Acta Math., 191 (2003), 1–54

Determination of the spectral gap for Kac's master equation and related stochastic evolution

by

E. A. CARLEN

Georgia Institute of Technology
Atlanta, GA, U.S.A.

M. C. CARVALHO

Georgia Institute of Technology
Atlanta, GA, U.S.A.

and

M. LOSS

Georgia Institute of Technology
Atlanta, GA, U.S.A.

Contents

1. Introduction

We derive sharp bounds on the rate of relaxation to equilibrium for two models of random collisions connected with the Boltzmann equation, as well as several other stochastic evolutions of a related type. In fact, there is a fairly broad class of models to which the methods used here may be applied. The starting point is a model due to Mark Kac [10] of random energy-preserving "molecular collisions", and its analysis provides the pattern for the analysis of all of the models discussed here, including a more physically realistic model of random energy- and momentum-conserving collisions. However, since

The first and third authors were partially supported by U.S. N.S.F. Grant DMS 00-70589. The second author was on leave from Department of Mathematics, Faculty of Sciences, University of Lisbon, and was partially supported by FCT PRAXIS XXI and TMR ERB-FMRX CT97 0157.

the features of the Kac model have motivated the method of analysis presented here, we begin by introducing it.

The Kac model represents a system of N particles in one dimension evolving under a random collision mechanism. It is assumed that the spatial distribution of the particles is uniform, so that the state of the system is given by specifying the N velocities $v_1, v_2, ..., v_N$. The random collision mechanism under which the state evolves is that at random times T_j, a "pair collision" takes place in such a way that the total energy

$$E = \sum_{k=1}^{N} v_k^2 \tag{1.1}$$

is conserved. Since only a pair of one-dimensional velocities is involved in each collision, there are just two degrees of freedom active, and if the collisions were to conserve both energy and momentum, the only possible non-trivial result of a collision would be an exchange of the two velocities. Since Kac sought a model in which the distribution of the velocities would equilibriate over the energy surface specified by (1.1), he dropped the requirement of momentum conservation, and retained only energy conservation.

With energy conservation being the only constraint on a pair collision, the kinematically possible "post-collisional" velocities when particles i and j collide, v_i^* and v_j^*, are of the form

$$v_i^*(\theta) = v_i \cos(\theta) + v_j \sin(\theta) \quad \text{and} \quad v_j^*(\theta) = -v_i \sin(\theta) + v_j \cos(\theta) \tag{1.2}$$

where, of course, v_i and v_j are the pre-collisional velocities, and $\theta \in (-\pi, \pi]$.

To specify the evolution, consider it first in discrete time, collision by collision. Let

$$\vec{v}(k) = (v_1(k), v_2(k), ..., v_N(k)) \tag{1.3}$$

denote the state of the system just after the kth collision. Evidently, $\vec{v}(k)$ is a random variable with values in $S^{N-1}(\sqrt{E})$, the sphere in $\mathbf{R}^N$ of radius $\sqrt{E}$, where E is the energy. Let ϕ be any continuous function on $S^{N-1}(\sqrt{E})$. We will specify the collision mechanism by giving a formula for computing the conditional expectation of $\phi(\vec{v}(k+1))$ given $\vec{v}(k)$, which defines the one-step Markov transition operator Q through

$$Q\phi(\vec{v}) = \mathrm{E}\{\phi(\vec{v}(k+1)) \mid \vec{v}(k) = \vec{v}\}. \tag{1.4}$$

In the collision process to be modeled, the pair $\{i, j\}$, $i<j$, of molecules that collide is to be selected uniformly at random. Then the velocities v_i and v_j are updated by

choosing an angle θ, and letting (1.2) define the post-collisional velocities. Let $\varrho(\theta)$ be a probability density on the circle, i.e,

$$\int_{-\pi}^{\pi} \varrho(\theta)\, d\theta = 1, \tag{1.5}$$

and take ϱ to be the probability density for the outcome that the collision results in post-collisional velocities $v_i^*(\theta)$ and $v_j^*(\theta)$ as in (1.2).

The *one-step transition operator* Q for this process is defined as follows: For any continuous function ϕ on $S^{N-1}(\sqrt{E})$,

$$Q\phi(\vec{v}) = \binom{N}{2}^{-1} \sum_{i<j}^{N} \int_{-\pi}^{\pi} \varrho(\theta)\phi(v_1, v_2, ..., v_i^*(\theta), ..., v_j^*(\theta), ..., v_N)\, d\theta. \tag{1.6}$$

In terms of the process described above, $\mathrm{E}\{\phi(\vec{v}(k+1)) \,|\, \vec{v}(k)=\vec{v}\} = Q\phi(\vec{v})$.

The expression for Q can be simplified if for each $i<j$ we let $R_{i,j}(\theta)$ denote the rotation in $\mathbf{R}^N$ that induces a clockwise rotation in the (v_i, v_j)-plane through an angle θ, and fixes the orthogonal complement of this plane. Then $R_{i,j}(\theta)\vec{v}$ is the post-collisional velocity vector corresponding to the pre-collisional velocity vector $\vec{v}$, and (1.6) can be rewritten as

$$Q\phi(\vec{v}) = \binom{N}{2}^{-1} \sum_{i<j}^{N} \int_{-\pi}^{\pi} \varrho(\theta)\phi(R_{i,j}(\theta)\vec{v})\, d\theta. \tag{1.7}$$

Let $\mathcal{H}_{N,E}$ denote the Hilbert space of square-integrable functions ϕ on the sphere $S^{N-1}(\sqrt{E})$ equipped with the normalized uniform measure $d\mu_N$. Let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the inner product and norm on $\mathcal{H}_{N,E}$. It is clear from (1.7) that Q is an average over isometries on $\mathcal{H}_{N,E}$, and hence is a contraction, i.e., $\|Q\phi\|_2 \leqslant \|\phi\|_2$, and it is clear that $Q1=1$.

We now require that $\varrho(\theta)=\varrho(-\theta)$, so that Q is self-adjoint on $\mathcal{H}_{N,E}$. We also require that ϱ be continuous and strictly positive at $\theta=0$. The reason for this is that for any $\phi \in \mathcal{H}_{N,E}$,

$$2(\langle \phi, \phi \rangle - \langle \phi, Q\phi \rangle) = \binom{N}{2}^{-1} \sum_{i<j}^{N} \int_{-\pi}^{\pi} \varrho(\theta) \left[\int_{S^{N-1}(\sqrt{E})} (\phi(\vec{v}) - \phi(R_{i,j}(\theta)\vec{v}))^2\, d\mu_N \right] d\theta.$$

Under our conditions, the right-hand side vanishes if and only if for every sufficiently small θ, $\phi(\vec{v})=\phi(R_{i,j}(\theta)\vec{v})$ for almost every $\vec{v}$. This happens if and only if ϕ is constant. Thus, $\langle \phi, Q\phi \rangle = \|\phi\|_2$ if and only if ϕ is constant, so that 1 is an eigenvalue of Q of multiplicity one. This can be summarized in this context by saying that Q is ergodic.

Because Q is self-adjoint, it updates the probability density f_k for $\vec{v}$ as well. Indeed, for any test function ϕ,

$$\int_{S^{N-1}(\sqrt{E})} \phi(\vec{v}) f_{k+1}(\vec{v})\, d\mu_N = \mathrm{E}\,\phi(\vec{v}(k+1)) = \mathrm{E}(\mathrm{E}\{\phi(\vec{v}(k+1)\,|\,\vec{v}(k)\}) = \mathrm{E}\,Q\phi(\vec{v}(k))$$
$$= \int_{S^{N-1}(\sqrt{E})} \phi(\vec{v})\, Qf_k(\vec{v})\, d\mu_N,$$

which of course means that $Qf_k = f_{k+1}$.

One passes to a continuous time description by letting the waiting times between collisions become continuously distributed random variables. To obtain a Markov process, the distribution of these waiting times must be memoryless, and hence exponential. Therefore, fix some parameter $\tau_N > 0$, and define the Markovian semigroup G_t, $t > 0$, by

$$G_t f = e^{-t/\tau_N} \sum_{k=0}^{\infty} \frac{(t/\tau_N)^k}{k!} Q^k f = e^{(t/\tau_N)(Q-I)} f,$$

which gives the evolution of the probability density for $\vec{v}$, continuously in the time t.

It remains to specify the dependence of τ_N on N. Let $T^{(N)}$ denote the waiting time between collisions in the N-particle model. Suppose that the waiting time for any given particle to undergo a collision is independent of N, which corresponds roughly to adjusting the size of the container with N so that the particle density remains constant. Suppose also that these waiting times are all independent of one another, which *should* be more or less reasonable for a gas of many particles. (See Kac [10] for further discussion.) Then we would have $\Pr\{T_j^{(N)} > t\} = \Pr\{T_j^{(1)} > t\}^N$, or $e^{-t/\tau_N} = e^{-Nt/\tau_1}$. That is, $\tau_N = \tau_1/N$. Changing the time scale, we put $\tau_1 = 1$ and hence $\tau_N = 1/N$. Therefore, the semigroup is given by $G_t = e^{tN(Q-I)}$. For any initial probability density f_0, $f(\vec{v}, t) = G_t f_0(\vec{v})$ solves *Kac's master equation*

$$\frac{\partial}{\partial t} f(\vec{v}, t) = N(Q-I) f(\vec{v}, t), \tag{1.8}$$

which is the evolution equation for the model in so far as we are concerned with the probability density $f(\vec{v}, t)$ for the velocities at time t, and not the velocities $\vec{v}(t)$ themselves, which are random variables.

Because of the ergodicity, if f_0 is any initial probability density for the process, it is clear that $\lim_{t\to\infty} G_t f_0 = 1$. The question is how fast this relaxation to the invariant density 1 occurs. To quantify this, define

$$\lambda_N = \sup\{\langle f, Qf\rangle \,|\, \|f\|_2 = 1,\ \langle f, 1\rangle = 0\}. \tag{1.9}$$

Since Q is a self-adjoint contraction, its spectrum necessarily lies in the interval $[-1,1]$. Also, since Q commutes with the unitary change of scale that relates $\mathcal{H}_{N,E}$ and $\mathcal{H}_{N,E'}$ for two different values, E and E', of the energy, the spectrum of Q, and λ_N in particular, is independent of E.

We have already observed that 1 is an eigenvalue of Q of multiplicity one, so $\lambda_N \leqslant 1$. If $\lambda_N < 1$, there is a "gap" in the spectrum of Q. The spectral gap for $N(Q-I)$ is then

$$\Delta_N = N(1-\lambda_N). \tag{1.10}$$

This quantity is of interest in quantifying the rate of relaxation of $G_t f_0$ to 1 since for any square-integrable initial probability density f_0, as an easy consequence of the spectral theorem,

$$\|G_t(f_0-1)\|_2 \leqslant e^{-t\Delta_N}\|f_0-1\|_2.$$

Mark Kac, who introduced this operator and process [10] in 1956, observed that for each fixed l, the subspace of spherical harmonics of degree l in $S^{N-1}(\sqrt{E})$ is an invariant subspace under Q. (This is especially clear from (1.7) since if ϕ is in such a subspace, then so is $\phi \circ R_{i,j}(\theta)$ for each pair $i<j$ and each angle θ.) Since each of these subspaces is finite-dimensional, Q has a pure point spectrum. He remarks that it is not even evident that $\Delta_N > 0$ for all N, much less that there is a lower bound independent of N. (As Diaconis and Saloff-Coste noted in [5], Q is not compact.) He nonetheless conjectured that

$$\liminf_{N\to\infty} \Delta_N = C > 0. \tag{1.11}$$

Kac's conjecture in this form, for the special case $\varrho = 1/2\pi$ considered explicitly by Kac, was recently proved by Janvresse [9] using Yau's martingale method [13], [14]. Her proof gives no information on the value of C. One result, already proved in [3], is that in the case $\varrho = 1/2\pi$,

$$\Delta_N = \frac{1}{2}\,\frac{N+2}{N-1}, \tag{1.12}$$

and hence

$$\liminf_{N\to\infty} \Delta_N = \frac{1}{2}. \tag{1.13}$$

The result (1.12) has also been obtained by Maslen in unpublished work, using entirely different methods.([1]) Some account of Maslen's results can be found in a paper [5] by Diaconis and Saloff-Coste in which it is shown that $\Delta_N \geqslant C/N^2$ for the Kac model as well as a natural generalization of it in which the sphere S^{N-1} is replaced by the special

([1]) *Note added in proof.* Since this paper was written, the work of Maslen has been written up and published in: Maslen, D. K., The eigenvalues of Kac's master equation. *Math. Z.*, 243 (2003), 291–331.

orthogonal group $SO(N)$. Our method gives exact results in this case too, as we shall see.

Maslen's approach was based on the representation theory of the group $SO(N)$, and does not seem to extend to more general cases, such as a non-uniform density $\varrho(\theta)$, or to momentum-conserving collisions. According to his thesis advisor, Persi Diaconis, this is one reason it was never published. We will comment further on the relation of our paper to previous work, especially [9], [13], [14] and [5], in §3 where we carry out our analysis of the Kac model, and in §6 where we analyze the $SO(N)$-variant of the Kac model.

Kac did not explicitly conjecture (1.13), only (1.11), though he discussed motivations for his conjecture that do suggest (1.13). In particular, he was motivated by a connection between the many-particle evolution described by the master equation (1.8), and a model Boltzmann equation, and he did rigorously establish the following connection: For each integer k, $1\leqslant k\leqslant N$, let π_k be the kth coordinate projection on $S^{N-1}(\sqrt{E})$; i.e.,

$$\pi_k(v_1, v_2, ..., v_N) = v_k. \tag{1.14}$$

Given a probability density f on $S^{N-1}(\sqrt{E})$, define its kth *single-particle marginal with respect to Lebesgue measure*, $[f]_{(k)}(v)$, by

$$\int_{S^{N-1}(\sqrt{E})} \phi(v_k) f(\vec{v})\, d\mu_N = \int_{\mathbf{R}} \phi(v)[f]_{(k)}(v)\, dv$$

for all continuous functions ϕ on $\left[-\sqrt{E}, \sqrt{E}\right]$. It is natural to consider initial data f_0 for the Kac master equation that is invariant under permutation of particle coordinates since this property is preserved by the evolution. For such a density f_0, $[f_0]_k = [f_0]_1$ for all k. Because $[G_t f_0]_1$ contains much of the information in $G_t f_0$ that is physically relevant, it is natural to seek an equation for $[G_t f_0]_1$.

Kac showed that with $E=N$ (or just proportional to N), if a sequence of initial densities $f_0^{(N)}$ on $S^{N-1}(\sqrt{E})$ satisfies a certain symmetry and independence property that he called "molecular chaos", and if furthermore

$$g(v) = \lim_{N\to\infty} \left[f_0^{(N)}\right]_1(v)$$

exists in $L^1(\mathbf{R})$, then so does $g(v,t) = \lim_{N\to\infty} \left[G_t f_0^{(N)}(v)\right]_1$, and $g(v,t)$ satisfies the *Kac equation*

$$\frac{\partial}{\partial t} g(v,t) = 2\int_{-\pi}^{\pi} \left(\int_{\mathbf{R}} [g(v^*(\theta),t)g(w^*(\theta),t) - g(v,t)g(w,t)]\, dw \right) \varrho(\theta)\, d\theta. \tag{1.15}$$

The fact that there is a quadratic non-linearity on the right is due to the fact that the underlying many-particle dynamics is generated by pair collisions. The factor of 2 on

the right-hand side comes from the 2 in the the normalization factor $2/N(N-1)$ in the definition of Q, (1.6). The N is absorbed by the factor of N in $N(Q-I)$, the generator of G_t, and the $N-1$ is absorbed by summing over all the $N-1$ particles with which the first particle can collide.

Kac's limit theorem provides a direct link between the linear but many-particle master equation (1.8) and the one-variable but non-linear Kac equation (1.15). Kac's proposal was that one should be able to obtain quantitative results about the behavior of the master equation, and from these, deduce quantitative results on the Kac equation (1.15). Specifically, he was concerned with following this route to results on the rate of relaxation to equilibrium for solutions of (1.15).

It is easy to see that for any $\beta>0$,

$$m_\beta(v)=\sqrt{\frac{\beta}{2\pi}}\,e^{-\beta v^2/2} \tag{1.16}$$

is a steady-state solution of the Kac equation (1.15). (In the context of kinetic theory, the Gaussian density in (1.16) is known as the *Maxwellian* density with temperature $1/\beta$.) Indeed, as is well known, m_β is the limit of the single-particle marginal on $S^{N-1}\big(\sqrt{N/\beta}\,\big)$ as N tends to infinity. Kac wanted to show that for any reasonable initial data $g(v)$, the Kac equation had a solution $g(v,t)$ with $\lim_{t\to\infty} g(v,t)=m_\beta(v)$ where $\int_{\mathbf{R}} v^2 g(v)\,dv=1/\beta$. Indeed, he wanted to show that this convergence took place *exponentially fast*, and he boldly conjectured that one could prove this exponential convergence for the master equation from whence (1.15) came. At the time Kac wrote his paper, very little was known about the non-linear Boltzmann equation, Carleman's 1933 paper [2] being one of the few mathematical studies. Given the difficulties inherent in dealing directly with the non-linear equation, his suggested approach via the master equation was well motivated, though unfortunately he did not succeed himself in obtaining quantitative relaxation estimates by this route, and other workers choose to directly investigate the non-linear equation.

Evidence for the conjectured exponential convergence came from linearizing the Kac equation about the steady-state solutions m_β. The resulting generator of the linearized Kac equation can be written in terms of averages of Mehler kernels, as shown in [12], and so all of the eigenfunctions are Hermite polynomials (as Kac had observed in §9 of [10]). The eigenvalue corresponding to the nth-degree Hermite polynomial, $n\geqslant 1$, is then readily worked out to be (see [12])

$$2\int_{-\pi}^{\pi}(\sin^n(\theta)+\cos^n(\theta)-1)\varrho(\theta)\,d\theta. \tag{1.17}$$

The eigenvalue is zero for $n=2$, corresponding to conservation of energy. As we have indicated, Kac actually only considered the special case in which ϱ was uniform; i.e.,

$\varrho(\theta)=1/2\pi$. In this case, the eigenvalues are -2 for n odd, and are monotonically decreasing toward -2 for n even. Thus, the eigenvalue corresponding to the fourth-degree Hermite polynomial determines the spectral gap for the linearization of (1.15) in this case. *The fact that this gap is* $\frac{1}{2}$ *is consistent with* (1.13), *and bears out Kac's intuition that there is a close quantitative connection between his master equation* (1.8) *and the Kac equation* (1.15).

