# BOUNDS FOR KAC'S MASTER EQUATION

Persi Diaconis
Department of Mathematics
Stanford University
Stanford, CA, 94305

Laurent Saloff–Coste
CNRS, Toulouse France
and
Department of Mathematics
Cornell University
Ithaca, NY, 14853
e–mail: lsc@math.cornell.edu

Abstract.

Mark Kac considered a Markov Chain on the $n$–sphere based on random rotations in randomly chosen coordinate planes. This same walk was used by Hastings on the orthogonal group. We show that the walk has spectral gap bounded below by $c/n^3$. This and curvature information are used to bound the rate of convergence to stationarity.

Research partially supported by Nato Grant CRG 950686 and by NSF Grants DMS-9802855 and DMS-9504379

## 1. Introduction.

On Euclidean space $\mathbb{R}^n$ consider the rotation

$$g_{ij}(\theta) = \begin{pmatrix} 1 & 0 & & \dots & & 0 \\ 0 & \ddots & & & & \\ & & c & & s & \\ \vdots & & & \ddots & & \vdots \\ & & -s & & c & \\ & & & & \ddots & 0 \\ 0 & & & \dots & 0 & 1 \end{pmatrix} \tag{1.1}$$

where all the entries on the diagonal are equal to 1 except for the $(i,i)$ and $(j,j)$ entries that are equal to $c = \cos(\theta)$, and all the off-diagonal entries are 0 except for the $(i,j)$ and $(j,i)$ entries that are respectively $+s$ and $-s$ with $s = \sin(\theta)$, $0 \le \theta < 2\pi$. This represents a clockwise rotation by $\theta$ in the $i$, $j$ plane, $1 \le i < j \le n$. We consider the random walk on the orthogonal group $SO(n)$ generated by repeatedly multiplying by $g_{ij}(\theta)$ for $1 \le i < j \le n$ chosen uniformly and $\theta$ chosen uniformly in $[0, 2\pi)$. Call this measure $Q$ and let $Q^{*k}$ be the $k^{\text{th}}$ convolution power. Let $U$ denote the uniform distribution (Haar measure) on $SO(n)$. Our main result shows that $Q^{*k}$ is close to $U$ for $k$ of order $n^5 \log n$.

**Theorem 1.** *The random rotations measure $Q$ on $SO(n)$ satisfies*

$$|Q^{*k}(f) - U(f)| \le 7\sqrt{n}\Big(1 - \frac{1}{60n^3}\Big)^{2k/\left(\binom{n}{2}+2\right)}$$

*for $f$ any bounded Lipschitz function of norm at most* 1.

In Section 5, we prove a better though less explicit result showing convergence after $n^4 \log n$ steps.

**First motivation.** The present problem arose as part of Mark Kac's study of Boltzmann's derivation of a basic equation of kinetic theory (1956), (1959, pg.109–132). Kac simplified the problem to an n-particle system in one dimension. Assuming the positions are in equilibrium, he studied the velocities $(v_1, v_2, \dots, v_n)$. Kac assumed that only the total energy $v_1^2 + v_2^2 + \cdots + v_n^2 = n\sigma^2$ is conserved (hence the restriction to the sphere).

In Kac's model, the particles exchange energy as follows: at the times of a Poisson processes with rate $n\lambda$, a pair of indices $(i,j)$ is chosen at random and the velocities $v_i, v_j$ are changed by

$$(v_i, v_j) \to (v_i \cos(\theta) + v_j \sin(\theta), -v_i \sin(\theta) + v_j \cos(\theta))$$

with $\theta$ chosen uniformly in $[0, 2\pi)$. This gives rise to the operator $H_t = e^{-n\lambda t(I-Q)}$ on $L^2$ of the $n$-sphere with

$$Qf(V) = \frac{1}{2\pi\binom{n}{2}} \sum_{i<j} \int_0^{2\pi} f(g_{ij}(\theta)V)d\theta \tag{1.2}$$

where $V = (v_1, \ldots, v_n)$ and $g_{ij}(\theta)$ is the rotation defined at (1.1).

If an initial density $\phi(V, 0)$ is assumed on the sphere then the process at time $t$ has density $\phi(V, t) = H_t\phi(V, 0)$. Differentiating shows that $\phi(V, t)$ satisfies Kac's master equation

$$\frac{\partial\phi(V,t)}{\partial t} = -n\lambda(I-Q)\phi(V,t) = \frac{n\lambda}{2\pi\binom{n}{2}} \sum_{i<j} \int_0^{2\pi} [\phi(g_{ij}(\theta)V,t) - \phi(V,t)]d\theta. \tag{1.3}$$

Of course this is linear in $\phi$. To get an analog of the non–linear Boltzmann equation, Kac studied the marginal distribution of the first coordinate $v_1$, call this $f_1^n(v, t)$, and of the first two coordinates $f_2^n(v, w, t)$. **Assuming** the sequence of initial densities $\phi^n(V, 0)$ is symmetric in $(v_1, \ldots, v_n)$ and varies with $n$ so that the marginals approximately factor:

$$f_2^n(v, w, 0) \sim f_1^n(v, 0)f_1^n(w, 0)$$

Kac proved what has come to be called "propagation of chaos"

$$f_2^n(v, w, t) \sim f_1^n(v, t)f_1^n(w, t).$$

Plugging this in (1.3) gives Kac analog of **Boltzmann's equation**

$$\begin{aligned}\frac{\partial f(x,t)}{\partial t} = C \int_{-\infty}^{\infty} \int_0^{2\pi} \big\{ & f(x\cos\theta + y\sin\theta, t)f(-x\sin\theta + y\cos\theta, t) \\ & - f(x,t)f(y,t)\big\} d\theta dy.\end{aligned}$$

Kac left many details of the derivation vague. Among these is a bound for the second eigenvalue of the basic operator $n(I - Q)$ with $Q$ as in (1.2). Kac comments that it depends on $n$ and conjectures that it is bounded away from zero, uniformly in $n$. This corresponds to a bound on the second eigenvalue of $Q$ of the form

$$1 - \frac{\text{const}}{n}.$$

The argument for Theorem 1 gives the lower bound $1 - 1/(60n^3)$ and the upper bound $1 - 2/n$. Exactly determining the gap would be useful in pushing Kac's attempt to justify Boltzmann's proof of the $H$-theorem: entropy of the marginal density $f_1(v, t)$ is decreasing in $t$. Of course we know this for the entropy of the joint distribution density $\phi(V, t)$ and Kac (1956, pg.185-186) shows that if $\phi(V, t) \sim \prod f_1(v_i, t)$ in a suitable sense then the desired result transfers to $f_1$.

The stochastic dynamics underlying Kac formulation is used as an algorithm for studying solutions of Boltzmann's original equation. Indeed, following Kac (1956, Sec.2), Grunbaum (1971) gave an appropriate stochastic dynamics for the spatially homogeneous version of Boltzmann's original equation. This is further developed by Uchiyama (1988). Méléard (1996) points out that these stochastic processes are essentially the same as algorithms of Bird (1976) and Nambu (1983) for solving Boltzmann's original equation. See Perthame (1994) for more of this. Thus, spectral gaps and results of the present paper correspond to running time bounds for these algorithms applied to Kac's equation. Kac's paper has given rise to a fair sized literature on "propagation of chaos". Useful surveys are in Méléard (1996) and Sznitman (1991) and the thesis of Gottlieb (1998). There is also some literature on Kac's equation (1.3). See McKean (1966), Grünbaum (1972), Desvillettes (1995), Carlen et al (1997) and Méléard (1996). A good overall survey on Boltzmann's equation is Cercignani et al (1994).

**Second motivation.** The same random walk acting on all of $SO(n)$ was suggested by Hastings (1970) as a simple way of generating an approximately random rotation. In his paper, Hastings reports some numerical studies when $n = 50$. He used the random walk to estimate the average value of a function $f$: $J_f = \int_{SO(n)} f(m)dm$ by $\widetilde{J}_f = \frac{1}{N}\sum_{i=1}^{N} f(m_i)$. For example, if $f(m) = m_{11}^2 + m_{22}^2 + \cdots + m_{nn}^2$ it is known that $J_f = 1$. Using $N = 1000$, starting the walk at the identity, Hastings obtained $\widetilde{J}_f = 3.5 \pm 1.5$. He supposed this poor estimate was due to the starting place and showed empirically that if the walk is started more "centrally" (e.g., at a real version of the discrete Fourier matrix) satisfactory estimates were obtained.

We note that the walk analyzed here is an example of what statisticians call the Gibbs sampler (See e.g., Smith and Roberts (1993)): to sample from a vector distribution, pick a few random coordinates, freeze the rest, and sample from the correct distribution on the chosen coordinates given the frozen coordinates. The Gibbs sampler is also known as the heat bath or Glauber dynamics. To generate from the uniform distribution on the sphere these algorithms pick two coordinates at random and then choose from the conditional distribution given the rest. This is just Kac's walk! Our theorem thus gives one of very few available examples of a rate of convergence result for the Gibbs sampler.

One further motivation for the careful study of the present example is to begin the extension of the geometric theory of Markov chains developed in [7,8,38] from finite to continuous state spaces. There has been some previous study of rates of convergence of random walk on compact groups. Diaconis and Shahshahani (1986), Rosenthal (1994) and Porod (1995, 1996 a,b) study the walk on $O(n)$ generated by random reflexions. This walk is constant on conjugacy classes so character theory can be used to bound convergence.

One difficulty with Kac's walk is that the convolution $Q^{*k}$ is not absolutely continuous with respect to Haar measure. There is positive probability that all the $g_{ij}$ chosen have the same value $(i, j)$. Thus $Q^{*k}$ does not converge in $L^2$. This blocks the usual route used in [7, 8, 38] of bounding total variation convergence by $L^2$ convergence. We are able to prove total variation convergence (Corollary 2.1) but the argument only shows convergence after order $4^{n^2}$ steps.

The arguments developed in the present paper use a factorization of Haar measure to allow piecewise continuous paths to be chosen between points of $SO(n)$. It then applies comparison inequalities, much as in [7,8], to prove spectral gap bounds. The operator $Q$ is far from compact: In Section 3 we find eigenvalues with infinite multiplicity.

Section 2 gives a careful description of the factorization of Haar measure that we use. Basically, the Euler angles of a randomly chosen element in $SO(n)$ are independent beta variates. Section 3 contains the spectral gap estimates. It also gives results for $\theta$ chosen from a non–uniform distribution at each stage and for the walk driven by uniform rotations in planes corresponding to consecutive coordinates $(i, i+1)$, $1 \leq i \leq n-1$. Section 4 reviews needed geometric tools (Ricci curvature, diameter and volume growth) on $SO(n)$. The quantitative bound on the dual bounded Lipschitz rate of weak convergence using a spectral gap estimate may be of general interest. Section 5 shows how one can take full advantage of comparison inequalities to obtain improved rates of convergence for random walk on group, much as in [7].

It is straightforward to extend the analysis to a parallel walk on the unitary group $U(n)$. This may be of interest in connection with quantum computing. Randomly choosing a pair of coordinates and multiplying by a random element of $U(2)$ can be studied as a model of noisy quantum circuits. See Aharonov and Ben-Or (1997) and Shor (1996).

