



Any orthonormal basis in high dimension is uniformly distributed over the sphere

Sheldon Goldstein^a, Joel L. Lebowitz^a, Roderich Tumulka^b and Nino Zanghì^c

^a*Departments of Mathematics and Physics, Rutgers University, Hill Center, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA.*

E-mail: oldstein@math.rutgers.edu; lebowitz@math.rutgers.edu

^b*Department of Mathematics, Rutgers University, Hill Center, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA.*

E-mail: tumulka@math.rutgers.edu

^c*Dipartimento di Fisica, Università di Genova and INFN sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy. E-mail: zanghi@ge.infn.it*

Received 11 February 2015; revised 5 November 2015; accepted 15 November 2015

Abstract. Let \mathbb{X}^d be a real or complex Hilbert space of finite but large dimension d , let $\mathbb{S}(\mathbb{X}^d)$ denote the unit sphere of \mathbb{X}^d , and let u denote the normalized uniform measure on $\mathbb{S}(\mathbb{X}^d)$. For a finite subset B of $\mathbb{S}(\mathbb{X}^d)$, we may test whether it is approximately uniformly distributed over the sphere by choosing a partition A_1, \dots, A_m of $\mathbb{S}(\mathbb{X}^d)$ and checking whether the fraction of points in B that lie in A_k is close to $u(A_k)$ for each $k = 1, \dots, m$. We show that if B is any orthonormal basis of \mathbb{X}^d and m is not too large, then, if we randomize the test by applying a random rotation to the sets A_1, \dots, A_m , B will pass the random test with probability close to 1. This statement is related to, but not entailed by, the law of large numbers. An application of this fact in quantum statistical mechanics is briefly described.

Résumé. Soit \mathbb{X}^d un espace de Hilbert réel ou complexe de dimension finie mais grande d et soit $\mathbb{S}(\mathbb{X}^d)$ la sphère unité de \mathbb{X}^d , on note u la mesure uniforme normalisée sur $\mathbb{S}(\mathbb{X}^d)$. Pour un sous ensemble fini B de $\mathbb{S}(\mathbb{X}^d)$, nous pouvons tester s'il est approximativement uniformément distribué sur la sphère en choisissant une partition A_1, \dots, A_m de $\mathbb{S}(\mathbb{X}^d)$ et en vérifiant si la fraction des points dans B qui se trouvent dans A_k est proche de $u(A_k)$ pour tout $k = 1, \dots, m$. Nous montrons que si B est n'importe quelle base orthonormée de \mathbb{X}^d et que si m n'est pas trop grand, alors si on randomise le test en appliquant une rotation aléatoire aux ensembles A_1, \dots, A_m , l'ensemble B va passer le test avec une probabilité proche de 1. Ce résultat est relié à la loi des grands nombres. Une application de ce résultat en mécanique statistique quantique est décrite brièvement.

MSC: 60F05; 82B10; 28C10

Keywords: Law of large numbers; Haar measure on the orthogonal or unitary groups; Asymptotics in high dimension; Irreducible representations of the orthogonal or unitary groups; random orthonormal basis

1. Introduction

Let \mathbb{X}^d be a real or complex Hilbert space of finite but large dimension d , let $\mathbb{S}(\mathbb{X}^d)$ be the unit sphere in \mathbb{X}^d , and let $u = u_{\mathbb{S}(\mathbb{X}^d)}$ denote the uniform probability measure (i.e., normalized surface area) over $\mathbb{S}(\mathbb{X}^d)$. Given a large number of points on $\mathbb{S}(\mathbb{X}^d)$, we may ask whether these points are approximately uniformly distributed over $\mathbb{S}(\mathbb{X}^d)$. When we are given an orthonormal basis of \mathbb{X}^d , this provides us with d points on $\mathbb{S}(\mathbb{X}^d)$, which may at first seem like too small a number, given that $\mathbb{S}(\mathbb{X}^d)$ has real dimension $d - 1$ or $2d - 1$, for rendering meaningful the question whether these points are approximately uniformly distributed. However, the question is meaningful in a suitably coarse-grained sense of “approximately uniform,” viz., in the sense that for a partition A_1, \dots, A_m of $\mathbb{S}(\mathbb{X}^d)$ with $m \ll d$, the number of points in A_k , divided by d , is close to $u(A_k)$.

One version of our result asserts that for a *random* orthonormal basis $\{b_1, \dots, b_d\}$ with distribution $u_{\text{ONB}(\mathbb{X}^d)}$ (the uniform (normalized) measure over all orthonormal bases of \mathbb{X}^d , see below for more details), the empirical distribution on $\mathbb{S}(\mathbb{X}^d)$ of b_1, \dots, b_d is approximately uniform relative to the partition A_1, \dots, A_m with probability close to 1. Needless to say, for every fixed orthonormal basis $\{b_1, \dots, b_d\}$ there exist partitions A_1, \dots, A_m for which the number of basis vectors in A_k , divided by d , is not at all close to $u(A_k)$; for example, $A_1 = \{b_1, \dots, b_d\}$ and $A_2 = \mathbb{S}(\mathbb{X}^d) \setminus A_1$.

Our result as just formulated follows once we have it for $m = 2$, i.e., for partitions consisting merely of a set A and its complement. It therefore suffices to focus on the simpler statement that for any Borel set $A \subseteq \mathbb{S}(\mathbb{X}^d)$ and a $u_{\text{ONB}(\mathbb{X}^d)}$ -distributed orthonormal basis,

$$\mathbb{P}\left(\frac{1}{d}\#\{i \in \{1 \dots d\} : b_i \in A\} \approx u(A)\right) \approx 1, \quad (1.1)$$

where $\#S$ denotes the number of elements of a finite set S .

Here is a different way of phrasing our result. A good concept of “approximately uniformly distributed” should be invariant under rotations (or unitary transformations) of \mathbb{X}^d ; thus, if we claim of one orthonormal basis that it is approximately uniformly distributed, we should make this claim of *every* orthonormal basis. So let us regard $\{b_1, \dots, b_d\}$ as fixed and randomize A instead by considering the uniform distribution over all sets A' congruent to A . To this end, let G be the orthogonal group or unitary group of \mathbb{X}^d , depending on whether \mathbb{X}^d is real or complex, let u_G be the normalized uniform measure (i.e., the Haar measure) over G , and let R be a u_G -distributed random element of G . Our test set will be $A' = R(A)$ (i.e., a random rotation of A). Our result is that for every orthonormal basis $\{b_1, \dots, b_d\}$ and every Borel set $A \subseteq \mathbb{S}(\mathbb{X}^d)$,

$$\mathbb{P}\left(\frac{1}{d}\#\{i \in \{1 \dots d\} : b_i \in R(A)\} \approx u(A)\right) \approx 1. \quad (1.2)$$

1.1. Precise formulation

Definition 1. Let $\varepsilon, \delta > 0$. We say of a finite set $B \subseteq \mathbb{S}(\mathbb{X}^d)$ that it is ε - δ -uniform on $\mathbb{S}(\mathbb{X}^d)$ iff¹ for every Borel set $A \subseteq \mathbb{S}(\mathbb{X}^d)$,

$$\mathbb{P}\left(\left|\frac{\#(B \cap R(A))}{\#B} - u(A)\right| \leq \delta\right) \geq 1 - \varepsilon. \quad (1.3)$$

Theorem 1 (Version 1). For every $\varepsilon, \delta > 0$ and every $d \geq 4$ with $d \geq \delta^{-2}\varepsilon^{-1}$, every orthonormal basis B in $\mathbb{X}^d = \mathbb{R}^d$ or \mathbb{C}^d is ε - δ -uniform on $\mathbb{S}(\mathbb{X}^d)$.

Somewhat sloppily, we sometimes regard a basis as ordered (i.e., as a d -tuple) and sometimes as unordered (i.e., as a set). It does not matter which point of view we assume, and there are no bad consequences of switching the point of view; we call a d -tuple ε - δ -uniform if the corresponding set (obtained by forgetting the order) is.

We use the notation $Y \sim \mu$ for saying that the random variable Y has distribution μ . For example, in (1.3), $R \sim u_G$.

Version 2 of Theorem 1 (see below) provides an alternative formulation in terms of a *random* orthonormal basis. The uniform distribution $u_{\text{ONB}(\mathbb{X}^d)}$ can be defined as the distribution of the random orthonormal basis B obtained from a fixed orthonormal basis B_0 by applying a random rotation $R \sim u_G$, $B = R(B_0)$. The distribution of B is, in fact, independent of the choice of B_0 . Alternatively, a $u_{\text{ONB}(\mathbb{X}^d)}$ -distributed basis $\{b_1, \dots, b_d\}$ can be constructed as follows: Choose b_1 with distribution u from $\mathbb{S}(\mathbb{X}^d)$; let b_1^\perp denote the orthogonal complement of b_1 in \mathbb{X}^d , and $\mathbb{S}(b_1^\perp)$ the unit sphere in that subspace; choose b_2 uniformly in $\mathbb{S}(b_1^\perp)$; then choose b_3 uniformly in $\mathbb{S}(\{b_1, b_2\}^\perp)$; and so on. Theorem 1 can easily be seen to be equivalent to the following.

¹This definition possesses a natural generalization to measures instead of finite sets: We say of a normalized measure μ on (the Borel σ -algebra of) $\mathbb{S}(\mathbb{X}^d)$ that it is ε - δ -uniform on $\mathbb{S}(\mathbb{X}^d)$ iff for every Borel set $A \subseteq \mathbb{S}(\mathbb{X}^d)$, $\mathbb{P}(|\mu(R(A)) - u(A)| \leq \delta) \geq 1 - \varepsilon$. The definition for a finite set B then corresponds to the measure $\mu = (\#B)^{-1} \sum_{b \in B} \delta_b$ with δ_b the point mass at b , i.e., $\mu(A') = \#(B \cap A')/\#B$.

