The Annals of Applied Probability
2012, Vol. 22, No. 4, 1712–1727
DOI: 10.1214/11-AAP810
© Institute of Mathematical Statistics, 2012

TOTAL VARIATION BOUND FOR KAC'S RANDOM WALK

BY YUNJIANG JIANG[1]

Stanford University

We show that the classical Kac's random walk on $(n-1)$-sphere S^{n-1} starting from the point mass at e_1 mixes in $\mathcal{O}(n^5(\log n)^3)$ steps in total variation distance. The main argument uses a truncation of the running density after a burn-in period, followed by $\mathcal{L}^2$ convergence using the spectral gap information derived by other authors. This improves upon a previous bound by Diaconis and Saloff-Coste of order $\mathcal{O}(n^{2n})$.

1. Introduction. Consider n particles on $\mathbb{R}$ making random pairwise collisions, in such a way that the total kinetic energy is conserved. Since there is randomness involved, the situation is typically modeled by a Markov chain. Two natural questions are how would the particles be distributed in equilibrium and whether such equilibrium distribution is unique. And once these are answered, one would also like to know how long it takes for the particles to reach this equilibrium distribution. Of course these questions would depend on the mathematical models we choose to describe the system.

Mark Kac proposed the following toy model of one-dimensional Boltzmann gas dynamics that captures the above description (for historical development, see [3, 5]): For the n particles on $\mathbb{R}$, we can represent their velocities $(v_1, \dots, v_n)$ as a point on the unit sphere S^{n-1} after normalization so that

$$\sum_{i=1}^{n} v_i^2 = 1.$$

Conservation of kinetic energy (assuming 0 potential energy) in the gas dynamics is equivalent to $(v_1(t), \dots, v_n(t))$ staying on S^{n-1} for all $t \geq 0$. We will not introduce momentum conservation in our model, because that will force the collision to be inelastic (see second paragraph below), and reduces the model to a discrete Markov chain such as the random transposition walk on S_n. But when the particles live in $\mathbb{R}^3$, momentum conservation becomes quite interesting (see [4]). The technique in this paper might be applicable to that model as well.

Each time there is a collision, it occurs with probability 1 between no more than two particles, which corresponds to choosing two distinct coordinate directions x_i, x_j and rotating S^{n-1} along the 2-plane $x_i \wedge x_j$ by some angle θ. Notice that

Received December 2009; revised February 2011.

[1]Research supported in part by an NSF graduate fellowship.

MSC2010 subject classifications. 60-XX.

Key words and phrases. Markov chain mixing time, orthogonal group, Kac random walk, interacting particle systems.

$\sum_k v_k^2 = 1$ both before and after the collision, since the sum $v_i^2 + v_j^2$ is not affected by the rotation along the i, j plane and all the other velocities stay the same.

By disregarding the position information of the particles (which have to be confined in some compact domain, for example S^1, else they will eventually run off to infinity), each collision occurs between any pair of the particles with equal probability $\frac{1}{\binom{n}{2}}$. The rotation angle θ can be chosen from some distribution on $[0, 2\pi)$, which physically is a measure of the elasticity of the collision; for example, inelastic collision in $\mathbb{R}$ will correspond to a distribution of θ, that is, a delta measure concentrated at π. In this paper, we will assume that θ is uniformly distributed on $[0, 2\pi)$.

Thus we obtain a discrete-time Markov chain on S^{n-1} with transition kernel given by, for $f : S^{n-1} \to \mathbb{R}$ continuous, and $x \in S^{n-1}$,

$$(Kf)(x) = \frac{1}{\binom{n}{2}} \sum_{i \neq j}^{n} \int_0^{2\pi} f(R(i, j; \theta)x) \frac{1}{2\pi} \, d\theta, \tag{1}$$

where $R(i, j; \theta)$ denotes the rotation along the oriented $i \wedge j$ plane by the angle θ, and $R(i, j; \theta)x$ signifies the usual action of $SO(n)$ on S^{n-1}. By transposing, K defines a map from the set of probability measures on S^{n-1} to itself, since $K(1) = 1$.