In fact, as we shall see, in the case considered by Kac and some other cases as well, λ_N is an eigenvalue of Q of multiplicity one, and $Qf_N(\vec{v})=\lambda_N f_N(\vec{v})$ for

$$f_N(v_1,\ldots,v_N)=\sum_{j=1}^{N}(v_j^4-\langle 1,v_j^4\rangle). \tag{1.18}$$

If $E=N$ and g_N is defined by $P_1(f_N)=g_N\circ\pi_N$, then

$$\lim_{N\to\infty} g_N(v)=m_1(v)h_{(4)}(v), \tag{1.19}$$

where $h_{(4)}$ is the fourth-degree Hermite polynomial for the standard unit variance Gaussian measure on $\mathbf{R}$. (This is fairly evident, but will be fully evident in view of the formula for P_1 given in §2.) Thus, the correspondence between the spectral gaps extends to a correspondence between the eigenfunctions too.

McKean [12] and Grünbaum [7], [8] have further investigated these issues. In particular, McKean conjectured that reasonable solutions of (1.15) should relax to the Gaussian stationary solutions of the same energy in L^1 at the exponential rate $e^{-t/2}$ corresponding to the spectral gap in the linearized equation. He proved this for nice initial data but with exponential rate e^{-tc} where c is an explicit constant, but about an order of magnitude smaller than $\frac{1}{2}$. Later, in [4] this result was established with almost the sharp rate, i.e., $e^{-[1/2-\varepsilon]t}$ for nice initial data. See the papers for precise statements, but note that all of this is in the case $\varrho=1/2\pi$. (The results are stated differently in [12] and [4], which use a different time scale so that the factor of 2 in (1.15) is absent.)

If one expects that the linearized version of (1.15) is a good guide to the behavior solutions of (1.15), one might guess that (1.17) provides a good guide to the relaxation properties of solutions of (1.15). This would suggest that in the case in which ϱ is uniform, the slowest mode of relaxation corresponds to initial data of the form $m_1(v)(1+\varepsilon h_{(4)}(v))$ for small ε.

If one further believed that the non-linear Kac equation (1.15) is a good guide to behavior of solutions of Kac's master equation, then one might guess that the slowest mode of relaxation for the master equation is a symmetric fourth-degree polynomial, at least for uniform ϱ. Such a line of reasoning suggests (1.18) as a candidate for the

slowest mode of relaxation for Kac's master equation. This turns out to be correct, as we have indicated, and this shows how well-constructed the Kac model is: A great deal of information is washed out and lost whenever one passes from the N-particle distribution function $f(\vec{v})$ to its single-particle marginal distribution $g(v)$. In general, there would be no reason to expect that the slowest mode of decay for the master equation would not be lost in passing to the marginal. Here, it nevertheless is true.

Indeed, it is easy to see that f_N is in fact an eigenfunction of Q. We shall see that for many choices of ϱ, f_N is the optimizer in (1.9). This correspondence between Kac's master equation (1.8) and the linearized version of the Kac equation (1.15) is a full vindication of Kac's conjectures. It also shows that his model is free of extraneous detail at the microscopic level; what happens at the microscopic level described by the master equation is what happens at the level described by (1.15).

We conclude the introduction by briefly stating our results for the Kac model itself, and then describing the structure of the paper. The key result in our analysis of the Kac model is the following theorem which reduces the variational problem (1.9) to a much simpler, purely geometric, one-dimensional problem:

THEOREM 1.1. *For all* $N \geqslant 3$,

$$\Delta_N \geqslant (1-\varkappa_N)\Delta_{N-1} \tag{1.20}$$

where

$$\varkappa_N = \sup\left\{ \frac{\langle g\circ\pi_1, g\circ\pi_2\rangle}{\|g\circ\pi_1\|_2} \;\middle|\; g\in C(\mathbf{R}),\ \langle g\circ\pi_1, 1\rangle = 0 \right\}. \tag{1.21}$$

Notice first of all that g is a function of a single variable—in contrast to (1.9), (1.21) is a one-dimensional variational problem. Also notice that (1.21) doesn't involve ϱ, or otherwise directly refer to Q.

The bound in Theorem 1.1 implies that $\liminf_{N\to\infty} \Delta_N \geqslant \prod_{j=3}^{\infty}(1-\varkappa_j)\Delta_2$. Since the necessary and sufficient condition for the infinite product to be non-zero is that

$$\sum_{j=3}^{\infty} \varkappa_j < \infty, \tag{1.22}$$

proving that the Kac conjecture in the form (1.11) is reduced to the problem of proving the summability of $\varkappa_j$ and the strict positivity of Δ_2.

The second part is easy, since for two particles, Q is an operator on functions on S^1. Indeed,

$$\langle f, Qf\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\psi)f(\psi-\theta)\varrho(\theta)\,d\theta\,d\psi, \tag{1.23}$$

and writing this in terms of Fourier series leads to

$$\lambda_2 = \sup_{k\neq 0}\left\{\int_{-\pi}^{\pi} \varrho(\theta)\cos(k\theta)\,d\theta\right\}. \tag{1.24}$$

By the Riemann–Lebesgue lemma, $\lambda_2<1$, and hence $\Delta_2=2(1-\lambda_2)>0$.

As for the summability of $\varkappa_N$, note from (1.21) that $\varkappa_N$ is a measure of the dependence of the coordinate functions on the sphere. Note also that, like λ_N, $\varkappa_N$ is independent of E. This is for the exact same reason: K commutes with the unitary operator effecting a change of scale. For present purposes, choose $E=N$ so that the marginal distribution of (v_1, v_2) induced by μ_N is

$$\frac{|S^{N-3}|}{N|S^{N-1}|}\left(1-\frac{v_1^2+v_2^2}{N}\right)^{(N-4)/2} dv_1\,dv_2.$$

As N tends to infinity, this tends to

$$\frac{1}{2\pi}e^{-(v_1^2+v_2^2)/2}\,dv_1\,dv_2,$$

and under this limiting measure, the two coordinate functions v_1 and v_2 are independent. Hence for any admissible trial function g in (1.21),

$$\begin{aligned}\lim_{N\to\infty}\langle g\circ\pi_1, g\circ\pi_2\rangle &= \frac{1}{2\pi}\int_{\mathbf{R}^2} g(v_1)g(v_2)e^{-(v_1^2+v_2^2)/2}\,dv_1\,dv_2 \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} g(v_1)e^{-v_1^2/2}\,dv_1\,\frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} g(v_2)e^{-v_1^2/2}\,dv_2 \\ &= \lim_{N\to\infty}\langle g\circ\pi_1, 1\rangle\langle g\circ\pi_2, 1\rangle = 0,\end{aligned} \tag{1.25}$$

which implies that $\lim_{N\to\infty}\varkappa_N=0$, without, however, showing how fast. This is the last time in our discussion that it is of use to choose E proportional to N. *Henceforth we set* $E=1$.

In fact, it is not hard to compute $\varkappa_N$ exactly:

THEOREM 1.2. *For all* $N\geqslant 3$,

$$\varkappa_N = \frac{3}{N^2-1}.$$

Since this is summable, (1.22) holds, and so the Kac conjecture is proved. But Theorem 1.2 tells us much more than just (1.22). One can exactly solve the recurrence relation in Theorem 1 with $\varkappa_N=3/(N^2-1)$. As we shall see, this leads to

THEOREM 1.3. *For all* $N \geqslant 2$,

$$\Delta_N \geqslant \frac{1-\lambda_2}{2} \frac{N+2}{N-1}. \tag{1.26}$$

Moreover, this result is sharp for the case considered by Kac, i.e., constant density ϱ, *in which case* $\lambda_2 = 0$, *and more generally whenever*

$$\int_{-\pi}^{\pi} \varrho(\theta) \cos(k\theta)\, d\theta \leqslant \int_{-\pi}^{\pi} \varrho(\theta) \cos(4\theta)\, d\theta \tag{1.27}$$

for all $k \neq 0$. *In all of these cases,* λ_N *has multiplicity one, and the corresponding eigenfunction is*

$$\sum_{k=1}^{N} (v_k^4 - \langle v_k^4, 1 \rangle). \tag{1.28}$$

The division of our results on the original Kac model into Theorems 1.1, 1.2 and 1.3 of course reflects the steps in the method by which they are obtained. However, it also reflects a point of physical relevance, namely *that* $\varkappa_N$ *is completely independent of* $\varrho(\theta)$. The complicated details of the collision mechanism do not enter into $\varkappa_N$. Rather, they enter our estimate for Δ_N *only* through the value of the *two-particle* gap $\Delta_2 = 2(1-\lambda_2)$. Once this is computed, there is a *purely geometric* relation between the values of the gap for different values of N. The fact that there should be such a simple and purely geometric relation between the values of the gap for different values of N is a very interesting feature of the Kac model which expresses the strong sense in which it is a binary collision model.

The paper is organized as follows: In §2 we identify the general features of the Kac model that enable us to prove Theorem 1.1. We then introduce the notion of a *Kac system*, which embodies these features, and prove the results that lead to analogs of Theorem 1.1 for general Kac systems. This provides a convenient framework for the analysis of a number of models, as we illustrate in the next four sections. §3 is devoted to the Kac model itself, and contains the proofs of Theorems 1.1, 1.2 and 1.3. §4 is devoted to the analysis of the master equation for physical, three-dimensional, momentum- and energy-conserving Boltzmann collisions. §5 is devoted to a shuffling model that has been studied in full detail by Diaconis and Shahshahani [6]. We include this here because it can be viewed as the Kac model with momentum conservation, and is very simple. (We hasten to add that Diaconis and Shahshahani do much more for this model than compute the spectral gap.) Then in §6 we treat another generalization of the Kac model, this time in the direction of greater complexity: The $SO(N)$-model of Maslen, Diaconis and Saloff-Coste [5]. Finally, in §7 we show that the quartic eigenfunction (1.18) is indeed the gap eigenfunction for a wide range of non-uniform densities ϱ that violate (1.27).

2. General features

The Kac model introduced in the previous section has the following general features that are shared by all of the models discussed here:

Feature 1. For each $N>1$ there is measure space $(X_N, \mathcal{S}_N, \mu_N)$, with μ_N a probability measure, on which there is a measure-preserving action of Π_N, the symmetric group on N letters. We denote

$$\mathcal{H}_N = L^2(X_N, \mu_N). \tag{2.1}$$

We think of X_N as the "N-particle phase space" or "N-particle state space", and the action of Π_N as representing "exchange of particles". In the Kac model, X_N is S^{N-1}, $\mathcal{S}_N$ is the Borel field, and μ_N is the rotation-invariant probability measure on $S^{N-1}=X_N$. A permutation $\sigma\in\Pi_N$ acts on X_N through

$$\sigma(v_1, v_2, ..., v_N) = (v_{\sigma(1)}, v_{\sigma(2)}, ..., v_{\sigma(N)}).$$

Feature 2. There is another measure space $(Y_N, \mathcal{T}_N, \nu_N)$ and there are measurable maps $\pi_j: X_N \to Y_N$ for $j=1, 2, ..., N$ such that for all $\sigma\in\Pi_N$, and each j,

$$\pi_j \circ \sigma = \pi_{\sigma(j)}. \tag{2.2}$$

Moreover, for each j, and all $A\in\mathcal{T}_N$,

$$\nu_N(A) = \mu_N(\pi_j^{-1}(A)). \tag{2.3}$$

We denote

$$\mathcal{K}_N = L^2(Y_N, \nu_N). \tag{2.4}$$

We think of $\pi_j(x)$ as giving the "state of the jth particle when the N-particle system is in state x". For example, in the Kac model, we take

$$\pi_j(v_1, v_2, ..., v_N) = v_j \in [-1, 1], \tag{2.5}$$

and thus we take $Y_N=[-1, 1]$. In this case, Y_N does not depend on N, and it may seem strange to allow the single-particle state space itself to depend on N. However, the methods we use here permit this generality, and some of the examples considered here require it.

Notice that once Y_N and the π_j are given, ν_N is specified through (2.3). In the Kac model we therefore have

$$\nu_N(v) = \frac{|S^{N-2}|}{|S^{N-1}|}(1-v^2)^{(N-3)/2}\,dv. \tag{2.6}$$

Feature 3. For each $N \geqslant 3$ and each $j=1,2 \ldots, N$, there is a map

$$\phi_j : X_{N-1} \times Y_N \to X_N \tag{2.7}$$

so that

$$\pi_j(\phi_j(x,y)) = y \tag{2.8}$$

for all $j=1,\ldots,N$ and all $(x,y) \in X_{N-1} \times Y_N$. Moreover, ϕ_j has the property that for all $A \in \mathcal{S}_N$,

$$[\mu_{N-1} \otimes \nu_N](\phi_j^{-1}(A)) = \mu_N(A), \tag{2.9}$$

or equivalently, for all bounded measurable functions f on X_N, all $1 \leqslant j \leqslant N$,

$$\int_{X_N} f \, d\mu_N = \int_{Y_N} \left[\int_{X_{N-1}} f(\phi_j(x,y)) \, d\mu_{N-1}(x) \right] d\nu_N(y). \tag{2.10}$$

In the Kac model case, for any $\tilde{v} \in X_{N-1} = S^{N-2}$ and any $v \in Y_N = [-1,1]$ we put

$$\phi_N(\tilde{v}, v) = \left(\sqrt{1-v^2}\, \tilde{v}, v\right), \tag{2.11}$$

and $\phi_j = \sigma_{j,N} \circ \phi_N$, where $\sigma_{j,N}$ is the pair permutation interchanging j and N. In this case, (2.9) is easily verified.

So far, none of the features we have considered involve the dynamics. That is, the first three features are purely kinematical. The fourth feature brings in the Markov transition operator Q. We do not make the dependence of Q on N explicit in our notation, since this will always be clear from the context.

Feature 4. For each $N \geqslant 2$, there is a self-adjoint and positivity-preserving operator Q on $\mathcal{H}_N$ such that $Q1=1$. These operators are related to one another by the following: For each $N \geqslant 3$, each $j=1,2,\ldots,N$, and each square-integrable function f on X_N,

$$\langle f, Qf \rangle_{\mathcal{H}_N} = \frac{1}{N} \sum_{j=1}^{N} \int_{Y_N} \langle f_{j,y}, Q f_{j,y} \rangle_{\mathcal{H}_{N-1}} \, d\nu_N(y) \tag{2.12}$$

where for each j and each $y \in Y_N$,

$$f_{j,y}(\cdot) = f(\phi_j(\cdot, y)). \tag{2.13}$$

It is easily verified that the Kac model possesses this feature.

Definition. A *Kac system* is a system of probability spaces $(X_N, \mathcal{S}_N, \mu_N)$ and $(Y_N, \mathcal{T}_N, \nu_N)$ for $N \in \mathbf{N}$, $N \geqslant 2$, together with, for each N, maps π_j and ϕ_j, $j = 1, 2, ..., N$, a measure-preserving action of Π_N on $(X_N, \mathcal{S}_N, \mu_N)$, and a Markov transition operator Q on $\mathcal{H} = L^2(X_N, \mu_N)$, related to one another in such a way that they possess all of the properties specified in Features 1 through 4 above.

In analyzing the spectral gaps of the operators Q in Kac systems, certain other operators related to conditional expectations will play a central role, as indicated in the previous section. Suppose that $(X_N, \mathcal{S}_N, \mu_N)$, $(Y_N, \mathcal{T}_N, \nu_N)$, π_j and ϕ_j are defined and related as specified above.

For each $j = 1, 2, ..., N$, let P_j be the orthogonal projection onto the subspace of $\mathcal{H}_N$ consisting of functions of the form $g \circ \pi_j$ for some $g \in \mathcal{K}_N$. In probabilistic language, $P_j f$ is the conditional expectation of f given π_j; i.e., $P_j f = \mathrm{E}\{f \mid \pi_j\}$, with the expectation taken according to μ_N.

The features of a Kac system provide useful formulas for the P_j: With $y = \pi_j(x)$ and $f_{j,y}$ given by (2.13),

$$P_j f(x) = g(\pi_j(x)) \quad \text{where } g(y) = \int_{X_{N-1}} f_{j,y}(\tilde{x})\, d\mu_{N-1}(\tilde{x}). \tag{2.14}$$

In terms of these projections, define

$$P = \frac{1}{N} \sum_{j=1}^{N} P_j, \tag{2.15}$$

which is clearly a positive contraction on $\mathcal{H}_N$. This operator plays a fundamental role in what follows. Define λ_N by

$$\lambda_N = \sup\{\langle f, Qf \rangle_{\mathcal{H}_N} \mid \|f\|_{\mathcal{H}_N} = 1 \text{ and } \langle f, 1 \rangle_{\mathcal{H}_N} = 0\}, \tag{2.16}$$

and then

$$\Delta_N = N(1 - \lambda_N).$$

In this section we derive a recursion relation

$$\Delta_N \geqslant \left[\frac{N}{N-1}(1 - \mu_N)\right] \Delta_{N-1} \tag{2.17}$$

for all $N \geqslant 3$, where

$$\mu_N = \sup\{\langle f, Pf \rangle_{\mathcal{H}_N} \mid \|f\|_{\mathcal{H}_N} = 1 \text{ and } \langle 1, f \rangle_{\mathcal{H}_N} = 0\}. \tag{2.18}$$

Now suppose that all of the projections π_j were independent random variables. In analytic terms this means that $P_iP_jf=\int f\,d\mu_N$, $i\neq j$. Then, evidently, the spectrum of P would be $\{0,1/N,1\}$ and $\mu_N=1/N$, so that

$$\left[\frac{N}{N-1}(1-\mu_N)\right]=1.$$

This would imply that Δ_N is a non-decreasing function of N, which is too much to hope for. However, to the extent that the π_j are *almost* independent, we may hope to find

$$\mu_N\leqslant\frac{1}{N}+\frac{N-1}{N}\gamma_N \tag{2.19}$$

with γ_N rapidly decreasing in N. Inserting (2.19) in (2.17), we obtain

$$\Delta_N\geqslant(1-\gamma_N)\Delta_{N-1}. \tag{2.20}$$

As observed in the introduction, this will imply that $\liminf_{N\to\infty}\Delta_N>0$ provided $\Delta_2>0$ and that $\sum_{j=3}^{\infty}\gamma_j<\infty$.