Theorem 1 (Version 2). For every $\varepsilon, \delta > 0$ and every $d \geq 4$ with $d \geq \delta^{-2}\varepsilon^{-1}$, the following is true: Let $\mathbb{X}^d = \mathbb{R}^d$ or \mathbb{C}^d and $B \sim u_{\text{ONB}(\mathbb{X}^d)}$. For every Borel set $A \subseteq \mathbb{S}(\mathbb{X}^d)$,

$$\mathbb{P}\left(\left|\frac{\#(B \cap A)}{d} - u(A)\right| \leq \delta\right) \geq 1 - \varepsilon. \quad (1.4)$$

Corollary 1. For every $\varepsilon, \delta > 0$, every $m \in \mathbb{N}$, and every $d \geq 4$ with $d \geq m\delta^{-2}\varepsilon^{-1}$, the following is true: Let $\mathbb{X}^d = \mathbb{R}^d$ or \mathbb{C}^d and $B \sim u_{\text{ONB}(\mathbb{X}^d)}$. For every partition A_1, \dots, A_m of $\mathbb{S}(\mathbb{X}^d)$ consisting of Borel sets,

$$\mathbb{P}\left(\forall k \in \{1 \dots m\} : \left|\frac{\#(B \cap A_k)}{d} - u(A_k)\right| \leq \delta\right) \geq 1 - \varepsilon. \quad (1.5)$$

In the *real* case $\mathbb{X}^d = \mathbb{R}^d$, the statements refer to orthonormal bases of both orientations (“left-handed” and “right-handed”). Theorem 1 and Corollary 1 remain true when restricted to just one orientation, provided $d \geq 2m\delta^{-2}\varepsilon^{-1}$. After all, if 99% of all orthonormal bases have a property p and 50% of all orthonormal bases are left-handed, then at least 98% of all left-handed orthonormal bases must have the property p .

It is perhaps not surprising that the basis vectors are uniformly distributed, as their orthogonality will have the “repulsive” effect that no two of them can be close to each other. On the other hand, one might have expected that in order to obtain a uniformly distributed set, one has to use $\{\pm b_1, \dots, \pm b_d\}$, while the basis vectors $\{b_1, \dots, b_d\}$ alone tend (one might have expected) to clump on one side of the sphere, as they all lie on a cone around $b_1 + \dots + b_d$; however, when d is large then the opening angle of this cone, $2\arccos(d^{-1/2})$, is approximately $\pi - 2/\sqrt{d}$ and thus close to π (or 180°), so not very clumped after all.

The following version of Theorem 1 expresses the theorem in terms of test functions $\varphi : \mathbb{S}(\mathbb{X}^d) \rightarrow \mathbb{R}$ rather than test sets $A \subseteq \mathbb{S}(\mathbb{X}^d)$. Let $\mathbb{E}_\mu(f)$ and $\text{Var}_\mu(f)$ denote the mean and variance of the function $f : \Omega \rightarrow \mathbb{R}$ relative to the probability measure μ on Ω .

Theorem 1 (Version 3). For every $\varepsilon, \delta > 0$, every $d \geq 4$ with $d \geq 2\delta^{-2}\varepsilon^{-1}$, every orthonormal basis $\{b_1, \dots, b_d\}$ of $\mathbb{X}^d = \mathbb{R}^d$ or \mathbb{C}^d , $R \sim u_G$, and every test function $\varphi \in L^2(\mathbb{S}(\mathbb{X}^d), u, \mathbb{R})$,

$$\mathbb{P}\left(\left|\frac{1}{d} \sum_{j=1}^d \varphi(R(b_j)) - \mathbb{E}_u(\varphi)\right| \leq \delta \sqrt{\text{Var}_u(\varphi)}\right) \geq 1 - \varepsilon. \quad (1.6)$$

With the same methods as in this paper, one can perhaps show also that, in high dimension d , the action of the rotation (resp., unitary) group G on the unit sphere is weakly mixing, i.e., that for $R \sim u_G$ and any two measurable sets $A, B \subseteq \mathbb{S}(\mathbb{X}^d)$, $\mathbb{P}(u(A \cap R(B)) \approx u(A)u(B)) \approx 1$.

A physical application of our result is outlined in Section 3.

1.2. Comparison to known results

Theorem 1 can be regarded as a *typicality theorem*, i.e., as a statement about the *typical* behavior of something, here of orthonormal bases or sets congruent to a given set $A \subseteq \mathbb{S}(\mathbb{X}^d)$. Well-known examples of typicality theorems about spheres in high dimension include the following statements: (i) that in high dimension, most of the area of a sphere is near the equator, (ii) that in high dimension, most of the volume of the unit ball is near the surface.

Theorem 1 is similar to an instance of the *law of large numbers*, i.e., of the statement that if X_1, \dots, X_n are independent identically distributed (i.i.d.) random variables then for sufficiently large n their empirical distribution is arbitrarily close to the distribution of X_1 with probability arbitrarily close to 1. Suppose b_1, \dots, b_d were *independent* u -distributed random vectors on $\mathbb{S}(\mathbb{X}^d)$, and let K_i be (in every realization) 1 or 0 depending on whether $b_i \in A$ or not. Then the K_i are i.i.d. random variables with distribution $\mathbb{P}(K_i = 1) = u(A)$, $\mathbb{P}(K_i = 0) = 1 - u(A)$, and the law of large numbers implies that (1.4) holds for sufficiently large d . Now in the situation of Version 2 of Theorem 1, b_1, \dots, b_d are *not* independent (since they have to be exactly orthogonal to each other), but they are approximately independent in the following sense: if we pick two independent random (uniformly distributed) vectors x, y on $\mathbb{S}(\mathbb{X}^d)$

with large d , then they are anyhow, with high probability, approximately orthogonal. (Indeed, it follows from symmetry considerations that their inner product $\langle x|y \rangle$ has expectation $\mathbb{E}\langle x|y \rangle = 0$ and variance $\mathbb{E}|\langle x|y \rangle|^2 = 1/d$, so $\langle x|y \rangle$ will typically be small like $1/\sqrt{d}$.) So Version 2 of Theorem 1 can be regarded as saying that the weak dependence between the basis vectors b_i does not disturb the relation (1.4) provided by the law of large numbers. Other results about how weak their dependence is can be found in [12–14].

Known theorems about *uniformity* (or *equidistribution*) are usually rather different in character from our result. One type of theorem asserts that a certain sequence x_n of points (e.g., $x_n = n\alpha \bmod 1$ for irrational α) is uniformly distributed as $n \rightarrow \infty$ in some set (e.g., the unit interval) [27]; another type concerns how uniformly certain paths (e.g., billiard trajectories) fill the space they are in [1,2]. Another circle of questions closer to our result, described in [2, Section 2], concerns quantifying how uniformly distributed a set $\{x_1, \dots, x_n\}$ in (say) the unit interval $[0, 1]$ is by comparing, for some test function $\varphi : [0, 1] \rightarrow \mathbb{R}$, $n^{-1} \sum_{i=1}^n \varphi(x_i)$ to $\int_0^1 dx \varphi(x)$. If the x_i are chosen at random (independently with uniform distribution), then the difference is (with high probability) of order $n^{-1/2}$ for any $\varphi \in L^2([0, 1])$. However, if the x_i are evenly spaced, $x_i = i/n$, and φ is sufficiently smooth, then the difference is of order n^{-1} or even smaller (see [2, Section 2] for a discussion). Thus, in a certain sense, some sets $\{x_1, \dots, x_n\}$ are *very* uniform. Our result can be expressed by saying that, for a random orthonormal basis $\{b_1, \dots, b_d\}$, $d^{-1} \sum_{i=1}^d \varphi(b_i)$ is (with high probability) close to the mean of φ for any $\varphi \in L^2(\mathbb{S}(\mathbb{X}^d))$ with typical error of order at most $d^{-1/2}$, see (1.6); we leave open the question whether, for sufficiently smooth functions, the error is smaller than that. The question of testing whether a certain distribution on the unit sphere is approximately uniform was also raised in [4].

Further facts that are somewhat related to our result come from the field of *geometric probability*. Wendel [26] considered X_1, \dots, X_n independent u -distributed on $\mathbb{S}(\mathbb{R}^d)$ and computed the probability that there exists a hemisphere containing all n points. A result described in [22, p. 326] concerns random rotations R_1, \dots, R_n in $\mathbb{X}^d = \mathbb{R}^3$ that are independent u_G -distributed and provides a formula, for arbitrary convex sets $A, A' \subseteq \mathbb{S}(\mathbb{R}^3)$, for the probability that $A' \cap R_1(A) \cap \dots \cap R_n(A) \neq \emptyset$. Further similar results (and open problems) are described in [22,23].

The phenomenon of *concentration of measure* [16,17], which can occur in a space \mathbb{Y} equipped with both a metric and a measure, refers to the situation that most points $y \in \mathbb{Y}$ (in terms of the measure) are close (in terms of the metric) to a certain set that is small in terms of the measure. For example, this occurs for $\mathbb{Y} = \mathbb{S}(\mathbb{R}^d)$ in high dimension d , where most points are close to the equator. As a consequence known as Levy's lemma [17, p. 6], every 1-Lipschitz function $f : \mathbb{S}(\mathbb{R}^d) \rightarrow \mathbb{R}$ (i.e., with $|f(\mathbf{x}) - f(\mathbf{y})| \leq \text{distance}(\mathbf{x}, \mathbf{y})$) is almost constant, i.e., is close to its median (or mean, for that matter) at most points. Theorem 1 is somewhat similar, as it asserts (in Version 2) that the function f_A on the set $\text{ONB}(\mathbb{X}^d)$ of orthonormal bases of \mathbb{X}^d defined by $f_A(B) = d^{-1} \#(B \cap A)$ is almost constant for every A . (More generally, every function f on $\text{ONB}(\mathbb{X}^d)$ of the form $f(b_1, \dots, b_d) = \sum_{i=1}^d \varphi(b_i)$ with $\varphi \in L^2(\mathbb{S}(\mathbb{X}^d))$ is almost constant, in the sense expressed in Version 3 of Theorem 1.)