Since the Lie group $SO(n)$ acts on itself, one can also define Kac's walk $\tilde{K}$ on $SO(n)$, given on test functions by

$$(\tilde{K} f)(A) = \frac{1}{\binom{n}{2}} \sum_{i \neq j}^{n} \int_0^{2\pi} f(R(i, j; \theta)A) \frac{1}{2\pi} \, d\theta, \tag{2}$$

where A is any element of $SO(n)$.

It is easy to check that U_{n-1}, the uniform distribution on S^{n-1}, is a stationary distribution for K: for each summand $K_{i,j}$ (without $\frac{1}{\binom{n}{2}}$ in (1)), we have

$$\begin{aligned}
U_{n-1}(K_{i,j} f) &= \int_{S^{n-1}} (K_{i,j} f)(x) U_{n-1}(dx) \\
&= \int_{S^{n-1}} \left(\int_0^{2\pi} f(R(i, j; \theta)x) \frac{1}{2\pi} \, d\theta \right) U_{n-1}(dx) \\
&= \int_{S^{n-1}} \frac{1}{2\pi} \left(\int_0^{2\pi} f(x) \, d\theta \right) U_{n-1}(R(i, j; -\theta) \, dx) \\
&= \int_{S^{n-1}} f(x) U_{n-1}(dx)
\end{aligned}$$

using a change of variable formula and the fact that U_{n-1} is invariant under rotations. This establishes that $U_{n-1} K_{i,j} = U_{n-1}$ for all $i \neq j$. Thus their average $U_{n-1} K = U_{n-1}$ as well.

By a similar argument, or more generally from the theory of random walks on compact groups, we also deduce that the Haar measure is the stationary distribution of $\tilde{K}$.

We further claim that the Markov chain defined by K is aperiodic because once a point is reached, it can be reached in the next step with positive probability density for any rotation. It is also irreducible since along a sequence of rotations $(i_1 \wedge i_2, \dots, i_k \wedge i_{k+1})$ that form a connected spanning graph in K_n, the complete graph on n vertices, one can transport any point on S^{n-1} to any other point with positive probability density; such sequence of rotations certainly occur with positive probability. In fact, by a slightly more involved argument using Hurwitz factorization of $SO(n)$ in terms of Givens' rotations [5], one can show that Kac's random walk on $SO(n)$ is also irreducible, which certainly implies irreducibility on S^{n-1} since the latter is a projection of the former. Furthermore, both chains are recurrent because the state space is compact. Thus by convergence theory of Harris chains, we know that with any initial distribution μ on S^{n-1},

$$\lim_{l \to \infty} \mu K^l(A) - U_{n-1}(A) = 0$$

uniformly in $A \subset S$. This implies convergence in total variation distance by definition.

Using the $\mathcal{L}^2$ theory of discrete-time Markov chains, it can be shown that if the starting distribution μ is in $\mathcal{L}^2(S^{n-1}, U_{n-1})$, then we get the following convergence bound:

$$\|\mu K^l - U_{n-1}\|_{\mathrm{TV}} < \|\mu - 1\|_{\mathcal{L}^2}\left(1 - \frac{1}{2n}\right)^l$$

by the result in [3] and [8], which show that the spectral gap of K is given by $\frac{n+2}{2n(n-1)}$ for $n \geq 2$. See also [6] for an earlier Martingale argument to get $\Omega(1/n)$ spectral gap bound, and [1, 4] for generalizations.

If the initial distribution μ does not have an $\mathcal{L}^2$ density with respect to U_{n-1}, then direct application of the $\mathcal{L}^2$ theory above provides no information. The best result for the rate of convergence when the initial distribution is, say, concentrated at one point is given in [5], where it was shown that at most $\mathcal{O}(n^{2n}\log(\varepsilon^{-1}))$ steps are required to get within ε close to U_{n-1} in total variation distance. The $\mathcal{L}^2$ theory gives a mixing time of $\mathcal{O}(2n\log(\varepsilon^{-1}))\|\mu\|_{\mathcal{L}^2}$.