The spectrum of P turns out to be closely related to the spectrum of a relatively simple operator K in the single-particle space $\mathcal{K}_N$. Define a contraction K on $\mathcal{K}_N$ by

$$(Kg)\circ\pi_N=P_N(g\circ\pi_{N-1}). \tag{2.21}$$

Note that $Kg(y)$ is the conditional expectation of $g\circ\pi_2$ given that $\pi_1=y$. That is,

$$Kg(y)=\mathrm{E}\{g\circ\pi_N\,|\,\pi_{N-1}=y\}. \tag{2.22}$$

In concrete examples, it is easy to deduce an explicit formula for K from (2.21) and (2.14) or directly from (2.22). By the permutation symmetry, specifically the invariance of μ_N and (2.2),

$$P_i(g\circ\pi_j)=(Kg)\circ\pi_i\quad\text{for all } i\neq j. \tag{2.23}$$

Combining (2.13), (2.14) and (2.21), we obtain

$$Kg(y)=\int_{X_{N-1}}g(\pi_{N-1}(\phi_N(\tilde{x},y)))\,d\mu_{N-1}(\tilde{x}), \tag{2.24}$$

which provides an explicit form for the operator K. For example, in the case of the Kac model we obtain

$$\begin{aligned}Kg(v)&=\int_{X_{N-1}}g\big(\sqrt{1-v^2}\,w_{N-1}\big)\,d\mu_{N-1}(w)=\int_{-1}^{1}g\big(\sqrt{1-v^2}\,w\big)\,d\nu_{N-1}(w)\\&=\frac{|S^{N-3}|}{|S^{N-2}|}\int_{-1}^{1}g\big(\sqrt{1-v^2}\,w\big)(1-w^2)^{(N-4)/2}\,dw\end{aligned} \tag{2.25}$$

from (2.6) and (2.24).

Since K is a contraction, its spectrum lies in $[-1,1]$. Define numbers $\varkappa_N$ and β_N by

$$\varkappa_N = \sup\{\langle g, Kg\rangle_{\mathcal{K}_N} \mid \|g\|_{\mathcal{K}_N}=1 \text{ and } \langle 1, g\rangle_{\mathcal{K}_N}=0\} \tag{2.26}$$

and

$$-(N-1)\beta_N = \inf\{\langle g, Kg\rangle_{\mathcal{K}_N} \mid \|g\|_{\mathcal{K}_N}=1\}. \tag{2.27}$$

(The factor of $-(N-1)$ in the definition of β_N is included with (2.19) in mind.)

THEOREM 2.1. *Given any Kac system, let P and K be defined by* (2.15) *and* (2.21). *Let μ_N, $\varkappa_N$ and β_N be defined by* (2.18), (2.26) *and* (2.27), *respectively.*

Then, either $\mu_N=0$ or

$$\mu_N = \max\left\{\frac{1}{N}+\frac{N-1}{N}\varkappa_N, \frac{1}{N}+\frac{N-1}{N}\beta_N\right\}. \tag{2.28}$$

In case $\varkappa_N>\beta_N$, and $\varkappa_N$ is an eigenvalue of K, then μ_N is an eigenvalue of P, and both eigenvalues have the same multiplicity. In fact, the map

$$h \mapsto \left(\frac{1}{N(1+(N-1)\varkappa_N)}\right)^{1/2} \sum_{j=1}^{N} h\circ\pi_j \tag{2.29}$$

is an isometry from the $\varkappa_N$-eigenspace of K in $\mathcal{K}_N$ onto the μ_N-eigenspace of P in $\mathcal{H}_N$.

Proof. Since P is self-adjoint, it suffices to consider trial functions in the range of P. Therefore, suppose that $f=Pg$, $\langle g, 1\rangle_{\mathcal{H}_N}=0$. Then with h_j defined by $Nh_j\circ\pi_j=P_jg$, $j=1\ldots N$, we have

$$f=\sum_{j=1}^{N} h\circ\pi_j.$$

A simple calculation yields

$$\begin{aligned}\|f\|^2_{\mathcal{H}_N} &= \sum_{j=1}^{N}\|h_j\|^2_{\mathcal{K}_N} + \sum_{\substack{i,j=1\\ i\neq j}}^{N}\langle h_i, Kh_j\rangle_{\mathcal{K}_N} \\ &= \sum_{j=1}^{N}\langle h_j, (I-K)h_j\rangle_{\mathcal{K}_N} + \sum_{i,j=1}^{N}\langle h_i, Kh_j\rangle_{\mathcal{K}_N}.\end{aligned} \tag{2.30}$$

Now introduce $\tilde{h}=N^{-1}\sum_{j=1}^{N} h_j$ and $m_j=h_j-\tilde{h}$. Evidently,

$$\sum_{j=1}^{N} m_j = 0. \tag{2.31}$$

Then the result (2.30) can be written as

$$\|f\|^2_{\mathcal{H}_N} = N\langle \tilde{h}, (I+(N-1)K)\tilde{h}\rangle_{\mathcal{K}_N} + \sum_{j=1}^{N} \langle m_j, (I-K)m_j\rangle_{\mathcal{K}_N}. \tag{2.32}$$

The easiest way to see this is to introduce the vector

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix}$$

and the $(N\times N)$-block matrices

$$\begin{bmatrix} I & K & K & \dots & K \\ K & I & K & \dots & K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & K & K & \dots & I \end{bmatrix} = K \begin{bmatrix} I & I & I & \dots & I \\ I & I & I & \dots & I \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I & I & I & \dots & I \end{bmatrix} + (I-K) \begin{bmatrix} I & 0 & 0 & \dots & 0 \\ 0 & I & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I \end{bmatrix}. \tag{2.33}$$

Then, in the obvious sense of the dot product, (2.30) can be written

$$\|f\|^2_{\mathcal{H}_N} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix} \cdot \begin{bmatrix} I & K & K & \dots & K \\ K & I & K & \dots & K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & K & K & \dots & I \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix}.$$

Because of (2.31),

$$\begin{bmatrix} I & I & I & \dots & I \\ I & I & I & \dots & I \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I & I & I & \dots & I \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_N \end{bmatrix} = 0,$$

and (2.32) easily follows.

In the same way, one computes that

$$N\langle f, Pf\rangle_{\mathcal{H}_N} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix} \cdot \begin{bmatrix} I & K & K & \dots & K \\ K & I & K & \dots & K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K & K & K & \dots & I \end{bmatrix}^2 \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix}.$$

This reduces to

$$N\langle f, Pf\rangle_{\mathcal{H}_N} = N\langle \tilde{h}, (I+(N-1)K)^2\tilde{h}\rangle_{\mathcal{K}_N} + \sum_{j=1}^{N} \langle m_j, (I-K)^2 m_j\rangle_{\mathcal{K}_N}. \tag{2.34}$$

It is clear from (2.32) and (2.34) that $\langle f, Pf\rangle_{\mathcal{H}_N}/\|f\|^2_{\mathcal{H}_N}$ equals the greater of

$$\sup\{\langle g,(I+(N-1)K)g\rangle_{\mathcal{K}_N} \mid \|g\|_{\mathcal{K}_N}=1 \text{ and } \langle g,1\rangle_{\mathcal{K}_N}=0\}$$

and

$$\sup\{\langle g,(I-K)g\rangle_{\mathcal{K}_N} \mid \|g\|_{\mathcal{K}_N}=1\}.$$

The identity (2.28) follows easily from this as the definition of β_N. The final assertion is now easily checked. □

The next theorem provides the recurrence (2.17).

THEOREM 2.2. *Given any Kac system, let P, μ_N and λ_N be defined by* (2.15), (2.18) *and* (2.16), *respectively. Then*

$$\lambda_N \leqslant \lambda_{N-1}+(1-\lambda_{N-1})\mu_N. \tag{2.35}$$

Moreover, there is equality in (2.35) *if and only if the suprema in* (2.16) *and* (2.18) *are attained at a common function f_N.*

Proof. We start from (2.12), taking any function $f\in\mathcal{H}_N$ satisfying the conditions imposed in (2.16):

$$\begin{aligned}
\langle f, Qf\rangle_{\mathcal{H}_N} &= \frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}\langle f_{j,y}, Qf_{j,y}\rangle_{\mathcal{H}_{N-1}}\,d\nu_N(y)\\
&= \frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}\langle [f_{j,y}-P_jf(y)]+P_jf(y), Q([f_{j,y}-P_jf(y)]+P_jf(y))\rangle_{\mathcal{H}_{N-1}}\,d\nu_N(y)\\
&= \frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}\langle [f_{j,y}-P_jf(y)], Q[f_{j,y}-P_jf(y)]\rangle_{\mathcal{H}_{N-1}}\,d\nu_N(y)\\
&\quad +\frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}|P_jf(y)|^2\,d\nu_N(y),
\end{aligned}$$

since each $P_jf(y)$ is constant on X_{N-1} and so on $\mathcal{H}_{N-1}$, $QP_jf(y)=P_jf(y)$, and

$$\langle [f_{j,y}-P_jf(y)], P_jf(y)\rangle_{\mathcal{H}_{N-1}}=0.$$

But

$$\frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}|P_jf(y)|^2\,d\nu_N(y)=\langle f, Pf\rangle_{\mathcal{H}_N},$$

and hence

$$\langle f, Qf\rangle_{\mathcal{H}_N} = \frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}\langle [f_{j,y}-P_jf(y)], Q[f_{j,y}-P_jf(y)]\rangle_{\mathcal{H}_{N-1}}\, d\nu_N(y) + \langle f, Pf\rangle_{\mathcal{H}_N}. \tag{2.36}$$

Now since $\langle [f_{j,y}-P_jf(y)], 1\rangle_{\mathcal{H}_{N-1}}=0$ for each y and j,

$$\begin{aligned}\langle [f_{j,y}-P_jf(y)], Q[f_{j,y}-P_jf(y)]\rangle_{\mathcal{H}_{N-1}} &\leqslant \lambda_{N-1}\|f_{j,y}-P_jf(y)\|^2_{\mathcal{H}_{N-1}}\\ &= \lambda_{N-1}(\|f_{j,y}\|^2_{\mathcal{H}_{N-1}} - |P_jf(y)|^2_{\mathcal{H}_{N-1}}).\end{aligned}$$

Averaging over j and integrating over y,

$$\frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}\langle [f_{j,y}-P_jf(y)], Q[f_{j,y}-P_jf(y)]\rangle_{\mathcal{H}_{N-1}}\, d\nu_N(y) \leqslant \lambda_{N-1}(\|f\|^2_{\mathcal{H}_N} - \langle f, Pf\rangle_{\mathcal{H}_N}).$$

From this and (2.36), (2.35) follows, since f itself is an admissible trial function for μ_N. The final statement is an evident consequence of the proof of (2.35). □

COROLLARY 2.3. *With $\varkappa_N$ and β_N defined as in* (2.26) *and* (2.27), *define*

$$\Delta_N = N(1-\lambda_N). \tag{2.37}$$

Then

$$\Delta_N \geqslant (1-\max\{\varkappa_N, \beta_N\})\Delta_{N-1} \tag{2.38}$$

for all $N\geqslant 3$, and hence for all $N>2$,

$$\Delta_N \geqslant \prod_{j=3}^{N}(1-\max\{\varkappa_j, \beta_j\})\Delta_2. \tag{2.39}$$

Proof. This follows directly from (2.28), (2.35) and (2.37). □

We see that a sufficient condition for $\liminf_{N\to\infty}\Delta_N>0$ is $\Delta_2>0$ and

$$\prod_{N=3}^{\infty}(1-\max\{\varkappa_N, \beta_N\}) > 0.$$

Assuming that $\max\{\varkappa_N, \beta_N\}<1$ for all $N>3$, this last condition is of course satisfied whenever

$$\sum_{N=3}^{\infty}\max\{\varkappa_N, \beta_N\} < \infty.$$

3. Analysis of the Kac model

The Kac model, with $(X_N, \mathcal{S}_N, \mu_N)$ being S^{N-1} equipped with its rotation-invariant probability measure and

$$Qf(v_1, v_2, ..., v_N) = \binom{N}{2}^{-1} \sum_{i<j}^{N} \int_{-\pi}^{\pi} \varrho(\theta) f(v_1, v_2, ..., v_i^*(\theta), ..., v_j^*(\theta), ..., v_n)\, d\theta,$$

was the basic motivating example for the definition of a Kac system made in the previous section, where the rest of the elements of the system, namely the action of Π_N, the spaces $(Y_N, \mathcal{T}_N, \nu_N)$, and the maps π_j and ϕ_j, have all been specified.

All that remains to be done before we apply the results of §2 is to compute the spectrum of K. There are a number of ways that this can be done. The method presented here is the one that most readily adapts to the case of three-dimensional momentum-conserving collisions, which we treat in the next section. In a later section we shall use a more group-theoretic approach when we discuss the generalization of the Kac walk to $SO(N)$.

THEOREM 3.1. *There is a complete orthonormal set $\{g_n\}$, $n \geqslant 0$, of eigenfunctions of K where g_n is a polynomial of degree n and the corresponding eigenvalue α_n is zero if n is odd, and if $n=2k$, α_n is given by*

$$\alpha_{2k} = (-1)^k \frac{|S^{N-3}|}{|S^{N-2}|} \int_0^{\pi} (1-\sin^2(\theta))^k \sin^{N-3}(\theta)\, d\theta. \tag{3.1}$$

In particular,

$$\begin{aligned} \alpha_2 &= -\frac{1}{N-1}, \\ \alpha_4 &= \frac{3}{N^2-1}, \\ \alpha_6 &= -\frac{15}{(N-1)(N+1)(N+3)}, \\ \alpha_8 &= \frac{105}{(N-1)(N+1)(N+3)(N+5)} \end{aligned} \tag{3.2}$$

and $|\alpha_{2k+2}| < |\alpha_{2k}|$ for all k. Hence for the Kac model,

$$\max\{\varkappa_N, \beta_N\} = \varkappa_N = \frac{3}{N^2-1}.$$

Proof. We have already deduced an explicit form (2.25) for K in the previous section. We note that by an obvious change of variable, we may rewrite it as

$$Kg(v) = \frac{|S^{N-3}|}{|S^{N-2}|} \int_0^{\pi} g\big(\sqrt{1-v^2}\cos(\theta)\big) \sin^{N-3}(\theta)\, d\theta.$$

The right-hand side is clearly an even function of v. The operator K evidently annihilates all odd functions. Hence we may assume that g is even.

Further, since $\left(\sqrt{1-v^2}\right)^{2k}=(1-v^2)^k$ is a polynomial of degree $2k$ in v, we see that the space of polynomials of degree $2n$ or less is invariant under K for all n. This implies that the eigenvectors are even polynomials, and that there is exactly one such eigenvector for each degree $2k$.

Now let g_{2k} be the eigenvector that is a polynomial of degree $2k$, and let α_{2k} be the corresponding eigenvalue. We may normalize g_{2k} so that the leading coefficient is 1, and we then have

$$g_{2k}=v^{2k}+h(v)$$

where $h(v)$ is an even polynomial in v of degree no more than $2k-2$. Thus

$$\alpha_{2k}v^{2k}+\alpha_{2k}h(v)=\alpha_{2k}g_{2k}=Kg_{2k}=Kv^{2k}+Kh(v).$$

This implies that

$$Kv^{2k}=\alpha_{2k}v^{2k}+\text{lower order.}$$

The identity (3.1) now follows directly from the formula for K, the recurrence relation

$$\int_0^\pi \sin^n(\theta)\,d\theta=\frac{n-1}{n}\int_0^\pi \sin^{n-2}(\theta)\,d\theta, \tag{3.3}$$

and the fact that $K1=1$. Observe that the leading coefficient of v in $(1-v^2)^k$ is $(-1)^k$.

The final, and crucial, point is the monotonicity $|\alpha_{2k+2}|<|\alpha_{2k}|$ for all k. Without such a monotonicity property, it can be very difficult to identify the spectral gap even if one has an explicit formula for each of the eigenvalues. We will encounter such a problem in §4. Here, we are fortunate:

$$(1-\sin^2(\theta))^k>(1-\sin^2(\theta))^{k+1}$$

for all k and almost all θ. From this and (3.1), the assertion easily follows. □

It is evident from (3.2) that, using the notation of Theorem 2.1, $\varkappa_N=3/(N^2-1)$ and $\beta_N=1/(N-1)^2$. Hence, for $N\geqslant 3$, $\varkappa_N>\beta_N$, and Theorems 1.1 and 1.2 are now proved.

In order to prove Theorem 1.3, it is necessary to determine Δ_2. But in (1.24) we have already determined λ_2, and since $\Delta_2=2(1-\lambda_2)$, it follows that

$$\Delta_2=2\inf_{k\neq 0}\left\{\int_{-\pi}^{\pi}(1-\cos(k\theta))\,\varrho(\theta)\,d\theta\right\}.$$

By the Riemann–Lebesgue lemma, $\lambda_2<1$, and so in any case $\Delta_2>0$. In the case Kac considered, Q is just the projection onto the constants and $\lambda_2=0$ so that $\Delta_2=2$.

