

New tests of uniformity on the compact classical groups as diagnostics for weak-* mixing of Markov chains

AMIR SEPEHRI

Department of Statistics, Sequoia Hall, Stanford, CA 94305, USA. E-mail: asepehri@stanford.edu

This paper introduces two new families of non-parametric tests of goodness-of-fit on the compact classical groups. One of them is a family of tests for the eigenvalue distribution induced by the uniform distribution, which is consistent against all fixed alternatives. The other is a family of tests for the uniform distribution on the entire group, which is again consistent against all fixed alternatives. The construction of these tests heavily employs facts and techniques from the representation theory of compact groups. In particular, new Cauchy identities are derived and proved for the characters of compact classical groups, in order to accommodate the computation of the test statistic. We find the asymptotic distribution under the null and general alternatives. The tests are proved to be asymptotically admissible. Local power is derived and the global properties of the power function against local alternatives are explored.

The new tests are validated on two random walks for which the mixing-time is studied in the literature. The new tests, and several others, are applied to the Markov chain sampler proposed by Jones, Osipov and Rokhlin [*Proc. Natl. Acad. Sci.* **108** (2011) 15679–15686], providing strong evidence supporting the claim that the sampler mixes quickly.

Keywords: Cauchy identity; goodness-of-fit; mixing-diagnostics for Markov Chains; non-parametric hypothesis testing; random rotation generators; representation theory of compact groups; spectral analysis

1. Introduction

Recent work of Jones, Osipov and Rokhlin [22] suggested a Markov chain on the orthogonal group that is supposedly used to sample from the uniform distribution. They prescribe a particular number of steps after which the chain is mixed, resulting in a fast random rotation generator which is at the core of several successful randomized data analysis algorithms. Examples include approximate algorithms for highly over-determined linear regression (Rokhlin and Tygert [37]), low-rank matrix approximation (Liberty et al. [28]), and very high dimensional nearest neighbor analysis (Jones, Osipov and Rokhlin [22]). The new sampler could offer a significant reduction in computational cost compared to the best exact algorithm in the literature (see Section 1.1). In applications where multiplication of a random orthogonal matrix with many vectors is needed, the new sampler is much faster than conventional random rotation generators.

It is desirable to have outputs that are approximately uniformly distributed. This is not just a mere theoretical preference; it is a matter of practical importance. In fact, as discussed in Observation 5.1 in the supplementary material Sepehri [41], the performance of the approximate nearest neighbor algorithm was improved by using a uniform sampler compared to non-uniform

samplers (the approximate nearest neighbor algorithm is sketched in Section 3 of the supplementary material Sepehri [41]). Therefore, one needs to investigate the mixing properties of the new sampler. Unfortunately, due to complex construction of the new sampler, analytical study of the mixing-time seems to be impractical. This paper suggests to numerically study the mixing-time of the new sampler using statistical tests of goodness-of-fit.

There is a sizable literature on goodness-of-fit testing on non-Euclidean spaces. Major work has been devoted to the development of goodness-of-fit tests on the circle and sphere (see Rayleigh [36], Ajne [2], Beran [7], Watson [45–47], Wellner [48]). The literature on goodness-of-fit testing for the orthogonal group has been limited to three dimensions; two commonly used tests for three dimensional rotations are Downs’ generalization of the Rayleigh test (Downs [15]) and Prentice’s generalization of Giné’s G_n test (Prentice [35], Giné [16]). For a more detailed review of the literature see Mardia and Jupp [29]. In an important development in high dimensional setting, Coram and Diaconis [10] proposed a family of statistical tests for the eigenvalue distribution induced from the Haar measure on the unitary group, $U(n)$. Their tests are relatively easy to compute and consistent against all fixed alternatives. One of the new tests in this paper was inspired by the tests of Coram and Diaconis [10].

This paper settles the question about the mixing-time of the new sampler using statistical tests. Various known tests are applied (see Sections 1.1 and 2), confirming that the new sampler mixes quickly. New tests are introduced (see Sections 3 and 4) and validated using the benchmark examples of Section 1.2. The new tests are applied to the new sampler and the results are compared to other tests in Section 5. The results are in agreement with the claim that the new sampler mixes quickly, that is, after a given number of steps. Local properties, including local power, of the new tests are studied in Section 6. Similar tests are stated for the other compact groups in Section 7. A further test based on the properties of trace is presented in Section 6 of the supplementary material Sepehri [41].

1.1. Pseudorandom orthogonal transformations

In their recent work, Jones, Osipov and Rokhlin [22] proposed a pseudorandom orthogonal matrix generator which consecutively applies two dimensional rotations in coordinate planes, preconditioned using a Fourier type matrix. This is formally described below. Suppose n , M_1 , M_2 are positive integers. Define a pseudorandom n -dimensional orthogonal transformation Θ as a composition of $M_1 + M_2 + 1$ orthogonal operators

$$\Theta = \left(\prod_{i=1}^{M_1} Q_i \cdot P_i \right) \cdot F^n \cdot \left(\prod_{j=1}^{M_2} Q_j P_j \right). \quad (1)$$

Each P_i and P_j is a uniformly distributed $n \times n$ permutation matrix, independent of others. That is, each P_i corresponds to a permutation p_i of $\{1, \dots, n\}$ and P_i acts on vectors as follows

$$(P_i v)_j = v_{p_i(j)}.$$

Each Q_j is defined as

$$Q_j = Q_{n-1,j} \cdot Q_{n-2,j} \cdots Q_{1,j},$$

where $Q_{l,j}$ is a uniform two dimensional rotation in the plane generated by the l th and $l + 1$ th coordinates. That is, $(Q_{l,j}v)_i = v_i$ for $i \neq l, l + 1$ and

$$\begin{aligned}(Q_{l,j}v)_l &= \cos \theta_{l,j} v_l + \sin \theta_{l,j} v_{l+1}, \\ (Q_{l,j}v)_{l+1} &= -\sin \theta_{l,j} v_l + \cos \theta_{l,j} v_{l+1},\end{aligned}$$

where $\theta_{l,j}$ is a uniform number in $[0, 2\pi]$. All Q_i and Q_j are independent of each other. Lastly, the linear operator F^n is defined as follows. Let $d = \lfloor \frac{n}{2} \rfloor$ and T be the following $d \times d$ matrix:

$$T_{k,l} = \frac{1}{\sqrt{d}} \exp \left[-\frac{2\pi i (k-1)(l-1)}{d} \right].$$

Define $Z : \mathbb{R}^{2d} \rightarrow \mathbb{C}^d$ as

$$[Z(x)]_l = x_{2l-1} + ix_{2l}.$$

For n even, define F^n as

$$F^n = Z^{-1} \cdot T \cdot Z. \quad (2)$$

If n is odd, F^n fixes the last coordinate of x , and F^{n-1} defined in (2) is applied to the first $n - 1$ coordinates. The cost of applying Θ to vector $x \in \mathbb{R}^n$ is of order $O(n(\log n + M_1 + M_2))$, because the cost of applying the operator F^n is $O(n \log n)$ and each operator $Q_j \cdot P_j$ costs $O(n)$. It is claimed in Jones, Osipov and Rokhlin [22] that if $M_1 + M_2 = O(\log n)$, then the distribution of Θ is close to the uniform distribution on the set of all $n \times n$ orthogonal matrices. This makes the new sampler much faster than the state of the art Subgroup Algorithm of Diaconis and Shahshahani [14], which is an $O(n^3)$ algorithm for generating uniform $n \times n$ rotation matrices. However, it remain to be investigated whether the distribution of the output is close to the uniform distribution. Throughout the paper, the ‘mixing time’ or the number of steps required for the Jones–Osipov–Rokhlin sampler to mix refers to the quantity $M_1 + M_2$.

1.2. Benchmark examples

Two benchmark examples of random walks on $SO(n)$ and their mixing properties are used as sanity check for the tests considered in this paper.

1. *Kac’s random walk.* The standard *Kac’s random walk* $\{O_k\}$ on $SO(n)$ is the defined as follows:

$$O_{k+1} = R_{i,j}^{(k)}(\theta) O_k,$$

where $R_{i,j}^{(k)}(\theta)$ is an elementary rotation with angle θ in the plane generated by the i th and j th coordinate axes, where $\{i, j\}$ is uniformly chosen among all pairs from $\{1, \dots, n\}$ and θ uniformly random in $[0, 2\pi)$. This walk was introduced as part of Kac’s effort to

simplify Boltzmann's proof of the H-theorem (Kac [25]) and Hastings's simulations of random rotations (Hastings [19]). Convergence of the Kac's random walk has been studied by various authors in different senses. In the current discussion, the focus is on convergence in Wasserstein distance which metrizes the weak convergence; for a review of the literature see Pak and Sidenko [32], Oliveira [31], Pillai and Smith [33]. The best known bound on the mixing-time in Wasserstein distance is obtained by Oliveira [31], providing an upper bound of order $n^2 \log n$ on the mixing-time which is at most a factor $\log n$ away from optimal. For $n = 51$, which is the case studied numerically in this paper, $n^2 \log n \approx 10\,000$, but the constants are not known and the actual mixing time could be much smaller of larger than this value.

2. *Product of random reflections.* As described in Diaconis [11], the following random walk on $O(n)$ arose in a telephone encryption problem. At each step, the current orthogonal matrix is multiplied by a random reflection, a matrix of the form $I - 2u^T u$ for a uniform unit vector $u \in \mathbb{S}^{n-1}$.

The mixing-time for this chain has been studied carefully in Diaconis and Shahshahani [13], Porod [34], Rosenthal [39], proving that $\frac{1}{2}n \log n + cn$ steps are necessary and sufficient for convergence of the reflection walk to the uniform distribution in total variation distance. In fact, Porod [34] gives explicit lower- and upper-bounds for the total variation distance between this Markov chain and the Haar measure as a function of c , but the bounds are not tight. For $n = 51$, $\frac{1}{2}n \log n \approx 100$.

2. Some already known tests

Two important tests for uniformity on $SO(3)$, the Rayleigh's test and the Gine's test, are reviewed in this section. The application to the examples of the previous section is demonstrated.

2.1. Rayleigh's test

Perhaps the first test of uniformity on $SO(3)$ was introduced by Rayleigh [36]. Given data $g_1, \dots, g_N \in SO(3)$ define

$$T_R = 3N \operatorname{tr}(\bar{g}^T \bar{g}),$$

where

$$\bar{g} = \frac{1}{N} \sum_{i=1}^N g_i.$$

The Rayleigh's test for uniformity rejects for large values of T_R . This can be directly generalized to the higher dimensional case. For any $n \in \mathbb{N}$ and $g_1, \dots, g_N \in SO(n)$ define

$$T_R = nN \operatorname{tr}(\bar{g}^T \bar{g}),$$

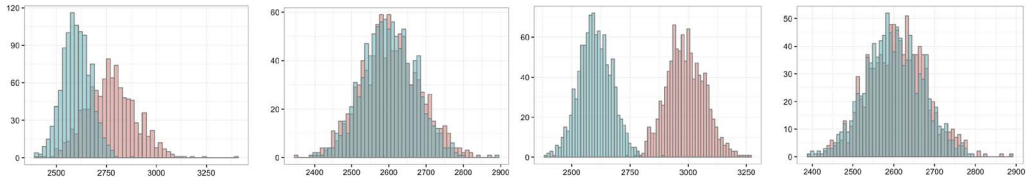


Figure 1. Histograms of Rayleigh's statistic under Setup 2.1. The alternatives are (left to right) the Kac's walk after 100 steps, after 150 steps, the product of 90 random reflections, and 140 random reflections.

where

$$\bar{g} = \frac{1}{N} \sum_{i=1}^N g_i.$$

The Rayleigh's test was applied to the benchmark examples of the previous section under the following setup.