If we measure convergence of K or $\tilde{K}$ in terms of other probability metrics, most notably $\mathcal{L}^1$ or $\mathcal{L}^2$ transportation cost, then the available convergence rate results are much better. Using comparison techniques, it was shown in [5] that $\mathcal{O}(n^4 \log n)$ steps suffice for Kac's walk on $SO(n)$ to get arbitrarily close to stationarity in $\mathcal{L}^1$ transportation distance, which metrizes weak convergence. This was improved in [10] to an upper bound of $\mathcal{O}(n^{2.5}\log n)$, using a coupling argument. Since the standard projection map $\pi : SO(n) \to S^{n-1}$ can only decrease Riemannian distance, all the transportation mixing time results for $SO(n)$ are also valid for S^{n-1}. This is of course true for total variation mixing as well, but unfortunately we cannot obtain polynomial total variation mixing time for the walk on $SO(n)$.

These suggest that polynomial time mixing should also be true for total variation distance, since there is nothing pathological about the walk. The main difficulty in the analysis lies in that the distribution of the walk at any finite time step will never have a finite $\mathcal{L}^2$ density with respect to the Lebesgue measure on S^{n-1} if we start with the point mass. In the following section, however, we will show that by some removing the singular set of the density after some burn-in period, and using the fact that total variation distance between two measures decreases under the evolution of a Markov chain, one can still essentially use the spectral gap to obtain a polynomial bound on the total variation mixing time. More explicitly, we have the following theorem.

THEOREM 1.1. *Let K denote the Markov kernel for Kac's random walk on the $(n-1)$-sphere, $S^{n-1} \subset \mathbb{R}^n$, let U denote the uniform distribution on S^{n-1}, and let δ_{e_1} denote the probability measure concentrated at the point $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^n$. Then*

$$\|\delta_{e_1} K^t - U\|_{\mathrm{TV}} \le \varepsilon$$

for $t > cn^5 (\log n)^3 \log \varepsilon^{-1}$, where c is a constant that does not depend on n.

REMARK. 1. For a fixed ε, the proof we give below produces a bound with an additional factor of $\log \varepsilon^{-2}$ for the mixing time. Now for general Markov chains on any state space, we have the following sub-multiplicative property ([7], Section 4.4):

$$\bar{d}(s+t) \le \bar{d}(s)\bar{d}(s)$$

for $\bar{d}(s) := \sup_{\mu,\nu} \|\mu K^s - \nu K^s\|_{\mathrm{TV}}$ and $d(t) := \sup_{\mu} \|\mu K^t - \pi\|_{\mathrm{TV}} \le \bar{d}(t) \le 2d(s)$. We deduce that $d(tk) \le (2d(t))^k$, hence $\tau_{\mathrm{mix}}(\varepsilon) \le \log_2(1/\varepsilon)\tau_{\mathrm{mix}}(1/4)$, that is, the additional factor can be removed.

2. Very recently, I learned that Aaron Smith [11] came up with a coupling argument based on Wasserstein contraction that gets the correct order $\mathcal{O}(n \log n)$ of total variation mixing time for the Gibbs sampler on the n-simplex. Since Kac's walk on the sphere is in fact a Gibbs sampler on the n-simplex if one squares the coordinates, at least if one starts with a measure symmetric under the transform $\vec{x} \to -\vec{x}$, his argument presumably gives the same result here as well. But I believe the argument presented here is of independent interest, especially in comparison analysis, for which transportation mixing time bound might not be available.

3. As mentioned above, we are unable to get any polynomial mixing time result for Kac's walk on $SO(n)$. But in fact, even for the induced walk on the Grassmanian space, $SO(n)/SO(n-k)$ where $k \ge 2$, polynomial mixing is beyond reach at the moment. The difficulty of applying the present technique is that the support of the running distribution cannot be confined into nice submanifolds of the state space for $k \ge 2$, thus an induction based on the dimension of the support does not work.