In preliminary work, David Maslin (1999) has determined that the spectral gap of Kac's walk equals

$$\frac{1}{2n} + \frac{3}{2n(n-1)}$$

with multiplicity $n(n-1)(n+6)(n+1)/24$. His argument makes heavy use of representation theory of $SO(n)$. E. Janvresse (1999) has also obtained a bound of the from $c/n$ for the walk on the sphere by a different method.

**Acknowledgement.** We thank Eric Carlen for telling us about Kac's work and its development. We thank David Maslin for keeping us informed about his progress on Kac's problem.

## 2. A Factorization of Haar Measure.

This section gives a probabilistic interpretation to Hurwitz' (1897) construction of Haar measure on $SO(n)$.

Let $g_i(\theta) = g_{i-1,i}(\theta)$, $2 \leq i \leq n$. These rotations act on a column vector $[x_1, \dots, x_n]^t$ by

$$g_i x = [x_1, \dots, cx_{i-1} + sx_i, -sx_{i-1} + cx_i, \dots, x_n]^t.$$

Choosing $c = \frac{\pm x_{i-1}}{\sqrt{x_i^2 + x_{i-1}^2}}$, $s = \frac{\pm x_i}{\sqrt{x_i^2 + x_{i-1}^2}}$ (same sign in each) results in a vector with $i^{\text{th}}$ coordinate zero. A succession of such rotations can be used to bring a given $m \in SO(n)$ to diagonal form. Suppose e.g. that $n$ is 4. Then, with obvious notation

$$\begin{array}{cccc} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{array} \xrightarrow{g_4} \begin{array}{cccc} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{array} \xrightarrow{g_3} \begin{array}{cccc} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{array} \xrightarrow{g_2} \begin{array}{cccc} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{array}$$

$$\xrightarrow{g_4'} \begin{array}{cccc} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{array} \xrightarrow{g_3'} \begin{array}{cccc} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{array} \xrightarrow{g_4''} \begin{array}{cccc} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{array}$$

The final matrix is orthogonal and so all off diagonal entries must be zero and all diagonal entries must be $\pm 1$. By using the free choice of sign in $g_i$, the final matrix may be taken as the identity. Thus

$$m = (g_4^t g_3^t g_2^t)(g_4^{'t} g_3^{'t})(g_4^{''t}).$$

Clearly this generalizes so that any element of $SO(n)$ can be represented as

$$m = (g_n^1 g_{n-1}^1 \cdots g_2^1)(g_n^2 \cdots g_3^2) \cdots (g_n^{n-2} g_{n-1}^{n-2}) g_n^{n-1}.$$

Hurwitz discovered that a uniform probability distribution on $SO(n)$ (now called Haar measure) can be derived by giving an appropriate product measure to the $\{g_j^i\}$ above. This may be seen by an elementary argument. Recall the gamma density on $[0, \infty)$, $\gamma_a(x) = \Gamma(a)^{-1} e^{-x} x^{a-1}$. The following facts from a first probability course are useful.

**Lemma 2.1.** *If $X_1, X_2, \dots, X_n$ are independent with $X_i$ having a $\gamma_{a_i}$ distribution then*

*(1)* $\dfrac{X_1}{X_1+X_2}, \dfrac{X_1+X_2}{X_1+X_2+X_3}, \dots, \dfrac{X_1+\cdots+X_{n-1}}{X_1+\cdots+X_n}$, $X_1+\cdots+X_n$ *are independent with* $\dfrac{X_1+\cdots+X_i}{X_1+\cdots+X_{i+1}}$ *having density* $\beta(A,B;x) = \dfrac{\Gamma(A)}{\Gamma(A)\Gamma(B)} x^{A-1}(1-x)^{B-1}$ *on* $[0,1]$ *for* $A = a_1 + \cdots + a_i$, $B = a_{i+1}$ *and* $S_n = X_1 + \cdots + X_n$ *having density* $\gamma_{a_1+\cdots+a_n}$.

*(2)* $S_n$ *is independent of the vector* $\left(\dfrac{X_1}{S_n}, \dots, \dfrac{X_n}{S_n}\right)$.

**Lemma 2.2.** *Let $Z_1, Z_2, \dots, Z_n$ be independent standard Gaussian random variables. Then*

*(1)* $\Big(\frac{Z_1}{\sqrt{Z_1^2+\cdots+Z_n^2}}, \dots, \frac{Z_n}{\sqrt{Z_1^2+\cdots+Z_n^2}}\Big)$ *is uniformly distributed on the $n$–sphere with first coordinate having density*

$$\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}(1-x^2)^{\frac{n-3}{2}} \quad on \quad [-1,1].$$

*(2) Let $W_1, \dots, W_{n+1}$ be standard Gaussian variables, independent of each other and of $Z_1, \dots, Z_n$ in (1). Let*

$$A = \frac{W_1}{\sqrt{W_1^2+\cdots+W_{n+1}^2}}, \ B = \sqrt{\frac{W_2^2+\cdots+W_{n+1}^2}{W_1^2+\cdots+W_{n+1}^2}} = \sqrt{1-A^2}.$$

*Then*

$$A, B\Big(\frac{Z_1}{\sqrt{Z_1^2+\cdots+Z_n^2}}, \dots, \frac{Z_n}{\sqrt{Z_1^2+\cdots+Z_n^2}}\Big)$$

*is uniformly distributed on the $(n+1)$–sphere.*

Proof: Property (1) is a classical fact following from the invariance of the Gaussian product density $e^{-\frac{1}{2}(z_1^2+\cdots+z_n^2)}/(2\pi)^{\frac{n}{2}}$ under rotations. For (2), observe that $Z_1^2$ is a scale multiple of a $\gamma_{1/2}$ variables. Squaring $A$, $B$, $\frac{Z_1}{\sqrt{Z_1^2+\cdots+Z_n^2}}$, the sum $W_1^2+\cdots+W_{n+1}^2$ is independent of all the ratios. Multiplying through, we have

$$W_1^2, (W_2^2+\cdots+W_{n+1}^2)\cdot\Big(\frac{Z_1^2}{Z_1^2+\cdots+Z_n^2}, \dots, \frac{Z_n^2}{Z_1^2+\cdots+Z_n^2}\Big).$$

The last $n$–components are distributed as a vector of independent scaled $\gamma_{1/2}$ variates using Lemma 2.1 (2). So the ratio of the square roots of the entries to the square root of the sum of all the entries is uniform on the $(n+1)$–sphere by Lemma 2.1 (1). □

The next result puts the pieces together to give a probabilistic version of Hurwitz (1897). For $2 \le j \le n$ fixed, consider rotations of the form

$$g_j = \begin{pmatrix} 1 & 0 & \dots & & & & \\ 0 & \ddots & & & & & \\ \vdots & & & & & & \\ & & x & y & & & \\ & & -y & x & & & \\ & & & & & & \vdots \\ & & & & & \ddots & 0 \\ & & & & \dots & 0 & 1 \end{pmatrix}$$

where all diagonal entries are equal to 1 except for the $(j-1,j-1)$ and $(j,j)$ entries which are equal to $x \in [0,1]$ and all off-diagonal entries are equal to zero except for the $(j-1,j)$ and $(j,j-1)$ which are equal respectively to $y = \sqrt{1-x^2}$ and $-y$. Let $\nu_j$ be the measure supported by these rotations and such that, under $\nu_j$, $x$ has the distribution of the first coordinate of a point uniformly chosen on the $n+2-j$–sphere.

**Proposition 2.1.** *Let $\{G_j^i\}$, $1 \le i < j \le n$, be independent random matrices in $SO(n)$ with $\{G_j^i\}_{i=1}^{j-1}$ having common distribution $\nu_j$. Then*

$$(G_n^1 G_{n-1}^1 \cdots G_2^1) \cdots (G_n^{n-2} G_{n-1}^{n-2}) G_n^{n-1}$$

*is uniformly distributed on $SO(n)$.*

Proof: The idea of the proof is simple. First, in $\mathbb{R}^3$, if a uniform rotation in the $(x,y)$ plane is followed by an independent rotation taking the $z$ axis to a uniform direction, the result is uniform on $SO(3)$. In general, if $e_1, \dots, e_n$ is the standard basis for $\mathbb{R}^n$ and $\{N_i\}_{i=2}^n$ are independent random matrices in $SO(n)$ with $N_2$ fixing $e_1, \dots, e_{n-2}$ and uniform in the $(n-1,n)$ plane, $\dots$, $N_i$ fixing $e_1, \dots, e_{n-1}$ and taking $e_{n-i+1}$ to a uniform vector in span $e_{n-1+1}, \dots, e_n$. Then the product $N_n N_{n-1}, \dots, N_2$ is uniformly distributed on $SO(n)$. This is given a formal proof in [10].

To finish the proof, we argue that

$$N_i = G_n^{n-(i-1)} \cdots G_{n-(i-2)}^{n-(i-1)}, \quad 2 \le i \le n,$$

have the required properties. Proceed by induction. Dropping the superscript, $G_n$ has form

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & a' & b' \\ & & -b' & a' \end{pmatrix}$$

with $a'$, $b'$ chosen uniformly on the 2–sphere. $G_{n-1}$ has form

$$\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & a & b & 0 \\ & & -b & a & 0 \\ & & 0 & 0 & 1 \end{pmatrix}$$

with $a$ distributed as $\dfrac{Z_1}{\sqrt{Z_1^2+Z_2^2+Z_3^2}}$ and $b = \sqrt{1-a^2}$. The product $G_n G_{n-1}$ has the last three elements of the third column $(a, -ba', bb')$. From Lemma 2.2, this is uniform on the three-sphere. The product $G_n G_{n-1} G_{n-2}$ has ones on the diagonal down to the $(n-5, n-5)$ place and the last four entries of column $n - 4$ uniform on the four-sphere. Continuing, $G_n G_{n-1} \cdots G_2$ has its first column uniformly distributed on the $n$–sphere. □

Remark: Using standard characterizations of beta and gamma random variables we can prove a converse: if $G \in SO(n)$ is chosen from Haar measure then, almost surely, the factorization into rotations as above is uniquely defined and the terms $G_j^i$ are independent with distributions specified by Proposition 2.1.

As a corollary of Proposition 2.1 we show that the random rotation chain of Theorem 1 satisfies a Döblin condition and thus converges to the uniform distribution in total variation norm. This gives a remarkably poor bound but, up to minor improvements, it is the best we know.