It remains to solve the recurrence relation (1.20). Notice that

$$1-\varkappa_N = \frac{(N-2)(N+2)}{(N-1)(N+1)}. \tag{3.4}$$

The product of these terms collapses and

$$\prod_{j=3}^{N} \frac{(j-2)(j+2)}{(j-1)(j+1)} = \frac{1}{4}\,\frac{N+2}{N-1}. \tag{3.5}$$

It then follows from (2.39) of Corollary 2.3 that

$$\Delta_N \geqslant \frac{1}{4}\,\frac{N+2}{N-1}\,\Delta_2 = \frac{1}{4}\,\frac{N+2}{N-1}\,2(1-\lambda_2) = \frac{1-\lambda_2}{2}\,\frac{N+2}{N-1}. \tag{3.6}$$

Now, we inquire into the sharpness of this result. By Theorems 2.1 and 3.1,

$$Pf_N = \mu_N f_N \tag{3.7}$$

if and only if f_N has the form $f_N = \sum_{j=1}^{N} g_N \circ \pi$ and $Kg_N = (3/(N^2-1))g_N$. That is, (3.7) holds exactly when, up to a multiple,

$$f_N(\vec{v}) = \sum_{j=1}^{N} (v_j^4 - \langle 1, v_j^4 \rangle).$$

(When doing the computation, bear in mind that on the sphere, any multiple of $\sum_{j=1}^{N} v_j^2$ is a constant.) By the last part of Theorem 2.2, the bound obtained in Theorem 1.3 can only be sharp if $Qf_N = \lambda_N f_N$ for each N. Hence it is natural to compute Qf_N. The result is contained in the next lemma.

LEMMA 3.2. *For* $f_N(\vec{v}) = \sum_{j=1}^{N} (v_j^4 - \langle 1, v_j^4 \rangle)$,

$$Qf_N = \left(1 - \frac{2\gamma(N+2)}{N(N-1)}\right) f_N \tag{3.8}$$

where

$$\gamma = \frac{1}{4}\left(1 - \int_{-\pi}^{\pi} \cos(4\theta)\varrho(\theta)\,d\theta\right). \tag{3.9}$$

Proof. This is a straightforward calculation. □

Clearly, for the original Kac model, with ϱ uniform, $\gamma = \frac{1}{4}$, and so (3.8) implies that $\Delta_N = N(1-\lambda_N)$ is no larger than $(N+2)/2(N-1)$. Since for the original Kac model

$\lambda_2=0$, this upper bound on Δ_N coincides with the lower bound in (3.6), and hence (3.6) is sharp in this case.

In fact, the upper bound on Δ_N provided by Lemma 3.2 coincides with the lower bound in (3.6) whenever $f_2(v_1,v_2)=v_1^4+v_2^4-\frac{3}{4}$ is such that

$$Qf_2=\lambda_2 f_2. \tag{3.10}$$

Writing $v_1=\cos(\theta)$ and $v_2=\sin(\theta)$, we have

$$f(\cos(\theta),\sin(\theta))=\tfrac{1}{4}\cos(4\theta).$$

Hence (3.10) certainly holds whenever (1.27) holds. Finally, the fact that under the condition (1.27), f_N is, up to a multiple, the only eigenfunction of Q with eigenvalue λ_N follows directly from Theorem 3.1, which says that $\varkappa_N$ has multiplicity one, and Theorem 2.1. This completes the proof of Theorem 1.3.

We shall show in §7 of this paper that actually in a wide range of circumstances,

$$Qf_N=\lambda_N f_N$$

for all N sufficiently large, even if this is false for, say, $N=2$. Thus in a great many cases Lemma 3.2 provides the precise value of λ_N, and hence Δ_N, for large N. However, before returning to analyze the Kac model in this detail, we proceed to give several more examples of Kac systems.

Having explained how our exact determination of the gap for Kac's original model works it is appropriate to compare this approach with Janvresse's [9] application of Yau's martingale method [13], [14] to the same problem. There are similarities between our analysis and Yau's method, in that Yau's martingale method uses induction on N, correlation estimates, and the same conditional expectation operators P_j. There are, however, significant differences, as indicated by the difference between Janvresse's estimate and our exact calculation.

First, in Yau's method the spectrum of the operators P_j is estimated not in $\mathcal{H}_{N,E}$, but in the Hilbert space whose inner product is $\langle h,(I-Q)h\rangle$, the so-called Dirichlet form space associated to Q. This means that the details of the dynamics enter (through Q) at each stage of the induction, while in our approach purely geometric estimates, as described in Theorem 1.1, relate Δ_N to Δ_{N-1}.

Second, Yau's method was designed to handle problems without the permutation symmetry that is present in the class of models considered here. The method just described makes full use of this symmetry. As an example, using this symmetry, we need only to produce spectral estimates on P, the average of the P_j. That the inductive argument presented here makes full use of this permutation symmetry is one source of its incisiveness in this class of problems.

4. Analysis of the Boltzmann collision model

Consider now a pair of identical particles with velocities v_i and v_j in $\mathbf{R}^3$. Now we will require that the collisions conserve momentum as well as energy. These are four constraints on six variables, and hence the set of all kinematically possible collisions is two-dimensional. It may be identified with S^2 as follows: For any unit vector ω in S^2, define

$$v_i^*(\omega) = v_i + (\omega \cdot (v_j - v_i))\,\omega, \tag{4.1}$$

$$v_j^*(\omega) = v_j - (\omega \cdot (v_j - v_i))\,\omega. \tag{4.2}$$

Now specify N velocities $\vec{v} = (v_1, v_2, ..., v_N)$ before the collision with

$$\sum_{j=1}^{N} |v_j|^2 = E \quad \text{and} \quad \sum_{j=1}^{N} v_j = 0. \tag{4.3}$$

The random collision mechanism is now that we pick a pair i, j, $i<j$, uniformly at random, and then pick an ω in S^2 at random, and the post-collisional velocities then become

$$((v_1, ..., v_i^*(\omega), ..., v_j^*(\omega), ..., v_N).$$

We then define the one-step transition operator Q by

$$Q_f(\vec{v}) = \binom{N}{2}^{-1} \sum_{i<j}^{N} \int_{S^2} f(v_1, v_2, ..., v_i^*(\omega), ..., v_j^*(\omega), ..., v_n)\, b(\omega \cdot (v_i - v_j)/|v_i - v_j|)\, d\omega, \tag{4.4}$$

where b is a non-negative function on $[-1, 1]$ so that

$$2\pi \int_0^{\pi} b(\cos(\theta)) \sin(\theta)\, d\theta = 1.$$

The function b puts a weight on the choice of ω so as to determine the relative likelihood of various scattering angles. This definition differs from the corresponding definition for the Kac model chiefly through the more complicated formulas (4.1) and (4.2) parameterizing three-dimensional momentum-conserving collisions. We begin the analysis of this Boltzmann collision model by specifying the structure needed to display it as a Kac system.

By choice of scales and coordinates, we may assume that

$$\sum_{j=1}^{N} |v_j|^2 = 1 \quad \text{and} \quad \sum_{j=1}^{N} v_j = 0 \tag{4.5}$$

both hold initially, and hence for all time. Thus our state space X_N is the set of all vectors

$$\vec{v} = (v_1, v_2, ..., v_N) \in \mathbf{R}^{3N}$$

satisfying the constraints in (4.5). We equip X_N with its Borel field and the metric and uniform probability measure inherited from its natural embedding in $\mathbf{R}^{3N}$. The symmetric group Π_N acts on X_N as follows: for $\sigma \in \Pi_N$,

$$\sigma(v_1, v_2, ..., v_N) = (v_{\sigma(1)}, v_{\sigma(2)}, ..., v_{\sigma(N)}).$$

This action is clearly measure preserving. We note that X_N is geometrically equivalent to the unit sphere S^{3N-4} in $\mathbf{R}^{3N-3}$, but apart from identifying normalization factors in our probability measures, this identification is not conducive to efficient computation because any embedding in $\mathbf{R}^{3N-3}$ obscures the action of the symmetric group.

To identify the single-particle state space Y_N, note that

$$\sup\{|v_N|^2 \mid (v_1, v_2, ..., v_N) \in X_N\} = \frac{N-1}{N}. \tag{4.6}$$

To see this, fix v_N and observe that $\sum_{j=1}^{N-1} v_j = -v_N$ due to the momentum constraint in (4.5). To maximize $|v_N|$, we must minimize the energy in the first $N-1$ particles. However, by convexity it is clear that

$$\inf\left\{ \sum_{j=1}^{N-1} |v_j|^2 \;\middle|\; \sum_{j=1}^{N-1} v_j = -v_N \right\}$$

is attained at

$$(v_1, v_2, ..., v_{N-1}) = -\frac{1}{N-1}(v_N, v_N, ..., v_N),$$

which leads directly to (4.6).

In short, the momentum constraint prevents all of the energy from belonging to a single particle, and so each v_j lies in the ball of radius $\sqrt{(N-1)/N}$ in $\mathbf{R}^3$. (While this is true for $N=2$, this case is somewhat special. For $N=2$, $v_2 = -v_1$ and so $|v_2| = 1/\sqrt{2}$, rather than $|v_2| \leqslant 1/\sqrt{2}$.)

We could take Y_N to be the ball of radius $\sqrt{(N-1)/N}$ in $\mathbf{R}^3$, for $N \geqslant 3$, which would then depend on N. However, certain calculations will work out more simply if we rescale and take Y_N to be the unit ball in $\mathbf{R}^3$, independent of N. Therefore, we define, for $N \geqslant 3$,

$$Y_N = \{v \in \mathbf{R}^3 \mid |v| \leqslant 1\}$$

and let $\mathcal{T}_N$ be the corresponding Borel field. We take Y_2 to be the unit sphere in $\mathbf{R}^3$. We are then led to define $\pi_j: X_N \to Y_N$ by

$$\pi_j(v_1, v_2, ..., v_N) = \left(\frac{N}{N-1}\right)^{1/2} v_j. \tag{4.7}$$

The measure ν_N is now determined through (2.3), but before deducing an explicit formula for it, we introduce the maps $\phi_j: X_{N-1} \times Y_N \to X_N$, through which this formula is readily determined.

Consider any fixed $N \geqslant 3$, so that X_{N-1} is non-empty. Fix a point

$$\vec{w} = (w_1, w_2, ..., w_{N-1}) \in X_{N-1},$$

and a point $v \in Y_N$. In order that we have

$$\pi_N(\phi_N(\vec{w}, v)) = v,$$

the Nth component of $\phi_N(\vec{w}, v)$ must be $\sqrt{(N-1)/N}\, v$. Now observe that for any $\alpha \in \mathbf{R}$,

$$\vec{v} = (v_1, v_2, ..., v_N) = \left(\alpha w_1 - \frac{1}{\sqrt{N^2-N}} v, ..., \alpha w_{N-1} - \frac{1}{\sqrt{N^2-N}} v, \sqrt{\frac{N-1}{N}}\, v\right)$$

satisfies $\sum_{j=1}^N v_j = 0$, and

$$\sum_{j=1}^N |v_j|^2 = \alpha^2 + |v|^2,$$

since $\sum_{j=1}^{N-1} |w_j|^2 = 1$ and $\sum_{j=1}^{N-1} w_j = 0$. Therefore, define

$$\alpha^2(v) = 1 - |v|^2 \tag{4.8}$$

and

$$\begin{aligned} &\phi_N((w_1, w_2, ..., w_{N-1}), v) \\ &\quad = \left(\alpha(v) w_1 - \frac{1}{\sqrt{N^2-N}} v, ..., \alpha(v) w_{N-1} - \frac{1}{\sqrt{N^2-N}} v, \sqrt{\frac{N-1}{N}}\, v\right), \end{aligned} \tag{4.9}$$

and we have that $\phi_N: X_{N-1} \times Y_N \to X_N$. For $j = 1, ..., N-1$, let $\sigma_{j,N}$ be the pair permutation exchanging j and N, and define $\phi_j = \sigma_{j,N} \circ \phi_N$. We now show that with these definitions (2.9) holds, and in the process, obtain an explicit formula for ν_N.

LEMMA 4.1. *For $N \geqslant 3$, the measure ν_N induced on Y_N through* (2.3) *for the Boltzmann collision model is*

$$d\nu_N(v) = \frac{|S^{3N-7}|}{|S^{3N-4}|}(1-|v|^2)^{(3N-8)/2}\,dv. \tag{4.10}$$

In the case $N=2$, ν_2 is the uniform probability measure on $S^2=Y_2$. Moreover, for these measures ν_N, and with ϕ_j defined as above, (2.9) *holds for the Boltzmann collision model for all $N \geqslant 3$.*

Proof. The measure μ_N is defined through the natural embedding of X_N in $\mathbf{R}^{3N}$, and hence it is advantageous to consider the tangent spaces to X_N as subspaces of $\mathbf{R}^{3N}$. Making this identification, a vector $\vec{\xi}=(\xi_1, \xi_2, ..., \xi_N)$ is tangent to X_N at $\vec{v}=(v_1, v_2, ..., v_N) \in X_N$ provided

$$\sum_{j=1}^{N} \xi_j \cdot v_j = 0 \quad \text{and} \quad \sum_{j=1}^{N} \xi_j = 0. \tag{4.11}$$

Likewise, a vector $\vec{\eta}=(\eta_1, \eta_2, ..., \eta_{N-1})$ is tangent to X_{N-1} at $\vec{w}=(w_1, w_2, ..., w_{N-1}) \in X_{N-1}$ provided

$$\sum_{j=1}^{N-1} \eta_j \cdot w_j = 0 \quad \text{and} \quad \sum_{j=1}^{N-1} \eta_j = 0. \tag{4.12}$$

And finally, it is clear that the tangent space at any point v of Y_N is $\mathbf{R}^3$.

Now let

$$(\phi_N)_*: T_*(X_{N-1}) \times T_*(Y_N) \to T_*(X_N)$$

be the tangent bundle map induced by ϕ_N. One easily computes the derivatives and finds that for a tangent vector $(\vec{\eta}, 0)$ at $(\vec{w}, v)$,

$$(\phi_N)_*(\vec{\eta}, 0) = (\alpha(v)\vec{\eta}, 0). \tag{4.13}$$

Likewise, for a tangent vector $(\vec{0}, u)$ at $(\vec{w}, v)$,

$$\begin{aligned}(\phi_N)_*(\vec{0}, u) \\ = \left(-\frac{v \cdot u}{\alpha(v)} w_1 - \frac{1}{\sqrt{N^2-N}} u, ..., -\frac{v \cdot u}{\alpha(v)} w_{N-1} - \frac{1}{\sqrt{N^2-N}} u, \sqrt{\frac{N-1}{N}}\, u\right).\end{aligned} \tag{4.14}$$

Now let ζ_X be any vector of the type in (4.13), and let ζ_Y be any vector of the type in (4.14). Obviously

$$\langle \zeta_X, \zeta_Y \rangle = 0 \tag{4.15}$$

where the inner product is the standard inner product in $\mathbf{R}^{3N}$. Moreover,

$$\langle \zeta_X, \zeta_X \rangle = \alpha^2(v) \langle \vec{\eta}, \vec{\eta} \rangle \tag{4.16}$$

where the inner product on the right is the standard one in $\mathbf{R}^{3N-3}$. The determinant of the matrix corresponding to the quadratic form q_X given by

$$\vec{\eta} \mapsto \alpha^2(v) \langle \vec{\eta}, \vec{\eta} \rangle$$

is

$$\det(q_X) = \alpha^{2(3N-7)}(v) \tag{4.17}$$

since X_{N-1} is $(3N-7)$-dimensional. Finally,

$$\langle \zeta_Y, \zeta_Y \rangle = \frac{(v \cdot u)^2}{\alpha^2(v)} + |u|^2, \tag{4.18}$$

and the determinant of the matrix corresponding to the quadratic form q_Y given by

$$u \mapsto \frac{(v \cdot u)^2}{\alpha^2(v)} + |u|^2$$

is

$$\det(q_Y) = 1 + \frac{|v|^2}{\alpha^2(v)} = \frac{1}{\alpha^2(v)}. \tag{4.19}$$

Now let $\tilde{\mu}_N$ and $\tilde{\mu}_{N-1}$ denote the *unnormalized* measures on X_N and X_{N-1} given by the Riemannian structures induced by their natural Euclidean embeddings. If $(x_1, \ldots, x_{3N-7})$ is any set of coordinates for X_{N-1}, and if (y_1, y_2, y_3) are the obvious Euclidean coordinates for Y_N, then these induce, through ϕ_N, a system of coordinates on X_N. (Since X_{N-1} is a sphere, up to a set of measure zero, one chart of coordinates suffices.) The volume element $d\tilde{\mu}_N(x, y)$ in these coordinates can now be expressed in terms of the volume element $d\tilde{\mu}_{N-1}(x)$ using (4.15), (4.17) and (4.19):

$$d\tilde{\mu}_N(x, y) = \alpha^{3N-7}(v)\, d\tilde{\mu}_{N-1}(x) \frac{1}{\alpha(v)}\, dy.$$

Since we know that

$$\int_{X_N} d\tilde{\mu}_N = |S^{3N-4}|,$$

we easily deduce from this that for all continuous functions f on X_N,

$$\int_{X_N} f(v)\, d\mu_N = \frac{|S^{3N-7}|}{|S^{3N-4}|} \int_{Y_N} \left[\int_{X_{N-1}} f \circ \phi_N\, d\mu_{N-1} \right] (1-|v|^2)^{(3N-8)/2}\, dv. \tag{4.20}$$

Finally, suppose that f has the form $f = g \circ \pi_N$ for some continuous function g on Y_N, $N \geqslant 3$. Then evidently $f \circ \phi_N(\vec{w}, v) = g(v)$ everywhere on $X_{N-1} \times Y_N$, and hence by the definition (2.3) and (4.20),

$$\int_{Y_N} g\, d\nu_N = \int_{X_N} f\, d\mu_N = \frac{|S^{3N-7}|}{|S^{3N-4}|} \int_{Y_N} g(v)(1-|v|^2)^{(3N-8)/2}\, dv. \tag{4.21}$$

Hence we see that (4.10) holds, and hence that (2.9) holds for the Boltzmann collision model. □

LEMMA 4.2. *The Boltzmann collision model consisting of*

$$(X_N, \mathcal{S}_N, \mu_N),\ (Y_N, \mathcal{T}_N, \nu_N),\ \pi_j,\ \phi_j, \quad j=1,\dots,N,$$

and Q as specified in this section constitute a Kac system as defined in §2.