4. Another line of research is concerned with entropy mixing time of Kac's random walk (see [2] and references therein). In order for entropy distance to go down to zero, the starting measure has to have a density with finite relative entropy with respect to the uniform measure on S^{n-1}. It is not clear whether starting at a point mass the chain will have finite entropy in finite time.

2. Bounding the total variation distance. This section gives bounds on the convergence rate of Kac's random walk on S^{n-1} starting at a standard basis vector e_i, in total variation distance.

Recall the total variation distance between two probability measures μ and ν on the same probability space $(S, \mathcal{S})$ is defined by the following variational quantity:

$$\|\mu - \nu\|_{\mathrm{TV}} = 2 \sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)|,$$

where $\mathcal{S}$ is the σ-algebra on S.

Alternatively, total variation has the variational characterization in terms of bounded functions:

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{f : \|f\|_\infty \le 1} |\mu(f) - \nu(f)|.$$

This will be used to show the weakly contracting property of Markov chains under total variation distance below.

Let A_k be the event that at the kth step of the walk, every pair of coordinates has been used. Then we have

$$P(A_k^c) := \eta_k < \binom{n}{2}\left(1 - \frac{1}{\binom{n}{2}}\right)^k.$$

Conditioning on this event, we have the following two claims:

CLAIM 1. *The density $g := \frac{d\mu'_k}{dU_{n-1}}$ of the resulting distribution μ'_k of the conditioned random walk with respect to the uniform distribution on S^{n-1} satisfies the following bound*:

$$g(x) \le \Bigl|\min_{1 \le i \le n} x_i\Bigr|^{-n} \left(\sum_{i=1}^n (-\log|x_i|)^k\right) C^k \prod_{m=1}^k m! \tag{3}$$

$$\le C^k k^{k^2} \Bigl|\min_{1 \le i \le n} x_i\Bigr|^{-n} \Bigl(-\log\Bigl|\min_{1 \le i \le n} x_i\Bigr|\Bigr)^k =: C(n,k) \tag{4}$$

for some fixed absolute constant C.

CLAIM 2. *For $k > -n^2 \log n \log \varepsilon$, and $\varepsilon < n^{-3}$, the set $H_\varepsilon := \{x \in S^{n-1} : |x_i| < \varepsilon$ for some $i\}$ satisfies the following bound on its probability under the A_k-conditional distribution*:

$$\mu'_k(H_\varepsilon) \le \varepsilon^{1/8}. \tag{5}$$

Let us first show how claims 1 and 2 lead to a polynomial time convergence rate for Kac's walk under total variation norm. Let μ_k be the distribution on S^{n-1} after k steps of the random walk, and let μ'_k be μ_k conditional on A_k, that is, for $B \subset S^{n-1}$,

$$\mu'_k(B) = P(\delta_{e_1} R^k \in B | A_k),$$

where R is the one-step transition kernel of Kac's random walk.

Then we have

$$\|\mu'_k - \mu_k\|_{\mathrm{TV}} < \eta_k < \binom{n}{2}\left(1 - \frac{1}{\binom{n}{2}}\right)^k. \tag{6}$$

To check this, let $B \subset S^{n-1}$ be Lebesgue measurable. Then we have

$$\begin{aligned}\mu_k(B) &= P(\delta_{e_1} R^k \in B | A_k) P(A_k) + P(\delta_{e_1} R^k \in B | A_k^c) P(A_k^c) \\ &\le \mu'_k(B) + \eta_k.\end{aligned}$$

This implies

$$\mu_k(B) - \mu'_k(B) \le \eta_k.$$

On the other hand, since

$$\mu'_k(B) = \frac{P(\{\delta_{e_1} R^k \in B\} \cap A_k)}{P(A_k)},$$

we also get

$$\frac{\mu_k(B)}{1 - \eta_k} > \mu'_k(B)$$

which gives

$$\mu_k(B) > \mu'_k(B) - \eta_k \mu'_k(B)$$

hence

$$\mu'_k(B) - \mu_k(B) < \eta_k$$

which establishes (6).