Setup 2.1. The sample size is $N = 200$ and the dimension is $n = 51$. Each test statistic is computed on 1000 independent repetitions. This setup will be used throughout the paper. All the histograms in this paper are illustrated in blue under the null and in red under the alternatives.

Figure 1 illustrates the histograms of the Rayleigh's statistics computed on the product of random reflections and Kac's walk, with that corresponding to the uniform distribution overlaid. Throughout the paper, in all similar figures the color blue corresponds to the Haar distributed samples and the color red corresponds to the alternative.

The Rayleigh's test does not seem to have remarkable powerful on either examples. In particular, the Rayleigh's test fails to reject the null hypothesis even after 150 steps of the Kac's walk. The Anderson–Darling p -value for the 1000 values of the Rayleigh's statistic after different number of steps are given in Tables 1 and 2 for the product of random reflections and the Kac's walk.

Applying Rayleigh's test to samples generated by the new sampler of Jones, Osipov and Rokhlin [22] provides no evidence for departure from uniformity, after only one iteration. The p -value of the Anderson–Darling test under Setup 2.1 is 0.35.

Table 1. p -values corresponding to Rayleigh's test on iterated random reflections

# of steps	50	75	90	100	110	125	140	150	175	200
A–D test	$\ll 1e-32$	$\ll 1e-32$	$\ll 1e-32$	$\ll 1e-32$	$\ll 1e-32$	$4.1e-16$	0.03	0.37	0.84	0.39

Table 2. p -values corresponding to Rayleigh’s test on Kac’s walk

# of steps	100	150	200	250	300	350	400	450	500
A–D test	$\ll 1e-32$	0.08	0.58	0.23	0.23	0.86	0.70	0.82	0.89

2.2. Gine’s test

Another important test of uniformity on $SO(3)$ was introduced by Giné [16]. Given data $g_1, \dots, g_N \in SO(3)$, define

$$T_G = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sqrt{\text{tr}(I - g_i^T g_j)}.$$

Giné’s test rejects for large values of T_G . It is consistent against all fixed alternatives on $SO(3)$, but not in any higher dimensions. The corresponding test was carried out on the benchmark examples and the new sampler. Giné’s tests seems to be more powerful than the Rayleigh’s test on these examples. Histograms of the values of the Giné’s statistic are illustrated in Figure 2.

Applying the Giné’s test to the new sampler provides no evidence for departure from the null. The Anderson–Darling p -values after one and two iterations of the sampler are 0.58 and 0.45, respectively.

3. Tests based on eigenvalues

The new sampler has passed all the tests considered in the previous sections. In this section, various new tests based on the eigenvalues are introduced and applied to the benchmark examples as well as the new sampler.

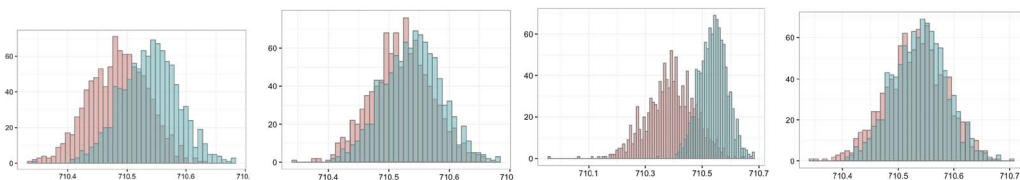


Figure 2. Histogram of Giné’s statistic under Setup 2.1. The alternatives are (left to right) product of 110 random reflections, 125 random reflections, the Kac’s walk after 100 steps, and 150 steps.

3.1. A test based on exponential families

The joint density of the eigenvalues of a uniformly random $g \in SO(2n + 1)$ is given by Weyl [49], page 224, as

$$f(e^{\pm i\theta_1}, \dots, e^{\pm i\theta_n}) \propto \prod_i \sin^2\left(\frac{\theta_i}{2}\right) \prod_{i < j} |\cos \theta_i - \cos \theta_j|^2, \quad (3)$$

where $(1, e^{\pm i\theta_1}, \dots, e^{\pm i\theta_n})$ are eigenvalues of g . By a change of variables $x_i = \cos \theta_i$, the density, in terms of $(x_1, \dots, x_n) \in [-1, 1]^n$, becomes

$$f(x_1, \dots, x_n) \propto \prod_{i < j} |x_i - x_j|^2 \prod_i \frac{\sqrt{1 - x_i}}{\sqrt{1 + x_i}}. \quad (4)$$

The density f can be embedded in the following exponential family:

$$f_{\gamma, \alpha, \beta}(x_1, \dots, x_n) \propto \prod_{i < j} |x_i - x_j|^{2\gamma} \prod_i (1 - x_i)^{\alpha-1} (1 + x_i)^{\beta-1}. \quad (5)$$

The normalizing constant is given by Selberg's integral (Mehta [30], pg. 320, eqn. (17.5.9)), as

$$\begin{aligned} & \int_{[-1, 1]^n} \prod_{i < j} |x_i - x_j|^{2\gamma} \prod_i (1 - x_i)^{\alpha-1} (1 + x_i)^{\beta-1} dx_1 \dots dx_n \\ &= 2^{\gamma n(n-1) + n(\alpha + \beta - 1)} \prod_{j=0}^{n-1} \frac{\Gamma(1 + \gamma + j\gamma) \Gamma(\alpha + j\gamma) \Gamma(\beta + j\gamma)}{\Gamma(1 + \gamma) \Gamma(\alpha + \beta + \gamma(n + j - 1))}. \end{aligned}$$

The density (4) is the special case of (5) for $\gamma_0 = 1$, $\alpha_0 = 3/2$, and $\beta_0 = 1/2$. We abuse the notation to denote by $f_{1, \frac{3}{2}, \frac{1}{2}}$ both densities (3) and (4). Recall that the eigenvalues of a uniform orthogonal matrix g are placed quite regularly on the unit circle. For example, the trace of g is approximately normal with mean zero and variance one (Diaconis and Mallows [12]); whereas, for uniformly distributed points on the unit circle, say in conjugate pairs, because of the Law of Large Number and the central limit theorem, the sum has mean zero and variance of order $O(n)$. In particular, the magnitude of the sum is of order $O(\sqrt{n})$. The family of densities $f_{\gamma, \alpha, \beta}$ models the regularity of the configuration of the points on the circle. The case of $\gamma = 0, \alpha = \beta = 1/2$ of (5) corresponds to θ_i being independent uniform on $[0, \pi]$. As γ tends to infinity the points $e^{i\theta_j}$ become evenly placed on the semicircle. This is illustrated in Figure 3.

Testing for the Haar measure in this parametric family translates to

$$H_0 : (\gamma, \alpha, \beta) = (\gamma_0, \alpha_0, \beta_0) \quad \text{vs} \quad H_1 : (\gamma, \alpha, \beta) \neq (\gamma_0, \alpha_0, \beta_0). \quad (6)$$

Unfortunately, there is no uniformly most powerful test available in this setting. However, the problem fits into the framework of Asymptotically Normal Experiments of Le Cam. This allows for construction of an asymptotically *maximin* optimal test as follows. Define $T_1(x) =$

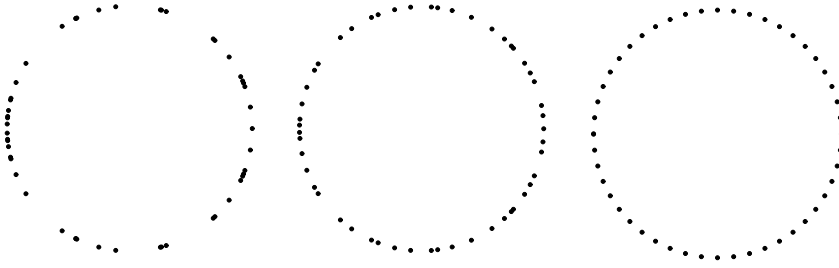


Figure 3. Left: 25 uniform points on the circle and their conjugates, middle: eigenvalues of a uniform 51×51 orthogonal matrix, right: 51 evenly placed points on the circle. The sum of all complex numbers is approximately: 7.8 on the left, 0.78 in the middle, and 0 on the right.

$2 \sum_{i < j} \log |x_i - x_j|$, $T_2(x) = \sum_i \log(1 - x_i)$, and $T_3(x) = \sum_i \log(1 + x_i)$. Then (T_1, T_2, T_3) is a sufficient statistic for the exponential family $f_{\gamma, \alpha, \beta}$. That is,

$$f_{\gamma, \alpha, \beta}(x) = \exp(\gamma T_1(x) + (\alpha - 1)T_2(x) + (\beta - 1)T_3(x) - \mathcal{A}(\gamma, \alpha, \beta)),$$

where $\mathcal{A}(\gamma, \alpha, \beta) = \log(2^{\gamma n(n-1) + n(\alpha + \beta - 1)} \prod_{j=0}^{n-1} \frac{\Gamma(1 + \gamma + j\gamma)\Gamma(\alpha + j\gamma)\Gamma(\beta + j\gamma)}{\Gamma(1 + \gamma)\Gamma(\alpha + \beta + \gamma(n + j - 1))})$. Then, given data $x^{(1)}, \dots, x^{(N)}$ the likelihood is

$$\prod_m f_{\gamma, \alpha, \beta}(x^{(m)}) = \exp(\gamma T_1^{(N)} + (\alpha - 1)T_2^{(N)} + (\beta - 1)T_3^{(N)} - N\mathcal{A}(\gamma, \alpha, \beta)),$$

where $T_i^{(N)} = \sum_m T_i(x^{(m)})$. By standard asymptotic theory, under the null, $T^{(N)} = \frac{1}{N}(T_1^{(N)}, T_2^{(N)}, T_3^{(N)})$ is approximately normal with mean $\mu = \nabla \mathcal{A}(\gamma_0, \alpha_0, \beta_0)$ and covariance matrix $\Sigma = \nabla^2 \mathcal{A}(\gamma_0, \alpha_0, \beta_0)/N$. The test that rejects for large values of

$$T = (T^{(N)} - \mu)^T \Sigma^{-1} (T^{(N)} - \mu)$$

is asymptotically maximin optimal (Lehmann and Romano [27], Theorem 13.5.5). Moreover, μ and Σ can be computed using the recurrence relations and series expansions for digamma and trigamma functions (Abramowitz and Stegun [1], pp. 258–259). For 51×51 orthogonal matrices, $n = 25$, and $(\gamma_0, \alpha_0, \beta_0) = (1, 3/2, 1/2)$ they are

$$\mu \approx (-329.70, -14.73, -19.92),$$

$$\Sigma \approx \begin{bmatrix} 5.67712 \times 10^{-2} & -1.40363 \times 10^{-2} & -3.61067 \times 10^{-4} \\ -1.40363 \times 10^{-2} & 1.52093 \times 10^{-2} & -3.46523 \times 10^{-3} \\ -3.61067 \times 10^{-4} & -3.46523 \times 10^{-3} & 3.96834 \times 10^{-2} \end{bmatrix}.$$

We applied this test on two samples of size $N = 200$ and 1000 ; data was generated using only one step of the pseudorandom sampler (1) of Section 1.1. The test statistic T evaluated to 3.84 and 1.34, respectively. Under the null, T is approximately χ_3^2 distributed. The corresponding p -values are 0.72 and 0.28, respectively; there is no evidence for departure from uniformity.