A fact related to Theorem 1 and concentration of measure is *Raz's lemma* [15,18,21], which roughly asserts the following: Let A be a subset of $\mathbb{S}(\mathbb{R}^d)$, and let $1 \ll k < d$. For most k -dimensional subspaces $U \subseteq \mathbb{R}^d$, $u_U(A \cap U) \approx u(A)$, where u_U is the normalized uniform measure on $\mathbb{S}(U)$.

1.3. Ideas of proof

Our proof of Theorem 1 is based on Theorem 2 below. Let $\text{Var}(Y)$ (and $\text{Cov}(X, Y)$) denote the variance (covariance) of the random variable Y (variables X and Y).

Theorem 2. *Let $d \geq 4$, let $\mathbb{X}^d = \mathbb{R}^d$ or $\mathbb{X}^d = \mathbb{C}^d$, let $\{b_1, \dots, b_d\} \sim u_{\text{ONB}(\mathbb{X}^d)}$, and let $\varphi \in L^2(\mathbb{S}(\mathbb{X}^d), u, \mathbb{R})$. Then*

$$|\text{Cov}(\varphi(b_1), \varphi(b_2))| \leq \frac{1}{d-1} \text{Var}_u(\varphi). \quad (1.7)$$

The estimate is sharp in the sense that for every d there exists a φ for which equality holds.

Theorem 2 expresses the fact that the b_i are *weakly* correlated. If they were independent, the covariance of $\varphi(b_1)$ and $\varphi(b_2)$ would be zero; since each b_i has distribution u , $\text{Var}(\varphi(b_i)) = \text{Var}_u(\varphi)$, and (1.7) states that the correlation coefficient of $\varphi(b_1)$ and $\varphi(b_2)$ is small (viz., no greater than $1/(d-1)$).

The proof of Theorem 1 (say, in Version 3) proceeds by noting that the random quantity $d^{-1} \sum_{i=1}^d \varphi(R(b_i))$ has expectation equal to the mean of φ and showing that it has small variance. The variance of a sum $\sum \varphi(R(b_i))$ is

the sum of the variances $\text{Var}(\varphi(R(b_i)))$ plus the sum of the covariances $\text{Cov}(\varphi(R(b_i)), \varphi(R(b_j)))$ for $i \neq j$; the variances can be computed, and the covariances can be estimated using Theorem 2. Chebyshev's inequality then yields Theorem 1.

The proof of Theorem 2 is, in turn, based on Theorem 3 below. Let \mathbf{x}^\perp denote the subspace orthogonal to $\mathbf{x} \in \mathbb{X}^d$,

$$\mathbf{x}^\perp = \{\mathbf{y} \in \mathbb{X}^d : \langle \mathbf{x} | \mathbf{y} \rangle = 0\}, \quad (1.8)$$

$\mathbb{S}(\mathbf{x}^\perp)$ the unit sphere in that subspace, and $u_{\mathbb{S}(\mathbf{x}^\perp)}$ the normalized uniform measure over that sphere. In the following, we use the double factorial notation

$$n!! = \begin{cases} 1 \cdot 3 \cdot 5 \cdots (n-2) \cdot n & \text{if } n \text{ is odd,} \\ 2 \cdot 4 \cdot 6 \cdots (n-2) \cdot n & \text{if } n \text{ is even,} \end{cases} \quad (1.9)$$

and $0!! = 1$.

Theorem 3. Suppose $d \geq 4$ and, again, $\mathbb{X}^d = \mathbb{R}^d$ or $\mathbb{X}^d = \mathbb{C}^d$. The equation

$$(T\psi)(\mathbf{x}) = \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) \psi(\mathbf{y}), \quad (1.10)$$

defines a bounded, self-adjoint operator $T : \mathcal{H} \rightarrow \mathcal{H}$ on the (∞ -dimensional, complex) Hilbert space $\mathcal{H} = L^2(\mathbb{S}(\mathbb{X}^d), u, \mathbb{C})$. Furthermore, T has pure point spectrum, and its eigenvalues are: for $\mathbb{X}^d = \mathbb{R}^d$,

$$0, 1, \text{ and } (-1)^{\ell/2} \frac{(\ell-1)!!(d-3)!!}{(\ell+d-3)!!} \quad \text{for } \ell = 2, 4, 6, \dots, \quad (1.11)$$

and for $\mathbb{X}^d = \mathbb{C}^d$,

$$0, 1, \text{ and } (-1)^\ell \binom{\ell+d-2}{\ell}^{-1} \quad \text{for } \ell = 1, 2, 3, \dots \quad (1.12)$$

For both $\mathbb{X}^d = \mathbb{R}^d$ or $\mathbb{X}^d = \mathbb{C}^d$, the largest absolute eigenvalue of T is 1, with a 1-dimensional eigenspace formed by the constant functions, and the second largest absolute eigenvalue of T is $1/(d-1)$.

The operator T is related to the Radon transformation, the differences being that one integrates only over the unit sphere, and that the only hyperplanes considered are those passing through the origin. In [15], this operator is called the *spherical Radon transformation*.

A result very similar to Theorem 2 is Theorem 5.2 in [15], which, however, neither implies nor is implied by our Theorem 2. To facilitate the comparison, we can express (1.7) in terms of the operator T introduced in (1.10) as

$$\left| \langle \varphi | T\varphi \rangle - \left(\int \varphi \right)^2 \right| \leq \frac{1}{d-1} \left\| \varphi - \int \varphi \right\|_{L^2}^2 \quad (1.13)$$

(with $\int \varphi = \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \varphi(\mathbf{x})$), and in fact it follows from Theorem 3 that

$$\left| \langle \chi | T\varphi \rangle - \left(\int \chi \right) \left(\int \varphi \right) \right| \leq \frac{1}{d-1} \left\| \chi - \int \chi \right\|_{L^2} \left\| \varphi - \int \varphi \right\|_{L^2}. \quad (1.14)$$

Theorem 5.2 in [15] provides a bound for the left-hand side of (1.14) in terms of the L^∞ norms of χ and φ in case these norms are not too large.

In Section 2, we provide proofs of Theorems 1–3 and Corollary 1; our proofs make repeated use of the rotational/unitary symmetry of the problem. In Section 3, we briefly outline a physical application discussed in detail in [8].

2. Proofs

2.1. Proof of Theorems 1–2 and Corollary 1 from Theorem 3

Proof of Version 1 of Theorem 1 from Version 2. Suppose Version 2 is true. Fix $\varepsilon, \delta > 0$, let $d \geq 4$ and $d \geq \delta^{-2}\varepsilon^{-1}$, and let B_0 be any fixed orthonormal basis in $\mathbb{X}^d = \mathbb{R}^d$ or \mathbb{C}^d . Then a random orthonormal basis B with distribution $u_{\text{ONB}(\mathbb{X}^d)}$ can be thought of as obtained from B_0 by applying a random rotation, $B = R^{-1}B_0$ with $R \sim u_G$ (which implies $R^{-1} \sim u_G$). Then

$$\#(B_0 \cap R(A)) = \#(R^{-1}(B_0) \cap A) = \#(B \cap A), \quad (2.1)$$

so (1.3) is equivalent to (1.4). \square

Proof of Corollary 1 from Version 2 of Theorem 1. Let $d \geq 4$, and let $E_{k,d}$ denote the event that

$$\left| \frac{\#(B \cap A_k)}{d} - u(A_k) \right| \leq \delta. \quad (2.2)$$

Version 2 of Theorem 1, with ε replaced by ε/m , yields that for any $d \geq m\delta^{-2}\varepsilon^{-1}$, $\mathbb{P}(E_{k,d}) \geq 1 - \varepsilon/m$, and thus $\mathbb{P}(E_{1,d} \cap \dots \cap E_{m,d}) \geq 1 - \varepsilon$. \square

Proof of Version 2 of Theorem 1 from Version 3. Let φ be the indicator function of A . Then φ lies in $L^2(\mathbb{S}(\mathbb{X}^d), u, \mathbb{R})$, has mean $u(A)$ and variance

$$\text{Var}_u(\varphi) = u(A)(1 - u(A)) \leq \frac{1}{4}. \quad (2.3)$$

If we think of $B \sim u_{\text{ONB}(\mathbb{X}^d)}$ again as obtained by applying a random rotation $R \sim u_G$ to a fixed orthonormal basis $B_0 = \{b_1, \dots, b_d\}$, then

$$\#(B \cap A) = \#(R(B_0) \cap A) = \sum_{j=1}^d \varphi(R(b_j)). \quad (2.4)$$

Thus, if we replace δ in Version 3 by 2δ , we obtain that Version 2 is true for $d \geq \frac{1}{2}\delta^{-2}\varepsilon^{-1}$, and in particular for $d \geq \delta^{-2}\varepsilon^{-1}$. (We dropped the factor $\frac{1}{2}$ in Version 2 for the sake of simplicity.) \square

