density function proportional to $(1-x^2)^{\alpha}$.

Suppose that m is even and take $\alpha = m/2-1$. Then the $(2n)^{\text{th}}$ moment of the distribution with probability density function proportional to $(1-x^2)^{m/2-1}$ is

$$\frac{i(n,\alpha)}{i(0,\alpha)} = \frac{(2n-1)!!(2\alpha+1)!!}{(2n+2\alpha+1)!!} = \frac{(2n-1)!!(m-1)!!}{(2n+m-1)!!},$$

which we recognize as the $(2n)^{\text{th}}$ moment of $\mathrm{Re}[M_n]$ using equation (11). As the probability density function we are considering is even, its odd moments vanish; the odd moments of $\mathrm{Re}[M_n]$ vanish as well. Thus we have shown agreement of every moment between $\mathrm{Re}[M_n]$ and the random variable with probability density function proportional to $(1-x^2)^{m/2-1}$. This proves Theorem 26, for all even m.

We now need to consider odd m, which correspond to half-integer α. To compute $i(n,\alpha)$ for half-integer α, it suffices to determine $i(n, 1/2)$. Making a trigonometric substitution, this is

$$\int_{-1}^{1} x^{2n}\sqrt{1-x^2}\,\mathrm{d}x = 2\int_0^{\pi/2} \cos^{2n}\theta \sin^2\theta\,\mathrm{d}\theta = 2\int_0^{\pi/2}\cos^{2n}\theta\,\mathrm{d}\theta - 2\int_0^{\pi/2}\cos^{2n+2}\theta\,\mathrm{d}\theta.$$

The Wallis cosine formula tells us the values of these integrals; using this, we find

$$i(n,1/2) = 2\cdot\left(\frac{(2n-1)!!}{(2n)!!}\cdot\frac{\pi}{2} - \frac{(2n+1)!!}{(2n+2)!!}\cdot\frac{\pi}{2}\right) = \pi\cdot\left(\frac{(2n+2)(2n-1)!! - (2n+1)!!}{(2n+2)!!}\right) = \frac{\pi(2n-1)!!}{(2n+2)!!}.$$

Applying the recurrence relation (12), we get the general formula for half-integer α,

$$i(n,\alpha) = \frac{2\alpha \cdot 2(\alpha-1)\cdots 3}{(2n+1)(2n+3)\cdots(2n+2\alpha-2)} \cdot \frac{\pi(2n+2\alpha-2)!!}{(2n+2\alpha+1)!!} = \frac{\pi(2\alpha)!!(2n-1)!!}{(2n+2\alpha+1)!!}.$$

Taking $n=0$ gives us the normalization that we should use for the probability density function proportional to $(1-x^2)^\alpha$. Thus $i(n,\alpha)/i(0,\alpha)$ gives us the $(2n)^{\text{th}}$ moment of this distribution. (Of course, all odd moments vanish, as the density function is even.) Computing these moments with $\alpha = m/2-1$, we find

$$\frac{i(n,\alpha)}{i(n,0)} = \frac{(2n-1)!!(2\alpha+1)!!}{(2n+2\alpha+1)!!} = \frac{(2n-1)!!(m-1)!!}{(2n+m-1)!!}.$$

We have arrived at the same expression as we had for even m; as in that case, this shows that $\text{Re}[M_n]$ has probability density function proportional to $(1-x^2)^{m/2-1}$.

We have proven Theorem 26 in all cases, which was the goal of this section.

References

[1] Banica, T. (2009). "The Orthogonal Weingarten Formula in Compact Form." arXiv:0906.4694.

[2] Dumitriu, I. et al (2008). "MOPS: Multivariate Orthogonal Polynomials (symbolically)." arXiv:math-ph/0409066.

[3] Eaton, M. and Muirhead, R. (2006). "On Powers of a Random Orthogonal Matrix." Presentation slides, http://web.mit.edu/sea06/agenda/talks/Muirhead.pdf.

[4] Eaton, M. and Muirhead, R. (2008). "A Decomposition Result for the Haar Distribution on the Orthogonal Group." arXiv:0807.2598.

[5] Eaton, M. and Muirhead, R. (2009). "The 'North Pole Problem' and Random Orthogonal Matrices." *Stat. and Prob. Let.*, Vol. 79, pp. 1878–1883.

[6] Gorin, T. (2002). "Integrals of Monomials Over the Orthogonal Group." *J. Math. Phys.*, Vol. 43, pp. 3342–3351.

[7] James, A.T. (1960). "The Distribution of the Latent Roots of the Covariance Matrix." *Ann. Math. Statist.*, Vol. 31, pp. 151–158.

[8] James, A.T. (1961). "Zonal Polynomials of the Real Positive Definite Symmetric Matrices." *Ann. Math.*, Vol. 74, pp. 456–469.

[9] Kakavand, H. (2005). "On Patterns in Eigenvalues of Unitary Matrices." Research note. http://www.stanford.edu/~hossein/PatternsEigenvalues.pdf.

[10] Koev, P. (2007). "Haar Orthogonality for any α." Research note.

[11] Marzetta, T. et al (2002). "Structured unitary space-time autocoding constellations." *IEEE Trans. on Inf. Theory*, Vol. 48, pp. 942–950.

[12] Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory.*

[13] Rains, E. (1997). "High Powers of Random Elements of Compact Lie Groups." *Prob. Theory Relat. Fields*, Vol. 107, pp. 219–241.