Table 3. p -values corresponding to product of random reflections

$k =$	100	125	130	135	140	150
$N = 200$	0	0.0002	0.001	0.055	0.09	0.20
$N = 1000$	0	0	0	0.0001	0.03	0.42

To further explore the performance of T , it was applied to the benchmark examples. Tables 3 and 4 show the p -values corresponding to T applied to the benchmark examples, suggesting that T detects the cut-off to some extent. In particular, it seems to be more powerful than both Gine’s and Rayleigh’s tests.

3.2. A family of consistent tests for $f_{1,\frac{3}{2},\frac{1}{2}}$

Given data $x^{(1)}, \dots, x^{(N)}$ in $[-1, 1]^n$, this section introduces a family of tests $T_z^{(N)}(x^{(1)}, \dots, x^{(N)})$, for a parameter $0 < z < 1$, that are invariant under the natural symmetries of $f_{1,\frac{3}{2},\frac{1}{2}}$ and are consistent against all alternatives. The asymptotic distributions under null and alternative are also available. The construction is given for a general hypothesis testing problem in the following section; It is then carried out for $f_{1,\frac{3}{2},\frac{1}{2}}$.

3.2.1. Spectral tests on general spaces

Let \mathcal{X} be a Polish space and μ a probability measure on \mathcal{X} . Consider the standard non-parametric goodness-of-fit testing problem: given independent and identically distributed observations $x_1, \dots, x_N \in \mathcal{X}$ from a probability measure ν on \mathcal{X} , test if $\nu = \mu$. A general test based on spectral techniques can be constructed as follows.

Let $\mathcal{L}^2(\mathcal{X}, \mu)$ be the space of square μ -integrable functions on \mathcal{X} . Assume that $\mathcal{L}^2(\mathcal{X}, \mu)$ is separable with a countable orthonormal basis $\{f_i \mid i \geq 0\}$, with $f_0 = 1$. For example, it suffices to assume that \mathcal{X} is compact. Theorem 13.8 in Thomson, Bruckner and Bruckner [43] gives a condition for L^2 to be separable in the general setting. Define the empirical measure of $\{x_i\}$ as $\nu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$, and the Fourier coefficients of ν_N as

$$\widehat{\nu}_N(i) = \int_{\mathcal{X}} f_i(x) \nu_N(dx) = \frac{1}{N} \sum_{n=1}^N f_i(x_n).$$

Table 4. p -values corresponding to Kac’s walk

$k =$	150	200	225	240	250	300
$N = 200$	0	0	0.001	0.004	0.19	0.40
$N = 1000$	0	0	0	0	0.00001	0.65

Then, under the null, $\widehat{v}_N(i) \rightarrow 0$ as n grows to infinity, for $i > 0$. This property characterizes μ ; that is, if $x_1, x_2, \dots \in \mathcal{X}$ are i.i.d. draws from ν and $\widehat{v}_N(i) \rightarrow 0$ for all $i > 0$ then, $\mu = \nu$. This property can be used to construct tests of fit for μ . By the central limit theorem

$$\sqrt{N}\widehat{v}_N(i) \rightarrow \mathcal{N}(0, 1).$$

Thus, $N|\widehat{v}_N(i)|^2$ is asymptotically χ_1^2 distributed. For a sequence of weights $\mathbf{c} = (c_1, c_2, \dots)$, define

$$T_{\mathbf{c}} = N \sum_i c_i |\widehat{v}_N(i)|^2.$$

Assuming that $T_{\mathbf{c}}$ converges to a finite value, a test that rejects for large values of $T_{\mathbf{c}}$ can be used for testing $H_0 : \nu = \mu$. Many well-known classical tests can be constructed in this manner. The most important example is the celebrated Neyman's smooth test for uniformity on the unit interval. Neyman's test uses Legendre polynomials as the orthonormal basis. Under mild conditions and the assumption $c_i > 0$ the test based on $T_{\mathbf{c}}$ is consistent against all alternatives, and has various desired statistical properties. There is a vast literature on properties of tests of this form; we do not attempt to review the literature since it is considered classical nowadays.

There are two main challenges in using the above machinery in a general problem: (1) finding an orthonormal basis for \mathcal{L}^2 , (2) computing $T_{\mathbf{c}}$. In his celebrated paper, Giné [16] gave a solution for the first challenge for the testing problem with μ being the invariant measure on a compact Riemannian manifold M ; this is sketched below.

Let Δ be the Laplace–Beltrami operator (Laplacian) of M acting on the space of Schwartz functions by duality. Denote by E_k the k th invariant eigenspace of Δ with eigenvalue σ_k . Let $\{f_i^k\}_{i=0}^{\dim E_k}$ be an orthonormal basis for E_k . Then, $\{f_i^k \mid k \geq 0, 1 \leq i \leq \dim E_k\}$ is an orthonormal basis for $\mathcal{L}^2(M, \mu)$. Note that the hypothesis testing problem

$$H_0 : \nu = \mu \quad \text{vs} \quad H_1 : \nu \neq \mu$$

is invariant under natural symmetries of M . Therefore, by the Hunt–Stein theorem (Lehmann and Romano [27], page 331), one only needs to consider invariant tests. Giné [16] suggested the test, called Sobolev test, based on

$$T_N^{\boldsymbol{\alpha}}(\mathbf{x}) = N \sum_{k=1}^{\infty} \alpha_k \sum_{f_i \in E_k} \left[\int_M f_i d\nu_N(\mathbf{x}) \right]^2,$$

for a sequence of weights $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots)$ such that $\sup |\alpha_k \sigma_k^s| < \infty$ for some $s > \frac{1}{2} \dim M$. Note that the weights depend only on the eigenspaces; this ensures that the test is invariant. Giné [16] studied statistical properties of the Sobolev tests; he derived the null and alternative distribution and investigated local optimality properties following Beran [7].

Although these tests have been successful in practice, usually substantial non-trivial work is required to carry out the details for any particular example of interest. Giné [16] carried out the program for the circle, sphere, and the projective plane, recovering many known examples in

the literature and introducing new tests of uniformity. Several authors have studied and derived Sobolev tests for different examples including circular and directional data, tests of symmetry, and unitary eigenvalues; see Prentice [35], Wellner [48], Jupp and Spurr [23,24], Hermans and Rasson [20], Baringhaus [5], Sengupta and Pal [40], Coram and Diaconis [10].

Regarding the second challenge, note that

$$\begin{aligned} T_c &= N \sum_k c_k |\widehat{v}_N(k)|^2 \\ &= \frac{1}{N} \sum_{i,j=1}^N \sum_k c_k f_k(x_i) f_k(x_j). \end{aligned}$$

Therefore, it suffices to have a way of computing

$$K(x, y) = \sum_k c_k f_k(x) f_k(y). \quad (7)$$

To compute $K(\cdot, \cdot)$, Giné [16,17] suggested partial answers for the Sobolev tests, based on Zonal functions; there still remains the challenge to find a closed form for T_N^α , or to compute it effectively for a general Riemannian manifold M . The paper resolves this issue for compact classical groups and a class of weight sequences $\{c_k\}$.

3.2.2. Spectral tests for $f_{1,3/2,1/2}$

The following facts and notation will be used throughout this paper. For each partition λ , with at most n parts, of an arbitrary non-negative integer, there exists an integer d_λ and a map π_λ from $SO(2n+1)$ to the set of all $d_\lambda \times d_\lambda$ matrices with the following properties. For $g, h \in SO(2n+1)$, one has $\pi(g \cdot h) = \pi(g)\pi(h)$; the set of all matrix coordinates $\{\pi_\lambda^{i,j} \mid \lambda, 1 \leq i, j \leq d_\lambda\}$ is an orthonormal basis for $\mathcal{L}^2(SO(2n+1))$. Moreover, if $\chi_\lambda(g) = \text{tr}(\pi_\lambda(g))$, then $\{\chi_\lambda\}$ is an orthonormal basis for $\mathcal{L}^2(f_{1,3/2,1/2})$. These are standard facts from representation theory of Lie groups. A brief introduction is given in Section 1 of the supplementary material Sepehri [41]. For a textbook treatment see Bump [9], Goodman and Wallach [18].

Given independent observations $g_1, \dots, g_N \in SO(2n+1)$, define the Fourier coefficient corresponding to λ as

$$\widehat{\chi}_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \chi_\lambda(g_i).$$

For $0 < z < 1$ define the test statistics $T_z^{(N)}$ as

$$T_z^{(N)} = N \sum_{\lambda \neq 0} z^{|\lambda|} |\widehat{\chi}_N(\lambda)|^2, \quad (8)$$

where $|\lambda| = \lambda_1 + \dots + \lambda_n$ is the sum of the parts of the partition λ and sum is over all partitions of all positive integers with at most n parts.

To use $T_z^{(N)}$ in practice, a closed form expression for the kernel given in (7),

$$K_z(g_i, g_j) = \sum_{\lambda \neq 0} z^{|\lambda|} \chi_\lambda(g_i) \chi_\lambda(g_j),$$

would yield a closed form expression for $T_z^{(N)}$.

Coram and Diaconis [10] used the closed form expression for $K_z(g, h)$, given by the Cauchy identity for the Schur functions, to build a test for the eigenvalue distribution induced by the Haar measure on the unitary group. The test statistic $T_z^{(N)}$ is an analogue for the orthogonal group of their test.

In the case of the orthogonal group there was no closed form for K_z available in the literature. Motivated by the testing problem under study in the present paper, the author derived Cauchy identities for all of the compact classical groups.

Proposition 3.1 (Cauchy identity for $SO(2n+1)$, Theorem 2.1 in Sepehri [41]). *Let $g, h \in SO(2n+1)$ have eigenvalues equal to $(1, e^{\pm i\theta_1}, \dots, e^{\pm i\theta_n})$ and $(1, e^{\pm i\phi_1}, \dots, e^{\pm i\phi_n})$, respectively. Then,*

$$K_z(g, h) = \frac{(1-z)^n \det \left[\frac{(1+z)^2 + 2z(\cos \theta_k + \cos \phi_l)}{(1+z^2)^2 - 4(z+z^3) \cos \theta_k \cos \phi_l + 2z^2(\cos 2\theta_k + \cos 2\phi_l)} \right]_{k,l}}{(4z)^{\binom{n}{2}} \prod_{i < j} (\cos \theta_i - \cos \theta_j) \prod_{i < j} (\cos \phi_i - \cos \phi_j)} - 1. \quad (9)$$

Despite the complicated appearance of the formula (9), it is relatively easy to compute if the dimension is not too large, offering a way to compute $T_z^{(N)}$.

The test based on $T_z^{(N)}$ was applied to the benchmark examples and the new sampler. It is indeed more powerful than all tests considered in the previous sections on both examples. Figure 4 illustrates the histogram of the values of $T_z^{(N)}$ under Setup 2.1. The 5%-level test based on $T_{1/2}^{(200)}$ has power equal to 0.64 against the product of 140 random reflections; the power drops to 0.30 after 150 steps. Similarly, it has power equal to 0.93 against the Kac's walk after 250 steps, which drops to 0.25 after 300 steps.