Proof of Version 3 of Theorem 1 from Theorem 2. Let $\varepsilon, \delta > 0$, and let $R_j := R(b_j)$, so that $\{R_1, \dots, R_d\} \sim u_{\text{ONB}(\mathbb{X}^d)}$. Then

$$f(R) := \frac{1}{d} \sum_{j=1}^d \varphi(R_j) \quad (2.5)$$

has mean (since each R_j is u -distributed)

$$\mathbb{E}f(R) = \frac{1}{d} \sum_{i=1}^d \mathbb{E}\varphi(R_i) = \mathbb{E}_u(\varphi) \quad (2.6)$$

and variance (since the R_j are exchangeable)

$$\text{Var}(f(R)) = \frac{1}{d^2} \left(\sum_{i=1}^d \text{Var}(\varphi(R_i)) + \sum_{\substack{i,j=1 \\ i \neq j}}^d \text{Cov}(\varphi(R_i), \varphi(R_j)) \right) \quad (2.7)$$

$$= \frac{1}{d^2} (d \operatorname{Var}(\varphi(R_1)) + (d^2 - d) \operatorname{Cov}(\varphi(R_1), \varphi(R_2))) \quad (2.8)$$

$$\leq \frac{1}{d} \operatorname{Var}_u(\varphi) + \frac{d-1}{d} |\operatorname{Cov}(\varphi(R_1), \varphi(R_2))| \quad (2.9)$$

$$\leq \frac{2}{d} \operatorname{Var}_u(\varphi) \quad (2.10)$$

by (1.7) of Theorem 2 for $d \geq 4$. Chebyshev's inequality (see, e.g., [3, p. 65]) asserts that for any random variable X ,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \eta) \leq \frac{1}{\eta^2} \operatorname{Var}(X). \quad (2.11)$$

Setting $X = f(R)$ and $\eta = \delta \sqrt{\operatorname{Var}_u(\varphi)}$, we obtain that

$$\mathbb{P}(|f(R) - \mathbb{E}_u(\varphi)| \geq \delta \sqrt{\operatorname{Var}_u(\varphi)}) \leq \frac{1}{\delta^2 \operatorname{Var}_u(\varphi)} \operatorname{Var}(f(R)) \quad (2.12)$$

$$\leq \frac{2}{\delta^2 d}, \quad (2.13)$$

which yields (1.6) if $2/(\delta^2 d) \leq \varepsilon$, thus proving Version 3 of Theorem 1. \square

Proof of Theorem 2 from Theorem 3. Since (1.7), when true, will remain true if φ is changed by adding a constant, we can assume without loss of generality that φ has mean 0. Thus,

$$\operatorname{Var}_u(\varphi) = \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) |\varphi(\mathbf{x})|^2 = \|\varphi\|_{\mathcal{H}}^2. \quad (2.14)$$

Let $|1\rangle$ denote the constant 1 function in \mathcal{H} . The property of mean 0 can be expressed as $\langle 1|\varphi\rangle = 0$, or $\varphi \in |1\rangle^\perp$.

We can think of the joint distribution of b_1 and b_2 as follows: b_1 is chosen uniformly on $\mathbb{S}(\mathbb{X}^d)$, then b_2 is chosen uniformly on $\mathbb{S}(b_1^\perp)$. Thus,

$$\operatorname{Cov}(\varphi(b_1), \varphi(b_2)) = \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) \varphi(\mathbf{x}) \varphi(\mathbf{y}) \quad (2.15)$$

$$= \langle \varphi | T \varphi \rangle \quad (2.16)$$

by Theorem 3, with $\langle \cdot | \cdot \rangle$ the inner product in \mathcal{H} . Since, by Theorem 3, T is self-adjoint, and $\mathbb{C}|1\rangle$ is the eigenspace of T with eigenvalue 1, its orthogonal complement $|1\rangle^\perp$ is mapped by T to itself. Since φ lies in $|1\rangle^\perp$, we have by the Cauchy–Schwarz inequality that

$$|\langle \varphi | T \varphi \rangle| \leq \|\varphi\| \|T \varphi\| \quad (2.17)$$

$$\leq \|T\|_{|1\rangle^\perp} \|\varphi\|^2 \quad (2.18)$$

with $\|T\|_{|1\rangle^\perp}$ the operator norm of T on $|1\rangle^\perp$. By Theorem 3 again, T has pure point spectrum, so the operator norm of T on $|1\rangle^\perp$ is the largest absolute non-1 eigenvalue, which is $1/(d-1)$. Equality holds in (2.17) and (2.18) when φ is an associated eigenfunction. We thus have (1.7) for $d \geq 4$, including the statement that equality holds in (1.7) for suitable φ , viz., for the eigenfunction with absolute eigenvalue $1/(d-1)$. \square

2.2. Proof of Theorem 3 in the real case

For the proof of Theorem 3, we need Lemma 2 below, for which we offer two different proofs, the first of which is based on Lemma 1 below. Let S_ℓ denote the group of permutations of $\{1, \dots, \ell\}$.

Lemma 1. Let $d \geq 3$ and $\ell \geq 1$. Suppose the rank- ℓ tensor $A \in (\mathbb{R}^d)^{\otimes \ell}$ is symmetric,

$$A_{i_1 \dots i_\ell} = A_{i_{\sigma(1)} \dots i_{\sigma(\ell)}} \quad \forall \sigma \in S_\ell, \quad (2.19)$$

and invariant under (the obvious action of) the orthogonal group $O(d)$,

$$\sum_{j_1 \dots j_\ell=1}^d M_{i_1 j_1} \dots M_{i_\ell j_\ell} A_{j_1 \dots j_\ell} = A_{i_1 \dots i_\ell} \quad \forall M \in O(d). \quad (2.20)$$

If ℓ is odd then $A = 0$, and if ℓ is even then A is a multiple of \tilde{A} given by the symmetrization of $\delta_{i_1 i_2} \delta_{i_3 i_4} \dots \delta_{i_{\ell-1} i_\ell}$,

$$\tilde{A}_{i_1 \dots i_\ell} = \frac{1}{\ell!} \sum_{\sigma \in S_\ell} \delta_{i_{\sigma(1)} i_{\sigma(2)}} \delta_{i_{\sigma(3)} i_{\sigma(4)}} \dots \delta_{i_{\sigma(\ell-1)} i_{\sigma(\ell)}}, \quad (2.21)$$

where δ_{ij} is the Kronecker symbol (unit matrix).

Proof. The lemma can be translated into a statement about homogeneous polynomials $P(x_1, \dots, x_d)$ of degree ℓ ; the translation is based on writing such a polynomial in the form

$$P(x_1, \dots, x_d) = \sum_{i_1, \dots, i_\ell=1}^d A_{i_1 \dots i_\ell} x_{i_1} \dots x_{i_\ell}, \quad (2.22)$$

where the coefficients $A_{i_1 \dots i_\ell}$ can be taken to be symmetric under permutation of the indices. Lemma 1 is thus equivalent to the following:

Suppose the homogeneous polynomial $P(x_1, \dots, x_d)$ of degree² ℓ is $O(d)$ -invariant. If ℓ is odd then $P = 0$, and if ℓ is even then P is a multiple of $(x_1^2 + \dots + x_d^2)^{\ell/2}$.

To see this, note that since P is $O(d)$ -invariant, its restriction to $\mathbb{S}(\mathbb{R}^d)$ must be constant. Since P is homogeneous of degree ℓ , it must be of the form $c|\mathbf{x}|^\ell$, where $|\mathbf{x}| = \sqrt{x_1^2 + \dots + x_d^2}$, and c is a constant. If ℓ is odd, then invariance under the matrix $M \in O(d)$ with entries $M_{ij} = -\delta_{ij}$ implies that $P = -P$, so $P = 0$. \square

Lemma 2. Suppose $d \geq 3$ and $\ell \geq 1$. Let P be a homogeneous real polynomial of degree ℓ in d variables,

$$P(x_1, \dots, x_d) = \sum_{i_1 \dots i_\ell=1}^d C_{i_1 \dots i_\ell} x_{i_1} \dots x_{i_\ell} \quad (2.23)$$

with a symmetric tensor C . Then the average of P over the unit sphere is

$$\int_{\mathbb{S}(\mathbb{R}^d)} u(d\mathbf{x}) P(\mathbf{x}) = \begin{cases} 0 & \text{if } \ell \text{ odd,} \\ \alpha_{\ell,d} \sum_{i_1 \dots i_{\ell/2}=1}^d C_{i_1 i_2 i_3 \dots i_{\ell/2}} & \text{if } \ell \text{ even} \end{cases} \quad (2.24)$$

with

$$\alpha_{\ell,d} = \frac{(\ell-1)!!(d-2)!!}{(\ell+d-2)!!}. \quad (2.25)$$

First proof of Lemma 2. By linearity, the average of P must be

$$\sum_{i_1 \dots i_\ell=1}^d A_{i_1 \dots i_\ell} C_{i_1 \dots i_\ell} \quad (2.26)$$

²Although a polynomial of degree ℓ is usually taken to be non-zero, we include here the possibility $P = 0$.

with

$$A_{i_1 \dots i_\ell} = \int_{\mathbb{S}(\mathbb{R}^d)} u(d\mathbf{x}) x_{i_1} \cdots x_{i_\ell}. \quad (2.27)$$

The tensor A is symmetric and $O(d)$ -invariant. By Lemma 1, $A = 0$ for odd ℓ and

$$A = \alpha_{\ell,d} \tilde{A} \quad (2.28)$$

for even ℓ , with \tilde{A} as in Lemma 1 and some constant $\alpha_{\ell,d}$. Thus, for even ℓ , the average of P is given by (2.26) with A replaced by $\alpha_{\ell,d} \tilde{A}$. Since C is symmetric, this value is equal to

$$\alpha_{\ell,d} \sum_{i_1 \dots i_\ell=1}^d \delta_{i_1 i_2} \delta_{i_3 i_4} \cdots \delta_{i_{\ell-1} i_\ell} C_{i_1 \dots i_\ell}. \quad (2.29)$$

(That is, it is not necessary to symmetrize the product of the δ s, since C is symmetric.) This proves (2.24).