**Corollary 2.1.** *The convolution $Q^{*k}$ of Theorem 1 converges to Haar measure on $SO(n)$ in total variation. Indeed*

$$\|Q^{*k} - U\|_{TV} \le (1-c)^{\lfloor k/\binom{n}{2} \rfloor} \quad \textit{with} \quad c = 4^{-n^2} n^{-n}.$$

Proof: We claim that $Q^{*\binom{n}{2}}(A) \ge cU(A)$ for all Borel sets $A$. This Döblin condition implies the result (see e.g. Kloss (1959)). To prove the claim observe that the chance that the first $\binom{n}{2}$ steps of the walk pick rotations in the exact coordinates used for the factorization of Haar measure in Proposition 2.1 is $1/\binom{n}{2}^{\binom{n}{2}}$. For this component of $Q^{*\binom{n}{2}}$ the density of the corresponding random matrix is

$$\Pi f_n^{i,j}(x_{ij})$$

with the product ranging over the chosen coordinates and

$$f_n(x) = \frac{(1-x^2)^{-1/2}}{\pi} \ge \frac{1}{\pi}.$$

Proposition 2.1 gives the density of Haar measure as

$$\Pi f_{n_{ij}}^{i,j}(x_{ij})$$

for the same coordinates but different densities. There are $n-1$ terms of the Haar density with $n_{ij} = n$. Cancel these from both sides. Bound the remaining density factors of the component of $Q^{*\binom{n}{2}}$ below by $1/\pi^{\binom{n-1}{2}}$. To bound the remaining factors of the density of Haar measure above, use Lemma 2.2 (1) for $k \ge 3$

$$f_{n+2-k}(x) = \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{k-1}{2})}(1-x^2)^{\frac{k-3}{2}} \le \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{k-1}{2})} \le \frac{k}{2\sqrt{\pi}}.$$

There are $n-k+1$ terms in the product for a given $k$, $3 \le k \le n$. This shows that the remaining factors of the density of Haar measure are bounded above by

$$\left(\frac{1}{2\sqrt{\pi}}\right)^{\binom{n-1}{2}} 3^{n-2} \cdot 4^{n-3} \cdots n.$$

Combining bounds gives the result with

$$c > \left(\frac{1}{2\pi^{3/2}}\right)^{\binom{n-1}{2}} (3^{n-2} \cdot 4^{n-3} \cdots n)^{-1} > 4^{-n^2} n^{-n}. \qquad \square$$

The next corollary uses part of the factorization to represent the measure $Q_\theta$ on $SO(n)$ which rotates by a fixed angle $\theta$ in a randomly chosen two–dimensional space. Formally, let $R_\theta$ be the $n \times n$ matrix with the $2 \times 2$ block $\begin{matrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{matrix}$ in the upper left hand corner. Let $Q_\theta$ be the probability distribution corresponding to $MR_\theta M^{-1}$ where $M$ is chosen from Haar measure. Thus $Q_\theta$ is uniformly distributed on the conjugacy class containing $R_\theta$. Repeated convolutions of $Q_\theta$ were studied by Rosenthal (1994). We will use his results to get bounds on the spectrum of $Q$ in the next section.

**Corollary 2.2.** *With notation as in Proposition 2.1, the measure $Q_\theta$ is the probability distribution of $TR_\theta T^{-1}$ where $T = [(G_n^1 G_{n-1}^1 \cdots G_2^1))(G_n^2 G_{n-1}^2 \cdots G_3^2)]$.*

Proof: The argument for Proposition 2.1 shows that $(G_n^1 \cdots G_2^1)$ has columns $V_1, V_2 \cdots V_n$ with $V_1$ uniformly distributed. Similarly, $(G_n^2 \cdots G_3^2)$ has form

$$\begin{pmatrix} 1 & 0 & \cdots & \\ 0 & W_2 & \cdots & W_n \\ \vdots & & & \end{pmatrix}$$

with the column $W_2$ uniformly distributed. The product $T$ of these two has first two columns $V_1, W_{22}V_2 + W_{32}V_3 + \cdots + W_{n2}V_n$. Now $V_2, V_3, \ldots, V_n$ is an orthonormal basis for $V_1^\perp$ and so the second column is uniformly distributed in this space. Thus the first two columns of the product are uniformly distributed two–plane. By direct computation, the matrix $TR_\theta T^{-1}$ only depends on the first two columns and so has the same distribution as $MR_\theta M^{-1}$. $\square$

Remark: Similar factorizations hold for the unitary and symplectic group. Factorizations also hold for finite groups generated by reflexions (e.g. the symmetric group). Details and further applications can be found in [6]. These factorizations can be used exactly as in Section 3 below to give bounds on the eigenvalues of associated random walks. For example, on the symmetric group, the parallel to Kac's walk is the walk generated by random transposition.

## 3. Spectral bounds.

**3A. Introduction.** For $1 \le i < j \le n$, let $\mu_{ij}$ be the push–forward of the measure $f(x) = \frac{(1-x^2)^{-1/2}}{\pi}$ on $[-1,1]$ under the map

$$x \mapsto \begin{pmatrix} 1 & 0 & \dots & & & & \\ 0 & \ddots & & & & & \\ \vdots & & & & & & \\ & & x & y & & & \\ & & -y & x & & & \\ & & & & & & \vdots \\ & & & & & \ddots & 0 \\ & & & & \dots & 0 & 1 \end{pmatrix}, \quad y = \sqrt{1-x^2},$$

where the $x$'s are in position $(i,i)$ and $(j,j)$ and $y$ (resp. $-y$) is in position $(i,j)$, (resp $(j,i)$). This corresponds to rotation by a uniform angle in the $(i,j)$ plane. Let

$$Q = \frac{1}{\binom{n}{2}} \sum_{i<j} \mu_{ij}. \tag{3.1}$$

Then $Q$ is a symmetric probability measure on $SO(n)$. It acts on the real vector space $L^2(SO(n))$ via $Qf(x) = \int f(xy^{-1})Q(dy) = \int f(xy)Q(dy)$. Because of symmetry, $Q$ is a bounded self-adjoint operator on $L^2$. It has real spectrum contained in $[-1,1]$. In this section we bound the spectral gap, that is, the norm of $Q$ acting on $L_0^2 = \{f \in L^2 : \int f dx = 0\}$.

We will show that for all $f \in L_0^2$

$$\|f\|_2^2 \le A\langle (I-Q)f, f\rangle, \ \|f\|_2^2 \le A\langle (I+Q)f, f\rangle, \text{ for } A = 60n^3. \tag{3.2}$$

For $f \in L_0^2$ with $\|f\|_2^2 = 1$, (3.2) implies that

$$\langle Qf, f\rangle = 1 - \langle (I-Q)f, f\rangle \le 1 - \frac{1}{A}, \ \langle Qf, f\rangle = -1 + \langle (I+Q)f, f\rangle \ge -1 + \frac{1}{A}.$$

Now, an elementary argument in Riesz–Nagy ((1960), Sec. 9.2) shows $\|Q\|_{0,2\to 2} = \max(-m, M)$ with $m = \min_{\|f\|_{0,2}=1} \langle Qf, f\rangle$, $M = \max_{\|f\|_{0,2}=1} \langle Qf, f\rangle$. Thus (3.2) proves

**Theorem 3.1.** *The probability measure $Q$ of (3.1) on $SO(n)$ satisfies*

$$\|Q - U\|_{2\to 2} = \|Q\|_{0,2\to 2} \le 1 - \frac{1}{60n^3}. \tag{3.3}$$

The argument for (3.2) is by comparison with a measure $\widetilde{Q}$ on $SO(n)$. This $\widetilde{Q}$ results from rotating by $\pi$ in a randomly chosen plane. Formally

(3.4)

$\widetilde{Q}$ is the distribution of $M^{-1}RM$ on $SO(n)$ where $M$ is Haar distributed and $R$ is a diagonal matrix with two minus ones and $n-2$ ones on the diagonal.

Another description of $\widetilde{Q}$ is given by Corollary 2.2. This $\widetilde{Q}$ is uniformly distributed on the conjugacy class containing $R$. Its spectral behavior can thus be obtained by character theory on $SO(n)$. This was done by Rosenthal (1994) whose results are described in Lemma 3.2. In Section 3B we show that for every $f \in L^2$

$$\langle (I-\widetilde{Q})f,f\rangle \leq 16n^2\langle (I-Q)f,f\rangle,\ \langle (I+\widetilde{Q})f,f\rangle \leq 16n^2\langle (I+Q)f,f\rangle.$$

These comparison inequalities are proved using the factorizations of Section 2. In Section 3C we show that $Q$ has the eigenvalue $1-\frac{2}{n}$ with infinite multiplicity. In Section 3D we describe some variants of the measure $Q$ to which the present techniques apply.

We conclude this section by proving two needed lemmas. The first is for functions on the circle $S^1$.

**Lemma 3.1.** *For any $h \in L^2(S^1)$, and any probability measure $\nu$ on $S^1$*

*(a)* $\langle (I-\nu)h,h\rangle \leq \iint (h(x)-h(xy))^2 dxdy$

*(b)* $\langle (I+\nu)h,h\rangle \leq \iint (h(x)+h(xy))^2 dxdy.$

Proof: Acting by convolution on $S^1$, $\nu$ has spectrum in $[-1,1]$. Thus $0 \leq \langle (I-\nu)h,h\rangle \leq 2\|h\|_2^2$. Further, if $\overline{h} = \int h(x)dx$,

$$\langle (I-\nu)(h-\overline{h}),h\rangle = \langle (I-\nu)h,h\rangle$$

and $\langle (I-\nu)(h-\overline{h}),\overline{h}\rangle = 0$. Hence

$$\langle (I-\nu)h,h\rangle = \langle (I-\nu)(h-\overline{h}),h-\overline{h}\rangle \leq 2\|h-\overline{h}\|_2^2 = \iint (h(x)-h(xy))^2 dxdy.$$

The proof of (b) is similar. □

The second lemma gives a sharp bound on the spectral gap for the measure $\widetilde{Q}$. This leans heavily on work of Rosenthal (1994).

**Lemma 3.2.** *For $n \geq 2$, $\widetilde{Q}$ defined at (3.4) and any $f \in L_0^2$*

$$\|f\|_2^2 \leq a\langle (I-\widetilde{Q})f, f\rangle, \ \ \|f\|_2^2 \leq a\langle (I+\widetilde{Q})f, f\rangle \ \textit{with } a = \frac{4n}{15}.$$

Proof: Rosenthal (1994) determined all the eigenvalues of $\widetilde{Q}$ by character theory. The different eigenvalues are indexed by $n$–tuples $a_1 < a_2 < \cdots < a_m$ with $a_i$ integers or half integers. For definiteness, we treat the case where $n = 2m+1$ is odd, so $a_i - \frac{1}{2} \in \{0, 1, 2, \dots\}$. These $a_i$ index the irreducible representations. For example, the trivial representation corresponds to $\frac{1}{2}, \frac{3}{2}, \dots, m - \frac{1}{2}$ and the $n$–dimensional representation corresponds to $a^* = \left(\frac{1}{2}, \frac{3}{2}, \dots, m-\frac{3}{2}, m+\frac{1}{2}\right)$. Rosenthal shows there is an eigenvalue $\beta(\mathbf{a})$ of $\widetilde{Q}$ for each such $m$–tuple given by

$$\beta(\mathbf{a}) = \frac{(2m-1)!}{2^{2m-1}} \sum_{j=1}^{m} \frac{(-1)^{a_j - j + \frac{1}{2}}}{a_j \prod_{i=1}^{j-1}(a_j^2 - a_i^2) \prod_{i=j+1}^{m}(a_i^2 - a_j^2)}. \tag{3.5}$$

These eigenvalues have multiplicity the square of the dimension of the corresponding irreducible representation which is given by a similar formula. We do not need to consider these multiplicities. As $\mathbf{a}$ varies, these are all the eigenvalues of $\widetilde{Q}$.