Next recall that a Markov kernel is weakly contracting in total variation norm because if f is a bounded continuous function on the state space with

$$\|f\|_\infty \le 1,$$

then $Rf(x) = \int R(x, dy) f(y)$ satisfies the same $\mathcal{L}^\infty$ bound, hence

$$(\mu R - \nu R)(f) = (\mu - \nu)(Rf) \le \|\mu - \nu\|_{\mathrm{TV}}.$$

Thus by the triangle inequality we just need to bound $\|\mu'_k R^l - U_{n-1}\|_{\mathrm{TV}}$ from now on, where U_{n-1} denotes the uniform distribution on S^{n-1}, and at the end add η_k to the resulting bound.

Next we modify μ'_k to a different distribution ν_k as follows. We define ν_k in terms of its density with respect to U_{n-1}.

On the set H_ε^c,

$$\frac{d\nu_k}{dU_{n-1}} := \frac{d\mu'_k}{dU_{n-1}}.$$

On the set H_ε, we let its density be a constant equal to the mass of H_ε under μ'_k divided by its mass under U_{n-1}, which is what's needed for ν_k to be a probability distribution on S^{n-1}; we invoke Claim 1 above to get an upper bound on this constant:

$$\begin{aligned}\frac{d\nu_k}{dU_{n-1}} &\equiv \frac{\mu'_k(H_\varepsilon)}{U_{n-1}(H_\varepsilon)} \\ &< \frac{\varepsilon^{1/4}}{\varepsilon(\Gamma(n/2))/(\Gamma((n-1)/2)\Gamma(1/2))} \\ &< \frac{\varepsilon^{1/4}}{\varepsilon\sqrt{(n-2)/2\pi}} \\ &< \varepsilon^{-3/4}\sqrt{\frac{2\pi}{n-2}}.\end{aligned}$$

In the computation above we used two ingredients. First we used that

$$\frac{\Gamma(n/2)}{\Gamma((n-1)/2)} > \sqrt{\frac{n-2}{2}} \tag{7}$$

which follows from log convexity of the Γ function. Since $\frac{1}{2}(\log\Gamma(n) + \log\Gamma(n-1)) > \log\Gamma(n-1/2)$, we get

$$\frac{\Gamma(n)}{\Gamma(n-1/2)} > \frac{\Gamma(n-1/2)}{\Gamma(n-1)},$$

which implies (7) above. By incrementing n by $1/2$, we also get a reverse inequality of the form

$$\frac{\Gamma(n/2)}{\Gamma((n-1)/2)} < \sqrt{\frac{n-1}{2}}. \tag{8}$$

This will be useful later when we bound $U(H_\varepsilon)$ in the proof of Claim 2.

The second ingredient is the formula for the coordinate marginal density for the uniform distribution on the sphere (see [5] but with a small typo, namely by n-sphere they meant $(n-1)$-sphere):

$$\frac{d}{da}\mathbb{P}_U(x_1 \in [-1, a]) = \frac{\Gamma((n+1)/2)}{\Gamma(1/2)\Gamma(n/2)}(1-a^2)^{(n-2)/2}, \tag{9}$$

where P_U denotes uniform distribution on S^{n-1}.