Remark 3.2. The test statistic $T_z^{(N)}$ has, by construction, a decomposition to approximately independent parts. When the test rejects the null hypothesis, it would be illuminating regarding

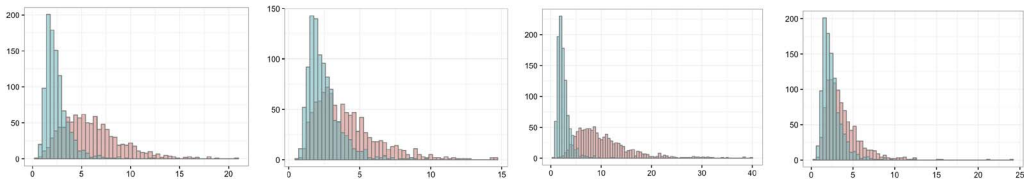


Figure 4. Histogram of $T_z^{(N)}$ under Setup 2.1 and $z = 0.5$. The alternatives are (left to right) the product of 140 random reflections, 150 random reflections, the Kac's random walk after 250 steps, and 300 steps.

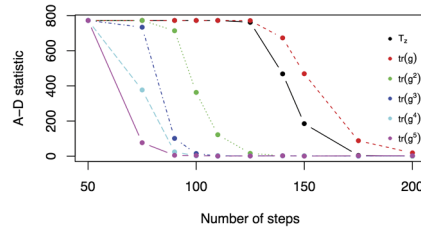


Figure 5. Values of the Anderson–Darling statistic for comparison of the uniform sample and the product of random reflections for different tests.

the nature of the departure from uniformity to see which of the components is larger than its typical values. This was investigated using lower order terms of the form

$$C_k^{(N)} = \frac{1}{N} \sum_{i=1}^N \text{tr}(g_i^k),$$

under Setup 2.1 and result is shown in Figures 5 and 6 for $k = 1, \dots, 5$. As it can be seen, the component corresponding to $\text{tr}(g)$ is the most significant against the product of random reflections. For the Kac’s walk, $\text{tr}(g^2)$ is the most significant and captures the cutoff more clearly compared to $T_z^{(N)}$. The nature of deviation differs for these examples.

The test based on $T_{1/2}^{(200)}$ applied to the new sampler, after a single iteration, provides no evidence for departure from the null; the Anderson–Darling p -value based on 1000 values of $T_{1/2}^{(200)}$ is equal to 0.35.

3.2.3. Asymptotic distribution under the null hypothesis

The representation (8) allows for derivation of the distribution under the null hypothesis of uniformity. This is the content of the following proposition.

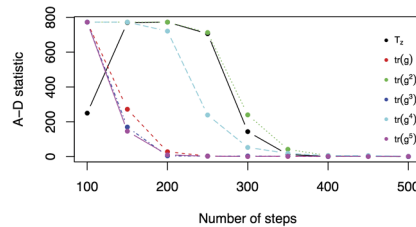


Figure 6. Values of the Anderson–Darling statistic for comparison of the uniform sample and the Kac’s walk for different tests.

Proposition 3.3. Assume g_1, \dots, g_N are independent draws from the uniform distribution on $SO(2n+1)$, and $T_z^{(N)}$ is defined as in (8). Then, for any $z \in (0, 1)$ one has

$$T_z^{(N)} \rightarrow T_z = \sum_{k=1}^{\infty} z^k \chi_{p(n,k)}^2 \quad (\text{as } N \rightarrow \infty).$$

On the right hand side, $p(n, k)$ is the number of partitions of k with at most n parts and random variables $\chi_{p(n,k)}^2$ are chi-square with $p(n, k)$ degrees of freedom which are mutually independent.

Proof. The CLT along with the orthogonality relations between the irreducible characters assert that $\sqrt{N} \widehat{\chi}_N(\lambda)$ converges to standard Gaussian variables for $\lambda \neq 0$, and the limiting variables are mutually independent. Let $\mathbb{D} = \{\mathbf{x} = (x_\lambda)_{\lambda \neq 0} \mid \sum_{\lambda \neq 0} z^{|\lambda|} x_\lambda^2 < \infty\}$. \mathbb{D} is a Hilbert space equipped with the inner-product

$$\langle \mathbf{x}, \mathbf{y} \rangle_z = \sum_{\lambda \neq 0} z^{|\lambda|} x_\lambda y_\lambda.$$

Define the map $\phi : \mathbb{D} \mapsto \mathbb{R}$ as $\phi(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle_z = \sum_{i \in \mathcal{I}} z^{|\lambda|} x_\lambda^2$; it is second order Hadamard directionally differentiable with derivative $\phi'_\mathbf{x}(h) = 2 \langle \mathbf{x}, h \rangle_z$ and the second derivative $\phi''_\mathbf{x}(h) = 2\phi(h)$. Let $\widehat{\chi}_N = (\widehat{\chi}_N(\lambda))_\lambda$. Then,

$$T_z^{(N)} = N\phi(\widehat{\chi}_N).$$

Van Der Vaart and Wellner [44], Theorem 1.4.8, asserts that weak convergence in a countable product space, to separable limiting variables, is determined by weak convergence of all finite-dimensional marginals. That means that $\sqrt{N} \widehat{\chi}_N$ converges weakly to a random element of \mathbb{D} , denoted by $Z = (Z_\lambda)_\lambda$ where Z_λ are independent normal variables. Since, $\sqrt{N}(\widehat{\chi}_N - 0) \rightarrow Z$, the second order Delta method (Römisch [38]) yields

$$N\phi(\widehat{\chi}_N) \rightarrow \phi(Z) = \sum_{\lambda \neq 0} z^{|\lambda|} Z_\lambda^2.$$

Collecting powers of z proves the proposition. □

Remark 3.4. As a consequence of Proposition 3.3, asymptotic expectation and variance of $T_z^{(N)}$ are given as

$$\mathbb{E}(T_z) = \sum_{k=1}^{\infty} z^k p(n, k) \quad \text{and} \quad \text{Var}(T_z) = 2 \sum_{k=1}^{\infty} z^{2k} p(n, k).$$

The right hand sides can be simplified using the identity (Andrews [3], Theorems 1.1 & 1.4),

$$\sum_{k=0}^{\infty} z^k p(n, k) = \prod_{i=1}^n \frac{1}{1 - z^i}.$$

Table 5. Mean and variance of T_z for $n = 25$

$z =$	0.5	0.8	0.9	0.99
Mean	2.46	291.45	402 914.7	2.844628×10^{25}
Variance	0.9047073	18.86372	870.2173	8.097291×10^{18}

We get

$$\mathbb{E}(T_z) = \prod_{i=1}^n \frac{1}{1-z^i} - 1 \quad \text{and} \quad \text{Var}(T_z) = 2 \left(\prod_{i=1}^n \frac{1}{1-z^{2i}} - 1 \right).$$

Table 5 shows the asymptotic mean and variance for 51×51 rotations, i.e. $n = 25$, and different values of z . Finite sample expectation and variance under Setup 2.1 and $z = 0.5$ are 2.48 and 1.26 respectively. Empirical quantiles are given in Table 6.

3.2.4. Distribution under fixed alternative hypotheses

The alternative distribution is given below.

Proposition 3.5. *Let F be a probability measure on $[0, \pi]^n$ which is different from $f_{1, \frac{3}{2}, \frac{1}{2}}$. Let $\theta^{(1)}, \dots, \theta^{(N)}$ be independent draws from F . Then, $T_z^{(N)}$ is asymptotically normal. In fact,*

$$\sqrt{N}(T_z^{(N)}/N - \mu) \rightarrow \mathcal{N}(0, \sigma^2) \quad \text{as } N \rightarrow \infty,$$

with $\mu = \int r^2(\theta) F(d\theta) - 1$ and $\sigma^2 = 4[\int (\int r(\theta) g(\theta, \phi) f_{1, \frac{3}{2}, \frac{1}{2}}(d\theta))^2 F(d\phi) - \mu^2]$, where g and r are defined as $r(\theta) = \int g(\theta, \phi) F(d\phi)$ and

$$g(\theta, \phi) = \frac{(1 - \sqrt{z})^n \det \left[\frac{(1 + \sqrt{z})^2 + 2\sqrt{z}(\cos \theta_k + \cos \phi_l)}{(1+z)^2 - 4\sqrt{z}(1+z)\cos \theta_k \cos \phi_l + 2z(\cos 2\theta_k + \cos 2\phi_l)} \right]_{k,l}}{(16z)^{\binom{n}{2}/2} \prod_{i < j} (\cos \theta_i - \cos \theta_j) \prod_{i < j} (\cos \phi_i - \cos \phi_j)}.$$

Proof. Proof follows from Proposition (4.6) of Giné [16] and is given in detail in Section 4 of the supplementary material Sepehri [41]. \square

Table 6. $P(T_z^{(N)} \leq w_p) = p$ based on 1000 Monte Carlo runs with $N = 200$, $n = 25$, $z = 0.5$

p	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
w_p	0.93	1.16	1.30	1.66	2.20	2.94	3.94	4.65	7.38

Remark 3.6. A direct consequence is that $T_z^{(N)}$ is consistent against all fixed alternatives; not only the limiting distribution differs, so does the scaling. In particular, $\mathbb{P}_\nu(T_z^{(N)} > c_{z,1-\alpha}) \rightarrow 1$ as N tends to infinity, for all alternatives ν .

4. Beyond the eigenvalues

Although the test based on $T_z^{(N)}$ proved successful in different examples, it failed to reject the null hypothesis against the alternative given by the new sampler of Jones, Osipov and Rokhlin [22], even after only one step of the sampler. To overcome this deficiency, it is needed, and natural, to resort to the properties beyond the eigenvalues. This section presents a test for the full Haar measure on $SO(2n+1)$ based on the machinery of Section 3.2.1.

Let $G = SO(2n+1)$. Given data $g_1, \dots, g_N \in G$ independently drawn from a measure ν on G , consider testing the null hypothesis $H_0 : \nu = \mu$, where μ is the uniform (Haar) measure. To construct a spectral test, use the orthonormal basis for $\mathcal{L}^2(G)$ given by the matrix coordinates of the irreducible representations; see Section 1 in Sepehri [41]. Define the Fourier component corresponding to λ as

$$\widehat{\pi}_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \pi_\lambda(g_i).$$

Note that $\widehat{\pi}_N(\lambda)$ is a $d_\lambda \times d_\lambda$ matrix. A test that rejects for large values of

$$U_{\mathbf{c}}^{(N)} = N \sum_{\lambda \neq 0} c_\lambda \|\widehat{\pi}_N(\lambda)\|_F^2 \quad (10)$$

can be used, for an arbitrary sequence of positive weights $\mathbf{c} = \{c_\lambda\}$. To find a closed form for $U_{\mathbf{c}}$, note that

$$\begin{aligned} \|\widehat{\pi}_N(\lambda)\|_F^2 &= \text{tr}(\widehat{\pi}_N(\lambda) \widehat{\pi}_N(\lambda)^*) \\ &= \text{tr}\left(\frac{1}{N^2} \sum_{i,j} \pi_\lambda(g_i) \pi_\lambda(g_j)^*\right) \\ &= \text{tr}\left(\frac{1}{N^2} \sum_{i,j} \pi_\lambda(g_i) \pi_\lambda(g_j^{-1})\right) \\ &= \text{tr}\left(\frac{1}{N^2} \sum_{i,j} \pi_\lambda(g_i g_j^{-1})\right) \\ &= \frac{1}{N^2} \sum_{i,j} \text{tr}(\pi_\lambda(g_i g_j^{-1})) \\ &= \frac{1}{N^2} \sum_{i,j} \chi_\lambda(g_i g_j^{-1}), \end{aligned}$$

where the third equality follows from the fact that π_λ is a unitary matrix. Therefore, one can write

$$U_c^{(N)} = \frac{1}{N} \sum_{i,j} \sum_{\lambda \neq 0} c_\lambda \chi_\lambda(g_i g_j^{-1}).$$

Using tools from representation theory of $SO(2n+1)$, a closed form expression for $\sum_\lambda c_\lambda \chi_\lambda(h)$ can be found for c of a particular form as follows; see Section 3 in the supplementary material Sepehri [41]. For arbitrary parameters $z, q \in (0, 1)$, there exists a set of positive weights $c_\lambda(z, q)$, given explicitly by equation (24) in the supplementary material Sepehri [41], such that

$$\sum_{\lambda \neq 0} c_\lambda(z, q) \chi_\lambda(g) = \frac{\prod_{k < j} (1 - z^2 q^{k+j-2}) \prod_k (1 + z q^{k-1})}{\prod_{k,j} (1 - z q^{k-1} e^{i\theta_j})(1 - z q^{k-1} e^{-i\theta_j})} - 1, \quad (11)$$

where $1, e^{\pm i\theta_1}, \dots, e^{\pm i\theta_n}$ are eigenvalues of g . This motivates the following definition.