The eigenvalues can be bounded by

$$|\beta(\mathbf{a})| \leq r(\mathbf{a}) = \frac{(2m-1)!}{2^{2m-1}} \sum_{j=1}^{m} \frac{1}{a_j \prod_{i=1}^{j-1}(a_j^2 - a_i^2) \prod_{i=j+1}^{m}(a_i^2 - a_j^2)}. \tag{3.6}$$

Obviously, for $\mathbf{a}$ corresponding to any non–trivial representation, $r(\mathbf{a})$ is largest when $\mathbf{a} = \mathbf{a}^*$ defined above. This $\mathbf{a}^*$ corresponds to the $n$–dimensional representation for which the eigenvalue is $\beta(\mathbf{a}^*) = \frac{1}{n}\mathrm{Tr}\,(R) = 1 - \frac{4}{n}$. Comparing (3.5) and (3.6) we have

$$r(a^*) = 1 - \frac{4}{n} + \frac{4}{4^n(n+\frac{1}{2})} \leq 1 - \frac{15}{4n}.$$

This $r(a^*)$ bounds the absolute value of the largest and smallest eigenvalues $\beta_+$, $\beta_-$. The claimed bounds for $I \pm \widetilde{Q}$ follow since these operators have largest and smallest eigenvalues $1-\beta_+$, $1-\beta_-$. □

**3B. Comparison Inequalities.** This Section proves (3.2) and so Theorem 3.1.

**Proposition 3.1.** *For the probabilities $Q$, $\widetilde{Q}$ on $SO(n)$ defined in (3.1), (3.4) and any $f \in L_0^2(SO(n))$*

$$\langle (I-\widetilde{Q})f, f\rangle \leq 16n^2\langle (I-Q)f, f\rangle.$$

Proof: The argument uses the factorization of Haar measure derived in Section 2. In the calculation below, squared differences are bounded by writing $y = y_1 \cdots y_k$ so

$$\begin{aligned}[f(x) - f(xy)]^2 &= [f(x) - f(xy_1 \dots y_k)]^2 \\ &= [(f(x) - f(xy_1)) + (f(x) - f(xy_1 y_2)) + \cdots \\ &\qquad + (f(xy_1 \dots y_{k-1}) - f(xy_1 \dots y_k))]^2 \\ &\leq k[(f(x) - f(xy_1))^2 + \cdots + (f(xy_1 \dots y_{k-1}) - f(xy_1 \dots y_k))^2].\end{aligned}$$

Integrating over $x$ in $SO(n)$ gives

$$\int (f(x) - f(xy))^2 dx \leq k \sum_{i=1}^{k} \int (f(x) - f(xy_i))^2 dx.$$

The factorization of Section 2 depends on an ordering of $\{1, 2, \dots, n\}$ and only involves rotations $G_i$ using pairwise adjacent coordinates. The measure $Q$ uses all pairs of coordinates. We symmetrize by conjugating by the permutation matrix corresponding to $\sigma$ in $S_n$. Write $g_{i,\sigma}(\theta) = g_{\sigma(i-1),\sigma(i)}(\theta)$ and $G_{i,\sigma}$ for the corresponding random element of $SO(n)$. Thus for any fixed $\sigma$

$$(G^1_{n,\sigma} \cdots G^1_{2,\sigma}) \cdots (G^{n-2}_{n,\sigma} G^{n-2}_{n-1,\sigma}) G^{n-1}_{n,\sigma} \tag{3.7}$$

is a uniformly distributed element of $SO(n)$. Further, with

$$T_\sigma = (G^1_{n,\sigma} \cdots G^1_{2,\sigma})(G^2_{n,\sigma} \cdots G^2_{3,\sigma}), \; R^\sigma = \sigma^{-1} \begin{pmatrix} -1 & & & & \\ & -1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \sigma$$

$$T_\sigma R^\sigma T_\sigma^{-1} \tag{3.8}$$

has distribution $\widetilde{Q}$. See Corollary 2.2.

Write $\nu^j_{k,\sigma}$ for the distribution of $G^j_{k,\sigma}$ (this distribution actually does not depend on $j$) and $\mu_{k,\sigma}$ for the distribution corresponding to a uniform rotation in coordinate plane $\sigma(k-1)$, $\sigma(k)$.

For any $f \in L^2_0$,

$$\begin{aligned}2\langle (I - \widetilde{Q}) f, f \rangle &= \iint (f(x) - f(xy))^2 dx \widetilde{Q}(dy) \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} \int \cdots \int (f(x) - f(xg^1_{n,\sigma} \cdots g^2_{3,\sigma} R^\sigma (g^2_{3,\sigma})^{-1} \\ &\qquad \cdots (g^1_{n,\sigma})^{-1})^2 dx \nu^1_{n,\sigma}(dg^1_{n,\sigma}) \cdots \nu^2_{3,\sigma}(dg^2_{3,\sigma}).\end{aligned}$$

Using the Cauchy–Schwarz inequality on the differences as above, $2\langle (I-\widetilde{Q})f,f\rangle$ is bounded above by

$$\frac{(4n-5)}{n!}\sum_{\sigma\in S_n}\int (f(x)-f(xR^{\sigma})^2 dx + 2\sum_{k=1}^{2}\sum_{\ell=k+1}^{n}\int (f(x)-f(xg))^2 dx\nu^k_{\ell,\sigma}(dg).$$

To complete the argument we show that in each term above the measure $\nu$ can be replaced by the measure $\mu$ corresponding to a uniform rotation in the chosen coordinate plane. To see this, fix a term

$$\int |f(x)-f(xg)|^2 dx d\nu^k_{\ell,\sigma}(g)$$

in the sum above. Factor dx into appropriate pieces in the order $\tau$ as in (3.7) where the permutation $\tau$ is chosen so that the right-most factor $G^{n-1}n,\tau$ is a (uniform) rotation in the desired coordinate plane $(\sigma(\ell-1),\sigma(\ell))$ (this is achieved by any $\tau$ such that $(\tau(n-1),\tau(n))=(\sigma(\ell-1),\sigma(\ell))$). Fixing the other coordinates define a function $\widetilde{f}$ on $S^1$ as $\widetilde{f}(z)=f(g^1_{n,\tau}\cdots g^{n-2}_{n,\tau}g^{n-2}_{n-1,\tau}g(z))$ with $g(z)$ in $SO(n)$ having $z$, $\sqrt{1-z^2}$ installed in the appropriate places. Using Lemma 3.1, the integral over $z$ with any measure $\nu$ is smaller than the integral with $z$ uniform. This also holds for the terms $R^{\sigma}$ (where $\nu$ is point mass). Thus $2\langle (I-\widetilde{Q})f,f\rangle$ is bounded above by

$$\frac{(4n-5)}{n!}\sum_{\sigma\in S_n}\int (f(x)-f(xg)^2 dx\mu_{2,\sigma}(dg)+2\sum_{k=1}^{2}\sum_{\ell=k+1}^{n}\int (f(x)-f(xg))^2 dx\mu_{\ell,\sigma}(dg).$$

As $\mu_{\ell,\sigma}=\mu_{ij}$ with $i=\sigma(\ell-1)$, $j=\sigma(\ell)$, we see that a given term

$$\int (f(x)-f(xg)^2 dx\mu_{ij}(dg)$$

appears at most $2(n-2)!+2\times 2(2n-3)(n-2)!\le 8(n-1)!$ times in the sum above. Finally, this yields the bound

$$\langle (I-\widetilde{Q})f,f\rangle \le \alpha\langle (I-Q)f,f\rangle$$

with

$$\alpha=\frac{(4n-5)8(n-1)!\binom{n}{2}}{n!}\le 16n^2.$$

□

The next result yields a lower bound for negative eigenvalues.

**Proposition 3.2.** *For the probabilities $Q$, $\widetilde{Q}$ on $SO(n)$ defined in (3.1), (3.4) and any $f \in L_0^2(SO(n))$*

$$\langle (I+\widetilde{Q})f, f\rangle \le 16n^2 \langle (I+Q)f, f\rangle.$$

Proof: The argument parallels the proof of Proposition 3.1 using a factorization of odd length $k$ for $y = y_1 \cdots y_k$ so that

$$\begin{aligned}[f(x)+f(xy)]^2 = [(f(x)+f(xy_1)) - (f(xy_1)+f(xy_1y_2)) + \cdots \\ + (f(xy_1\cdots y_{k-1}) + f(xy_1\cdots y_k))]^2 \\ \le k[(f(x)+f(xy_1))^2 + \cdots + (f(xy_1\cdots y_{k-1})+f(xy_1\cdots y_k))^2].\end{aligned}$$

The factorization (3.8) always has odd length. Now, proceed as in Proposition (3.1), using Lemma 3.1 b). □

**3C. Examples.** Section 3B gives upper bounds on the eigenvalues of $Q$ defined in (3.1) of form $\beta_i \le 1 - \frac{1}{60n^3}$. For the $n$–dimensional representation $\rho(m)$ we have

$$\widehat{Q}(\rho) = \frac{1}{\binom{n}{2}} \sum_{i<j} \widehat{\mu}_{ij}(\rho) = \Big(1-\frac{2}{n}\Big) I \tag{3.9}$$

where the last equality comes from computing $\widehat{\mu}_{ij}(\rho) = \int \rho(m)\mu_{ij}(dm)$. This is a diagonal matrix with zero entries at $(i,i)$, $(j,j)$ and ones elsewhere. Summing over $i,j$ gives (3.9). Since the $n$–dimensional representation appears $n$ times in the decomposition of $L_0^2$, $\big(1-\frac{2}{n}\big)$ appears as an eigenvalue with multiplicity at least $n^2$. The next result shows that it appears with infinite multiplicity.

**Proposition 3.3.** *On $SO(n)$, let $f(m) = m_{1,1}^k$ for $k$ odd. Then*

$$Qf(m) = \Big(1-\frac{2}{n}\Big) f(m) \quad \text{for all} \quad k = 1,3,5,\ldots.$$

Proof:

$$Qf(m) = \frac{1}{\binom{n}{2}} \sum_{i<j} \mu_{ij} f(m) = \frac{\binom{n-1}{2}}{\binom{n}{2}} f(m) + \frac{1}{\binom{n}{2}} \sum_{j=2}^{n} \mu_{1j} f(m) = \Big(1-\frac{2}{n}\Big) f(m).$$

Indeed, $\mu_{ij}$ with $2<i<n$ doesn't move the first coordinate and

$$\mu_{1j} f(m) = \int_0^{2\pi} (m_{11}\cos\theta - m_{1j}\sin\theta)^k d\theta = 0.$$

□

By similar fooling around with test functions we can find eigenvalues of form $\big(1-\frac{c}{n}\big)$ with infinite multiplicity and $c$ smaller than 2. Marc Kac conjectured and it has now been proved by D. Maslin (1999) and, independently, by E. Janvresse (1999), that all eigenvalues are smaller than $1-\frac{c}{n}$ for some universal $c$. Maslin's result applies to the walk on $SO(n)$ whereas Jeanvresse's is restricted to the sphere.