The total variation distance between μ'_k and ν_k is given simply by their total variation distance over the region H_ε, hence we have

$$\|\mu'_k - \nu_k\|_{\mathrm{TV}} \le \mu'_k(H_\varepsilon) + \frac{n\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)}\varepsilon \tag{10}$$

$$\le n^{3/2}\varepsilon + \varepsilon^{1/8}. \tag{11}$$

Thus by choosing ε sufficiently small, whose exact value we will determine in the end, we can make sure that μ'_k and ν_k are very close in total variation distance. And again by weak contractivity of Markov kernel, we now simply need to focus on bounding $\|\nu_k R^l - U_{n-1}\|_{\mathrm{TV}}$. Since ν_k has an $\mathcal{L}^2$ density with respect to U_{n-1}, we can use the spectral gap to bound the rate of convergence. First we bound the $\mathcal{L}^2(dU_{n-1})$ distance between ν_k and U_{n-1}:

$$\begin{aligned} &\|\nu_k - U_{n-1}\|_{\mathcal{L}^2(dU_{n-1})} \\ &\quad = \left(\int_{H_\varepsilon}\left|\frac{d\nu_k}{dU_{n-1}} - 1\right|^2 dU_{n-1} + \int_{H_\varepsilon^c}\left|\frac{d\nu_k}{dU_{n-1}} - 1\right|^2 dU_{n-1}\right)^{1/2}. \end{aligned} \tag{12}$$

Let us bound the two integrals separately.

For the first integral on the right-hand side of (12), we have

$$\begin{aligned} \int_{H_\varepsilon}\left|\frac{d\nu_k}{dU_{n-1}} - 1\right|^2 dU_{n-1} &\le \int_{H_\varepsilon}\left(\frac{d\nu_k}{dU_{n-1}}\right)^2 dU_{n-1} + U_{n-1}(H_\varepsilon) \\ &< \varepsilon^{-3/2}\frac{8\pi}{n-2}\varepsilon\frac{\Gamma(n/2)}{\Gamma((n-1)/2)\Gamma(1/2)} \\ &< 4\varepsilon^{-1/2}\sqrt{\frac{2\pi}{n-2}}. \end{aligned} \tag{13}$$

For the second integral, notice first that H_ε^c is the set of points on S^{n-1} for which all the coordinates are greater than ε. So Claim 2 tells us that the density $\frac{d\nu_k}{dU_{n-1}}$ over this region is bounded above by ε^{-n}, from which we immediately get the following bound:

$$\int_{H_\varepsilon^c}\left|\frac{d\nu_k}{dU_{n-1}} - 1\right|^2 dU_{n-1} < \varepsilon^{-2n} + 1. \tag{14}$$

Combining (13) and (14), we get, for $\varepsilon < \frac{1}{2}$ and $n > 2$, say, that

$$\|\nu_k - U_{n-1}\|_{\mathcal{L}^2(dU_{n-1})} \le 2\varepsilon^{-n}.$$

By the results in [3], we know that the spectral gap of the Kac kernel is $\frac{1}{2n}$, so we get

$$\begin{aligned} \|\nu_k R^l - U_{n-1}\|_{\mathrm{TV}} &\le \left\| \frac{d\nu_k}{dU_{n-1}} - 1 \right\|_{\mathcal{L}^2(dU_{n-1})} \left(1 - \frac{1}{2n}\right)^m \\ &\le 2\varepsilon^{-n}\left(1 - \frac{1}{2n}\right)^m. \end{aligned} \tag{15}$$

Finally, combining (6) (10) and (15), we get

$$\begin{aligned} \|\delta_{e_1} R^{k+l} - U_{n-1}\|_{\mathrm{TV}} \le{}& \binom{n}{2}\left(1 - \frac{1}{\binom{n}{2}}\right)^k + n^{3/2}\varepsilon + \varepsilon^{1/8} \\ &+ C^k k^{k^2} |\varepsilon|^{-n} (-\log \varepsilon)^k \left(1 - \frac{1}{2n}\right)^l. \end{aligned} \tag{16}$$

So it remains to minimize the right-hand side of (16) with respect to k and l.