To compute $\alpha_{\ell,d}$ for even ℓ , it suffices to compare one nonzero component of A and \tilde{A} , say $A_{11\dots 1}$ (the average of x_1^ℓ) and $\tilde{A}_{11\dots 1} = 1$. To compute $A_{11\dots 1}$, we use spherical coordinates (with $r = 1$), setting $x_1 = \cos \theta$. Let

$$c(d) = \begin{cases} 2 & \text{if } d \text{ odd,} \\ \pi & \text{if } d \text{ even} \end{cases} \quad (2.30)$$

and

$$g(n) = \prod_{k=1}^n c(k) = \underbrace{2\pi 2\pi 2 \cdots}_{n \text{ factors of } 2 \text{ or } \pi} = \begin{cases} 2^{(n+1)/2} \pi^{(n-1)/2} & \text{if } n \text{ odd,} \\ (2\pi)^{n/2} & \text{if } n \text{ even.} \end{cases} \quad (2.31)$$

We note [19] that for $n \geq 2$, we have

$$|\mathbb{S}(\mathbb{R}^n)| = \frac{n\pi^{n/2}}{\Gamma(n/2 + 1)} = \frac{g(n)}{(n-2)!!} \quad (2.32)$$

for the surface area of $\mathbb{S}(\mathbb{R}^n)$. Thus, for even ℓ ,

$$\alpha_{\ell,d} = \int_{\mathbb{S}(\mathbb{R}^d)} x_1^\ell u(d\mathbf{x}) \quad (2.33)$$

$$= \frac{1}{|\mathbb{S}(\mathbb{R}^d)|} \int_0^\pi d\theta \cos^\ell \theta \sin^{d-2} \theta |\mathbb{S}(\mathbb{R}^{d-1})| \quad (2.34)$$

$$= \frac{(d-2)!!}{c(d)(d-3)!!} c(d) \frac{(\ell-1)!!(d-3)!!}{(\ell+d-2)!!} \quad (2.35)$$

$$= \frac{(\ell-1)!!(d-2)!!}{(\ell+d-2)!!} \quad (2.36)$$

using

$$\int_0^\pi d\theta \sin^p \theta \cos^q \theta = c(p) \frac{(q-1)!!(p-1)!!}{(p+q)!!} \quad (2.37)$$

if q is even. This proves (2.25). \square

Second proof of Lemma 2. This proof is based on Gaussianization (this strategy was suggested to us by B. Collins). Let $\mathbf{Y} = (Y_1, \dots, Y_d)$ be a random vector consisting of d independent standard normal random variables, and let $Z = |\mathbf{Y}|$ and $\mathbf{X} = \mathbf{Y}/Z$; then \mathbf{X} and Z are independent, and $\mathbf{X} \sim u$. For $P(x_1, \dots, x_d) = x_1^{n_1} \cdots x_d^{n_d}$ with $n_1 + \cdots + n_d = \ell$,

$\mathbb{E}P(\mathbf{Y})$ is, on the one hand, equal to $\mathbb{E}P(\mathbf{XZ}) = \mathbb{E}[Z^\ell P(\mathbf{X})] = \mathbb{E}Z^\ell \mathbb{E}P(\mathbf{X})$ (where the last factor is the quantity we want to compute) and, on the other hand, equal to the product of the n_j th moments of the standard normal distribution; it is known that the n th moment is 0 if n is odd and $(n-1)!!$ if n is even. Thus, $\mathbb{E}P(\mathbf{X}) = 0$ for odd ℓ . Since $Z^2 \sim \chi^2(d)$, we have that for even ℓ , $\mathbb{E}Z^\ell$ is the $\ell/2$ th moment of the χ^2 -distribution with d degrees of freedom, which is known [5] to be $(d + \ell - 2)!! / (d - 2)!!$. Thus,

$$\int_{\mathbb{S}(\mathbb{R}^d)} u(d\mathbf{x}) x_1^{n_1} \cdots x_d^{n_d} = \begin{cases} 0 & \text{if any } n_j \text{ is odd,} \\ \frac{(d-2)!!(n_1-1)!! \cdots (n_d-1)!!}{(d+\ell-2)!!} & \text{if all } n_j \text{ are even,} \end{cases} \quad (2.38)$$

which is equivalent to Lemma 2. \square

Proof of Theorem 3 in the real case $\mathbb{X}^d = \mathbb{R}^d$. We first show that the expression (1.10) defining T is well defined for any $\psi \in L^2 = L^2(\mathbb{S}(\mathbb{R}^d), u, \mathbb{C})$. In fact, it is well defined for any $\psi \in L^1 = L^1(\mathbb{S}(\mathbb{R}^d), u, \mathbb{C})$. (Note $L^2 \subset L^1$ for a finite measure space such as $(\mathbb{S}(\mathbb{R}^d), u)$.) To see this, we use that $u(d\mathbf{x})u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) = u(d\mathbf{y})u_{\mathbb{S}(\mathbf{y}^\perp)}(d\mathbf{x})$; indeed, both equal the unique rotation invariant measure on the subset where $\mathbf{x} \perp \mathbf{y}$. Now it follows that for $\psi \in L^1$,

$$\int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \left| \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) \psi(\mathbf{y}) \right| \leq \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) |\psi(\mathbf{y})| \quad (2.39)$$

$$= \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{y}) \int_{\mathbb{S}(\mathbf{y}^\perp)} u_{\mathbb{S}(\mathbf{y}^\perp)}(d\mathbf{x}) |\psi(\mathbf{y})| \quad (2.40)$$

$$= \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{y}) |\psi(\mathbf{y})|, \quad (2.41)$$

so $T\psi$ is well defined almost everywhere, lies in L^1 again, has norm $\|T\psi\|_1 \leq \|\psi\|_1$, and is independent of the choice of representative in the equivalence class that is a vector in L^1 . Since $L^2 \subset L^1$, the integral formula (1.10) is well defined also for any L^2 function. To see that $T\psi \in L^2$ for $\psi \in L^2$, note that by the Cauchy–Schwarz inequality, $|\int \mu(dx) f(x)|^2 \leq \int \mu(dx) |f(x)|^2$ for any normalized measure μ , so

$$|T\psi(\mathbf{x})|^2 \leq \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) |\psi(\mathbf{y})|^2, \quad (2.42)$$

and thus

$$\int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) |T\psi(\mathbf{x})|^2 \leq \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) |\psi(\mathbf{y})|^2 \quad (2.43)$$

$$\leq \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{y}) \int_{\mathbb{S}(\mathbf{y}^\perp)} u_{\mathbb{S}(\mathbf{y}^\perp)}(d\mathbf{x}) |\psi(\mathbf{y})|^2 \quad (2.44)$$

$$\leq \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{y}) |\psi(\mathbf{y})|^2, \quad (2.45)$$

so $T\psi \in L^2$ for $\psi \in L^2$ and $\|T\psi\|_2 \leq \|\psi\|_2$, so T is bounded. To see that it is self-adjoint, note that

$$\langle T\psi | \chi \rangle = \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \left(\int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) \psi(\mathbf{y}) \right)^* \chi(\mathbf{x}) \quad (2.46)$$

$$= \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{x}) \int_{\mathbb{S}(\mathbf{x}^\perp)} u_{\mathbb{S}(\mathbf{x}^\perp)}(d\mathbf{y}) \psi^*(\mathbf{y}) \chi(\mathbf{x}) \quad (2.47)$$

$$= \int_{\mathbb{S}(\mathbb{X}^d)} u(d\mathbf{y}) \int_{\mathbb{S}(\mathbf{y}^\perp)} u_{\mathbb{S}(\mathbf{y}^\perp)}(d\mathbf{x}) \psi^*(\mathbf{y}) \chi(\mathbf{x}) \quad (2.48)$$

$$= \langle \psi | T\chi \rangle. \quad (2.49)$$

Next, observe that T is $O(d)$ -invariant,

$$U(M)TU(M)^{-1} = T \quad \forall M \in O(d), \quad (2.50)$$

where $(U(M)\psi)(\mathbf{x}) = \psi(M\mathbf{x})$. For $\ell = 0, 1, 2, \dots$, let \mathcal{A}_ℓ be the set of all harmonic homogeneous polynomials of degree ℓ in d variables; for $\ell = 0$ and 1 , \mathcal{A}_ℓ is just the set of homogeneous polynomials of degree ℓ , while for $\ell \geq 2$, the elements are of the form

$$P(x_1, \dots, x_d) = \sum_{i_1 \dots i_\ell = 1}^d C_{i_1 \dots i_\ell} x_{i_1} \dots x_{i_\ell} \quad (2.51)$$

with traceless symmetric C , i.e.,

$$\sum_{i=1}^d C_{i_1 \dots i_{\ell-2} ii} = 0 \quad (2.52)$$

for all $i_1, \dots, i_{\ell-2} \in \{1, \dots, d\}$. Let \mathcal{H}_ℓ be the set of the restrictions of the \mathcal{A}_ℓ functions to $\mathbb{S}(\mathbb{R}^d)$. The functions in \mathcal{H}_ℓ form the d -dimensional analog of the spherical harmonics. It is known (e.g., [24,25]) that the \mathcal{H}_ℓ are irreducible representation spaces of $O(d)$, that they are pairwise inequivalent representations, that they are mutually orthogonal in \mathcal{H} , and that together they span \mathcal{H} in the L^2 norm,

$$\mathcal{H} = \bigoplus_{\ell=0}^{\infty} \mathcal{H}_\ell. \quad (2.53)$$

From this it follows by Schur's lemma that T , since it is $O(d)$ -invariant, is a multiple of the identity on each \mathcal{H}_ℓ . Thus, T has pure point spectrum, and each eigenspace must be either one of the \mathcal{H}_ℓ or the sum of several of the \mathcal{H}_ℓ . (This observation was made before in [15].)

To compute the eigenvalue τ_ℓ of T on \mathcal{H}_ℓ , it suffices to consider any $P \in \mathcal{H}_\ell$ and compare the average of P over the equator $\mathbb{S}(\mathbb{R}^{d-1}) = \{\mathbf{x} \in \mathbb{S}(\mathbb{R}^d) : x_d = 0\}$, or $TP(0, 0, \dots, 1)$, with the value of P at the north pole, $P(0, \dots, 0, 1) = C_{ddd\dots d}$.