**3D. Variations and Remarks.** The methods of this section are fairly robust and give similar results for a variety of measures $Q$ on $SO(n)$.

Variation 1. For $1 \leq i < j \leq n$, let $\lambda_{ij}$ be a measure on $SO(n)$ with the property that for some $k$ and $c \geq 0$, $\lambda_{ij}^{*k} \geq c\mu_{ij}$ with $\mu_{ij}$ from (3.1). For example, $\lambda_{ij}$ may correspond to rotation in the $(i,j)$ plane through an angle uniformly chosen in $[-\frac{\pi}{4}, \frac{\pi}{4}]$. Let

$$Q_\lambda = \frac{1}{\binom{n}{2}} \sum_{i<j} \lambda_{ij}. \tag{3.10}$$

**Theorem 3.2.** *The probability measure $Q_\lambda$ of (3.10) satisfies*

$$\|Q_\lambda - U\|_{2\to 2} \leq 1 - \frac{c}{60k^2n^3}.$$

Proof: We may compare $Q_\lambda$ with $Q$ of (3.1) using the domination and Cauchy Schwarz

$$\begin{aligned}\iint (f(x) \pm f(xy))^2 dx\mu_{ij}(dy) &\leq \frac{1}{c} \iint (f(x) \pm f(xy_1y_2\cdots y_k))^2 dx\lambda_{ij}(dy_1)\cdots\lambda_{ij}(dy_k)\\ &\leq \frac{k^2}{c} \int (f(x) \pm f(xy))^2.\end{aligned}$$

This gives

$$\langle (I \pm Q)f, f\rangle \leq \frac{k^2}{c} \langle (I \pm Q_\lambda)f, f\rangle. \qquad \square$$

Variation 2. For $2 \leq i < n$, let $\mu_i$ correspond to uniform rotation in coordinates $(i-1, i)$. Let

$$\overline{Q} = \frac{1}{n-1} \sum_{i=2}^{n} \mu_i. \tag{3.11}$$

Thus $\overline{Q}$ corresponds to uniform rotation in pairwise adjacent coordinates. This might be an appropriate model for energy exchange of $n$ particles confined to a line. The following theorem gets a bound on the spectral gap of $\overline{Q}$.

**Theorem 3.3.** *The probability measure $\overline{Q}$ of (3.11) satisfies*

$$\|\overline{Q} - U\|_{2\to 2} \leq 1 - \frac{16}{3n^3}.$$

Proof: Use the factorization of Corollary 2.2 without further symmetrization to compare with the measure $\widetilde{Q}$ of (3.4). The argument for Proposition 3.1 gives

$$\begin{aligned}\langle (I-\widetilde{Q})f,f\rangle &\le (4n-5)\Big[\int (f(x)-f(xR)^2dx + 4\sum_{i=2}^{n}\int (f(x)-f(xg))^2 dx\nu_i(dg)\Big] \\ &\le 5(4n-5)(n-1)\langle (I-\overline{Q})f,f\rangle\end{aligned}$$

where $\nu_i$ is as in Proposition 2.1. Thus

$$\langle (I-\widetilde{Q})f,f\rangle \le 20n^2\langle (I-\overline{Q})f,f\rangle. \tag{3.12}$$

Now, Lemma 3.2 yields, for any $f\in L^2_0$,

$$\|f\|_2^2 \le \frac{16}{3}n^3\langle (I-\overline{Q})f,f\rangle.$$

□

An argument similar to that used in Proposition 3.2 shows

$$\|f\|_2^2 \le \frac{16}{3}n^3\langle (I+\overline{Q})f,f\rangle.$$

These results combine to prove the claim. □

Remarks: (1) We believe that the gap estimate $c/n^3$ from Theorem 3.3 is sharp: Kac's walk is somewhat analogous to random transposition on the symmetric group $S_n$ whereas the variant of Theorem 3.3 is analogous to random adjacent transposition. The spectral gap of random transposition is of order $1/n$ whereas that of random adjacent transposition is of order $1/n^3$ (See [7] and the references therein).

(2) In the arguments for Theorems 3.1–3.3 it is possible to avoid the use of character theory but get a slightly worse bound. For example, consider $Q$ defined at (3.1). Use the factorization of Proposition 2.1 to represent a uniform rotation as a product of $\binom{n}{2}$ rotations. Using Cauchy–Schwarz and then Lemma 3.1 along with symmetrization gives a bound of the form

$$\|f\|_2^2 \le 2\binom{n}{2}^2\langle (I-Q)f,f\rangle.$$

The comparison with $\widetilde{Q}$ and Rosenthal's sharp bounds on the eigenvalues of $\widetilde{Q}$ improve this by a factor of $n$.

## 4. Rates of Weak Convergence.

**4A. Introduction.** This section develops bounds for the rate of convergence of a random walk generated by a probability $Q$ on a compact group $G$ to stationarity. The bounds use the second eigenvalue and some geometric information about volume growth. Section B gives bounds for the dual bounded Lipschitz metric on probabilities. Section C gives bounds for discrepancy. Section D specializes the bounds to $SO(n)$ and the $n$–sphere. The main results are summarized in Corollary 4.1 and Theorems 4.2, 4.3. We hope that this material may be more generally useful. In the remainder of this introduction we set out our notation.

Let $G$ be a compact metrizable group with normalized Haar measure $dg$. Let $H$ be a closed subgroup and $\mathcal{X} = G/H = \{gH : g \in G\}$ be the associated homogeneous space ($G$ acts on the left of $\mathcal{X}$). For example, if $G = SO(n+1)$, $H = \{\text{id}\}$ (resp. $H = SO(n)$). Then $\mathcal{X} = SO(n+1)$ (resp. $\mathcal{X} = S^n$ the sphere in $\mathbb{R}^{n+1}$).

Consider a symmetric measure $Q$ on $G$ (so $Q(A) = Q(A^{-1})$ for Borel sets $A$). Define

$$Qf(g) = Q * f(g) = \int_G f(v^{-1}g)Q(dv).$$

Hence $Q$ is a Markov operator on $G$ which is self–adjoint on $L^2(G)$ and commutes with right translations.

Define a Markov kernel on $\mathcal{X}$ by setting

$$K(x, \overline{A}) = Q(AHu^{-1}) \quad \text{if} \quad x = uH, \quad \overline{A} = AH.$$

Let $K^\ell(x, dy) = K_x^\ell(dy)$ denote the distribution after $\ell$ steps.

Recall that there exists on $\mathcal{X}$ a unique $G$–invariant probability measure $m(dx) = dx$ such that

$$\int_G f(g)dg = \int_{\mathcal{X}} \int_H f(gh)dhdx$$

for any continuous function $f$ ($dh$ denotes the normalized Haar measure on $H$). It follows that, if $x = gH$ and $f(u) = f(uH)$, we have

$$Qf(g) = Kf(x) = \int_G f(v^{-1}x)Q(dv)$$

for any bounded measurable function $f : \mathcal{X} \to \mathbb{R}$. The symmetry assumption on $Q$ implies that $K$ is reversible with respect to $m$. We will work on $L^2 = L^2(\mathcal{X}, m)$ on which $K$ is self–adjoint.

Let $m : f \to mf$ denote the operator that sends any function $f$ to the constant function $mf(x) = \int f(y)dy$ and set

$$\beta = \sup_{f \in L^2} \left\{ \frac{\|(K-m)f\|_2}{\|f\|_2} \right\}$$

where $\|f\|_2^2 = \int_{\mathcal{X}} |f(y)|^2 dy$.

Assume that $\mathcal{X}$ carries a $G$ left–invariant distance $d = d_{\mathcal{X}}$. Let $B(x,r) \subset \mathcal{X}$ denote the corresponding balls. Let $\rho$ be the diameter of $\mathcal{X}$.

**4B. Bounded Lipschitz Functions.** Define the bounded–Lipschitz norm of a function $f$ by

$$\|f\|_{BL} = \|f\|_\infty + L(f) \quad \text{with} \quad L(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} .$$

The functions with $\|f\|_{BL} < \infty$ form a Banach algebra carefully discussed in Dudley (1976).

Consider the volume growth function $s \to V(s) = m(B(e, s))$ where $e$ is a base point on $\mathcal{X}$. By invariance of $d$ and of the measure $m$, the volume $V(r)$ does not depend on the choice of $e$. The volume growth functions for $SO(n)$ and $S^n$ are determined in Section 4C below.

**Theorem 4.1.** *Assume that there are positive $c$ and $n$ such that $V(r) \geq c(r/\rho)^n$ for $0 < r \leq \rho$. Then, the Markov chain $K^\ell$ on $\mathcal{X}$ defined in Section 4A satisfies*

$$\|(K^\ell - m)f\|_\infty \leq 3\rho c^{-1/(2+n)} \beta^{2\ell/(n+2)} L(f).$$

This theorem is proved by a sequence of lemmas. First observe that

$$\|(K^\ell - m)f\|_\infty = \|(K^\ell - m)(f - a)\|_\infty$$

for any constant $a$. It follows that we may assume that $f$ changes sign on $\mathcal{X}$. Now, if $f$ changes sign,

$$\|f\|_\infty \leq \rho L(f).$$

Set $\chi_r(x, y) = V(r)^{-1} \mathbf{1}_{B(x,r)}(y)$ and

$$f_r(x) = \chi_r f(x) = \int_{\mathcal{X}} f(y) \chi_r(x, y) dy = V(r)^{-1} \int_{B(x,r)} f(y) dy.$$

**Lemma 4.1.** *For any Lipschitz function $f$,*

$$\|f - f_r\|_\infty \leq r L(f).$$

Proof:

$$\left| f(x) - \int \chi_r(x, y) f(y) dy \right| \leq \int |f(x) - f(y)| \chi_r(x, y) dy \leq r L(f).$$

□

Let $T_g : f \to T_g f$ be defined by $T_g f(x) = f(gx)$ for any function $f$. Since the distance $d$ is invariant under the left action of $G$, we have $gB(x, r) = B(gx, r)$ and

$$\begin{aligned} T_g f_r(x) = f_r(gx) &= V(r)^{-1} \int_{B(gx,r)} f(y) dy = V(r)^{-1} \int_{gB(x,r)} f(y) dy \\ &= V(r)^{-1} \int_{B(x,r)} f(gy) dy = V(r)^{-1} \int_{B(x,r)} T_g f(y) dy. \end{aligned}$$

Hence $T_g \chi_r = \chi_r T_g$ for all $r > 0$ and $g \in G$. It follows that

$$K \chi_r = \chi_r K, \tag{4.1}$$

that is $\chi_r$ and $K$ commute.