Definition 4.1. Let $z, q \in (0, 1)$ be arbitrary parameters. Define the test statistic as

$$U_{z,q}^{(N)} = \frac{1}{N} \sum_{k,l=1}^N \left(\frac{\prod_{i < j} (1 - z^2 q^{i+j-2}) \prod_i (1 + z q^{i-1})}{\prod_{i,j} (1 - z q^{i-1} e^{i\theta_j^{k,l}})(1 - z q^{i-1} e^{-i\theta_j^{k,l}})} - 1 \right), \quad (12)$$

where $1, e^{\pm i\theta_1^{k,l}}, \dots, e^{\pm i\theta_n^{k,l}}$ are the eigenvalues of $g_k g_l^T$.

Remark 4.2. The test based on $U_{z,q}^{(N)}$ fits into the framework proposed by Giné [16]. In fact, the eigenfunction of the Laplace–Beltrami operator on G are exactly the matrix coordinates of the irreducible representations of G , for all compact classical groups.

The test based on $U_{z,q}^{(N)}$ was applied to the new sampler of Jones, Osipov and Rokhlin [22] and exhibited non-trivial power against the alternative distribution generated by a single iteration of the sampler. For $z = 0.2, q = 0.4, n = 25$, and $N = 200$ the histogram of 1000 values of $U_{z,q}^{(N)}$ is illustrated in Figure 7. The power of the 5%-level test is 0.17 which is non-trivial although not particularly high. However, this might well be a result of the small sample size. Note that the testing problem is a non-parametric test of goodness-of-fit in $\binom{51}{2} = 1275$ dimensions with only 200 observations. Note that recent work of Arias-Castro, Pelletier and Saligrama [4] suggest that in the non-parametric goodness-of-testing problem in d dimensions, the sample size N has to be exponential in d in order to have a test that discriminates against the alternatives of fixed distance from the null; that is, $\log(N)/d$ should be bounded below. In the case of $SO(51)$, $d = 1275$; even $N = 200\,000$ results in $\log N/d \approx 0.0095$. For a brief description of their results see Section 8 in the supplementary material Sepehri [41].

Under the same setup as above the 1000 values under the null and alternative were compared using Anderson–Darling test, where alternative was taken to be the output of the new sampler after different number of iterations. The result is shown in Figure 8 and Table 7. It appears that

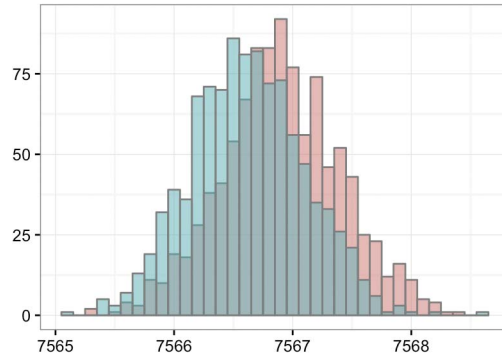


Figure 7. Histogram of $U_{z,q}^{(N)}$ under Setup 2.1, $z = 0.2$, and $q = 0.4$. The alternative is the distribution of the orthogonal matrices generated by a single iteration of the new sampler.

the distribution of the output is not close to uniform after only one iteration. The null hypothesis is rejected at 5% level, suggesting some departure from uniformity, for the distribution of the output after two iterations, but not beyond two steps. This is in agreement with the prescribed number of steps given in Jones, Osipov and Rokhlin [22] for the chain to mix.

Remark 4.3. Power of the test based on $U_{z,q}^{(N)}$ indeed depends on the choice of (z, q) . Given the form of $c_\lambda(z, q)$, the test with a larger value of z has more power against the alternatives which deviate from the uniform distributions in high frequencies. On the other hand, when the alternative differs from the uniform distribution on lower frequencies the test with smaller z is more powerful. Our preliminary numerical investigations suggest that the alternatives considered in the present paper fall into latter category.

Dependence on q is less clear, because the coefficient $c_\lambda(z, q)$'s relation to q is more complicated. Figure 9 illustrates $\log(c_\lambda(1, q))$ for different values of q and several partitions λ . c_λ is

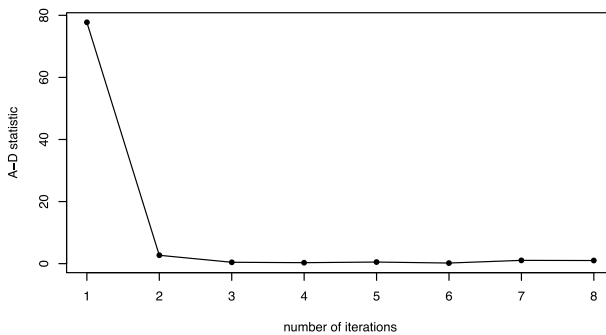


Figure 8. Values of the Anderson–Darling statistic versus the number of iterations of the new sampler.

Table 7. p -values corresponding to different iterations of the new sampler

# of iterations	1	2	3	4	5	6	7	8
A–D test	5.94e–43	3.92e–02	8.02e–01	9.25e–01	7.29e–01	9.90e–01	3.26e–01	3.52e–01

increasing in q but order of the coefficients for a fixed q depends on q which makes it hard to draw general conclusions.

4.1. Distribution under the null and alternative

The asymptotic distribution of $U_{z,q}^{(N)}$ under the null and fixed alternatives can be derived in a similar fashion to those of $T_z^{(N)}$.

Proposition 4.4 (Asymptotic null distribution). *Assume g_1, \dots, g_N are independent draws from the uniform distribution on $SO(2n + 1)$, and $z, q \in (0, 1)$. Then,*

$$U_{z,q}^{(N)} \rightarrow U_{z,q} = \sum_{\lambda \neq 0} \frac{c_\lambda(z, q)}{d_\lambda} \chi_{d_\lambda}^2,$$

where $d_\lambda = \chi_\lambda(I)$ is the dimension of the irreducible representation corresponding to λ , $c_\lambda(z, q)$ is as in (11), and the chi-square variables are mutually independent.

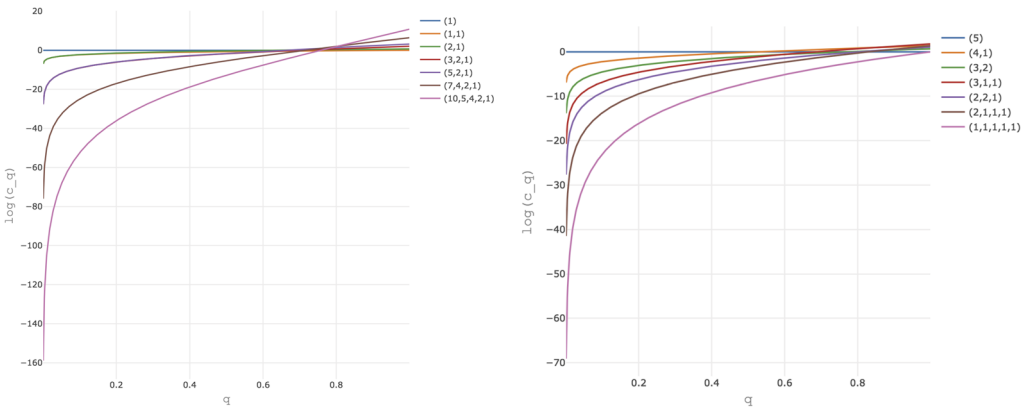


Figure 9. Logarithmic graph of $c_\lambda(1, q)$ against q for several λ . The left plot shows the graphs for various frequencies. The plot on the right shows the graphs for all $|\lambda| = 5$.

Proof. The statement follows from the orthogonality relations between the matrix-coordinates of the irreducible representations, the central limit theorem, and the fact that

$$\mathbb{E}[(\pi_{ij}^\lambda)^2] = 1/d_\lambda. \quad \square$$

The asymptotic distribution under the alternative is given as follows.

Proposition 4.5 (Asymptotic alternative distribution). *Let F be a distribution on $SO(2n+1)$ different from the uniform measure. Given data g_1, \dots, g_N independently drawn from F , $U_{z,q}^{(N)}$ is asymptotically normally distributed. In fact,*

$$\sqrt{N}(U_{z,q}^{(N)}/N - v) \rightarrow \mathcal{N}(0, \sigma^2) \quad \text{as } N \rightarrow \infty,$$

with $v = \int r^2(g)F(dg) - 1$ and $\sigma^2 = 4[\int r(g)u(g, h)\mu(dg)]^2 F(dh) - v^2]$, where u is defined below in (14) and r is defined as $r(g) = \int u(g, h)F(dh)$.

Proof. Proof follows from Proposition (4.6) of Giné [16] and the following lemma. The lemma is proved in Section 4 of the supplementary material Sepehri [41].

Lemma 4.6. *For $g_1, \dots, g_N \in SO(2n+1)$ one has*

$$U_{z,q}^{(N)} = \frac{1}{N} \int \left| \sum_{i=1}^N u(g_i, g) \right|^2 \mu(dg), \quad (13)$$

where u is defined through

$$u(g, h) = \sum_{\lambda \neq 0} \sqrt{d_\lambda c_\lambda(z, q)} \chi_\lambda(g^T h). \quad (14)$$

\square

Remark 4.7. A direct consequence is that $U_{z,q}^{(N)}$ is consistent against all fixed alternatives; not only the limiting distribution differs, so does the scaling.

Remark 4.8. The associate editor brought to our attention a recent work of Kerkyacharian, Nickl and Picard [26] that provides concentration inequalities and confidence band for a class of needlet density estimators on compact homogeneous manifolds, in particular on compact classical groups. Without getting into details, the following is a high level description of their approach. They introduce a needlet projection kernel $A_j(x, y)$ of order j for any non-negative integer j . Then, given observations x_1, \dots, x_n from a density f , they define a needlet density estimator through

$$f_n(j, y) = \frac{1}{n} \sum_{i=1}^n A_j(x_i, y).$$

Provided that f is bounded they prove the following concentration inequality.