By Lemma 2, the average of $P(\mathbf{x})$ with traceless C over the equator is 0 for odd ℓ , while for even $\ell \geq 2$ it is

$$\int_{\mathbb{S}(\mathbb{R}^{d-1})} u_{d-1}(d\mathbf{x}) P(\mathbf{x}) = \alpha_{\ell, d-1} \sum_{i_1 \dots i_{\ell/2} = 1}^{d-1} C_{i_1 i_1 i_2 i_2 \dots i_{\ell/2} i_{\ell/2}} \quad (2.54)$$

$$= -\alpha_{\ell, d-1} \sum_{i_2 \dots i_{\ell/2} = 1}^{d-1} C_{dd i_2 i_2 \dots i_{\ell/2} i_{\ell/2}} \quad (2.55)$$

$$= (-1)^2 \alpha_{\ell, d-1} \sum_{i_3 \dots i_{\ell/2} = 1}^{d-1} C_{dddd i_3 i_3 \dots i_{\ell/2} i_{\ell/2}} \quad (2.56)$$

$$= (-1)^{\ell/2} \alpha_{\ell, d-1} C_{dd\dots d}. \quad (2.57)$$

Thus, the eigenvalue of the operator T on \mathcal{H}_ℓ is

$$\tau_\ell = \begin{cases} 0 & \text{if } \ell \text{ odd,} \\ (-1)^{\ell/2} \alpha_{\ell, d-1} & \text{if } \ell \geq 2 \text{ even.} \end{cases} \quad (2.58)$$

We can now identify the largest absolute eigenvalues. Since, by (2.25),

$$\alpha_{\ell+2, d} = \frac{\ell+1}{\ell+d} \alpha_{\ell, d} < \alpha_{\ell, d}, \quad (2.59)$$

we have that

$$\max_{\ell=2,4,6,\dots} \alpha_{\ell,d} = \alpha_{2,d} = \frac{1}{d}. \quad (2.60)$$

Thus, 1 does not occur as an eigenvalue except for constant functions, and the largest absolute non-1 eigenvalue is $\alpha_{2,d-1} = 1/(d-1)$. \square

2.3. Proof of Theorem 3 in the complex case

Lemmas 3 and 4 provide the complex analogs of Lemmas 1 and 2. Lemma 4 is equivalent to Theorem 18 in [6]; it is proved there using Gaussianization and here in a different way using Lemma 3.

Lemma 3. *Let $d \geq 2$ and $\ell, \ell' \in \{0, 1, 2, \dots\}$. Suppose the rank- $(\ell + \ell')$ tensor $A \in (\mathbb{C}^d)^{\otimes(\ell+\ell')}$ is symmetric in the first ℓ indices and symmetric in the last ℓ' indices,*

$$A_{i_1 \dots i_\ell i'_1 \dots i'_{\ell'}} = A_{i_{\sigma(1)} \dots i_{\sigma(\ell)} i'_{\sigma'(1)} \dots i'_{\sigma'(\ell')}} \quad \forall \sigma \in S_\ell \quad \forall \sigma' \in S_{\ell'}, \quad (2.61)$$

and invariant under $U(d)$, acting in the obvious way on the first ℓ indices and in the conjugate way on the last ℓ' indices,

$$\sum_{j_1 \dots j_\ell, j'_1 \dots j'_{\ell'}=1}^d M_{i_1 j_1} \dots M_{i_\ell j_\ell} \overline{M}_{i'_1 j'_1} \dots \overline{M}_{i'_{\ell'} j'_{\ell'}} A_{j_1 \dots j_\ell j'_1 \dots j'_{\ell'}} = A_{i_1 \dots i_\ell i'_1 \dots i'_{\ell'}} \quad \forall M \in U(d). \quad (2.62)$$

If $\ell \neq \ell'$ then $A = 0$, and if $\ell = \ell'$ then A is a multiple of \tilde{A} given by the symmetrization of $\delta_{i_1 i'_1} \delta_{i_2 i'_2} \dots \delta_{i_\ell i'_\ell}$ in either the primed or the unprimed indices,

$$\tilde{A}_{i_1 \dots i_\ell i'_1 \dots i'_{\ell'}} = \frac{1}{\ell!} \sum_{\sigma \in S_\ell} \delta_{i_{\sigma(1)} i'_1} \delta_{i_{\sigma(2)} i'_2} \dots \delta_{i_{\sigma(\ell)} i'_\ell}. \quad (2.63)$$

Proof. Also this lemma can be translated into a statement about polynomials. The relevant polynomials to consider are the polynomials $P(z_1, \dots, z_d, \bar{z}_1, \dots, \bar{z}_d)$ homogeneous of degree ℓ in z and degree ℓ' in \bar{z} ; they can be thought of as complex polynomials in $2d$ complex variables, with the conjugates of z_1, \dots, z_d inserted as the last d variables; they can be written as

$$P(z_1, \dots, z_d, \bar{z}_1, \dots, \bar{z}_d) = \sum_{i_1, \dots, i_\ell, i'_1, \dots, i'_{\ell'}=1}^d A_{i_1 \dots i_\ell, i'_1 \dots i'_{\ell'}} z_{i_1} \dots z_{i_\ell} \bar{z}_{i'_1} \dots \bar{z}_{i'_{\ell'}}. \quad (2.64)$$

The pair (ℓ, ℓ') is called the *bi-degree*³ of P . Lemma 3 can then be paraphrased as:

Suppose the bi-homogeneous polynomial $P(z_1, \dots, z_d, \bar{z}_1, \dots, \bar{z}_d)$ of bi-degree (ℓ, ℓ') is $U(d)$ -invariant. If $\ell \neq \ell'$ then $P = 0$, and if $\ell = \ell'$ then P is a multiple of $(|z_1|^2 + \dots + |z_d|^2)^\ell$.

Considering (2.62) for $M = e^{i\theta} I$ with $\theta \in \mathbb{R}$ and I the identity matrix, we obtain that $e^{i(\ell-\ell')\theta} A = A$, so for $\ell \neq \ell'$ we have that $A = 0$ (and $P = 0$). Now assume $\ell = \ell'$. Since P is $U(d)$ -invariant, its restriction to $\mathbb{S}(\mathbb{C}^d)$ must be constant. Since $P(\mathbf{z}, \bar{\mathbf{z}})$ is real-homogeneous of degree 2ℓ , it must be of the form $c|\mathbf{z}|^{2\ell}$, where c is a complex constant and $|\mathbf{z}| = \sqrt{|z_1|^2 + \dots + |z_d|^2}$. \square

Lemma 4. *Suppose $d \geq 2$ and $\ell, \ell' \in \{0, 1, 2, \dots\}$. Let $P(\mathbf{z}, \bar{\mathbf{z}})$ be a bi-homogeneous polynomial of bi-degree (ℓ, ℓ') ,*

$$P(\mathbf{z}, \bar{\mathbf{z}}) = \sum_{i_1 \dots i_\ell, i'_1 \dots i'_{\ell'}=1}^d C_{i_1 \dots i_\ell, i'_1 \dots i'_{\ell'}} z_{i_1} \dots z_{i_\ell} \bar{z}_{i'_1} \dots \bar{z}_{i'_{\ell'}} \quad (2.65)$$

³Again, we include the possibility $P = 0$.

with a complex tensor C that is symmetric in the primed and in the unprimed indices. Then the average of P over the unit sphere is

$$\int_{\mathbb{S}(\mathbb{C}^d)} u(d\mathbf{z}) P(\mathbf{z}, \bar{\mathbf{z}}) = \begin{cases} 0 & \text{if } \ell \neq \ell', \\ \beta_{\ell,d} \sum_{i_1 \dots i_{\ell}=1}^d C_{i_1 \dots i_{\ell} i_1' \dots i_{\ell}'} & \text{if } \ell = \ell' \geq 1 \end{cases} \quad (2.66)$$

with

$$\beta_{\ell,d} = \binom{\ell + d - 1}{\ell}^{-1}. \quad (2.67)$$

Proof. By linearity, the average of P must be

$$\sum_{i_1 \dots i_{\ell}, i_1' \dots i_{\ell}'}^d A_{i_1 \dots i_{\ell}, i_1' \dots i_{\ell}'} C_{i_1 \dots i_{\ell}, i_1' \dots i_{\ell}'} \quad (2.68)$$

with

$$A_{i_1 \dots i_{\ell}, i_1' \dots i_{\ell}'} = \int_{\mathbb{S}(\mathbb{C}^d)} u(d\mathbf{z}) z_{i_1} \dots z_{i_{\ell}} \bar{z}_{i_1'} \dots \bar{z}_{i_{\ell}'} \quad (2.69)$$

The tensor A is symmetric in the first ℓ variables and symmetric in the last ℓ' variables, and $U(d)$ -invariant in the sense of (2.62). Lemma 3 now yields $A = \beta_{\ell,d} \tilde{A}$ and thus (2.66) except for the value of the constant $\beta_{\ell,d}$.