**Lemma 4.2.** *For any bounded function $f$,*

$$\|(K-m)^\ell f_r\|_\infty \le \beta^\ell V(r)^{-1/2}\|f\|_\infty.$$

Proof: We have

$$\begin{aligned}(K-m)^\ell f_r(x) &= (K-m)^\ell \chi_r f(x)\\ &= \chi_r(K-m)^\ell f(x)\\ &= \int_{\mathcal{X}}\int_{\mathcal{X}} \chi_r(x,z)(K^\ell(z,y)-1)f(y)dydz\\ &\le \Big(\int |\chi_r(x,z)|^2 dz\Big)^{1/2}\|(K-m)^\ell f\|_2\\ &\le V(r)^{-1/2}\beta^\ell\|f\|_2 \le V(r)^{-1/2}\beta^\ell\|f\|_\infty.\end{aligned}$$

□

We return to the proof of the Theorem. Fix $f$, and assume that $f$ changes sign on $\mathcal{X}$. Hence, $\|f\|_\infty \le \rho L(f)$. Write, for any $r>0$,

$$\begin{aligned}\|(K^\ell - m)f\|_\infty &\le \|(K-m)^\ell f_r\|_\infty + \|(K^\ell-m)(f-f_r)\|_\infty\\ &\le V(r)^{-1/2}\beta^\ell\|f\|_\infty + 2\|f-f_r\|_\infty\\ &\le (V(r)^{-1/2}\beta^\ell\rho + 2r)L(f).\end{aligned}$$

Observe also that $\|(K^\ell-m)f\|_\infty \le 2\|f\|_\infty \le 2\rho L(f)$. Thus, if we assume that $V(r)\ge c(r/\rho)^n$ for $0<r\le\rho$, it follows that

$$\|(K^\ell-m)f\|_\infty \le \rho[(c(r/\rho)^n)^{-1/2}\beta^\ell + 2r/\rho]L(f)$$

for all $r>0$. Picking $r$ so that

$$(r/\rho)^{(2+n)/2} = c^{-1/2}\beta^\ell$$

yields

$$\|(K^\ell-m)f\|_\infty \le 3\rho c^{-1/(2+n)}\beta^{2\ell/(n+2)}L(f).$$

□

We end this section by stating a version of Theorem 4.1 in terms of the dual bounded Lipschitz distance. Following Dudley (1966), define the dual bounded Lipschitz distance $D_*(\mu,\nu)$ between two probability measures $\mu$ and $\nu$ by

$$D_*(\mu,\nu) = \sup_{\|f\|_\infty + L(f)\le 1} |\mu(f)-\nu(f)|.$$

Dudley [11,12,13,14] shows that $D_*$ metrizes weak $*$ convergence of probability measures on $\mathcal{X}$. Further, if the Prohorov metric on probability measures is defined by

$$R(\mu,\nu)=\inf_{\varepsilon}\{\mu(A)\le\nu(A^{\varepsilon})+\varepsilon, \text{all Borel } A\},\ A^{\varepsilon}=\{y\in G:\exists\, x\in A: d(x,y)<\varepsilon\}.$$

Then $R(\mu,\nu)\le 2D_*(\mu,\nu)$.

Corollary 4.1 gives a bound for the rate of convergence in $D_*$ and $R$ distance in the presence of a bound on the spectral gap $\beta$.

**Corollary 4.1.** *Assume that $V(r)\ge c(r/\rho)^n$ for $0<r\le\rho$. Then, for every $x$, and $\ell$, the Markov chain $K^{\ell}$ on $\mathcal{X}$ defined in Section 4A satisfies*

$$D_*(K_x^{\ell},m)\le 3\rho c^{-1/(2+n)}\beta^{2\ell/(n+2)}.$$

**4C. Discrepancy.** Consider now the discrepancy distance associated to the metric $d$ defined by

$$D(\mu,\nu)=\sup_{x\in\mathcal{X},r>0}\{|\mu(B(x,r))-\nu(B(x,r))|\}.$$

Discrepancy is a standard measure of the rate of convergence in the metric theory of numbers. Kuipers and Nederiter (1974) is a book length treatment of techniques to bound discrepancy. Phillips–Lubotzky–Sarnak (1986) give discrepancy bounds for a random walk on the sphere. Su (1995) gives discrepancy rates for a variety of random walks on compact spaces. Some remarks comparing these results to our results appear at the end of this section.

**Theorem 4.2.** *Assume that $V(r)\ge c(r/\rho)^n$ for $0<r\le\rho$ and that for some $C>0$*

$$V(r+\varepsilon)-V(r-\varepsilon)\le C\varepsilon/\rho$$

*for all $r,\varepsilon>0$. Then, for every $x$ and $\ell$ the Markov chain $K^{\ell}$ defined in Section 4A satisfies*

$$D(K_x^{\ell},m)\le 2C^{n/(n+2)}c^{-1/(n+2)}\beta^{2\ell/(n+2)}.$$

<u>Proof</u>: Fix $\varepsilon,r>0$ and $y\in\mathcal{X}$. Set $B=B(y,r)$, $\varphi(z)=\mathbf{1}_B$ and $\varphi_1(z)=\mathbf{1}_{B(y,r-\varepsilon)},\varphi_2(z)=\mathbf{1}_{B(y,r+\varepsilon)}$. Recall that $\chi_{\varepsilon,z}(w)=V(\varepsilon)^{-1}\mathbf{1}_{B(z,\varepsilon)}(w)$. Viewing $\chi_\varepsilon$ has a Markov operator, we have

$$\chi_\varepsilon\varphi_1\le\varphi\le\chi_\varepsilon\varphi_2.$$

Furthermore

$$m(|\varphi-\chi_\varepsilon\varphi_i|)\le V(r-\varepsilon)-V(r+\varepsilon)\le C\varepsilon/\rho.$$

We consider two cases, depending on whether $K_x^{\ell}(B)\ge m(B)$ or not.

If $K_x^\ell(B) \geq m(B)$, then

$$\begin{aligned} |K_x^\ell(B) - m(B)| &= K^\ell \varphi(x) - m(B) \\ &\leq K^\ell \chi_\varepsilon \varphi_2(x) - m(B) \\ &\leq m(|\varphi - \chi_\varepsilon \varphi_2|) + |(K^\ell - m)\chi_\varepsilon \varphi_2(x)| \\ &\leq C\varepsilon/\rho + c^{-1/2}(\varepsilon/\rho)^{-n/2}\beta^\ell. \end{aligned}$$

The last inequality uses the volume hypotheses and Lemma 4.2.

In the case where $K_x^\ell(B) < m(B)$ the same inequality is obtained by using $\varphi_1$ instead of $\varphi_2$ in the argument. Hence

$$|K_x^\ell(B) - m(B)| \leq C\varepsilon/\rho + c^{-1/2}(\varepsilon/\rho)^{-n/2}\beta^\ell.$$

For $(\varepsilon/\rho)^{(n+2)/2} = C^{-1}c^{-1/2}\beta^\ell$ this yields

$$D(K_x^\ell, m) \leq 2C^{n/(n+2)}c^{-1/(n+2)}\beta^{2\ell/(n+2)}.$$

□

Remark: Su (1995) analyzes simple random walk on the circle $S^1$ taking steps $\pm\alpha$ for irrational $\alpha$. He bounds the rate of convergence to stationarity giving results that depend on the degree of irrationality of $\alpha$. The bounds use standard tools from uniform distribution mod(1): Leveque's inequality and the Erdös–Turan bound [25]. In the notation of this section, Leveque's inequality on $S^1$ gives $D(K_x^\ell, m) \leq C\beta^{2\ell/3}$ for a universal constant $C$. This also follows from Theorem 4.2. The Erdös–Turan bound gives $D(K_x^\ell, m) \leq C(\frac{1}{h} + \beta^\ell \log h)$ for any positive integer $h$. Optimizing in $h$ gives as a slight improvement the bound $C^1 \ell\beta^\ell$. On the sphere $S^2$, Lubotzky–Phillips–Sarnak (1986) show that discrepancy satisfies $D(K_x^\ell, m) \leq C\beta^{2\ell/3}$ for a universal $C$. Theorem 4.2 yields $D(K_x^\ell, m) \leq C\beta^{\ell/2}$. Their improved estimate leans heavily on special features available for dimension 2.

**4D. Examples.** To treat examples we will use the following results on volume growth for Riemannian manifolds. Gallot, Hulin, and LaFontaine (1990) is a useful reference for this material.

**Lemma 4.3.** (Bishop, Gromov; [15], p.133) *Let $M$ be a compact Riemannian manifold of dimension $n$ endowed with its distance function and its canonical Riemannian measure. Let $\mathcal{V}(x, r)$ denote the volume of the ball of radius $r > 0$ around $x \in M$. Let $\rho$ be the diameter of $M$. Assume that the Ricci curvature of $M$ is nonnegative. Then*

$$\frac{\mathcal{V}(x,r)}{\mathcal{V}(x,t)} \leq \frac{r^n}{t^n} \quad \textit{and} \quad \frac{\mathcal{V}(x,r) - \mathcal{V}(x,s)}{\mathcal{V}(x,t)} \leq \frac{r^n - s^n}{t^n}$$

*for all* $0 < t \le s \le r < +\infty$.

*In particular, if* $\mathcal{V} = \mathcal{V}(\rho)$ *is the volume of* $M$*, and* $V(x,r) = \mathcal{V}(x,r)/\mathcal{V}$,

$$V(x,r) \ge (r/\rho)^n \quad \textit{for all} \quad 0 < r \le \rho.$$

*Furthermore,*

$$V(x, r+\varepsilon) - V(x, r-\varepsilon) \le \frac{\omega_n n \rho^n}{\mathcal{V}} (\varepsilon/\rho)$$

*for all* $r, \varepsilon > 0$ *where* $\omega_n$ *is the volume of the unit ball in* $\mathbb{R}^n$.