Proposition 4.9 (Proposition 4 in Kerkycharian, Nickl and Picard [26]). *Let \mathbf{M} be a compact homogeneous manifold and suppose $f : \mathbf{M} \rightarrow [0, \infty)$ is bounded. We have for every $n \in \mathbb{N}$, every $j \in \mathbb{N}$, every $\Omega \subset \mathbf{M}$, and every $x \in \mathbf{M}$ that*

$$\mathbb{P}\left(\sup_{y \in \Omega} |f_n(j, y) - \mathbb{E}f_n(j, y)| \geq \sigma^R(\Omega, n, j, x)\right) \leq e^{-x},$$

where $\sigma^R(\Omega, n, j, x)$ depends on \mathbf{M} , n , j , Ω and x , and is defined explicitly in their paper.

In the case of f being the uniform distribution, $\mathbb{E}f_n(j, y) = f = \text{constant}$. Therefore, the concentration inequality above gives a confidence band for the density estimator $f_n(j, y)$. In particular, this confidence band can be used to construct **non-asymptotic** tests of goodness-of-fit for the uniform distribution on compact classical groups. The constants in the definition of $\sigma^R(\Omega, n, j, x)$ are computed explicitly for $SO(3)$ in the Supplementary material.

5. Numerical comparison of different tests

This section compares different tests discussed in this paper on the benchmark examples, with a particular focus on detection of the cutoff, as well as the new sampler. Setup 2.1 is considered ($n = 51$). Focus on the following four test: Rayleigh's test, Gine's test, $T_z^{(N)}$, and $U_{z,q}^{(N)}$. The numerical observations are summarized below.

The benchmark examples

Each test was computed 1000 times on the samples generated by the benchmark examples for different number of steps. Each of the 1000 simulations were based on $N = 200$ observations; $T_z^{(N)}$ was computed with $z = 0.5$ and $U_{z,q}^{(N)}$ with $z = 0.2$ and $q = 0.4$. The samples were generated using k steps for

$$k \in \{100, 150, 200, 250, 300, 350, 400, 450, 500\}$$

for the Kac's walk and

$$k \in \{50, 75, 90, 100, 110, 125, 140, 150, 175, 200\}$$

for product of random reflections. For each fixed number of steps the 1000 values were compared to those corresponding to the uniform distribution using the Anderson–Darling test. Figure 10 illustrates the values are plotted against the number of steps of the chain.

Figure 10 suggests that the Gine's test has the least power against the alternative generated by the Kac's walk among the four tests considered here. The Rayleigh's test and $U_{z,q}^{(N)}$ seem to perform similarly, indicating some evidence for a cutoff but, perhaps, earlier than it should possibly occur. The test based on $T_z^{(N)}$ outperforms the other three tests and provides evidence that a cutoff does not occur with less than 350 step, if it occurs at all.

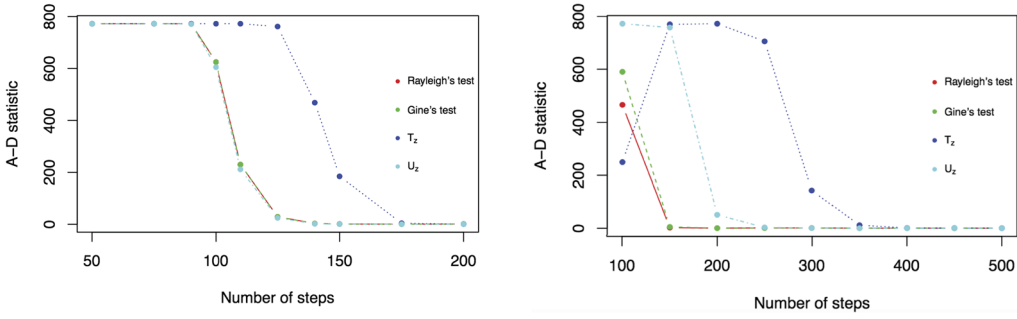


Figure 10. Values of the Anderson–Darling statistic for comparison of the uniform sample and the product of random reflections (left) and the Kac’s walk (right).

A similar but slightly different result holds for product of random reflection; the Rayleigh’s test has the least power, the Gine’s test and $U_{z,q}^{(N)}$ are qualitatively identical. Again, $T_z^{(N)}$ is superior to the other three tests; it picks up the occurrence of the cutoff and suggests that it might happen in around 175 steps.

The new sampler

The same procedure was repeated for $1 \leq k \leq 8$ iterations of the new sampler of Jones, Osipov and Rokhlin [22]. The result is shown in Figure 11.

The Rayleigh’s test, the Gine’s test, and the test based on $T_z^{(N)}$ exhibit no power against this alternative. Only $U_{z,q}^{(N)}$ has power against the alternative generated by a single iteration of the sampler. It also rejects the null hypothesis at 5%-level for the sample generated by two iterations of the sampler. It, too, has no power beyond two iterations. This is in agreement with the recommendation of Jones, Osipov and Rokhlin [22] on the number of iterations needed for mixing of the sampler.

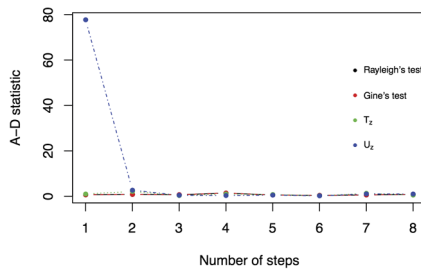


Figure 11. Values of the Anderson–Darling statistic for comparison of the uniform sample and the new sampler.

6. Asymptotic properties under local alternatives

This section studies properties of $T_z^{(N)}$ against local alternatives, that is, alternatives that approach the null as the sample size grows. The case of $U_{z,q}^{(N)}$ is similar. Section 7 in the supplementary material Sepehri [41] presents a brief introduction to Le Cam's theory of *asymptotically normal experiments*, which is the framework used here to study the local properties, as well as a detailed analysis of the local properties of the tests of this paper.

6.1. Power calculations under local alternatives

The following standard local hypothesis testing setting is considered here. Let $\{f(\cdot | \theta) | \theta \in \Omega\}$ be a Q.M.D. family of density functions with respect to the eigenvalue distribution induced by the Haar measure, where $\Omega \subset \mathbb{R}^k$ for a fixed k . Assume that $f(\cdot | \theta_0) = 1$, that is, θ_0 corresponds to $f_{1, \frac{3}{2}, \frac{1}{2}}$. Given data $x_1, \dots, x_N \sim f(\cdot | \theta)$, consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_0 + h/\sqrt{N}$ for a fixed $h \in \Omega$. Let $\ell(x | \theta) = \log f(x | \theta)$ be the *log-likelihood function*, $\eta(x | \theta) = \nabla_\theta \ell(x | \theta)$ the *score function*, and $\mathcal{I}(\theta) = -\mathbb{E}_\theta \nabla_\theta^2 \ell(x | \theta)$ the *Fisher information matrix* at θ . Let L_N denote the logarithm of the likelihood ratio of the data; Le Cam's first lemma (Lehmann and Romano [27], Theorem 12.2.3) asserts

$$\begin{aligned} L_N &= \sum_{i=1}^N \ell(x_i | \theta_0 + h/\sqrt{N}) - \ell(x_i | \theta_0) \\ &= \left(\frac{\sum_{i=1}^N \eta(x_i | \theta_0)}{\sqrt{N}} \right)^T h + \frac{1}{2} h^T \left(\frac{\sum_{i=1}^N \nabla_\theta^2 \ell(x_i | \theta_0)}{N} \right) h + o_p(1). \end{aligned} \quad (15)$$

Since the score function of QMD families is square-integrable, one has

$$\eta(\cdot | \theta_0) = \sum_{\lambda} \hat{\eta}(\lambda) \chi_{\lambda}(\cdot), \quad (16)$$

where $\hat{\eta}(\lambda)$ is the Fourier coefficient $\int \eta(x | \theta_0) \chi_{\lambda}(x) dx$ and the equality is interpreted in $\mathcal{L}^2(f_{1, \frac{3}{2}, \frac{1}{2}})$. Therefore,

$$L_N = \sum_{\lambda} \hat{\eta}(\lambda)^T h \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \chi_{\lambda}(x_i) \right) + \frac{1}{2} h^T \left(\frac{1}{N} \sum_{i=1}^N \nabla_\theta^2 \ell(x_i | \theta_0) \right) h + o_p(1).$$

As $N \rightarrow \infty$, using Law of Large Numbers and central limit theorem, one has

$$\frac{1}{N} \sum_{i=1}^N \nabla_\theta^2 \ell(x_i | \theta_0) \rightarrow -\mathcal{I}(\theta_0) \quad \text{and} \quad \frac{1}{\sqrt{N}} \sum_{i=1}^N \chi_{\lambda}(x_i) \rightarrow Z_{\lambda},$$

where Z_λ are independent standard normal variables. The joint limiting distribution of $(T_z^{(N)}, L_N)$ is

$$(T_z^{(N)}, L_N) \rightarrow (T_z, L),$$

where

$$T_z = \sum_{\lambda \neq 0} z^{|\lambda|} Z_\lambda^2 \quad \text{and} \quad L = \sum_{\lambda} (\widehat{\eta}(\lambda)^T h) Z_\lambda - \frac{1}{2} h^T \mathcal{I}(\theta_0) h. \quad (17)$$

Since $f(\cdot | \theta)$ is Q.M.D., Le Cam's third lemma (Lehmann and Romano [27], Theorem 12.3.3) implies that the limiting distribution of $T_z^{(N)}$ under $f(\cdot | \theta_0 + h/\sqrt{N})$ is given by the following characteristic function

$$\mathbb{E}_h e^{itT_z} = \mathbb{E}_0 [e^{itT_z} e^L].$$

Using (17) and the fact that $\{Z_\lambda\}$ are independent standard normal variables, one has

$$\begin{aligned} \mathbb{E}_h e^{itT_z} &= e^{-\frac{1}{2} h^T \mathcal{I}(\theta_0) h} \prod_{\lambda} \mathbb{E}_0 e^{itz^{|\lambda|} Z_\lambda^2 + (\widehat{\eta}(\lambda)^T h) Z_\lambda} \\ &= e^{\frac{1}{2} [-h^T \mathcal{I}(\theta_0) h + \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2]} \prod_{\lambda} \mathbb{E}_0 e^{itz^{|\lambda|} Z_\lambda^2 + (\widehat{\eta}(\lambda)^T h) Z_\lambda - \frac{1}{2} (\widehat{\eta}(\lambda)^T h)^2} \\ &= e^{\frac{1}{2} [-h^T \mathcal{I}(\theta_0) h + \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2]} \prod_{\lambda} \int_{\mathbb{R}} e^{itz^{|\lambda|} Z_\lambda^2 + (\widehat{\eta}(\lambda)^T h) Z_\lambda - \frac{1}{2} (\widehat{\eta}(\lambda)^T h)^2} \frac{e^{-\frac{1}{2} Z_\lambda^2}}{\sqrt{2\pi}} dZ_\lambda \\ &= e^{\frac{1}{2} [-h^T \mathcal{I}(\theta_0) h + \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2]} \prod_{\lambda} \int_{\mathbb{R}} e^{itz^{|\lambda|} Z_\lambda^2} \frac{e^{-\frac{1}{2} (Z_\lambda^2 - 2(\widehat{\eta}(\lambda)^T h) Z_\lambda + (\widehat{\eta}(\lambda)^T h)^2)}}{\sqrt{2\pi}} dZ_\lambda \\ &= e^{\frac{1}{2} [-h^T \mathcal{I}(\theta_0) h + \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2]} \prod_{\lambda} \int_{\mathbb{R}} e^{itz^{|\lambda|} Z_\lambda^2} \frac{e^{-\frac{1}{2} (Z_\lambda - (\widehat{\eta}(\lambda)^T h))^2}}{\sqrt{2\pi}} dZ_\lambda \\ &= e^{\frac{1}{2} [-h^T \mathcal{I}(\theta_0) h + \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2]} \prod_{\lambda} \mathbb{E} e^{itz^{|\lambda|} U_\lambda} \\ &= e^{\frac{1}{2} [-h^T \mathcal{I}(\theta_0) h + \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2]} \mathbb{E} e^{it \sum_{\lambda} z^{|\lambda|} U_\lambda} \\ &= \mathbb{E} e^{it \sum_{\lambda} z^{|\lambda|} U_\lambda}, \end{aligned}$$

where $U_\lambda = (Z_\lambda + \widehat{\eta}(\lambda)^T h)^2 \sim \chi_1^2((\widehat{\eta}(\lambda)^T h)^2)$ is a non-central chi-square variable on one degree of freedom with non-centrality parameter equal to $(\widehat{\eta}(\lambda)^T h)^2/2$. The last step holds because $h^T \mathcal{I}(\theta_0) h = \sum_{\lambda} (\widehat{\eta}(\lambda)^T h)^2$.