Proof. The properties not already established in Lemma 4.1 are now easily checked using (4.9). □

Now in order to apply the results of §2 to this Kac system, we need to determine the spectral properties of the operator K. The explicit form of K for the Boltzmann collision model is easily obtained from (2.24): For all functions g on Y_N, the unit ball in in $\mathbf{R}^3$, and all $N>3$,

$$\begin{aligned} Kg(v) &= \int_{X_{N-1}} g\left(\sqrt{\frac{N}{N-1}}\left(\sqrt{1-|v|^2}\,w_{N-1} - \frac{1}{\sqrt{N^2-N}}v\right)\right) d\mu_{N-1}(w) \\ &= \int_{Y_{N-1}} g\left(\sqrt{\frac{N}{N-1}}\sqrt{1-|v|^2}\sqrt{\frac{N-2}{N-1}}\,y - \frac{1}{N-1}v\right) d\nu_{N-1}(y) \qquad (4.22) \\ &= \frac{|S^{3N-10}|}{|S^{3N-7}|}\int_{|y|\leqslant 1} g\left(\frac{\sqrt{N^2-2N}}{N-1}\sqrt{1-|v|^2}\,y - \frac{1}{N-1}v\right)(1-|y|^2)^{(3N-11)/2}\,dy. \end{aligned}$$

(The restriction to $N>3$ is because (4.21) only gives us the right form for ν_N in this range. Indeed, $3N-11$ is negative for $N=3$. The correct analogs of (4.21) and (4.22) are easily worked out by the same sort of analysis. We do not do this here, as we do not need these formulas.)

Several properties of K are evident from (4.22). First, K commutes with rotations in $\mathbf{R}^3$. That is, if $R:\mathbf{R}^3\to\mathbf{R}^3$ is a rotation, then clearly

$$K(g\circ R) = (Kg)\circ R.$$

Because K commutes with rotations, we may restrict our search for eigenfunctions g of K to functions of the form

$$g(v) = h(|v|)|v|^l \mathcal{Y}_{l,m}(v/|v|)$$

for some function h on $\mathbf{R}_+$, and some spherical harmonic $\mathcal{Y}_{l,m}$.

Second, for each $n\geqslant 0$, K preserves the space of polynomials of degree n or less. To see this, notice that any monomial in $\sqrt{1-|v|^2}\,y$ that is of odd degree is annihilated when integrated against $(1-|y|^2)^{(3N-11)/2}\,dy$, and any even monomial in $\sqrt{1-|v|^2}\,y$ is a polynomial in v.

Combining these two observations, we see that K has a complete basis of eigenfunctions of the form

$$g_{n,l,m}(v) = h_{n,l}(|v|^2)|v|^l \mathcal{Y}_{l,m}(v/|v|) \tag{4.23}$$

where $h_{n,l}$ is a polynomial of degree n.

A third observation leads to an explicit identification of these polynomials and a formula for the eigenfunctions: Suppose that $Kg(v)=\lambda g(v)$. Let $\hat{e}$ be any unit vector in $\mathbf{R}^3$. Then since g is a polynomial and hence continuous,

$$\lim_{t\to 1} Kg(t\hat{e}) = \lim_{t\to 1} \frac{|S^{3N-10}|}{|S^{3N-7}|} \int_{Y_{N-1}} g\left(\frac{\sqrt{N^2-2N}}{N-1}\sqrt{1-t^2}\, w - \frac{t}{N-1}\hat{e}\right)(1-|w|^2)^{(3N-11)/2}\, dw$$

$$= g\left(-\frac{1}{N-1}\hat{e}\right),$$

since $K1=1$. Combining this with $Kg(v)=\lambda g(v)$, we have

$$\lambda g(\hat{e}) = g\left(-\frac{1}{N-1}\hat{e}\right). \tag{4.24}$$

Now consider any eigenfunction $g_{n,l,m}$ of the form given in (4.23), and let $\lambda_{n,l}$ be the corresponding eigenvalue, which will not depend on m. Then taking any $\hat{e}$ so that $\mathcal{Y}_{l,m}(\hat{e})\neq 0$, we have from (4.24) that

$$\lambda_{n,l} = \frac{h_{n,l}(1/(N-1)^2)}{h_{n,l}(1)}\left(-\frac{1}{N-1}\right)^l. \tag{4.25}$$

Finally, a fourth elementary observation identifies the polynomials $h_{n,l}$. For all distinct positive integers n and p, the eigenfunctions $g_{n,l,m}$ and $g_{p,l,m}$ are orthogonal in $\mathcal{K}_N$. Hence for each l, and for $n\neq p$,

$$\int_{|v|\leqslant 1} h_{n,l}(|v|^2)h_{p,l}(|v|^2)(1-|v|^2)^{(3N-8)/2}|v|^{2l}\, dv = 0.$$

Taking $r=|v|^2$ as a new variable, we have

$$\int_0^1 h_{n,l}(r)h_{p,l}(r)(1-r)^{(3N-8)/2} r^{l+1/2}\, dr = 0.$$

This is the orthogonality relation for a family of Jacobi polynomials in one standard form, and this identifies the polynomials $h_{n,l}$. A more common standard form, and one that is used in the sources to which we shall refer, is obtained by the change of variable $t=2r-1$, so that t ranges over the interval $[-1,1]$. Then for $\alpha,\beta>-1$, $J_n^{(\alpha,\beta)}(t)$ is the

orthogonal nth-degree polynomial for the weight $(1-t)^{\alpha}(1+t)^{\beta}$. Then with the variables t and $|v|^2$ related as above, i.e.,

$$t=2|v|^2-1, \tag{4.26}$$

we have

$$h_{n,l}(|v|^2)=J_n^{(\alpha,\beta)}(t) \tag{4.27}$$

for

$$\alpha=\tfrac{1}{2}(3N-8) \quad \text{and} \quad \beta=l+\tfrac{1}{2}. \tag{4.28}$$

The particular normalization of the Jacobi polynomials is irrelevant here, as we shall be concerned with ratios of the form $J_n^{(\alpha,\beta)}(t)/J_n^{(\alpha,\beta)}(1)$. Indeed, notice that from (4.26) when $|v|^2=1$, $t=1$, and when $|v|^2=1/(N-1)^2$, $t=-1+2/(N-1)^2$. Hence from (4.27) and (4.25), we see that

$$\lambda_{n,l}=\frac{J_n^{(\alpha,\beta)}(-1+2/(N-1)^2)}{J_n^{(\alpha,\beta)}(1)}\left(-\frac{1}{N-1}\right)^l. \tag{4.29}$$

We summarize this in the following lemma:

LEMMA 4.3. *Define the functions*

$$g_{n,l,m}(v)=h_{n,l}(|v|^2)|v|^l\mathcal{Y}_{l,m}(v/|v|),$$

$n\geqslant 0$, $l\geqslant 0$ *and* $-l\leqslant m\leqslant l$, *where the* $\mathcal{Y}_{l,m}$ *are an orthonormal family of spherical harmonics, and the* $h_{n,l}$ *are polynomials expressible in terms of the Jacobi polynomials through* (4.27). *Then*

$$\{g_{n,l,m}\,|\,n\geqslant 0,\ l\geqslant 0,\ -l\leqslant m\leqslant l\}$$

is a complete orthonormal basis of eigenfunctions of K. *Moreover, if* $\lambda_{n,l}$ *is the corresponding eigenvalue, then* (4.29) *holds.*

The problem of determining the spectral gap for K is thus reduced to the problem of determining the largest number of the form (4.29). The following integral representation of ratios of Jacobi polynomials, due to Koornwinder [11] (see also [1, pp. 31 ff.]), is useful in this regard.

For all $-1\leqslant x\leqslant 1$, all n and all $\alpha>\beta$,

$$\frac{J_n^{(\alpha,\beta)}(x)}{J_n^{(\alpha,\beta)}(1)}=\int_0^{\pi}\int_0^1\left[\tfrac{1}{2}(1+x-(1-x)r^2)+i\sqrt{1-x^2}\,r\cos(\theta)\right]^n dm_{\alpha,\beta}(r,\theta) \tag{4.30}$$

where

$$m_{\alpha,\beta}(r,\theta)=c_{\alpha,\beta}(1-r^2)^{\alpha-\beta-1}r^{2\beta+1}(\sin(\theta))^{2\beta}\,dr\,d\theta,$$

and $c_{\alpha,\beta}$ is a normalizing constant that makes $dm_{\alpha,\beta}$ a probability measure.

Notice from (4.28) that $\alpha>\beta$ exactly when $2l<3N-9$. Hence we define

$$l_0=\tfrac{1}{2}(3N-9). \tag{4.31}$$

For all $l<l_0$, we may use (4.30) to compute $\lambda_{n,l}$.

First, however, observe that

$$\begin{aligned}\left|\tfrac{1}{2}(1+x-(1-x)r^2)+i\sqrt{1-x^2}\,r\cos(\theta)\right|^2 \\ =\tfrac{1}{4}(1+x)^2+\tfrac{1}{4}(1-x)^2r^4+\tfrac{1}{2}(1-x^2)r^2\cos(2\theta)\leqslant 1,\end{aligned} \tag{4.32}$$

with equality exactly when $r=1$, and $\theta=0$ or π.

LEMMA 4.4. *For all $l<l_0(N)$, and all m,*

$$|\lambda_{n,l}|<\mu_{n,l}$$

where, with $x=-1+2/(N-1)^2$,

$$\mu_{n,l}=\int_0^\pi\int_0^1\left|\tfrac{1}{2}(1+x-(1-x)r^2)+i\sqrt{1-x^2}\,r\cos(\theta)\right|^n dm_{\alpha,\beta}(r,\theta)\left(\frac{1}{N-1}\right)^l.$$

Moreover, for each l, $n\mapsto\mu_{n,l}$ is monotone decreasing:

$$\mu_{k,l}<\mu_{j,l}\quad \textit{for all } k>j.$$

Proof. The monotonicity follows directly from (4.32), and the rest is a summary of the discussion above. □

We now proceed to calculate the eigenvalues for $l+\frac{1}{2}<\frac{1}{2}(3N-8)$ using (4.30) and (4.29). The case $n=0$ is trivial:

$$\lambda_{0,l}=(-1)^l\mu_{0,l}=\left(-\frac{1}{N-1}\right)^l \tag{4.33}$$

for all $l<l_0$. The monotonicity in Lemma 4.4 now guarantees that for all n and all $3\leqslant l<l_0$,

$$|\lambda_{n,l}|\leqslant\left(\frac{1}{N-1}\right)^3. \tag{4.34}$$

Next, it is straightforward to calculate $\lambda_{1,l}$ and $\lambda_{2,l}$ using (3.3) and the beta integral

$$\int_0^1(1-t)^{a-1}t^{b-1}\,dt=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

The results for $n=1$ and $n=2$ are

$$\lambda_{1,l} = \left[\varepsilon - (1-\varepsilon)\frac{2l+3}{3N-6}\right]\left(-\frac{1}{N-1}\right)^l \tag{4.35}$$

and

$$\lambda_{2,l} = \left[\varepsilon^2 - \frac{4l+10}{3N-6}\varepsilon(1-\varepsilon) + (1-\varepsilon)^2\frac{(2l+5)(2l+3)}{(3N-6)(3N-4)}\right]\left(-\frac{1}{N-1}\right)^l, \tag{4.36}$$

where

$$\varepsilon = \frac{1}{(N-1)^2}. \tag{4.37}$$

The eigenvalues

$$\lambda_{1,l} = -\left[\frac{1}{N-1} + \frac{2lN}{3(N-1)^2}\right]\left(-\frac{1}{N-1}\right)^l$$

are negative for l even, and hence irrelevant for calculating the gap of K. For N odd, they are no larger than $1/(N-1)^2$. Note that $\lambda_{2,0}$ is asymptotically $5/3N^2$ and otherwise $\lambda_{2,l}$ is $O(1/N^3)$.

Finally, a very simple computation provides a constant C independent of N so that

$$\mu_{3,0} \leqslant \frac{C}{N^3}, \quad \mu_{2,1} \leqslant \frac{C}{N^3} \quad \text{and} \quad \mu_{1,2} \leqslant \frac{C}{N^3}. \tag{4.38}$$

(The cases with n odd are most easily done through estimates on cases with even n, for example, since $\mu_{3,l} \leqslant \mu_{4,l}^{3/4}$ by Hölder's inequality, it suffices to show that $\mu_{4,l} = \mathcal{O}(1/N^4)$.) Therefore, again using the monotonicity from Lemma 4.4, the only values of (n,l) with $l<l_0$ such that $|\lambda_{n,l}|$ is of order $1/N^2$ or larger are those for which $n+l \leqslant 2$. By the computations above, we then have

$$\sup\{\lambda_{n,l} \mid n+l>0, l<l_0\} = \lambda_{2,0} \tag{4.39}$$

and

$$\inf\{\lambda_{n,l} \mid l<l_0\} = \lambda_{1,0}. \tag{4.40}$$

Regarding the restriction $l<l_0$ in (4.39) and (4.40), it is reassuring to note that for the largest value of l in this range, the corresponding eigenvalues are no larger than $(1/(N-1))^{(3N-11)/2}$. This suggests that a fairly crude bound on the part of the spectrum corresponding to $l \geqslant l_0$ will suffice to eliminate the restriction on l in (4.39) and (4.40). We shall show that this is the case.

For this purpose we need the integral kernel corresponding to the operator K. From (4.22) we have that for all $g \in \mathcal{K}_N$,

$$\langle g, Kg\rangle_{\mathcal{K}_N} = \frac{|S^{3N-10}|}{|S^{3N-4}|}\int_{|y|\leqslant 1}\int_{|v|\leqslant 1} g(v)g(u(y,v))(1-|y|^2)_+^{(3N-11)/2}(1-|v|^2)_+^{(3N-8)/2}\,dy\,dv \tag{4.41}$$

where

$$u(y,v)=\frac{\sqrt{N^2-2N}}{N-1}\sqrt{1-|v|^2}\,y-\frac{1}{N-1}v.$$

Here, $(x)_+$ denotes the positive part of the quantity x. Although the limits of integration make use of this notation redundant in (4.41), we are about to make a change of variables after which it is essential.

Under the change of variables $(y,v)\to(u(y,v),v)$ we find

$$1-|y|^2=\frac{1}{1-|v|^2}\left(1-\frac{N-1}{N}\left[\frac{N-1}{N-2}u^2+\frac{N-1}{N-2}v^2+\frac{2}{N-2}u\cdot v\right]\right),$$

and the Jacobian is such that

$$dy\,dv=\left(\frac{N^2-2N}{N^2-2N+1}\right)^{3/2}(1-|v|^2)^{3/2}\,du\,dv.$$

Introduce

$$p(u,v)=\frac{N-1}{N}\left[\frac{N-1}{N-2}u^2+\frac{N-1}{N-2}v^2+\frac{2}{N-2}u\cdot v\right].$$

Then, since the integrations in (4.41) can be extended to run over all of $\mathbf{R}^3$, but g is supported in the unit ball, this change of variables gives

$$\langle g,Kg\rangle_{\mathcal{K}_N}=\frac{|S^{3N-10}|}{|S^{3N-4}|}\left(\frac{N^2-2N}{N^2-2N+1}\right)^{3/2}\int_{|u|\leqslant 1}\int_{|v|\leqslant 1}g(v)g(u)(1-p(u,v))_+^{(3N-11)/2}\,du\,dv. \tag{4.42}$$

Introducing the kernel $K(u,v)$ by

$$K(u,v)=\frac{|S^{3N-10}|}{|S^{3N-4}|}\left(\frac{N^2-2N}{N^2-2N+1}\right)^{3/2}(1-p(u,v))_+^{(3N-11)/2}, \tag{4.43}$$

and letting $g_{n,l,m}$ be the normalized eigenfunctions introduced in Lemma 4.3,

$$\begin{aligned}\lambda_{n,l}&=\frac{1}{2l+1}\sum_{m=-l}^{l}\langle g_{n,l,m},Kg_{n,l,m}\rangle\\&=\int h(|u|)\,h(|v|)\left[\frac{1}{2l+1}\sum_{m=-l}^{l}\mathcal{Y}_{l,m}(u)\mathcal{Y}_{l,m}(v)\right]K(u,v)\,dv\,dw\\&=\int h_{n,l}(|u|^2)\,h_{n,l}(|v|^2)\,P_l((u\cdot v)/|u|\,|v|)\,K(v,w)\,dv\,dw\end{aligned} \tag{4.44}$$

where $P_l(\cos(\theta))$ is the Legendre polynomial of order l, and it is orthogonal to all other polynomials of degree strictly less than l.

Now suppose that $3N-11$ is even, which is the case when N is odd. Then, were it not for the the positive part taken in (4.43), $K(u,v)$ would be a polynomial of degree $\frac{1}{2}(3N-11)$ in $\cos(\theta)=(u\cdot v)/|u|\,|v|$. If $l\geqslant l_0$, then $l>\frac{1}{2}(3N-11)$, so for such values of l, the integral in (4.44) would vanish, and with it the eigenvalue $\lambda_{n,l}$.

The positive part in (4.43) prevents us from using this argument to conclude that $\lambda_{n,l}$ vanishes identically. However, we now show that for most values of $|u|$ and $|v|$, the positive part does not matter, and so $\lambda_{n,l}$ is negligibly small for $l\geqslant l_0$. To do this, introduce the quadratic forms

$$q_{\pm}(x,y)=\frac{N-1}{N}\left[\frac{N-1}{N-2}x^2+\frac{N-1}{N-2}y^2\pm\frac{2}{N-2}xy\right]$$

on $\mathbf{R}^2$. Notice that

$$1-q_+(|u|,|v|)\leqslant 1-p(u,v)\leqslant 1-q_-(|u|,|v|).$$

The eigenvalues of $q_{\pm}$ are, in both cases, $(N-1)/N$ and $(N-1)/(N-2)$. Hence

$$\frac{N-1}{N}(|u|^2+|v|^2)\leqslant q_{\pm}(|u|,|v|)\leqslant\frac{N-1}{N-2}(|u|^2+|v|^2). \tag{4.45}$$

Therefore,

$$1-\frac{N-1}{N-2}(|u|^2+|v|^2)\leqslant 1-p(u,v)\leqslant 1-\frac{N-1}{N}(|u|^2+|v|^2). \tag{4.46}$$

It follows that $(1-p(u,v))_+=0$ whenever $|u|^2+|v|^2\geqslant N/(N-1)$, and it follows that $(1-p(u,v))_+=1-p(u,v)$ whenever $|u|^2+|v|^2\leqslant(N-2)/(N-1)$. Define $A\subset\mathbf{R}^6$ by

$$A=\left\{(u,v)\in\mathbf{R}^6\;\middle|\;\frac{N-2}{N-1}<|u|^2+|v|^2<\frac{N}{N-1}\right\}. \tag{4.47}$$

Hence we may replace $K(u,v)$ in (4.44) by $K_A(u,v)=1_A(u,v)K(u,v)$.