**Example 1:** $G = SO(n+1)$, $H = SO(n)$, $\mathcal{X} = S^n \subset \mathbb{R}^{n+1}$ equipped with its canonical Riemannian distance. Let

$$\sigma_n = \mathrm{Vol}(S^n) = \frac{2\pi^{(n+1)/2}}{\Gamma((n+1)/2)}$$

be the volume of the unit sphere in $\mathbb{R}^{n+1}$. Recall that

$$\omega_n = \frac{\sigma_{n-1}}{n} = \frac{2\pi^{n/2}}{n\Gamma(n/2)}.$$

The diameter of $S^n$ is $\rho = \pi$. Corollary 4.1 and Theorem 4.2 yield

**Theorem 4.3.** *For the sphere* $S^n$*, the bounded Lipschitz distance* $D_*$ *satisfies*

$$D_*(K_x^\ell, m) \le 3\pi\beta^{2\ell/(n+2)}$$

*whereas the discrepancy* $D$ *satisfies*

$$D(K_x^\ell, m) \le \frac{2\pi^n \Gamma((n+1)/2)}{\Gamma(n/2)} \; \beta^{2\ell/(n+2)}.$$

**Example 2:** $\mathcal{X} = G = SO(n)$, $H = \{\mathrm{id}\}$. We need to fix an invariant metric on $SO(n)$ and compute the corresponding Riemannian volume $\mathcal{V}$ and diameter $\rho$. The dimension of $SO(n)$ is $N = \binom{n}{2}$. Up to scaling, there exists a unique bi–invariant metric. Because of bi–invariance, we need only specify the distance to the identity and $d(m, \mathrm{id})$ only depends on the eigenvalues $\theta_i$ of $m$. For $\theta \in [0, 2\pi]$, let $|\theta|_1 = \min(|\theta|, |\theta - 2\pi|)$. Then $d^2(m, \mathrm{id})$ is proportional to $\sum |\theta_i|_1^2$. The Ricci curvature of any bi-invariant metric on a compact Lie group is non-negative (in fact the sectional curvature is non-negative). See, e.g., Proposition 3.17 in [15]. To fix the scaling constant, recall that the Lie Algebra $L$ of $G = SO(n)$ can be identified with the space of real skew symmetric matrices with the exponential map given by

$$\exp : L \to G, \;\; M \to \exp(M) = e^M.$$

Let $\{E_{i,j} : 1 \le i < j \le n\}$ be the natural basis of $L$. Here $E_{i,j}$ denote the matrix with all entries equal to zero except the $(i,j)$ and $(j,i)$ entries which are respectively equal to 1 and $-1$. The usual Euclidean structure

$$\langle M, N\rangle = \sum_{i<j} M_{i,j} N_{i,j}$$

for which the above basis is orthonormal give rise to a bi–invariant Riemannian structure on $G$. For this Riemannian structure the volume form is given by

$$dg = \Big|\bigwedge_{i<j} g_j^t dg_i\Big|$$

where $g = (g_{i,j})_{1\le i,j\le n}$ and $g_i$ is the column vector $g_i = (g_{\ell,i})_{1\le \ell \le n}$. The volume $\mathcal{V}_n$ of $SO(n)$ is then equal to (recall $\sigma_i$ from Example 1)

$$\mathcal{V}_n = \prod_1^{n-1} \sigma_i = \frac{2^{n-1}\pi^{N/2}}{\Gamma(n/2)\Gamma((n-1)/2)\cdots\Gamma(3/2)\Gamma(1)} .$$

The diameter is equal to the diameter of a maximal torus which is

$$\rho_n = \pi k^{1/2} \quad \text{where} \quad n = 2k \quad \text{or} \quad n = 2k+1.$$

More generally, in terms of eigenvalues the distance to the identity is exactly $d^2(m,\mathrm{id}) = \sum |\theta_i|_1^2$ for this normalization.

There is a more canonical choice of bi-invariant Riemannian metric which is induced by the notion of Killing form. With the above notation the metric induced by the Killing form on $L$ is

$$B(M,N) = 2(n-2)\sum_{i<j} M_{i,j}N_{i,j} = (n-2)Tr(MN).$$

The Ricci curvature of the Killing form is $\frac{1}{4}B$ (see, e.g. Rothaus (1981) or [15]).

Here the constants $c$, $C$ in Theorems 4.1, 4.2 and Corollary 4.1 are given by $c = 1$ and

$$C = \frac{4\pi^N [n/2]^{N/2}\Gamma(n/2)\Gamma((n-1)/2)\cdots\Gamma(3/2)\Gamma(1)}{2^n\Gamma(N/2)} .$$

Recall that, for $t \ge 2$,

$$\sqrt{2\pi}(t-1)^{t-1/2}e^{-t+1}e^{(12(t-1)+1)^{-1}} \le \Gamma(t) \le \sqrt{2\pi}(t-1)^{t-1/2}e^{-t+1}e^{(12(t-1))^{-1}}.$$

Hence

$$\Gamma(n/2)\Gamma((n-1)/2)\cdots\Gamma(3/2)\Gamma(1) \le n(2\pi)^{n/2}[(n-2)/2]^{N/2}e^{(-N+n)/2},$$

$$\frac{\Gamma(n/2)\Gamma((n-1)/2)\cdots\Gamma(3/2)\Gamma(1)}{\Gamma(N/2)} \le \frac{n(N/2)^{1/2}e^{1+n/2}(2\pi)^{(n-1)/2}}{(n/2)^{N/2}}.$$

It follows that $C \le 10^{N+1}$ if $n \ge 3$. Applying these estimates we have

**Theorem 4.4.** *For the special orthogonal group $SO(n)$, $n \geq 3$, the bounded Lipschitz distance $D_*$ satisfies*

$$D_*(K_x^\ell, m) \leq 3\pi(n/2)^{1/2}\beta^{2\ell/(N+2)}$$

*whereas the discrepancy $D$ satisfies*

$$D(K_x^\ell, m) \leq 2 \times 10^{N+1}\beta^{2\ell/(N+2)}$$

*where $N = n(n-1)/2$.*

## 5. Improved bounds for compact groups.

This last section shows how to take full advantage of the comparison with a known random walk in controlling the bounded Lipschitz distance or discrepancy on a compact group. In the case of Kac's walk on $SO(n)$, the bounds obtained below improve by a factor of $n$ those of Theorem 4.4. The trick is to refine the comparison technique used in Section 4 to give bounds on all the eigenvalues (not just the spectral gap) and then use this additional information. Some care is needed. In our main example, the measure $Q$ has eigenvalues of infinite multiplicity while $\widetilde{Q}$ has all eigenvalues of finite multiplicity.

Let $G$ be a compact metrizable group equipped with its normalized Haar measure. Let $Q$ be a symmetric probability measure. As in Section 4A, we also view $Q$ as a convolution operator. $Q$ is then a self-adjoint operator on $L^2(G)$. Given a finite dimensional subspace $H$ of $L^2(G)$, define

$$\beta(H) = \inf\{\langle Qf, f\rangle : f \in H, \;\; \|f\|_2 = 1\}$$

and set

$$\beta_i = \sup\{\beta(H) : \mathrm{dim} H = i+1\}, \;\; i = 0, 1, 2, \ldots.$$

The $\beta_i$'s form a non-increasing sequence, $\beta_0 = 1$ and $\lim_{i\to\infty} = \beta$ is the top of the essential spectrum of $Q$. Note that this limit exists because the $\beta_i$'s are bounded below by $-1$.

We now repeat this construction for the negative eigenvalues by starting from the opposite end of the spectrum. Namely, define

$$\alpha(H) = \sup\{\langle Qf, f\rangle : f \in H, \;\; \|f\|_2 = 1\}$$

and set

$$\alpha_i = \inf\{\alpha(H) : \mathrm{dim} H = i\}, \;\; i = 1, 2, \ldots.$$

This time the $\alpha_i$'s form a non-decreasing sequence and we set $\lim_{i\to\infty} \alpha_i = \alpha$. This is the bottom of the essential spectrum of $Q$. Note that

$$\alpha \leq \beta$$

and that $\alpha = \beta = 0$ if $Q$ is a compact operator (e.g., when the measure $Q$ has an $L^2$ density).

Set

$$\gamma = \max\{-\alpha, \beta\}. \tag{5.1}$$

For any small $\epsilon \geq 0$, set

$$\Sigma_\epsilon(\ell) = \sum_{\substack{i:\, i>0 \\ \beta_i > \gamma+\epsilon}} |\beta_i|^{2\ell} + \sum_{i:\alpha_i < -\gamma-\epsilon} |\alpha_i|^{2\ell}. \tag{5.2}$$

In words, $\Sigma_\epsilon(\ell)$ is the sum of the power of the eigenvalues **lying outside the interval** $[-\gamma-\epsilon, \gamma+\epsilon]$ (excluding $\beta_0 = 1$).

Remark: For $\epsilon > 0$, $\Sigma_\epsilon(\ell)$ is always finite: it is a finite sum. The quantity $\Sigma_0(\ell)$ is infinite unless one of two cases arise:

(1) $\alpha = \beta = 0$ and $Q^{*\ell}$ has a density in $L^2(G)$, i.e., $Q^{2\ell}$ is trace class, in which case

$$\Sigma_0(\ell)^{1/2} = \|Q^\ell\|_{2\to\infty}$$

is also the $L^2$-norm of the density of $Q^{*\ell}$ w.r.t. Haar measure.

(2) $\gamma \neq 0$ **and** there are only finitely many eigenvalues lying outside the interval $[-\gamma, \gamma]$. In this second case, $\Sigma_0(\ell)$ is a finite sum and if $\gamma = \beta$ (resp. $-\gamma = \alpha$), $\beta$ (resp $\alpha$) is an eigenvalue of infinite multiplicity (if $\beta = -\alpha = \gamma$ at least one of them is an eigenvalue of infinite multiplicity, possibly both).

Keeping the notation of Section 4.B, we now can state a variant of Lemma 4.2.

**Lemma 5.1.** *For any $\epsilon \geq 0$ and for any bounded function $f$,*

$$\|(Q-m)^\ell f_r\|_\infty \leq \left(\Sigma_\epsilon(\ell)^{1/2} + (\gamma+\epsilon)^\ell V(r)^{-1/2}\right) \|f\|_\infty.$$

Proof: It suffices to prove the result for $\epsilon > 0$. The case $\epsilon = 0$ then follows by passing to the limit (if $\Sigma_0(\ell)$ is infinite, the limit inequality is trivial).

Fix $\epsilon > 0$. Let $Q = \int_{-1}^1 \lambda dE_\lambda$ be the spectral decomposition of $Q$. Define

$$Q_2 = \int_{[-\gamma-\epsilon, \gamma+\epsilon]} \lambda dE_\lambda$$

and $Q_1 = Q - Q_2$. Observe that $Q_1$ has density in $L^2(G)$ and that $Q_1, Q_2$ are bounded operators on $L^2(G)$ with $\|Q_2\|_{2\to2} = \gamma+\epsilon$. Observe also that $Q_1Q_2 = 0$. It follows that

$$(Q-m)^\ell = (Q_1-m)^\ell + Q_2^\ell$$

for any $\ell = 1, 2, \ldots$. Moreover,

$$\|(Q_1 - m)^\ell\|_{2\to\infty} = \Sigma_\epsilon(\ell)^{1/2}.$$

Thus, we have

$$\begin{aligned}(Q - m)^\ell f_r(x) &= (Q_1 - m)^\ell f_r(x) + Q_2^\ell \chi_r f(x) \\ &= (Q_1 - m)^\ell f_r(x) + \chi_r Q_2^\ell f(x) \\ &\le \|(Q_1 - m)^\ell f_r\|_\infty + \|\chi_r\|_2 \|Q_2^\ell f\|_2 \\ &\le \|(Q_1 - m)^\ell\|_{2\to\infty} \|f_r\|_2 + V(r)^{-1/2} \|Q_2^\ell\|_{2\to 2} \|f\|_2 \\ &\le \left(\Sigma_\epsilon(\ell)^{1/2} + V(r)^{-1/2}(\gamma + \epsilon)^\ell\right) \|f\|_\infty.\end{aligned}$$

Here we have used the obvious fact that $\|f_r\|_2 \le \|f\|_\infty$.

Using this Lemma, we obtain some improved versions of Theorem 4.1, 4.2.