To compute $\beta_{\ell,d}$, note that $\tilde{A}_{1 \dots 1 1 \dots 1} = 1$, so

$$\beta_{\ell,d} = A_{1 \dots 1 1 \dots 1} \quad (2.70)$$

$$= \int_{\mathbb{S}(\mathbb{C}^d)} u(d\mathbf{z}) |z_1|^{2\ell} \quad (2.71)$$

$$= \int_{\mathbb{S}(\mathbb{R}^{2d})} u(d\mathbf{x}) (x_1^2 + x_2^2)^{\ell} \quad (2.72)$$

$$= \sum_{k=0}^{\ell} \binom{\ell}{k} \int_{\mathbb{S}(\mathbb{R}^{2d})} u(d\mathbf{x}) x_1^{2k} x_2^{2(\ell-k)} \quad (2.73)$$

[using (2.24) for $P(x_1, \dots, x_{2d}) = x_1^{2k} x_2^{2(\ell-k)}$, which has $C_{i_1 \dots i_{2d}} = \binom{2\ell}{2k}^{-1}$ if $2k$ of the i_j are 1 and the others are 2, and $C_{i_1 \dots i_{2d}} = 0$ otherwise, so the sum in (2.24) has $\binom{\ell}{k}$ nonzero terms]

$$= \alpha_{2\ell, 2d} \sum_{k=0}^{\ell} \binom{\ell}{k}^2 \binom{2\ell}{2k}^{-1} \quad (2.74)$$

$$= \alpha_{2\ell, 2d} 4^{\ell} \binom{2\ell}{\ell}^{-1}, \quad (2.75)$$

where the last step can be obtained either from Gauss's theorem about the hypergeometric function ${}_2F_1(a, b; c; z)$ at $z = 1$ [11], or using Zeilberger's algorithm [28]. One easily verifies that (2.74) is equal to (2.67).

Alternatively, starting from (2.72), we can evaluate this integral by noting that for $\mathbf{x} = (x_1, \dots, x_D) = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \in \mathbb{R}^D$ with $\mathbf{x}^{(1)} = (x_1, \dots, x_n)$ and $\mathbf{x}^{(2)} = (x_{n+1}, \dots, x_{n+m})$ such that $n + m = D$, we have that for the D -dimensional volume measure,

$$d\mathbf{x} = d\mathbf{x}^{(1)} d\mathbf{x}^{(2)} \quad (2.76)$$

$$= r_1^{n-1} d\Omega^{(1)} dr_1 r_2^{m-1} d\Omega^{(2)} dr_2 \quad (2.77)$$

$$= \rho r_1^{n-1} r_2^{m-1} d\Omega^{(1)} d\Omega^{(2)} d\theta d\rho, \quad (2.78)$$

where $r_i = |\mathbf{x}^{(i)}|$, $d\Omega^{(i)}$ is the solid angle for $\mathbf{x}^{(i)}/r_i$, $\rho = |\mathbf{x}| = \sqrt{r_1^2 + r_2^2}$, $\cos \theta = r_1/\rho$ ($0 \leq \theta \leq \pi/2$), so $r_2 = \rho \sin \theta$. This yields, for $n = 2$ and $D = 2d$,

$$\beta_{\ell,d} = \int_{\mathbb{S}(\mathbb{R}^{2d})} u(d\mathbf{x}) (x_1^2 + x_2^2)^\ell \quad (2.79)$$

$$= \frac{1}{|\mathbb{S}(\mathbb{R}^{2d})|} \int_0^{\pi/2} d\theta \cos^{2\ell} \theta \cos \theta \sin^{2d-3} \theta 2\pi |\mathbb{S}(\mathbb{R}^{2d-2})| \quad (2.80)$$

$$= 2\pi \frac{(2d-2)!!g(2d-2)}{(2d-4)!!g(2d)} \int_0^{\pi/2} d\theta \sin^{2d-3} \theta \cos^{2\ell+1} \theta \quad (2.81)$$

$$= (2d-2) \frac{(2\ell)!!(2d-4)!!}{(2d+2\ell-2)!!} \quad (2.82)$$

$$= \frac{(2\ell)!!(2d-2)!!}{(2d+2\ell-2)!!} \quad (2.83)$$

using

$$\int_0^{\pi/2} d\theta \sin^p \theta \cos^q \theta = \frac{(q-1)!!(p-1)!!}{(p+q)!!} \quad (2.84)$$

for odd q . One easily verifies that (2.83) is equal to (2.67). \square

Proof of Theorem 3 in the complex case $\mathbb{X}^d = \mathbb{C}^d$. By the same reasoning as in the real case, involving (2.39)–(2.49), T must be self-adjoint and bounded. Clearly, it is $U(d)$ -invariant. For $\ell, \ell' \in \{0, 1, 2, \dots\}$, let $\mathcal{A}_{\ell\ell'}$ be the set of all harmonic bi-homogeneous polynomials $P(z_1, \dots, z_d, \bar{z}_1, \dots, \bar{z}_d)$ of bi-degree (ℓ, ℓ') ; for $\ell, \ell' \geq 1$, they are of the form

$$P(\mathbf{z}, \bar{\mathbf{z}}) = \sum_{i_1 \dots i_\ell, i'_1 \dots i'_{\ell'}=1}^d C_{i_1 \dots i_\ell i'_1 \dots i'_{\ell'}} z_{i_1} \dots z_{i_\ell} \bar{z}_{i'_1} \dots \bar{z}_{i'_{\ell'}} \quad (2.85)$$

with a tensor C that is symmetric in the sense of (2.61) and traceless in the sense that

$$\sum_{i=1}^d C_{i_1 \dots i_{\ell-1} i, i'_1 \dots i'_{\ell'-1} i} = 0 \quad (2.86)$$

for all $i_1, \dots, i_{\ell-1}, i'_1, \dots, i'_{\ell'-1} \in \{1, \dots, d\}$. Let $\mathcal{H}_{\ell\ell'}$ be the set of the restrictions of the $\mathcal{A}_{\ell\ell'}$ functions to $\mathbb{S}(\mathbb{C}^d)$. The functions in $\mathcal{H}_{\ell\ell'}$ form the complex analog of the spherical harmonics. It is known (e.g., [25, p. 296]) that the $\mathcal{H}_{\ell\ell'}$ are irreducible representation spaces of $U(d)$, that they are pairwise inequivalent representations [25, p. 296] that they are mutually orthogonal [25, p. 293], and that together they span \mathcal{H} in the L^2 norm [25, p. 294],

$$\mathcal{H} = \bigoplus_{\ell, \ell'=0}^{\infty} \mathcal{H}_{\ell\ell'}. \quad (2.87)$$

By the same reasoning as in the real case, each eigenspace of T must be either one $\mathcal{H}_{\ell\ell'}$ or the sum of several ones.

To compute the eigenvalue $\tau_{\ell\ell'}$ of T on $\mathcal{H}_{\ell\ell'}$, we consider any $P \in \mathcal{H}_{\ell\ell'}$ and compare the average of P over the equator $\mathbb{S}(\mathbb{C}^{d-1}) = \{\mathbf{z} \in \mathbb{S}(\mathbb{C}^d) : z_d = 0\}$, or $TP(0, 0, \dots, 1)$, with the value of P at the north pole, $P(0, \dots, 0, 1) = C_{d \dots d d \dots d}$.

By Lemma 4, the average of P over the equator is 0 for $\ell \neq \ell'$, while for $\ell = \ell' \geq 1$ it is

$$\int_{\mathbb{S}(\mathbb{C}^{d-1})} u_{d-1}(d\mathbf{z}) P(\mathbf{z}, \bar{\mathbf{z}}) = \beta_{\ell, d-1} \sum_{i_1 \dots i_\ell = 1}^{d-1} C_{i_1 \dots i_\ell, i_1 \dots i_\ell} \quad (2.88)$$

$$= -\beta_{\ell, d-1} \sum_{i_1 \dots i_{\ell-1} = 1}^{d-1} C_{i_1 \dots i_{\ell-1} d, i_1 \dots i_{\ell-1} d} \quad (2.89)$$

$$= (-1)^\ell \beta_{\ell, d-1} C_{dd \dots d}. \quad (2.90)$$

Thus, the eigenvalue of the operator T on $\mathcal{H}_{\ell\ell'}$ is

$$\tau_{\ell\ell'} = \begin{cases} 0 & \text{if } \ell \neq \ell', \\ (-1)^\ell \beta_{\ell, d-1} & \text{if } \ell = \ell' \geq 1. \end{cases} \quad (2.91)$$

For $\ell = \ell' = 0$, of course, $\mathcal{H}_{\ell\ell'} = \mathcal{H}_{00}$ consists of the constant functions, and the eigenvalue is $\tau_{00} = 1$. We can now identify the largest absolute eigenvalues. Since, by (2.67),

$$\beta_{\ell+1, d} = \frac{\ell+1}{\ell+d} \beta_{\ell, d} < \beta_{\ell, d}, \quad (2.92)$$

we have that

$$\max_{\ell=1,2,3,\dots} \beta_{\ell, d} = \beta_{1, d} = \frac{1}{d}. \quad (2.93)$$

Thus, 1 does not occur as an eigenvalue except for constant functions, and the largest absolute non-1 eigenvalue is $\beta_{1, d-1} = 1/(d-1)$. \square

3. Application

A physical application of our results, described in detail in [8], concerns quantum statistical mechanics, in particular the distribution of the wave function in thermal equilibrium.

3.1. Setup

Consider any quantum system S weakly coupled to another system B with a large (but finite) number of particles; B is called the “heat bath.” Suppose that the composite system $S \cup B$ is isolated, with Hilbert space

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B \quad (3.1)$$

and the Hamiltonian

$$H = H_S \otimes I_B + I_S \otimes H_B + H_{SB}, \quad (3.2)$$

where I denotes the identity operator, and the interaction term H_{SB} is assumed to be small and will be neglected for much of the reasoning. In correspondence to the physical assumption that $S \cup B$ is constrained to a finite volume of 3-space, we assume that H has pure point spectrum. Consider an energy interval $[E, E + \delta E]$ that is small on the macroscopic scale but large enough to contain many eigenvalues of H . Let the “micro-canonical” subspace \mathcal{H}_{mc} of \mathcal{H} be the spectral subspace corresponding to $[E, E + \delta E]$, i.e., \mathcal{H}_{mc} is spanned by the eigenvectors of H with eigenvalues between E and $E + \delta E$, and suppose that $S \cup B$ is in a pure state ψ in \mathcal{H}_{mc} . Without loss of generality, \mathcal{H}_{mc} , \mathcal{H}_S , and \mathcal{H}_B can be taken to have finite dimension, while $\dim \mathcal{H}_{mc}$ and $\dim \mathcal{H}_B$ should be large (like $\exp(10^{10})$). Most wave

functions $\psi \in \mathbb{S}(\mathcal{H}_{mc})$ (“most” relative to $u_{\mathbb{S}(\mathcal{H}_{mc})}$) represent states of thermal equilibrium. According to a fact known as “canonical typicality” [7,10,20], most $\psi \in \mathbb{S}(\mathcal{H}_{mc})$ are such that, for $\dim \mathcal{H}_S \ll \dim \mathcal{H}_{mc}$,

$$\rho_S^\psi \approx \rho_\beta, \quad (3.3)$$

where ρ_S^ψ denotes the reduced density matrix of S ,

$$\rho_S^\psi := \text{tr}_B |\psi\rangle\langle\psi|, \quad (3.4)$$

and ρ_β the “canonical” density matrix associated with inverse temperature $\beta = 1/kT$ (k = Boltzmann’s constant, T = temperature),

$$\rho_\beta := \frac{1}{Z} e^{-\beta H} \quad (3.5)$$

with $Z = \text{tr} e^{-\beta H}$; the value of β is determined by E and the sizes of S and B .