To draw a favorable conclusion from this we need only take into account the reference measure ν_N. It is convenient to absorb its density into the eigenfunctions. Define

$$\tilde{g}_{n,l,m}(u)=\left(\frac{|S^{3N-7}|}{|S^{3N-4}|}\right)^{1/2}(1-|u|^2)^{(3N-8)/2}g_{n,l,m}(u)$$

so that

$$\int_{|u|\leqslant 1}|\tilde{g}_{n,l,m}(u)|^2\,du=\|g_{n,l,m}\|^2_{\mathcal{K}_N}=1.$$

Define $\tilde{K}_A(u,v)$ by

$$\tilde{K}_A(u,v)=\left(\frac{N^2-2N}{N^2-2N+1}\right)^{3/2}\frac{|S^{3N-10}|}{|S^{3N-7}|}\,\frac{(1-p(u,v))_+^{(3N-11)/2}}{(1-|v|^2)^{(3N-8)/4}(1-|u|^2)^{(3N-8)/4}}\,1_A(u,v).$$

Then going back to (4.44), we have whenever N is odd and $l \geqslant l_0$,

$$\lambda_{n,l} = \frac{1}{2l+1} \sum_{m=-l}^{l} \int_{|u|\leqslant 1} \int_{|v|\leqslant 1} \tilde{g}_{n,l,m}(u)\, \tilde{g}_{n,l,m}(v)\, \widetilde{K}_A(u,v)\, du\, dv, \tag{4.48}$$

and so by (4.48) and the Schwarz inequality, whenever N is odd and $l \geqslant l_0$,

$$\lambda_{n,l}^2 \leqslant \int_{|u|\leqslant 1} \int_{|v|\leqslant 1} |\widetilde{K}_A(u,v)|^2\, du\, dv. \tag{4.49}$$

Now suppose that $(u,v) \in A$ and $|u| \geqslant |v|$. In this case $|v|^2 \leqslant \frac{3}{4}$, no matter what the value of $N \geqslant 3$. Also, by completing the square,

$$1 - p(u,v) \leqslant 1 - q_-(|u|,|v|) = (1-|u|^2) - \frac{(N-1)^2}{N(N-2)} \left(|v| - \frac{1}{N-1} |u| \right)^2.$$

Hence for such u and v,

$$\frac{(1-p(u,v))_+^{(3N-11)/2}}{(1-|v|^2)^{(3N-8)/4} (1-|u|^2)^{(3N-8)/4}} \leqslant \frac{(1-p(u,v))_+^{(3N-14)/4}}{\left(\frac{1}{4}\right)^{(3N-8)/4}},$$

and, from (4.46),

$$(1-p(u,v))^{(3N-14)/4} \leqslant \left(1 - \frac{N-1}{N} \frac{N-2}{N-1}\right)^{(3N-14)/4} = \left(\frac{2}{N}\right)^{(3N-14)/4}.$$

By symmetry in u and v, the same estimate holds when $(u,v) \in A$ and $|v| \geqslant |u|$. Hence

$$\int_{|u|\leqslant 1} \int_{|v|\leqslant 1} |\widetilde{K}_A(u,v)|^2\, du\, dv \leqslant \left(\frac{|S^{3N-10}|}{|S^{3N-7}|} \right)^2 \left(\frac{4\pi}{3} \right)^2 4^{3/4} \left(\frac{8}{N} \right)^{(3N-14)/4}.$$

This proves the following lemma:

LEMMA 4.5. *There is a finite integer N_0 such that for all odd integers $N \geqslant N_0$, and all $l \geqslant l_0$,*

$$|\lambda_{n,l}| \leqslant \left(\frac{1}{N}\right)^{N/2}.$$

On account of this result, increasing N_0 if need be, the condition $l < l_0$ in (4.39) and (4.40) may be dropped, and the estimates remain valid, for all odd integers N with $N \geqslant N_0$. Our next task is to remove the condition that N be odd.

LEMMA 4.6. *Let $\varkappa_N$ and β_N be defined for the Boltzmann collision model as in Theorem* 2.1. *Then for all $N>3$,*

$$|\beta_N| \leqslant |\beta_{N-1}|, \tag{4.50}$$

and for all N such that $\varkappa_{N-1}<\frac{1}{2}$,

$$\varkappa_N \leqslant \frac{\varkappa_{N-1}}{1-\varkappa_{N-1}}. \tag{4.51}$$

Proof. First let g satisfy $\|g\|_{\mathcal{K}_N}=1$ and $Kg=\varkappa_N g$. Then

$$\begin{aligned}\varkappa_N &= \langle g, Kg\rangle = \int_{X_N} g\circ\pi_1 g\circ\pi_2 \, d\mu_N \\ &= \int_{Y_N}\left[\int_{X_{N-1}} (g\circ\pi_1\circ\phi_N)(g\circ\pi_2\circ\phi_N)\, d\mu_{N-1}\right] d\nu_N.\end{aligned} \tag{4.52}$$

Now,

$$g\circ\pi_1\circ\phi_N(\vec{w}, v) = h_v\circ\pi_1(\vec{w})$$

where

$$h_v(y) = g\left(\frac{\sqrt{N^2-2N}}{N-1}\sqrt{1-|v|^2}\, y - \frac{1}{N-1} v\right).$$

Finally, let $\tilde{h}_v$ be given by

$$\tilde{h}_v\circ\pi_1 = h_v\circ\pi_1 - \int_{X_{N-1}} h_v\circ\pi_1 \, d\mu_{N-1} = h_v\circ\pi_1 - P_N(g\circ\pi_1) = h_v\circ\pi_1 - Kg\circ\pi_N.$$

Going back to (4.52) and using the variational definition of $\varkappa_{N-1}$ we have, much as in the proof of Theorem 2.2,

$$\begin{aligned}\varkappa_N &\leqslant \varkappa_{N-1}\int_{Y_N}\left[\int_{X_{N-1}} |h_v\circ\pi_1|^2 \, d\mu_{N-1}\right] d\nu_N(v) + (1-\varkappa_{N-1})\|Kg\|^2_{\mathcal{K}_N} \\ &= \varkappa_{N-1} + (1-\varkappa_{N-1})\varkappa_N^2.\end{aligned} \tag{4.53}$$

Since $\varkappa_{N-1}<1$, this last inequality may be written as $P_2(\varkappa_N)\geqslant 0$ where

$$P_2(x) = x^2 - \frac{1}{1-\varkappa_{N-1}} x + \frac{\varkappa_{N-1}}{1-\varkappa_{N-1}}.$$

The polynomial $P_2(x)$ has the roots $x=1$ and $x=\varkappa_{N-1}/(1-\varkappa_{N-1})<1$, and is negative between these two numbers. Since $\varkappa_N\leqslant 1$, (4.51) follows.

The proof of (4.50) is similar but simpler. Suppose that g satisfies $\|g\|_{\mathcal{K}_N}=1$ and $Kg=\tilde{\beta}_N g$ where

$$\tilde{\beta}_N = \inf\{\langle h, Kh\rangle \mid \|h\|_{\mathcal{K}_N}=1\}.$$

The analysis that lead to (4.53) now yields

$$\tilde{\beta}_N \geqslant \tilde{\beta}_{N-1}+(1-\tilde{\beta}_{N-1})\|Kg\|^2_{\mathcal{K}_N},$$

which certainly implies (4.50). □

We are finally ready to prove the analog of the original Kac conjecture for the Boltzmann collision model:

THEOREM 4.7. *For the Boltzmann collision model,*

$$\liminf_{N\to 0} \Delta_N > 0.$$

Proof. We choose N_0 large enough so that for all odd integers $N>N_0$,

$$\varkappa_N = \lambda_{2,0},$$

where $\lambda_{2,0}$ is specified in (4.36) and (4.37). We can do this since $\lambda_{2,0}\sim 5/3N^2$, and Lemma 4.6 tells us, increasing N_0 if need be, that

$$\varkappa_N \leqslant \frac{2}{N^2}$$

for all $N>N_0$. Now by Corollary 2.3,

$$\liminf_{N\to 0} \Delta_N > \prod_{j=N_0+1}^{\infty}\left(1-\frac{2}{j^2}\right)\Delta_{N_0}.$$

The infinite product is clearly strictly positive, and so it remains to verify that $\Delta_N>0$ for all N, and in particular for $N=N_0$.

This may as well be done by a compactness argument since we are not being specific about N_0. For $1\leqslant i<j\leqslant N$, define

$$R_{i,j}f(\vec{v}) = \int_{S^2} f(v_1, v_2, \ldots, v_i^*(\omega), \ldots, v_j^*(\omega), \ldots, v_n)\, b(\omega\cdot(v_i-v_j)/|v_i-v_j|)\, d\omega, \tag{4.54}$$

so that by (4.4),

$$Q_f(\vec{v}) = \binom{N}{2}^{-1}\sum_{i<j}^{N} R_{i,j}f(\vec{v}). \tag{4.55}$$

This operator is not compact. In the case where ω is selected uniformly, one easily sees that for any unit vector $\hat{e}\in\mathbf{R}^3$, and any odd integer k, $f_k(\vec{v})=(\pi_1(\vec{v})\cdot\hat{e})^k$ is an eigenfunction of Q, with a non-zero eigenvalue independent of k. In the case of the Kac model this was explicitly observed by Diaconis and Saloff-Coste, and this may have been clear to Kac when he remarked on the difficulty of showing that $\Delta_N>0$ for the original Kac model.

However, consider Q^{2N}. Observe from (4.54) and (4.55) that Q^{2N} is an average over monomials of degree $2N$ in the operators $R_{i,j}$. Each such monomial enters with the same positive weight, and each is a contraction on $\mathcal{H}_N$, since clearly each $R_{i,j}$ is a contraction on $\mathcal{H}_N$.

Now one such monomial is

$$A=(R_{1,2}R_{2,3}R_{3,4}\ldots R_{N,1})(R_{N,1}\ldots R_{4,3}R_{3,2}R_{2,1}),$$

which is positive. It follows that there is a positive number a so that

$$Q^{2N}=aA+(1-a)B,$$

where B is a self-adjoint contraction on $\mathcal{H}_N$. (B is the average over the remaining monomials.) Now it is easy to see that A is compact. Since it entails averages over each of the variables, it has a continuous kernel, and hence is Hilbert–Schmidt. Now

$$\begin{aligned}\sup\{\langle f,Q^{2N}f\rangle_{\mathcal{H}_N}\mid\|f\|_{\mathcal{H}_N}=1,\langle 1,f\rangle_{\mathcal{H}_N}=1\}&\\ =\sup\{\langle f,aA+(1-a)Bf\rangle\mid\|f\|_{\mathcal{H}_N}=1,\langle 1,f\rangle_{\mathcal{H}_N}=0\}&\\ \leqslant a\sup\{\langle f,Af\rangle\mid\|f\|_{\mathcal{H}_N}=1,\langle 1,f\rangle_{\mathcal{H}_N}=0\}+(1-a).&\end{aligned}$$

Now since A is compact, $\sup\{\langle f,Af\rangle_{\mathcal{H}_N}\mid\|f\|_{\mathcal{H}_N}=1,\langle 1,f\rangle_{\mathcal{H}_N}=0\}$ is attained. And clearly if f satisfies $\|f\|_{\mathcal{H}_N}=1$, $\langle 1,f\rangle_{\mathcal{H}_N}=0$ and $\langle f,Af\rangle_{\mathcal{H}_N}=1$, then $\|R_{k,k+1}f\|_{\mathcal{H}_N}=1=\|f\|_{\mathcal{H}_N}$, and this is impossible by our ergodicity assumptions. □

5. Analysis of a shuffling model

When momentum and energy are conserved for one-dimensional velocities, the only possibility is an exchange of velocities. Thus the Kac walk in this case is simply a walk on the permutations of $(v_1,v_2,\ldots,v_N)$, which, at least when all of these velocities are distinct, we may identify with a random walk on the permutation group Π_N. The corresponding walk has been thoroughly analyzed by Diaconis and Shahshahani [6], but we briefly discuss it in this section to illustrate several features of our approach. (In fact,

they estimate approach to uniformity in the total variation norm, for which they need, and derive, not only the spectral gap, but information on all of the eigenvalues and their multiplicities.)

Let $X_N=\Pi_N$, and let

$$Y_N=\{1,2,...,N\}. \tag{5.1}$$

For $\sigma\in\Pi_N$, define for $j=1,2,...,N$,

$$\pi_j(\sigma)=\sigma(j). \tag{5.2}$$

Let μ_N be normalized counting measure on X_N, so that ν_N is normalized counting measure on Y_N.

To define the transition function, fix a number p with $0<p<1$, which will represent the probability of "success" in a coin toss. Consider a deck of N distinct cards which are to be "shuffled" as follows: Pick a pair $i<j$ uniformly at random, and then toss a coin to generate independent Bernoulli variables with success probability p. If the result of the coin toss is success, exchange cards at the ith and jth positions from the top of the deck, and otherwise do nothing. This procedure is then repeated.

We can identify the state of the deck at each stage with the permutation σ which puts it in that order starting from a canonical "unshuffled" order. In these terms, the current state σ is updated by

$$\sigma\mapsto\sigma_{i,j}\sigma$$

where $\sigma_{i,j}$ is the pair permutation exchanging i and j, and fixing all else. If the result is not success, the current state σ is not altered. The one-step transition operator is clearly

$$Qf(\sigma)=\binom{N}{2}^{-1}\sum_{i<j}[pf(\sigma_{i,j}\sigma)+(1-p)f(\sigma)]. \tag{5.3}$$

To display this as a Kac system, define $\phi_N: X_{N-1}\times Y_N\to X_N$ by

$$\phi_N(\sigma,k)=\sigma_{k,N}\tilde{\sigma} \tag{5.4}$$

where $\tilde{\sigma}(j)=\sigma(j)$ for $j\leqslant N-1$, and $\tilde{\sigma}(N)=N$. Note that $\pi_N\circ\phi_N(\sigma,k)=\sigma_{k,N}(\tilde{\sigma}(N))=\sigma_{k,N}(N)=k$. We then define $\phi_j=\phi_N\circ\sigma_{j,N}$. It is clear that these maps are bijections, and since $\mu_{N-1}\otimes\nu_N$ is uniform counting measure on $X_{N-1}\times Y_N$, (2.9) is trivially true. Thus it is clear that this shuffling model is a Kac system.

Moreover it is easy to see that

$$Kg(i)=\sum_{j=1}^{N}K_{i,j}g(j) \tag{5.5}$$

where

$$K_{i,j} = \frac{1}{N-1}(1-\delta_{i,j}). \tag{5.6}$$

Hence K has the eigenvalues 1, with multiplicity one, and $-1/(N-1)$, with multiplicity $N-1$. Hence for this model, with $\varkappa_N$ and β_N as in Theorem 2.1,

$$-\varkappa_N = (N-1)\beta_N = \frac{1}{N-1} \tag{5.7}$$

and thus by Corollary 2.3,

$$\Delta_N \geqslant \prod_{j=3}^{N}\left(1-\frac{1}{(j-1)^2}\right)\Delta_2. \tag{5.8}$$

Again, this product collapses, and one finds

$$\prod_{j=3}^{N}\left(1-\frac{1}{(j-1)^2}\right) = \left(\frac{N}{N-1}\right)\frac{1}{2}. \tag{5.9}$$

Clearly Q_2 may be identified with the matrix

$$\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}, \tag{5.10}$$

and hence Q_2 has the eigenvalues 1 and $1-2p$. Hence $\lambda_2 = 1-2p$, and $\Delta_2 = 4p$. Combining this with (5.8) and (5.9), we have

$$\Delta_N \geqslant \frac{N}{N-1}2p. \tag{5.11}$$

To see that this result is sharp, we need to display an appropriate eigenfunction. We know from Theorem 2.2 that (5.11) can be sharp if and only if there is a function f_N satisfying both $Qf_N = \lambda_N f_N$ and $Pf_N = \mu_N f_N$. Theorem 2.1 then tells us that since $\beta_N > \varkappa_N$ for this problem, we get an eigenfunction of P with $Pf_N = \mu_N f_N$ from eigenfunctions h of K with $Kh = -1/(N-1)h$ through

$$f_N = h\circ\pi_i - h\circ\pi_j \quad \text{for } 1 \leqslant i < j \leqslant N \tag{5.12}$$

for some $i<j$. A tedious but straightforward computation, using $\sum_{j=1}^{N} h(j)=0$, which is equivalent to $Kh=-1/(N-1)h$, shows that

$$Qf_N = \left(1-\frac{2p}{N-1}\right)f_N.$$

This implies that $\lambda_N \geqslant 1-2p/(N-1)$ and hence

$$\Delta_N \leqslant \frac{N}{N-1}2p. \tag{5.13}$$

This leads to the following result:

THEOREM 5.1. *The binary shuffling model is a Kac system, and*

$$\Delta_N = \frac{N}{N-1} 2p. \tag{5.14}$$

Moreover,

$$N(I-Q)f = \left(\frac{N}{N-1} 2p\right) f$$

if and only if f has the form specified in (5.12) *for some function h on $\{1, 2, ..., N\}$ such that $\sum_{j=1}^{N} h(j)=0$. In particular, λ_N is an eigenvalue of Q of multiplicity $(N-1)^2$.*

Proof. The equality (5.14) follows from (5.11) and (5.13), and this identifies λ_N. We have shown above that every function f of the form (5.12) with $\sum_{j=1}^{N} h(j)=0$ satisfies $Qf=\lambda_N f$, and by Theorem 2.2, the converse holds as well since any such f must also satisfy $Pf=\mu_N f$, and this occurs only when f has the specified form. Finally, it is easily seen that the $N-1$ functions

$$h\circ\pi_1 - h\circ\pi_2, \quad h\circ\pi_2 - h\circ\pi_3, \quad h\circ\pi_3 - h\circ\pi_4, \quad ..., \quad h\circ\pi_{N-1} - h\circ\pi_N$$

are a basis for the span of the functions of the form specified in (5.12) whenever $\|h\|_K \neq 0$ and $\sum_{j=1}^{N} h(j)=0$. Also if h and $\tilde{h}$ are any two orthogonal eigenfunctions of K, $h\circ\pi_i - h\circ\pi_j$ is orthogonal to $\tilde{h}\circ\pi_k - \tilde{h}\circ\pi_l$ for all $i<j$ and $k<l$. Since the $-1/(N-1)$ has multiplicity $N-1$ as an eigenvalue of K, the final statement is now shown. □

Diaconis and Shahshahani actually devote most of their attention to the model in which the success probability p depends on N through $p=1-1/N$. The present methods are easily adapted to handle the case in which p depends on N. Let Q_r denote the transition operator defined in (5.3). Then clearly for two different success probabilities p and p',

$$Q_p = \frac{p}{p'} Q_{p'} + \left(\frac{p}{p'} - 1\right) I.$$

This may be used to take into account the effects of the N-dependence in p on λ_N.