**Theorem 5.1.** *Under the assumptions of Theorem 4.1, the bounded Lipschitz distance is bounded by*

$$D_*(Q^\ell, m) \le \Sigma_\epsilon(\ell)^{1/2} + 3\rho c^{-1/(2+n)}(\gamma + \epsilon)^{2\ell/(n+2)},$$

*for all $\epsilon \ge 0$.*

*Similarly, under the assumptions of Theorem 4.2, the discrepancy distance is bounded by*

$$D(K_x^\ell, m) \le \Sigma_\epsilon(\ell)^{1/2} + 2C^{n/(n+2)} c^{-1/(n+2)}(\gamma + \epsilon)^{2\ell/(n+2)},$$

*for all $\epsilon \ge 0$.*

Let us now illustrate how these results can be used for Kac's walk on $SO(n)$. Keep the notation of Section 3.2. Let $Q$, $\widetilde{Q}$ be the two probability measures on $SO(n)$ defined at (3.1), (3.4).

¿From the results of [36], all the eigenvalue of $\widetilde{Q}$ have finite multiplicity and the only accumulation point in the spectrum of $\widetilde{Q}$ is 0. In fact, $\widetilde{Q}^{*\ell}$ has a bounded density for $\ell$ large enough. Moreover, it is proved in [36] that there exist $B$ and $b > 0$ such that

$$\widetilde{\Sigma}_0(\ell) \le Be^{-bt} \tag{5.3}$$

for all $t > 0$ and $\ell$ such that

$$\ell \ge \frac{1}{8} n \log n + tn.$$

By Proposition 3.1 and the minimax principle we have, with obvious notation,

$$\beta_i \leq 1 - \frac{1-\widetilde{\beta}_i}{16n^2} \tag{5.4}$$

for $i = 0, 1, \ldots,$. ¿From the definition of $\beta$ and this inequality, it follows that

$$\beta \leq 1 - \frac{1}{16n^2}.$$

Similarly, Proposition 3.2 yields

$$\alpha_i \geq -1 + \frac{1+\widetilde{\alpha}_i}{16n^2} \tag{5.5}$$

for $i = 1, 2, \ldots,$ and

$$\alpha \geq -1 + \frac{1}{16n^2}.$$

This yields the following lemma.

**Lemma 5.2.** *Referring to the measures $Q, \widetilde{Q}$ on $SO(n)$ defined at (3.1), (3.4), we have*

$$\gamma \leq 1 - \frac{1}{16n^2}, \quad \widetilde{\gamma} = 0.$$

*Moreover*

$$\Sigma_\delta(\ell) \leq \widetilde{\Sigma}_0([\ell/32n^2])$$

*for*

$$\delta = 1 - \frac{1}{32n^2} - \gamma.$$

Proof: Only the second inequality need a further argument. The sum $\Sigma_\delta(\ell)$ only involves eigenvalues that fall outside $[-1 + (1/32n^2), 1 - (1/32n^2)]$. By (5.4)-(5.5), the corresponding eigenvalues of $\widetilde{Q}$ must be outside $[-1/2, 1/2]$. Thus,

$$\Sigma_\delta(\ell) \leq \sum_{i:\widetilde{\beta}_i \geq 1/2} \left(1 - \frac{1-\widetilde{\beta}_i}{16n^2}\right)^{2\ell} + \sum_{i:\widetilde{\alpha}_i \leq -1/2} \left(1 - \frac{1-|\widetilde{\alpha}_i|}{16n^2}\right)^{2\ell}.$$

By the elementary inequalities $\forall x \in (0, \infty),\ 1 - x \leq e^{-x}$ and $\forall x \in (1, 1/2),\ e^{-2x} \leq 1 - x$, we get

$$\Sigma_\delta(\ell) \leq \widetilde{\Sigma}_{1/2}([\ell/32n^2]).$$

□

Now, for the bounded Lipschitz distance $D_*$, Lemma 5.2, (5.3) and Theorem 5.1 yield (in applying Theorem 5.1, recall that $SO(n)$ has dimension $N = n(n+1)/2$, not $n$)

$$D_*(Q^\ell, m) \leq Be^{-bt} + 3\pi(n/2)^{1/2} e^{-\ell/(8n^3(n+1))}$$

for all $t > 0$ and $\ell$ such that

$$\ell \geq 2n^3 \log n + 16n^3 t.$$

It is easy to check that the dominant term is the last term and this gives convergence for $\ell$ of order $n^4 \log n$.

**Theorem 5.2.** *For Kac's walk on $SO(n)$ there exists a constant $B$ such that*

$$D_*(Q^\ell, m) \leq Be^{-t}$$

*for all $\ell, t > 0$ such that*

$$\ell \geq 4n^4 \log n + 8n^4 t$$

*whereas the discrepancy distance satisfy*

$$D(Q^\ell, m) \leq B' e^{-t}$$

*for all $\ell, t > 0$ such that*

$$\ell \geq 4n^6 + 8n^4 t.$$

Thanks to (3.12), the same result holds also for the walk $\overline{Q}$ defined at (3.11) with slightly different numerical constants. This is worth mentioning because it seems it would be hard to improve upon this result in the case of $\overline{Q}$.

## References

1. Aharonov, D. and Ben-Or, M. (1997), *Fault tolerant quantum computation with constant error*, Proc. $29^{th}$ S.T.O.C., Assoc. Comp. Mach. New-York, 176-188.
2. Bird, G. A. (1976), *Molecular Gas Dynamics*, Clarendon Press, Oxford.
3. Carlen, E., Gabetta, E. and Toscani, G. (1997), *Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas*, preprint, Dept. of Mathematics, Georgia Tech..
4. Cercignani C., Illmer R. and Pulvirinti M. (1994), *The mathematical theory of dilute gases*, Springer, New-York.
5. Desvillettes L. (1995), *About the regularizing properties of the non cutoff Kac equation*, Comm. math. Physics **168**, 416-440.
6. Diaconis, P. and Mallows, C. (1997), *Identities for normal variables arising from the classical groups*, Technical Report, Dept. of Mathematics, Cornell University.
7. Diaconis, P. and Saloff–Coste, L. (1993 a), *Comparison techniques for random walk on finite groups*, Ann. Probab. **21**, 2131–2156.
8. Diaconis, P. and Saloff–Coste, L. (1993 b), *Comparison theorems for reversible Markov chains*, Ann. Appl. Math. **3**, 696–730.
9. Diaconis, P. and Shahshahani, M. (1996), *Products of random matrices as they arise in the study of random walks on groups*, Contemporary Math. **50**, 183–195.
10. Diaconis, P. and Shahshahani, M. (1987), *The subgroup algorithm for generating uniform random variables*, Prob. Eng. Info. Sci. **1**, 15–32.
11. Dudley, R. (1966), *Convergence of Baire measures*, Studia Math. **27**, 251–268.
12. Dudley, R. (1968), *Distances of probability measures and random variables*, Ann. Math. Statist. **39**, 1563–1572.
13. Dudley, R. (1976), *Probabilities and Metrics, Lecture Notes Series 45*, Matematisk Institut Aarhus.
14. Dudley, R. (1989), *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove, Calif.
15. Gallot, S., Hulin, D. and LaFontaine, J. (1990), *Riemannian Geometry*, 2nd ed., Springer–Verlag, Berlin.
16. Gottlieb A. L. (1998), *Markov Transitions and Propagation of Chaos.*, Ph.D thesis, Department of Mathematics, University of California, Berkeley.

17. Grünbaum, F. A. (1971), *Propagation of chaos for the Boltzmann equation.*, Archive for Rational Mechanics and Analysis **42**, 323-345.
18. Grünbaum, A. (1972), *Linearization for the Boltzmann equation*, Trans. Amer. Math. Soc. **165**, 425–499.
19. Hastings, W. (1970), *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57**, 97–109.
20. Hurwitz, A. (1897), *Über die erzengung der invarianten durch integration*, Nach. Gesell. Wissen, Göttingen Math-Phys Klasse. 71-90. Reprinted in Hurwitz, A. (1963), Mathematische Werke (Vol. II, 546–564) Birkhäuser, Basel.
21. Janvresse E. (1999), *Bounds on random rotations on the sphere*, Preprint.
22. Kac, M. (1956), *Foundations of Kinetic Theory*, Proc. 3rd Berkeley Sympos. (J. Neymann, ed.) Vol. 3, pp. 171–197.
23. Kac, M. (1959), *Probability and Related Topics in Physical Science*, Wiley Interscience, N.Y.
24. Kloss, B. (1959), *Limiting distributions on bicompact topological groups*, Th. Prob. Appl. **4**, 237–270.
25. Kuipers, L. and Niederreiter, H. (1974), *Uniform Distribution of Sequences*, Wiley, N.Y.
26. Lubotzky, A., Phillips, R. and Sarnak, P. (1986), *Hecke operators and distributing points on the sphere I*, Comm. Pure Appl. Math. **39**, Supplement 1, S149–S186.
27. Maslin, D. (1999), *The eigenvalues of Kac's master equation*, Preprint, Department of Mathematics, Dartmouth..
28. McKean, H. (1966), *Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas*, Arch. Rational Mech. Anal **2**, 343–367.
29. Méléard S. (1996), *Asymptotic behavior of some interacting particle systems, Mc Kean-Vlasov and Boltzmann models.*, Lecture Notes in Math. 1627, Springer, New-York..
30. Nambu K. (1983), *Interrelations between various direct simulation methods for solving the Boltzmann equation*, J. Phys. soc. Japan **52**, 3382-3388.
31. Perthame B. (1994), *Introduction to the theory of random particles methods for Boltzmann equation*, In: Progress on Kinetic Theory, World Scientific, Singapore.
32. Porod, U. (1995), *$L_2$–lower bounds for a special class of random walks*, Probab. Th. Related Fields **101**, 277–289.
33. Porod, U. (1996 a), *The cut-off phenomenon for random reflections*, Ann. Prob. **24**, 74–99.
34. Porod, U. (1996 b), *The cut–off phenomenon for random reflections II: Complex and quaternionic cases*, Probab. Th. Related Fields **104**, 181–209.
35. Riesz, F., Nagy, B. (1960), *Functional Analysis*, New York.
36. Rosenthal, J. (1994), *Random rotations: Characters and random walks on $SO(n)$*, Ann. Probab. **22**, 398–423.
37. Rothaus, O. (1981), *Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities*, J. Func. Anal. **42**, 102–109.
38. Saloff–Coste, L. (1997), *Lectures on finite Markov chains*, Ecole d'ete de St. Flour, LNM 1665, 1997, Springer..
39. Shor, P. (1998), *Quantum computation*, Proc. I.C.M., Berlin 1998, **I**, 467-486.
40. Smith, A and Roberts, G. (1993), *Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with discussion)*, Jour. Roy. Stat. Soc., B, 55, 3-24..
41. Su, F. (1995), *Ph. D. Dissertation*, Dept. Mathematics, Harvard University.
42. Sznitman A-S. (1991), *Topics in Propagation of Chaos*, Springer Lecture Notes in Math, 1464, Springer, New York.
43. Uchiyama K. (1988), *Derivation of the Boltzmann equation from particle dynamics*, Hiroshima Math. J. **18**, 245-297.