3.2. Conditional wave function

As first pointed out in [9], it is also true for most $\psi \in \mathbb{S}(\mathcal{H}_{mc})$ that the “conditional wave function” ψ_S of system S (see below) has a probability distribution that depends only on H_S and β (and thus does not depend on H_B , H_{SB} if small enough, or on the details of ψ), called the “thermal equilibrium distribution of ψ_S .” This distribution is $\text{GAP}(\rho_\beta)$, the Gaussian Adjusted Projected measure with covariance operator ρ_β [9]. The mathematical proof [8] of this statement of “GAP typicality” is where Theorems 1 and 2 are useful.

To explain this further, we first elucidate the concept of “conditional wave function.” Given an orthonormal basis $\{b_1, \dots, b_d\}$ of \mathcal{H}_B and a vector $\psi \in \mathbb{S}(\mathcal{H})$, the conditional wave function ψ_S is a random vector in $\mathbb{S}(\mathcal{H}_S)$, obtained from ψ by means of the partial inner product,

$$\psi_S = \frac{1}{\mathcal{N}} \langle b_J | \psi \rangle_B, \quad (3.6)$$

with a random basis vector b_J , chosen with the Born-rule distribution

$$\mathbb{P}(J = j) = \|\langle b_j | \psi \rangle_B\|_{\mathcal{H}_S}. \quad (3.7)$$

(\mathcal{N} is a normalizing factor, and the partial inner product $\phi = \langle b | \psi \rangle_B$ is defined by the property $\langle \chi | \phi \rangle_{\mathcal{H}_S} = \langle \chi \otimes b | \psi \rangle_{\mathcal{H}}$.) Usually, ψ_S depends on ψ as well as on the basis $\{b_1, \dots, b_d\}$; however, in the special situation of thermal equilibrium, the distribution does not depend on the choice of basis, nor (as already mentioned) on ψ (except through H_S and β).

A key to proving GAP typicality is this statement: *If $\{b_1, \dots, b_d\}$ is a random orthonormal basis of \mathcal{H}_B then, for every $\psi \in \mathbb{S}(\mathcal{H})$, the distribution of ψ_S is close to $\text{GAP}(\rho_S^\psi)$ with probability near 1.* To prove this statement, two things are relevant: First, that when, for fixed ψ , the distribution of ψ_S on $\mathbb{S}(\mathcal{H}_S)$ is averaged over all orthonormal bases $\{b_1, \dots, b_d\}$, the result is $\text{GAP}(\rho_S^\psi)$. And second, the result of the present paper. That is because the distribution of ψ_S is actually of the form

$$\frac{1}{d} \sum_{i=1}^d \varphi(b_i), \quad (3.8)$$

where φ is a function on $\mathbb{S}(\mathcal{H}_B)$ that yields measures on $\mathbb{S}(\mathcal{H}_S)$ as values. Theorem 1 shows that the measure (3.8) will, with high probability, be close to its average

$$\int_{\mathbb{S}(\mathcal{H}_B)} u(d\mathbf{x}) \varphi(\mathbf{x}) = \text{GAP}(\rho_S^\psi), \quad (3.9)$$

as claimed.

Acknowledgements

We thank József Beck, Benoît Collins, Roe Goodman, Neil Sloane, and Doron Zeilberger for helpful discussions. We acknowledge support from the National Science Foundation [grant DMS-0504504 to Sheldon Goldstein; DMR 08-02120 to Joel L. Lebowitz], the Air Force Office of Scientific Research [grant AF-FA 49620-01-0154 to Joel L. Lebowitz], the John Templeton Foundation [grant 37433 to Sheldon Goldstein and Roderich Tumulka], the European Cooperation in Science and Technology [COST action MP1006 to Nino Zanghì], and Istituto Nazionale di Fisica Nucleare [to Nino Zanghì].

References

- [1] J. Beck. Deterministic approach to the kinetic theory of gases. *J. Stat. Phys.* **138** (2010) 160–269. [MR2594896](#)
- [2] J. Beck. Super-uniformity of the typical billiard path. In *An Irregular Mind: Szemerédi Is 70* 39–129. I. Bárány and J. Solymosi (Eds). Springer, Berlin, 2010. [MR2815601](#)
- [3] P. Billingsley. *Probability and Measure*. Wiley, New York, 1986. [MR0830424](#)
- [4] T. Cai, J. Fan and T. Jiang. Distributions of angles in random packing on spheres. *J. Mach. Learn. Res.* **14** (2013) 1837–1864. [MR3104497](#)
- [5] Chi-squared distribution. In *Wikipedia, the free encyclopedia*. Available at http://en.wikipedia.org/wiki/Chi-squared_distribution (accessed 6/8/2014).
- [6] T. Cioppa and B. Collins. Matrix units in the symmetric group algebra, and unitary integration. Preprint. Available at <http://arxiv.org/abs/1307.4766>.
- [7] J. Gemmer, G. Mahler and M. Michel. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems. Lecture Notes in Physics* **657**. Springer, Berlin, 2004. [MR2184466](#)
- [8] S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka and N. Zanghì. Universal probability distribution for the wave function of a quantum system entangled with its environment. *Comm. Math. Phys.* **342** (2016) 965–988. [MR3465436](#)
- [9] S. Goldstein, J. L. Lebowitz, R. Tumulka and N. Zanghì. On the distribution of the wave function for systems in thermal equilibrium. *J. Stat. Phys.* **125** (2006) 1193–1221. Available at <http://arxiv.org/abs/quant-ph/0309021>. [MR2282486](#)
- [10] S. Goldstein, J. L. Lebowitz, R. Tumulka and N. Zanghì. Canonical typicality. *Phys. Rev. Lett.* **96**, 050403 (2006). Available at <http://arxiv.org/abs/cond-mat/0511091>. [MR2204925](#)
- [11] Hypergeometric function. In *Wikipedia, the free encyclopedia*. Available at http://en.wikipedia.org/wiki/Hypergeometric_function (accessed 3/31/2014).
- [12] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.* **34** (4) (2006) 1497–1529. [MR2257653](#)
- [13] T. Jiang. Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles. *Probab. Theory Related Fields* **144** (1) (2009) 221–246. [MR2480790](#)
- [14] T. Jiang. The entries of Haar-invariant matrices from the classical compact groups. *J. Theoret. Probab.* **23** (4) (2010) 1227–1243. [MR2735744](#)
- [15] B. Klartag and O. Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *STOC '11: Proceedings of the 43rd ACM Symposium on the Theory of Computing* 31–40. ACM, New York, 2011. Available at <http://arxiv.org/abs/1009.3640>. [MR2931952](#)
- [16] M. Ledoux. *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, RI, 2001. [MR1849347](#)
- [17] V. D. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces. Lecture Notes in Mathematics* **1200**. Springer, Berlin, 1986. [MR0856576](#)
- [18] V. Milman and R. Wagner. Some remarks on a lemma of Ran Raz. In *Geometric Aspects of Functional Analysis* 158–168. V. Milman and G. Schechtman (Eds). *Lecture Notes in Mathematics* **1807**. Springer, Berlin, 2003. [MR2083396](#)
- [19] N-sphere. In *Wikipedia, the free encyclopedia*. Available at <http://en.wikipedia.org/wiki/N-sphere> (accessed 3/31/2014).
- [20] S. Popescu, A. J. Short and A. Winter. Entanglement and the foundation of statistical mechanics. *Nature Physics* **21** (11) (2006) 754–758.
- [21] R. Raz. Exponential separation of quantum and classical communication complexity. In *STOC '99: Proceedings of the 31st Annual ACM Symposium on the Theory of Computing* 358–367. ACM, New York, 1999. [MR1798056](#)
- [22] L. A. Santaló. *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading, MA, 1976. [MR0433364](#)
- [23] H. Solomon. *Geometric Probability*. SIAM, Philadelphia, 1978. [MR0488215](#)
- [24] Spherical harmonics. In *Wikipedia, the free encyclopedia*. Available at http://en.wikipedia.org/wiki/Spherical_harmonics (accessed 3/31/2014).
- [25] N. Ja. Vilenkin and A. U. Klimyk. *Representations of Lie Groups and Special Functions*. V. A. Groza and A. A. Groza (Transl.). Volume 2: *Class I Representations, Special Functions, and Integral Transforms*. Kluwer, Dordrecht, 1993. [MR1220225](#)
- [26] J. G. Wendel. A problem in geometric probability. *Math. Scand.* **11** (1962) 109–111. [MR0146858](#)
- [27] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.* **77** (3) (1916) 313–352. [MR1511862](#)
- [28] D. Zeilberger. The method of creative telescoping. *J. Symbolic Comput.* **11** (3) (1991) 195–204. [MR1103727](#)