6. The Kac walk on $SO(N)$

Let $SO(N)$ denote the group of orthogonal $(N\times N)$-matrices with unit determinant. In this section we consider a generalization of the original Kac model in which the state space is $SO(N)$ instead of S^{N-1}. This generalization was introduced by Diaconis and Saloff-Coste [5], and studied by Maslen as well, both in the case of "uniformly selected rotations", i.e., $\varrho(\theta)=1/2\pi$. To explain the nature of the underlying process, which these authors call the "Kac walk on $SO(N)$", we let $R_{i,j}(\theta)$ denote the same rotation in $\mathbf{R}^N$

that was used in (1.7), except now we identify it with the corresponding $(N\times N)$-matrix, and we will now consider our N-dimensional velocity vectors $\vec{v}$ as column vectors of an orthogonal matrix. Then multiplying $R_{i,j}(\theta)$ and the "pre-collisional velocity vector" $\vec{v}$ produces the "post-collisional velocity vector", just as in the original Kac model.

Given a continuous function f on $SO(N)$, define

$$Qf(G)=\binom{N}{2}^{-1}\sum_{i<j}\int_{-\pi}^{\pi}f(R_{i,j}(\theta)G)\varrho(\theta)\,d\theta \tag{6.1}$$

where $\varrho(\theta)$ satisfies the same conditions imposed on $\varrho(\theta)$ in the original Kac model.

The connection with the Kac walk on S^{N-1} becomes quite clear when one writes G in terms of its N columns, $G=[\vec{v}_1,\vec{v}_2,...,\vec{v}_N]$, since then

$$R_{i,j}(\theta)G=[R_{i,j}(\theta)\vec{v}_1,R_{i,j}(\theta)\vec{v}_2,...,R_{i,j}(\theta)\vec{v}_N].$$

Each of the $\vec{v}_j$ is an element of S^{N-1}, and it is clear from (1.7) that under the Kac walk on $SO(N)$, each column of G is a Markov process in its own right, and is in fact a copy of the original Kac walk on S^{N-1}. Therefore, if f depends on G only through the first column of G, which is an element of S^{N-1}, Qf coincides with what we would get by applying the operator Q for the Kac model to f considered as a function on S^{N-1}. In this sense the Kac walk on $SO(N)$ is a generalization, and indeed an extension, of the Kac walk on S^{N-1}.

This relation between the Kac walks on $SO(N)$ and S^{N-1} provides an immediate upper bound on the spectral gap for the walk on $SO(N)$: We see, by restricting the class of test functions to those that depend on G only through a single column, that the spectral gap for the Kac walk on $SO(N)$ cannot be larger than the spectral gap for the Kac walk on S^{N-1}.

In fact, as found by Maslen in the case in which $\varrho(\theta)$ is uniform, the two gaps actually coincide. In this section, we prove that the bound (3.6) holds for the Kac walk on $SO(N)$ when ϱ is not assumed to be uniform. The Kac walk on $SO(N)$ provides a good illustration of the methods of this paper in which the "single-particle space" depends on N. This example goes beyond our previous examples in other ways as well, as we shall see as soon as we begin with displaying it as a Kac system.

It turns out that it is most convenient to do this through consideration of the Kac walk on $O(N)$, the group of orthogonal $(N\times N)$-matrices. For a continuous function f on $O(N)$, we define Qf exactly as above, except that now G ranges over $O(N)$. We equip $O(N)$ with its Borel field and its normalized Haar measure μ_N. Then by our assumptions on ϱ, Q extends to be a self-adjoint Markovian contraction on $\mathcal{H}_N=$

$L^2(O(N), \mu_N)$. However, it is not ergodic; the nullspace of Q is two-dimensional, and spanned by 1 and det. Of course, on the subspace $L^2(SO(N), \mu_N)$, it is ergodic.

The reason for working in the non-ergodic $L^2(O(N), \mu_N)$-setting is that permutation symmetry plays an important role in our analysis. The natural action of Π_N on $O(N)$ is through interchange of rows. Note that this extends the action of Π_N on S^{N-1}, considered as the first column of G, that we used in our analysis of the original Kac model. Interchange of two rows of an element of $SO(N)$ of course changes the sign of the determinant, and so does not preserve $SO(N)$. An alternative is to conjugate elements of $SO(N)$ by the permutations; that is, to swap both rows and columns. This, however, complicates the construction of a Kac system for the model, and in any case, it is of some interest to show that the methods used here can be applied when there is more than one ergodic component. This said, we proceed with the construction of the Kac system.

The "N-particle space" $(X_N, \mathcal{S}_N, \mu_N)$ will of course be $O(N)$ equipped with its Borel field, and its normalized Haar measure is μ_N, as indicated above. Let $O(N)_+$ denote the component of $O(N)$ on which the determinant is positive, so that $O(N)_+$ is just $SO(N)$, and let $O(N)_-$ denote the component of $O(N)$ on which the determinant is negative.

For any permutation σ in Π_N, let P_σ denote the corresponding $(N \times N)$-permutation matrix. For G in $O(N)$, define $\sigma(G)$ by

$$\sigma(G) = P_\sigma G.$$

That is, σ acts on the matrix G by permuting its rows. Clearly this is a measure-preserving action of Π_N on $O(N)$.

We take the *single-particle space* $(Y_N, \mathcal{T}_N, \nu_N)$ to be S^{N-1} equipped with its Borel field and normalized rotation-invariant measure ν_N. For each $j = 1, 2, ..., N$, let $\mathbf{e}_j$ be the jth standard basis vector in $\mathbf{R}^N$, written as a row vector, so that for any $(N \times N)$-matrix A, $\mathbf{e}_j A$ is the jth row of A. We then define

$$\pi_j(G) = \mathbf{e}_j G.$$

That is, $\pi_j(G)$ is the jth row of G. It is clear that $\pi_j : X_N \to Y_N$ and that

$$\pi_j(\sigma(G)) = \pi_{\sigma(j)}(G)$$

for each j and G. So far, we have verified the first two features required of a Kac system.

The next steps in the construction of this Kac system are slightly more involved. We have to construct the maps $\phi_j : X_{N-1} \times Y_N \to X_N$, but since there is no canonical embedding of $O(N-1)$ into $O(N)$, they have to be constructed "by hand", using a convenient coordinate chart. (Just as with the original Kac model, the maps ϕ_j cannot

be continuous since $X_{N-1}\times Y_N$ and X_N just do not have the same topology. But just as in that case, we are only concerned with measure-theoretic properties of these mappings, and on a set of full measure they will be well behaved.)

Let

$$\vec{v}_- = (0, 0, ..., 0, -1)$$

denote the "south pole" in S^{N-1}. We may use the stereographic projection to identify $S^{N-1}\backslash\{\vec{v}_-\}$ with $\mathbf{R}^{N-1}$. At each point of $\mathbf{R}^{N-1}$ we of course have the standard orthonormal basis. The stereographic projection, which is conformal, carries this back to an orthogonal basis for the tangent space at the corresponding point in $S^{N-1}\backslash\{\vec{v}_-\}$. Normalizing these vectors, we thus obtain a smoothly varying frame of orthonormal vectors

$$\{\mathbf{u}_1(\vec{v}), \mathbf{u}_2(\vec{v}), ..., \mathbf{u}_{N-1}(\vec{v})\}$$

in $\mathbf{R}^N$ parameterized by $\vec{v}$ in $S^{N-1}\backslash\{\vec{v}_-\}$. For each $\vec{v}$, they form an orthonormal basis for the tangent space to S^{N-1} at $\vec{v}$.

Now for each $\vec{v}$ in $S^{N-1}\backslash\{\vec{v}_-\}$, define $U(\vec{v})$ to be the $((N-1)\times N)$-matrix whose jth row is $\mathbf{u}_j(\vec{v})$. Let the jth row of $U(\vec{v}_-)$ be $\mathbf{e}_j$. The map $v\mapsto U(\vec{v})$ is now defined on all of S^{N-1}, though of course it is discontinuous at the "south pole", $\vec{v}_-$. This, however, will not be a problem.

Now define the map

$$\phi_N : O(N-1)\times S^{N-1} \to O(N)$$

as follows: Given $\widetilde{G}$ in $O(N-1)$ and $\vec{v}$ in S^{N-1}, first form the $((N-1)\times N)$-matrix $\widetilde{G}U(\vec{v})$ with $U(\vec{v})$ as specified above. Because $\widetilde{G}$ is in $O(N-1)$ and the rows of $U(\vec{v})$ are orthonormal, the rows of $\widetilde{G}U(\vec{v})$ are orthonormal. Moreover, since the rows of $U(\vec{v})$ are a basis for the tangent space to S^{N-1} at $\vec{v}$, each one is orthogonal to $\vec{v}$. The rows of $\widetilde{G}U(\vec{v})$ are linear combinations of the rows of $U(\vec{v})$, and hence these too are orthogonal to $\vec{v}$. Therefore, if we form the $(N\times N)$-matrix

$$\begin{bmatrix} \widetilde{G}U(\vec{v}) \\ \vec{v} \end{bmatrix}$$

by adjoining $\vec{v}$ to $\widetilde{G}U(\vec{v})$ as the final row, we obtain an orthogonal matrix.

Next observe that the determinant of this matrix is just the determinant of $\widetilde{G}$. Indeed, it is clear that when $\vec{v}$ is the "north pole", so that $U(\vec{v})$ consists of the first $N-1$ rows of the $(N\times N)$-identity matrix, then the determinant of this matrix is simply the determinant of $\widetilde{G}$, which is either $+1$ or -1. Now as $\vec{v}$ varies in $S^{N-1}\backslash\{\vec{v}_-\}$,

$$\det\left(\begin{bmatrix} \widetilde{G}U(\vec{v}) \\ \vec{v} \end{bmatrix}\right)$$

varies continuously. Hence the value is just $\det(\widetilde{G})$ for all such $\vec{v}$. Continuity fails at the "south pole", but there it is again obvious by the special form of the definition of ϕ_N at $\vec{v}_-$ that still in this case the determinant is just that of $\widetilde{G}$. Hence the image of

$$(\widetilde{G},\vec{v})\mapsto\begin{bmatrix}\widetilde{G}U(\vec{v})\\ \vec{v}\end{bmatrix}=\phi_N(\widetilde{G},\vec{v})$$

does indeed lie in $O(N)$, and $\det(\phi_N(\widetilde{G},\vec{v}))=\det(\widetilde{G})$. It is also clear by construction that

$$\pi_N(\phi_N(\widetilde{G},\vec{v}))=\vec{v}$$

everywhere on $O(N-1)\times S^{N-1}$.

Finally, it remains to check that this map is well behaved with respect to the measures μ_{N-1}, ν_N and μ_N. Given a function f on $O(N)$, we may compute the average of f with respect to μ_N in two stages as follows: First compute the conditional expectation $g(\vec{v})$ where

$$g(\vec{v})=\mathrm{E}\{f(G)\,|\,\pi_N(G)=\vec{v}\}.$$

Then

$$\int_{O(N)}f(G)\,d\mu_N(G)=\int_{S^{N-1}}g(\vec{v})\,d\nu_N(\vec{v}).$$

Also, it is clear that we can compute the conditional expectation $\mathrm{E}\{f(G)|\pi_N(G)=\vec{v}\}$ by averaging over orbits generated by left multiplication of G by elements G' belonging to the subgroup of $O(N)$ consisting of orthonormal matrices whose final row is $\mathbf{e}_N$. This subgroup is just a copy of $O(N-1)$, and so

$$g(\pi_N(G))=\int_{O(N-1)}f(G'G)\,d\mu_{N-1}(G').$$

It follows directly from this that

$$\int_{O(N)}f(G)\,d\mu_N(G)=\int_{S^{N-1}}\left[\int_{O(N-1)}f(\phi_N(G',\vec{v}))\,d\mu_{N-1}(G')\right]d\nu_N(\vec{v}).$$

We define ϕ_j for $j=1,2,...,N-1$ in terms of ϕ_N and the pair permutations exchanging j and N, in the natural way. Clearly the analogs of the results just derived for ϕ_N hold for each ϕ_j as well. Thus, with these definitions, the third feature required of a Kac system is verified.

To complete the verification that the Kac walk on $O(N)$ can be made into a Kac system, we only need to verify that

$$\langle f,Qf\rangle_{\mathcal{H}_N}=\frac{1}{N}\sum_{j=1}^{N}\int_{Y_N}(\langle f_{j,\vec{v}},Qf_{j,\vec{v}}\rangle_{\mathcal{H}_{N-1}})\,d\nu_N(\vec{v})$$

where for each j and each $\vec{v}\in Y_N=S^{N-1}$,

$$f_{j,\vec{v}}(\cdot)=f(\phi_j(\cdot,\vec{v})).$$

This is clear, given the definition and computations just above.

We are not yet ready to apply Theorems 2.1 and 2.2 to this Kac system, since we must modify the definition of λ_N. We define

$$\lambda_N=\sup\{\langle f,Qf\rangle_{\mathcal{H}_N}\mid \|f\|_{\mathcal{H}_N}=1,\ \langle f,1\rangle_{\mathcal{H}_N}=0,\ \langle f,\det\rangle_{\mathcal{H}_N}=0\},$$

which differs from (2.16) due to the restriction that f be orthogonal to the determinant.

As a consequence, a *modification of the operators P_j is also required.* In the definition (2.14) used in examples with a single ergodic component, we averaged over all of X_{N-1}, or put differently, conditioned only on Y_N. Now we will condition on Y_N and the ergodic component. Given f in $\mathcal{H}_N$, and $j=1,2,...,N$, define g_+ and g_- on $Y_N=S^{N-1}$ as follows:

$$g_+(\vec{v})=2\int_{O(N-1)_+} f_{j,\vec{v}}(\widetilde{G})\,d\mu_{N-1}(\widetilde{G})$$

and

$$g_-(\vec{v})=2\int_{O(N-1)_-} f_{j,\vec{v}}(\widetilde{G})\,d\mu_{N-1}(\widetilde{G}).$$

The definitions are such that

$$g_\pm(\vec{v})=\mathrm{E}\{f\mid \pi_j=\vec{v}\text{ and }\det=\pm 1\}.$$

The factors of 2 are because $O(N-1)_\pm$ each accounts for exactly half of $O(N)$ by volume. We now define

$$P_jf(G)=\begin{cases} g_+(\pi_j(G)) & \text{if } \det(G)=+1,\\ g_-(\pi_j(G)) & \text{if } \det(G)=-1.\end{cases}$$

Since P_jf depends on G only through $\vec{v}=\pi_j(G)$, it is again convenient to abuse notation by writing $P_jf(\vec{v})$.

The point of the definitions is the following: Note that f in $\mathcal{H}_N$ satisfies both $\langle f,1\rangle_{\mathcal{H}_N}=0$ and $\langle f,\det\rangle_{\mathcal{H}_N}=0$ in case it satisfies both

$$\int_{O(N)_+} f(G)\,d\mu_N(G)=0\quad\text{and}\quad\int_{O(N)_-} f(G)\,d\mu_N(G)=0.$$

But in this case, by the definition of P_j,

$$\int_{O(N-1)_+}(f_{j,\vec{v}}(\widetilde{G})-P_jf(\vec{v}))\,d\mu_{N-1}(\widetilde{G})=0$$

and

$$\int_{O(N-1)_-} (f_{j,\vec{v}}(\widetilde{G}) - P_j f(\vec{v}))\, d\mu_{N-1}(\widetilde{G}) = 0,$$

for almost every $\vec{v}$. Hence $f_{j,\vec{v}}(\widetilde{G}) - P_j f(\vec{v})$ is, for almost every $\vec{v}$, orthogonal to both 1 and det on $O(N-1)$. This is the key requirement for the proof of Theorem 2.2 to hold with the modified definition of λ_N. We leave to the reader the easy verification of this. Of course the definitions of the quantities μ_N in (2.18), and $\varkappa_N$ in (2.26), have to be modified in the same way as was the definition of λ_N. With these modifications made, the analog of Theorem 2.1 holds as well, and again, as the proof is essentially the same, the verification of this is left to the reader. *We summarize this by saying that the presence of more than one ergodic component can be taken into account within the framework of ideas described in* §2 *by conditioning not just on the single-particle space, but on the ergodic components as well.* The present model is a case in point, which we choose to leave as an example rather than attempting a general formulation.

With these results in hand, our task is to compute the spectrum of the operator K for this system, which is still defined through (2.21). This is what is used in the proof of Theorem 2.1, though (2.24) no longer holds due to our modification of the definition of P_j.