Thus, the limiting distribution of $T_z^{(N)}$ under the alternative $\theta_0 + h/\sqrt{N}$ is

$$T_z \sim \sum_{\lambda \neq 0} z^{|\lambda|} U_\lambda, \quad (18)$$

where U_λ are as above.

The following proposition is an immediate consequence of the argument above.

Proposition 6.1. *Let $c_{z,1-\alpha}$ be the asymptotic rejection threshold for $T_z^{(N)}$. That is, using Proposition 3.3,*

$$\mathbb{P}\left(\sum_{k=1}^{\infty} z^k \chi_{p(n,k)}^2 > c_{z,1-\alpha}\right) = \alpha,$$

where $p(n, k)$ is the number of partitions of k into at most n parts and the chi-square variables are independent. The asymptotic power under the local alternative $\theta_0 + h/\sqrt{N}$ is

$$\beta(h) = \mathbb{P}\left(\sum_{\lambda \neq 0} z^{|\lambda|} U_\lambda > c_{z,1-\alpha}\right),$$

for U_λ as in (18).

Example 6.2. For $\theta \in \mathbb{R}$ let $f(x | \theta) \propto \exp(\theta \operatorname{tr}(x))$. Then, $\eta(x | 0) = \operatorname{tr}(x)$ and $\widehat{\eta}(\lambda) = 0$ for $\lambda \neq (1)$. The local power under θ/\sqrt{N} is

$$\beta(\theta) = \mathbb{P}\left(z \chi_1^2(\theta^2) + \sum_{k=2}^{\infty} z^k \chi_{p(n,k)}^2 > c_{z,1-\alpha}\right).$$

The results of this section can be extended to families with infinite-dimensional parameter space, under mild regularity conditions. See Example 7.13 in the supplementary material Sepehri [41] for an example.

6.2. Global asymptotic power function against local alternatives

It is well known that any test of goodness-of-fit exhibits poor power against local (contiguous) alternatives, except possibly in a finite number of directions. This section presents some results of this nature for $T_z^{(N)}$. For more details and statements in full generality, see Janssen [21].

6.2.1. Spectral decomposition of the power function

Consider the standard local hypothesis setup. For an arbitrary non-parametric unbiased test ϕ in the limiting Gaussian shift experiment, Janssen [21] has shown that the curvature of the power function admits a principal component decomposition.

Focus on $T_z^{(N)}$. Using the notation of Section 6.1, the asymptotic power against the local alternative $H_1 : \theta = \theta_0 + h/\sqrt{N}$ is given by Proposition 6.1 as

$$\beta(h) = \mathbb{P}\left(\sum_{\lambda \neq 0} z^{|\lambda|} U_\lambda > c_{z,1-\alpha}\right).$$

The rejection cutoff $c_{z,1-\alpha}$ is such that $\beta(\mathbf{0}) = \alpha$. Using Theorem 1 in Beran [6] one has the following second order Taylor expansion of $\beta(th)$ around $t = 0$

$$\beta(t \cdot h) = \alpha + \frac{t^2}{2} \sum_{\lambda} (\hat{\eta}(\lambda)^T h)^2 [G_\lambda(c_{z,1-\alpha}) - \alpha] + o(t^2),$$

where $G_\lambda(x) = \mathbb{P}(\sum_{\mu} z^{|\mu|} V_\mu > x)$, and V_μ is a χ_1^2 for $\mu \neq \lambda$, and V_λ is a χ_3^2 random variable. Therefore, the curvature of the power function around $t = 0$ is

$$a(h) = \langle T(h), h \rangle,$$

for the positive-definite bi-linear operator

$$T = \sum_{\lambda} [G_\lambda(c_{z,1-\alpha}) - \alpha] \hat{\eta}(\lambda) \hat{\eta}(\lambda)^T.$$

This readily gives a principal decomposition of the curvature, with principal components $\{\hat{\eta}(\lambda) \hat{\eta}(\lambda)^T\}$ and eigenvalues $G_\lambda(c_{z,1-\alpha}) - \alpha \geq 0$. For a fixed z and α , $G_\lambda(c_{z,1-\alpha}) - \alpha$ is a decreasing function of $|\lambda|$. Thus, the highest gain in power is against those alternatives that put most of the load on principal components for smaller $|\lambda|$. Theorem 2.1 in Janssen [21] implies that T is a Hilbert–Schmidt operator and $\|T\|^2 < 2\alpha(1 - \alpha)$. This implies that any test performs poor against all alternatives except for a finite dimensional space. More details and various other statements are given in Section 7 of the supplementary material Sepehri [41]. Local asymptotic relative efficiency and explicit bounds on the dimension of the subspace against which $T_z^{(N)}$ has power are also considered in Section 7.3.3 of the supplementary material Sepehri [41].

6.3. Asymptotic admissibility

This section argues that the new tests of this paper are asymptotically admissible in the following sense. As discussed in Section 6.1, there is a limiting hypothesis testing problem that captures the asymptotic properties of $T_z^{(N)}$ in the local hypothesis testing problem. That is, the limiting Gaussian process $\{Z_\lambda\}$, where under the limiting distribution corresponding to $h \in \Omega$, $Z_\lambda \sim \mathcal{N}(\hat{\eta}(\lambda)^T h, 1)$. The problem is to test $H_0 : h = 0$. The limiting test statistic is $T_z = \sum_{\lambda \neq 0} z^{|\lambda|} Z_\lambda^2$. The main result is based on the following definition and is presented below.

Definition 6.3 (Asymptotic admissibility). The sequence of test statistics $T_z^{(N)}$ is called asymptotically admissible if the limiting test T_z is admissible for the limiting hypothesis testing problem.

Corollary 6.4. *The limiting tests based on T_z is admissible. Therefore, the tests based on $\{T_z^{(N)}\}$ is asymptotically admissible.*

Proof. The proof is based on the following result of Birnbaum [8].

Lemma 6.5 (Strasser [42], Theorem 30.4). *Let $C \subset \mathbb{H}$ be a closed convex subset. Then,*

$$\phi(x) = \begin{cases} 1 & \text{if } x \notin C, \\ 0 & \text{if } x \in C, \end{cases}$$

is admissible for the testing problem $h = 0$ against $h \neq 0$ and is uniquely determined by its power function.

Note that the test based on T_z rejects for $\{\mathbf{Z} \in \mathbb{R}^\infty \mid \sum_\lambda z^{|\lambda|} Z_\lambda^2 > c_{z,1-\alpha}\}$. The set

$$C_z = \left\{ \mathbf{Z} \in \mathbb{R}^\infty \mid \sum_\lambda z^{|\lambda|} Z_\lambda^2 \leq c_{z,1-\alpha} \right\}$$

is clearly convex and closed. Thus the assertion follows from Lemma 6.5. □

A similar statement is true for $\{U_{z,q}^{(N)}\}$ and is omitted here.

7. Other compact groups

The compact classical groups fall into four general classes:

1. *Type A:* $U(n)$ and $SL(n)$.
2. *Type B:* $SO(2n+1)$.
3. *Type C:* $Sp(2n)$.
4. *Type D:* $SO(2n)$.

For each type the analogous tests to $T_z^{(N)}$ and $U_{z,q}^{(N)}$ are introduced in this section; only the definitions and explicit formulas are provided. Derivation of the asymptotic null and alternative distributions, and local power are identical to those for $SO(2n+1)$, hence omitted here. Note that each case requires particular facts and considerations from representation theory; details, derivations, and proofs are provided in Section 2 of the supplementary material Sepehri [41].

7.1. The test based on the eigenvalues

For the groups of type A, Coram and Diaconis [10] introduced a test which inspired the test based on $T_z^{(N)}$ of the present paper. The case of type B groups ($SO(2n+1)$) was discussed in Section 3.2.2. Type C and D are discussed below.

7.1.1. Type C

For $0 < z < 1$ define the test statistics $T_{C,z}^{(N)}$ as

$$T_{C,z}^{(N)} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N K_z^C(g_i, g_j),$$

where K_z^C is as in (7). A closed form for $K_z^C(g, h)$ can be found using the Cauchy identity for the symplectic group (Sepehri [41], Theorem 2.7) as follows:

$$K_z^C(g, h) = \frac{(1 - z^2)^m \det\left(\frac{1}{(1 - zx_i y_j)(1 - zx_i^{-1} y_j)(1 - zx_i y_j^{-1})(1 - zx_i^{-1} y_j^{-1})}\right)}{z^{(m)} \prod_{i < j} (y_i + y_i^{-1} - (y_j + y_j^{-1})) \prod_{i < j} (x_i + x_i^{-1} - (x_j + x_j^{-1}))} - 1,$$

where $\{x_i^{\pm}\}$ and $\{y_i^{\pm}\}$ are eigenvalues of g and h respectively.

7.1.2. Type D

For $0 < z < 1$ define the test statistics $T_{D,z}^{(N)}$ as

$$T_{D,z}^{(N)} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N K_z^D(g_i, g_j),$$

where K_z^D is as in (7). A closed form for $K_z^D(g, h)$ is given by the Cauchy identity for $SO(2n)$ (Sepehri [41], Theorem 2.9) as follows:

$$K_z^D(g, h) = \frac{\det\left(\frac{1}{1 - zx_i y_j} + \frac{1}{1 - zx_i^{-1} y_j} + \frac{1}{1 - zx_i y_j^{-1}} + \frac{1}{1 - zx_i^{-1} y_j^{-1}}\right)}{z^{(m)} \prod_{i < j} (y_i + y_i^{-1} - (y_j + y_j^{-1})) \prod_{i < j} (x_i + x_i^{-1} - (x_j + x_j^{-1}))} - 1,$$

where $\{x_i^{\pm}\}$ and $\{y_i^{\pm}\}$ are eigenvalues of g and h respectively.