However, it is clear from the definition of K in terms of conditional expectations, through (2.22), that

$$Kg(\vec{v}) = \int_{W\cdot\vec{v}=0} g(W)\, d\nu_{N-2}(W), \tag{6.2}$$

where the integral on the right is computed with respect to the uniform probability measure on S^{N-2}, identified with the subset of S^{N-1} consisting of those unit vectors W for which $W\cdot\vec{v}=0$, as indicated in the limits of integration. That is, the value of $Kg(\vec{v})$ is just the "equatorial average" of the values of g around the "equator" with respect to a pole running along $\vec{v}$.

It is immediately clear that K preserves the space of polynomials of any fixed degree d, and hence the eigenfunctions of K are the spherical harmonics on S^{N-1}.

The zonal spherical harmonics of degree d are those that depend on W only through $\mathbf{e}\cdot W$ for some fixed unit vector $\mathbf{e}$. Let $z_{d,\mathbf{e}}(W)$ denote the corresponding zonal spherical harmonic, and let $p_d(x)$ be the polynomial of degree d so that the zonal spherical harmonic $z_{d,\mathbf{e}}(W)$ on S^{N-1} is given by

$$z_{d,\mathbf{e}}(W) = p_d(\mathbf{e}\cdot W).$$

The normalization is fixed so that the *reproducing kernel* property holds:

$$\int_{S^{N-1}} h(W)\, z_{d,\mathbf{e}}(W)\, d\nu_{N-1}(W) = h(\mathbf{e})$$

for any spherical harmonic h of degree d on S^{N-1}. This means that

$$\|z_{d,\mathbf{e}}\|_2^2 = p_d(\mathbf{e}\cdot\mathbf{e}) = p_d(1).$$

Now fix $\vec{v}$, and let $h(W)$ be any spherical harmonic of degree d that is orthogonal to $z_{d,\vec{v}}$. Then

$$\int_{O(N-1)} h(\widetilde{G}W)\, d\mu_{N-1}(\widetilde{G}) = 0, \tag{6.3}$$

where $\widetilde{G}$ runs over those rotations of $\mathbf{R}^N$ that fix $\vec{v}$. This is because the left-hand side is a spherical harmonic of degree d that depends on W only through $\vec{v}\cdot W$. This means that it is a multiple of $z_{d,\vec{v}}$. However, since h was orthogonal to $z_{d,\vec{v}}$, so is the average, and hence the identity (6.3) is established. But comparing (6.2) and (6.3), we see that

$$Kg(\vec{v}) = \int_{O(N-1)} g(\widetilde{G}W_0)\, d\mu_{N-1}(\widetilde{G})$$

for any W_0 with $W_0\cdot\vec{v}=0$. Hence, under our assumptions on h, $Kh(\vec{v})=0$.

Now let g be any spherical harmonic of degree d. Let $P_{\vec{v}}g$ be defined by

$$P_{\vec{v}}g(W) = \frac{1}{p_d(1)}\left(\int_{S^{N-1}} g(W')z_{d,\vec{v}}(W')\, d\nu_{N-1}(W')\right) z_{d,\vec{v}}(W), \tag{6.4}$$

which is simply the orthogonal projection of g onto the span of $z_{d,\vec{v}}$. Evidently $g-P_{\vec{v}}g$ is a spherical harmonic of degree d that is orthogonal to $z_{d,\vec{v}}$, and hence by the above, $K(g-P_{\vec{v}}g)(\vec{v})=0$. It follows that

$$Kg(\vec{v}) = K(P_{\vec{v}}g)(\vec{v}).$$

But the right-hand side is easy to compute as clearly $Kz_{d,\vec{v}}=p_d(0)$. Now by (6.4) and the reproducing kernel property, we have

$$Kg(\vec{v}) = \frac{p_d(0)}{p_d(1)}g(\vec{v}).$$

Now it is possible to compute the ratios $p_d(0)/p_d(1)$ using generating functions, though it would not be so clear from this which value of d gives the largest ratio. However, none of this is necessary: If we fix any direction unit vector $\vec{e}$, and take any function ϕ on $[-1,1]$, we have

$$K(\phi(\vec{v}\cdot\vec{e})) = (K_{\mathrm{Kac}}\phi)(\vec{v}\cdot\vec{e})$$

where K_{Kac} is the operator K for the original Kac model, whose spectrum we have computed in Theorem 3.1. Hence the eigenvalues of K and K_{Kac} coincide and are provided by Theorem 3.1.

This solves the eigenvalue problem for K.

THEOREM 6.1. *Every spherical harmonic ϕ on S^{N-1} of degree d, considered as a function $\tilde{\phi}$ on $O(N)$ through $\tilde{\phi}(G)=\phi(G\vec{w})$ for any fixed vector $\vec{w}$ in S^{N-1}, is an eigenfunction of K, and the corresponding eigenvalue is $p_d(0)/p_d(1)$. These eigenvalues are exactly the eigenvalues of the operator K for the original Kac model that are given in Theorem* 3.1.

Therefore, (3.4) and (3.5) hold just as in §3, and we have once more that

$$\Delta_N \geqslant \frac{1}{4}\frac{N+2}{N-1}\Delta_2.$$

It remains to calculate Δ_2. But $O(2)$ is just two copies of S^1, and the same Fourier analysis argument described around (1.23) once more gives us

$$\lambda_2 = \sup_{k\neq 0}\left\{\int_{-\pi}^{\pi}\varrho(\theta)\cos(k\theta)\,d\theta\right\}$$

for the second largest eigenvalue of Q when $N=2$.

The generalization of Maslen's result to non-uniform ϱ now follows immediately from what has been said above.

Altogether, we have proved

THEOREM 6.2. *Formula* (3.6) *yields a lower bound for the spectral gap for the Kac walk on $SO(N)$ with the transition operator Q given by* (6.1) *for all $N\geqslant 2$. There is equality if and only if there is equality for the spectral gap for the transition operator Q_{Kac} of the corresponding Kac walk on S^{N-1}, as given in* (1.7) *with the same density ϱ. When ϱ is uniform, the multiplicity of the corresponding eigenvalue equals the dimension of the space of fourth-degree spherical harmonics on S^{N-1}.*

7. Analysis of maximizers for non-uniform $\varrho(\theta)$

We return to the Kac model on S^{N-1} with a non-uniform density $\varrho(\theta)$, so that Q is given by (1.7), and we examine the circumstances under which the quartic function f_N given in (1.18) is an optimizer for (1.9). Because of the very close relation of the Kac walk on $SO(N)$ to this model, as described in §6, our analysis is readily adapted to that model as well, though we shall not carry out the adaptation here.

According to Lemma 3.2, f_N is an eigenfunction of Q with the eigenvalue

$$1-\frac{2\gamma(N+2)}{N(N-1)} \tag{7.1}$$

where

$$\gamma = \frac{1}{4}\int(1-\cos(4\theta))\,\varrho(\theta)\,d\theta. \tag{7.2}$$

We therefore define Γ_N by

$$\Gamma_N = \frac{2\gamma(N+2)}{N-1} = N\left(1-\frac{\langle f_N, Qf_N\rangle}{\|f_N\|^2}\right), \tag{7.3}$$

so that if f_N happens to be a maximizer for (1.9), then $\Gamma_N = \Delta_N$, but otherwise $\Delta_N < \Gamma_N$. That is,

$$\Delta_N \leqslant \Gamma_N, \tag{7.4}$$

and there is equality in (7.4) if and only if f_N is a maximizer for (1.9).

Now the operator Q commutes with permutations, so the permutation-invariant functions f constitute an invariant subspace $\mathcal{S}$ of $\mathcal{H}_{N,E}$, and clearly this subspace includes f_N. It is the symmetric densities f that are relevant in Kac's limit theorem relating the master equation (1.8) and the Kac equation (1.15). We will therefore restrict our attention to this subspace, where the result is especially striking, and which is physically the most significant. We therefore define $\tilde{\lambda}_N$ to be the second largest eigenvalue of Q restricted to $\mathcal{S}$:

$$\tilde{\lambda}_N = \sup\{\langle f, Qf\rangle \mid f \in \mathcal{S}, \|f\|_2 = 1, \langle f, 1\rangle = 0\}, \tag{7.5}$$

and we define $\tilde{\Delta}_N = N(1-\tilde{\lambda}_N)$. Clearly

$$\tilde{\Delta}_N \leqslant \Gamma_N, \tag{7.6}$$

and there is equality in (7.6) if and only if f_N is a maximizer for (7.5).

Taking the symmetry constraints into account it is easy to compute $\tilde{\Delta}_2$ using (1.23), with the result that

$$\tilde{\Delta}_2 = 2 \min_{k \geqslant 1} \int (1-\cos(k\theta))\, \varrho(\theta)\, d\theta. \tag{7.7}$$

We see from (7.1) and (7.2) that if the supremum in (7.7) occurs at $k=4$, then

$$\tilde{\Delta}_2 = \Gamma_2. \tag{7.8}$$

It is easy to see that $\tilde{\Delta}_N$ and $\tilde{\Delta}_{N-1}$ are still related by the inequality proved in Theorem 1.1 for Δ_N and Δ_{N-1}:

$$\tilde{\Delta}_N \geqslant (1-\varkappa_N)\tilde{\Delta}_{N-1}. \tag{7.9}$$

We also know from Theorem 1.3 and the definition of Γ_N in terms of f_N that Γ_N solves this same recursion relation:

$$\Gamma_N = (1-\varkappa_N)\Gamma_{N-1}. \tag{7.10}$$

Notice that (7.8), (7.9) and (7.10) together imply that $\tilde{\Delta}_N \geqslant \Gamma_N$ for all $N \geqslant 2$. But from (7.6) we have $\tilde{\Delta}_N \leqslant \Gamma_N$ for all $N \geqslant 2$. Hence $\tilde{\Delta}_N = \Gamma_N$, and f_N is a maximizer.

We see that if f_2 is the maximizer for $N=2$, then f_N is a maximizer for all N. So far we are simply translating old results into the symmetric case, but we have relied less on explicit calculation in order to bring out the following point: Suppose that we had *any* sequence of admissible functions g_N for the variational problem (7.5), and we defined Γ_N by

$$\Gamma_N = N(1-\langle g_N, Qg_N\rangle). \tag{7.11}$$

Then if it happened that the Γ_N so defined satisfied the recurrence relation (7.10), and also satisfied (7.8), it would follow by simple comparison that $\tilde{\Delta}_N=\Gamma_N$ for all $N\geqslant 2$ and g_N would be a maximizer for (7.5). *All that was required of g_N is that* (7.11) *leads to a solution of* (7.10), *and that* (7.8) *holds.*

The following simple observation leads to further progress: Suppose that the minimum in (7.7) does not occur at $k=4$, and so (7.8) is false. But suppose that for some N_0, $\tilde{\Delta}_{N_0}=\Gamma_{N_0}$. That is, suppose that for $N=N_0$, f_N is a maximizer for (7.5). *Then f_N is a maximizer for* (7.5) *for all $N\geqslant N_0$, so that $\tilde{\Delta}_N=\Gamma_N$ for all $N\geqslant N_0$ and therefore*

$$\lim_{N\to\infty}\tilde{\Delta}_N = \lim_{N\to\infty}\Gamma_N = 2\gamma, \tag{7.12}$$

which is the result we would have gotten if the maximizer for $N=2$ had been quartic. Therefore, either (7.12) *holds, or else f_N is never a maximizer for* (7.5) *for any N.*

Now we know that f_N spans the second eigenspace of P corresponding to its second largest eigenvalue, where P is the operator whose second largest eigenvalue μ_N is the key to the recursion in Theorem 2.2. (Recall that P does not depend on ϱ.) *If for each N, the true maximizer h_N for* (7.5) *is orthogonal to f_N, which is the case whenever f_N is never a maximizer for* (7.5), *then we can replace μ_N in Theorem* 2.2 *by a smaller number $\tilde{\mu}_N$, and hence can replace $\varkappa_N$ in* (7.9) *by a smaller number $\tilde{\varkappa}_N$.* As we shall see, it turns out that this strictly smaller number $\tilde{\varkappa}_N$ is

$$\tilde{\varkappa}_N = \alpha_8(N) = \frac{105}{(N+5)(N+3)(N+1)(N-1)}, \tag{7.13}$$

where $\alpha(8)$ is an eigenvalue of the operator K as described in Theorem 3.1, while

$$\varkappa_N = \alpha(4) = \frac{3}{(N+1)(N-1)}.$$

In summary, we have two things working for us:

Either (7.9) *holds with a $\varkappa_N$ replaced by a strictly smaller number $\tilde{\varkappa}_N$ for all N, or else there is an N_0 so that $\tilde{\Delta}_N=\Gamma_N$ for all $N\geqslant N_0$.*

Now when $\varkappa_N$ is replaced by $\tilde{\varkappa}_N$ in (7.9), it leads to a *much better* lower bound for $\tilde{\Delta}_N$. But this improved lower bound cannot violate (7.6). If it does, it can only mean that the second alternative holds and not the first.

This argument leads to the following result:

THEOREM 7.1 (conditions for f_N to maximize for large N). *Assume that*

$$\tilde{\Delta}_2 > 0.45\Gamma_2. \tag{7.14}$$

Then for all N sufficiently large,

$$\tilde{\Delta}_N = \Gamma_N \tag{7.15}$$

and f_N is the corresponding eigenfunction.

Proof. All of the key ideas have been explained above, and it only remains to check the details. As we have seen, if (7.15) does not hold for all sufficiently large N, then f_N is orthogonal to the true gap eigenfunction h_N for all N, since Q is self-adjoint and f_N is always an eigenfunction.

This means that h_N is orthogonal to the constant function *and* to the function f_N. Now we repeat the induction argument in the proof of Theorem 2.2 once more, but for the constraint that h_N is orthogonal to both 1 and f_N. Under these new conditions, we obtain the recursion

$$\tilde{\lambda}_N \leqslant \tilde{\lambda}_{N-1} + (1-\tilde{\lambda}_{N-1})\tilde{\mu}_N$$

in place of (2.35), where $\tilde{\mu}_N$ is given by (2.18), except that now we require f to be orthogonal to both 1 and f_N. Again, Theorem 2.1 shows that $\tilde{\mu}_N$ can be computed in terms of the eigenvalues of K, with the result that it is $\alpha(8)$ that is now relevant, not $\alpha(4)$, due to the new constraint. (Recall that $\alpha_6(N)$ is negative, and so is irrelevant.) Hence

$$\tilde{\Delta}_N \geqslant \tilde{\Delta}_{N-1}(1-\alpha_8(N)). \tag{7.16}$$

The inequality (7.16) can be solved recursively to yield

$$\tilde{\Delta}_N \geqslant 90 \frac{\Gamma(N-1)\Gamma(N+7)\Gamma\big(N+3+i\sqrt{6}\big)\Gamma\big(N+3-i\sqrt{6}\big)}{\Gamma(N)\Gamma\big(5+i\sqrt{6}\big)\Gamma\big(5-i\sqrt{6}\big)\Gamma(N+6)\Gamma(N+4)\Gamma(N+2)}\tilde{\Delta}_2. \tag{7.17}$$

With the help of the relation

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$$

the limit as $N\to\infty$ of (7.17) can be computed and yields

$$\liminf_{N\to\infty} \tilde{\Delta}_N \geqslant \frac{3}{770}\frac{\sinh\big(\sqrt{6}\,\pi\big)}{\sqrt{6}\,\pi}\tilde{\Delta}_2 =: L\tilde{\Delta}_2.$$

This quantity has to be compared to 2γ and this shows that whenever $2\gamma < L\tilde{\Delta}_2$ (which is easily shown to be implied by (7.14)) there exists some finite N beyond which f_N is the gap eigenfunction. □

References

[1] ASKEY, R., *Orthogonal Polynomials and Special Functions.* SIAM Regional Conference Series in Applied Mathematics, 21. SIAM, Philadelphia, PA, 1975.

[2] CARLEMAN, T., Sur la théorie de l'équation intégrodifférentielle de Boltzmann. *Acta Math.*, 60 (1933), 91–146.

[3] CARLEN, E., CARVALHO, M. C. & LOSS, M., Many-body aspects of approach to equilibrium, in *Journées équations aux dérivées partielles* (Nantes, 2000), pp. XI:1–12; Also in *Séminaire: equations aux dérivées partielles*, 2000–2001, exp. n° XIX. École Polytechnique, Palaiseau, 2001.

[4] CARLEN, E. A., GABETTA, E. & TOSCANI, G., Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. *Comm. Math. Phys.*, 199 (1999), 521–546.

[5] DIACONIS, P. & SALOFF-COSTE, L., Bounds for Kac's master equation. *Comm. Math. Phys.*, 209 (2000), 729–755.

[6] DIACONIS, P. & SHAHSHAHANI, M., Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57 (1981), 159–179.

[7] GRÜNBAUM, F. A., Propagation of chaos for the Boltzmann equation. *Arch. Rational Mech. Anal.*, 42 (1971), 323–345.

[8] — Linearization for the Boltzmann equation. *Trans. Amer. Math. Soc.*, 165 (1972), 425–449.

[9] JANVRESSE, E., Spectral gap for Kac's model of Boltzmann equation. *Ann. Probab.*, 29 (2001), 288–304.

[10] KAC, M., Foundations of kinetic theory, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, Vol. III (Neyman, J., ed.), pp. 171–197. University of California Press, Berkeley and Los Angeles, 1956.

[11] KOORNWINDER, T. H., The addition formula for Jacobi polynomials, I. Summary of results. *Indag. Math.*, 34 (1972), 188–191.

[12] MCKEAN, H. P., JR., Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Arch. Rational Mech. Anal.*, 21 (1966), 343–367.

[13] YAU, H.-T., Logarithmic Sobolev inequality for lattice gases with mixing conditions. *Comm. Math. Phys.*, 181 (1996), 367–408.

[14] — Logarithmic Sobolev inequality for generalized simple exclusion processes. *Probab. Theory Related Fields*, 109 (1997), 507–538.

E. A. CARLEN
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
U.S.A.
carlen@math.gatech.edu

M. C. CARVALHO
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
U.S.A.
mcarvalh@math.gatech.edu

M. LOSS
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
U.S.A.
loss@math.gatech.edu

Received January 7, 2002