7.2. The test beyond the eigenvalues

A test similar to $U_{z,q}^{(N)}$ can be constructed for all compact groups. With abuse of notation, these tests all are denoted by $U_{z,q}^{(N)}$. The case of $SO(2n + 1)$ is already discussed in Section 4. The other cases are considered in this section. Only the definitions and explicit formulas are presented here. A detailed derivation and required facts from representation theory, as well as the proofs, are provided in Section 2 of the supplementary material Sepehri [41].

7.2.1. Type A

For $U(n)$, the test analogous to $U_{z,q}^{(N)}$ is defined as

$$U_{z,q}^{(N)} = \frac{1}{N} \sum_{i,j=1}^N \left(\prod_{l,k=1}^n \frac{1}{1 - zq^{l-1}y_k^{i,j}} - 1 \right),$$

where $y_1^{i,j}, \dots, y_n^{i,j}$ are the eigenvalues of $g_i^* g_j$.

7.2.2. Type C

Definition for groups of type C is as follows

$$U_{z,q}^{(N)} = \frac{1}{N} \sum_{k,l=1}^N \left(\frac{\prod_{i < j} (1 - z^2 q^{i+j-2})}{\prod_{i,j} (1 - zq^{i-1}y_j^{k,l})(1 - zq^{i-1}(y_j^{k,l})^{-1})} - 1 \right),$$

where $\{y_j^{k,l}, (y_j^{k,l})^{-1} \mid j = 1, \dots, n\}$ are the eigenvalues of $g_k g_l^T$.

7.2.3. Type D

Definition for groups of type D is as follows

$$U_{z,q}^{(N)} = \frac{1}{N} \sum_{k,l=1}^N \left(\frac{\prod_{i \leq j} (1 - z^2 q^{i+j-2})}{\prod_{i,j} (1 - zq^{i-1}y_j^{k,l})(1 - zq^{i-1}(y_j^{k,l})^{-1})} - 1 \right),$$

where $\{y_j^{k,l}, (y_j^{k,l})^{-1} \mid j = 1, \dots, n\}$ are the eigenvalues of $g_k g_l^T$.

8. Discussion

The current paper introduces and analyzes two new families of tests of uniformity on the compact classical groups. These tests are validated on two benchmark examples: the random walk of Kac and the products of random reflections. They exhibit satisfying agreement with the existing theory about the mixing-time of both random walks. The new tests, and several others, are applied to the new sampler of Jones, Osipov and Rokhlin [22]; all but one of the new tests failed to reject the null hypothesis of uniformity after any number of iterations of the new sampler. One of the new tests confirmed the prescribed number of steps to be used with the sampler in order to get approximately uniform outputs.

Acknowledgments

I am greatly indebted to my doctoral advisor, Persi Diaconis, for suggesting the problem and his continuing guidance and support. I thank Daniel Bump for discussions on representation

theory; Arun Ram for sharing his notes on the character theory and symmetric function theory; Sourav Chatterjee and David Siegmund for helpful comments and discussions; Matan Gavish for sharing his course report on numerical investigation of the mixing-time for a Markov chain sampler on the unitary group. I would like to express my special thanks to the associate editor and the anonymous referee because of their useful comments that improved this paper significantly. In particular, Remark 4.8 was the result of a suggestion by the associate editor.

This work is part of author's PhD dissertation at Stanford University.

Supplementary Material

Supplement to “New tests of uniformity on the compact classical groups as diagnostics for weak-* mixing of Markov chains” (DOI: [10.3150/18-BEJ1029SUPP](https://doi.org/10.3150/18-BEJ1029SUPP); .zip). We provide additional supporting material including background and proofs from representation theory, proofs of some of the results, introduction to Le Cam's theory, further derivation and analysis of local properties, as well as a test based on the trace. The motivating example is also reviewed.

References

- [1] Abramowitz, M. and Stegun, I.A. (1966). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. **55**. New York: Dover Publications. [MR0208797](#)
- [2] Ajne, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika* **55** 343–354. [MR0235662](#)
- [3] Andrews, G.E. (1998). *The Theory of Partitions*. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. [MR1634067](#)
- [4] Arias-Castro, E., Pelletier, B. and Saligrama, V. (2018). Remember the curse of dimensionality: The case of goodness-of-fit testing in arbitrary dimension. *J. Nonparametr. Stat.* **30** 448–471. [MR3794401](#)
- [5] Baringhaus, L. (1991). Testing for spherical symmetry of a multivariate distribution. *Ann. Statist.* **19** 899–917. [MR1105851](#)
- [6] Beran, R. (1975). Tail probabilities of noncentral quadratic forms. *Ann. Statist.* **3** 969–974. [MR0381122](#)
- [7] Beran, R.J. (1968). Testing for uniformity on a compact homogeneous space. *J. Appl. Probab.* **5** 177–195. [MR0228098](#)
- [8] Birnbaum, A. (1955). Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Ann. Math. Stat.* **26** 21–36. [MR0067438](#)
- [9] Bump, D. (2004). *Lie Groups*. *Graduate Texts in Mathematics* **225**. New York: Springer. [MR2062813](#)
- [10] Coram, M. and Diaconis, P. (2003). New tests of the correspondence between unitary eigenvalues and the zeros of Riemann's zeta function. *J. Phys. A* **36** 2883–2906. [MR1986397](#)
- [11] Diaconis, P. (2003). Patterns in eigenvalues: The 70th Josiah Willard Gibbs lecture. *Bull. Amer. Math. Soc. (N.S.)* **40** 155–178. [MR1962294](#)
- [12] Diaconis, P. and Mallows, C. (1986). On the trace of random orthogonal matrices. Unpublished Manuscript. Results Summarized in Diaconis (1990).
- [13] Diaconis, P. and Shahshahani, M. (1986). Products of random matrices as they arise in the study of random walks on groups. In *Random Matrices and Their Applications* (Brunswick, Maine, 1984). *Contemp. Math.* **50** 183–195. Providence, RI: Amer. Math. Soc. [MR0841092](#)

- [14] Diaconis, P. and Shahshahani, M. (1987). The subgroup algorithm for generating uniform random variables. *Probab. Engrg. Inform. Sci.* **1** 15–32.
- [15] Downs, T.D. (1972). Orientation statistics. *Biometrika* **59** 665–676. [MR0345334](#)
- [16] Giné, E.M. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3** 1243–1266. [MR0388663](#)
- [17] Giné, E.M. (1975). The addition formula for the eigenfunctions of the Laplacian. *Adv. Math.* **18** 102–107. [MR0377997](#)
- [18] Goodman, R. and Wallach, N.R. (2009). *Symmetry, Representations, and Invariants. Graduate Texts in Mathematics* **255**. Dordrecht: Springer. [MR2522486](#)
- [19] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. [MR3363437](#)
- [20] Hermans, M. and Rasson, J.-P. (1985). A new Sobolev test for uniformity on the circle. *Biometrika* **72** 698–702. [MR0817587](#)
- [21] Janssen, A. (1995). Principal component decomposition of non-parametric tests. *Probab. Theory Related Fields* **101** 193–209. [MR1318192](#)
- [22] Jones, P.W., Osipov, A. and Rokhlin, V. (2011). Randomized approximate nearest neighbors algorithm. *Proc. Natl. Acad. Sci.* **108** 15679–15686.
- [23] Jupp, P.E. and Spurr, B.D. (1983). Sobolev tests for symmetry of directional data. *Ann. Statist.* **11** 1225–1231. [MR0720267](#)
- [24] Jupp, P.E. and Spurr, B.D. (1985). Sobolev tests for independence of directions. *Ann. Statist.* **13** 1140–1155. [MR0803763](#)
- [25] Kac, M. (1959). *Probability and Related Topics in Physical Sciences* **1**. American Mathematical Soc.
- [26] Kerkycharian, G., Nickl, R. and Picard, D. (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Related Fields* **153** 363–404. [MR2925578](#)
- [27] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. [MR2135927](#)
- [28] Liberty, E., Woolfe, F., Martinsson, P.-G., Rokhlin, V. and Tygert, M. (2007). Randomized algorithms for the low-rank approximation of matrices. *Proc. Natl. Acad. Sci. USA* **104** 20167–20172. [MR2366406](#)
- [29] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics*. Chichester: Wiley. [MR1828667](#)
- [30] Mehta, M.L. (2004). *Random Matrices*, 3rd ed. *Pure and Applied Mathematics (Amsterdam)* **142**. Amsterdam: Elsevier/Academic Press. [MR2129906](#)
- [31] Oliveira, R.I. (2009). On the convergence to equilibrium of Kac’s random walk on matrices. *Ann. Appl. Probab.* **19** 1200–1231. [MR2537204](#)
- [32] Pak, I. and Sidenko, S. (2007). Convergence of Kac’s random walk. Preprint. Available at <http://www-math.mit.edu/~pak/research.html>.
- [33] Pillai, N.S. and Smith, A. (2016). On the mixing time of Kac’s walk and other high-dimensional Gibbs samplers with constraints. Preprint. Available at [ArXiv:1605.08122](https://arxiv.org/abs/1605.08122).
- [34] Porod, U. (1996). The cut-off phenomenon for random reflections. *Ann. Probab.* **24** 74–96. [MR1387627](#)
- [35] Prentice, M.J. (1978). On invariant tests of uniformity for directions and orientations. *Ann. Statist.* **6** 169–176. [MR0458721](#)
- [36] Rayleigh, L. (1880). XII. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philos. Mag.* **10** 73–78.
- [37] Rokhlin, V. and Tygert, M. (2008). A fast randomized algorithm for overdetermined linear least-squares regression. *Proc. Natl. Acad. Sci. USA* **105** 13212–13217. [MR2443725](#)

- [38] Römisch, W. (2005). Delta method, infinite dimensional. *Encyclopedia of Statistical Sciences*.
- [39] Rosenthal, J.S. (1994). Random rotations: Characters and random walks on $SO(N)$. *Ann. Probab.* **22** 398–423. [MR1258882](#)
- [40] Sengupta, A. and Pal, C. (2001). On optimal tests for isotropy against the symmetric wrapped stable-circular uniform mixture family. *J. Appl. Stat.* **28** 129–143. [MR1819907](#)
- [41] Sepehri, A. (2019). Supplement to “New tests of uniformity on the compact classical groups as diagnostics for weak-* mixing of Markov chains.” DOI:[10.3150/18-BEJ1029SUPP](#).
- [42] Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. *De Gruyter Studies in Mathematics* **7**. Berlin: de Gruyter. [MR0812467](#)
- [43] Thomson, B.S., Bruckner, J.B. and Bruckner, A.M. (2008). *Elementary Real Analysis*. Available at [ClassicalRealAnalysis.com](#).
- [44] van der Vaart, A.W. and Wellner, J.A. (1996). Weak convergence. In *Weak Convergence and Empirical Processes* 16–28. New York: Springer. [MR1385671](#)
- [45] Watson, G.S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48** 109–114. [MR0131930](#)
- [46] Watson, G.S. (1962). Goodness-of-fit tests on a circle. II. *Biometrika* **49** 57–63. [MR0138179](#)
- [47] Watson, G.S. (1967). Another test for the uniformity of a circular distribution. *Biometrika* **54** 675–677. [MR0221649](#)
- [48] Wellner, J.A. (1979). Permutation tests for directional data. *Ann. Statist.* **7** 929–943. [MR0536498](#)
- [49] Weyl, H. (1939). *The Classical Groups. Their Invariants and Representations*. Princeton, NJ: Princeton Univ. Press. [MR0000255](#)

Received June 2017 and revised February 2018