Suppose our target total variation distance is 3δ. Then we can simply divide 3δ into three equal parts and bound each summand in (16) by δ. We look at each summand below:

Bounding the first summand yields

$$\binom{n}{2}\left(1 - \frac{1}{\binom{n}{2}}\right)^k < \delta \quad \Rightarrow \quad k > (-\log\delta + 2\log n)\binom{n}{2}.$$

So it suffices to take

$$k > n^2 \log n \log \frac{1}{\delta}. \tag{17}$$

Bounding the second summand $\varepsilon^{1/8} + n^{3/2}\varepsilon < \delta$, it suffices to have $\varepsilon^{1/8} < \delta/2$ and $n^{3/2}\varepsilon < \delta/2$, which gives

$$\varepsilon < \tfrac{1}{2}\delta^8 n^{-3/2}.$$

But taking $\varepsilon = n^{-3}\delta^8$ certainly fulfills that, which will affect the bound on l in the third summand:

$$C^k k^{k^2} |\varepsilon|^{-n} (-\log\varepsilon)^k \left(1 - \frac{1}{n}\right)^l < \delta$$

implies we need l greater than

$$\begin{aligned} &2n\bigl(-\log\delta + k\log C + k^2 \log k - n\log\varepsilon + k\log\log(\varepsilon^{-1})\bigr) \\ &\quad < n\bigl(-\log\delta + k\log C \\ &\qquad + n^4(\log n)^2(\log\delta)^2\bigl(2\log n + \log\log n + \log\log(\delta^{-1})\bigr) \\ &\qquad\qquad + n(-8\log\delta + 3\log n) + k\log\log\varepsilon^{-1}\bigr) \\ &\quad < C' n^5 (\log n)^3 (\log\delta)^3 \end{aligned}$$

for some constant C'.

Clearly l dominates k, so it requires a total of $C'n^5(\log n)^3(\log\delta)^3$ steps to bring the running distribution of Kac's random walk to be 3δ close to its stationary distribution on the unit sphere S^{n-1}.

Finally we prove the two claims introduced in the beginning.

3. Proof of Claim 1. Starting at the delta mass at e_1, an admissible sequence of rotations in A_k will distribute it over the entire S^{n-1} with positive probability everywhere provided that $P(A_k) > 0$, that is, for sufficiently large k. This will certainly be the case if $k \geq -n^2 \log n \log\delta$ for $-\log\delta > 2$. So we will look at the conditional probability density given that the walk has taken a sequence of steps in A_k, and we will estimate the density growth from step $j-1$ to j, up to step k.

Observe that at step $j-1$, $j \leq k$, the support of the running distribution is a subsphere of S^{n-1}. Without loss of generality, we call this subsphere S^m. Denote by u_j, v_j the axes that span the plane along which the rotation γ_j takes place.

The way γ_j affects the previous running distribution can be classified into three cases:

1. $u_j, v_j \notin S^m$, in which case the running distribution remains unchanged.

2. $u_j, v_j \in S^m$, in which case the support after the rotation is still on S^m, but the density might change.

3. $u_j \in S_m$, $v_j \notin S_m$, in which case the support of the running distribution grows to be a sphere with one dimension higher than S^m, denoted without loss of generality S^{m+1}.

Case 1 clearly does not increase the density of the running distribution, because the rotation does not take S^m outside itself and for $\theta \in [0, 2\pi]$, the density at $(x_1, \dots, x_{m+1}, \dots, (u_j^2 + v_j^2)^{1/2}\cos\theta, \dots, (u_j^2 + v_j^2)^{1/2}\sin\theta, \dots, x_n)$ with respect to U_m only depends on the first $m+1$ coordinates, which means that averaging over θ uniformly in $[0, 2\pi]$ remains the same.

To understand Case 3, first observe that there can be at most n such steps in the history of the Kac walk. So if we can show each type 3 rotation increases the density by at most $|\min_{1\leq i\leq n} x_i|^{-1}$, then the factor $|\min_{1\leq i\leq n} x_i|^{-n}$ would be taken care of. This is the content of the following lemma.

LEMMA 3.1. *Assuming the running density $h_m(x_1, \dots, x_{m+1})$ with respect to U_m after step $j-1$ is bounded by $g_m(x_1, \dots x_{m+1})$, and that without loss of generality $u_j = x_{m+1}$, $v_j = x_{m+2}$, then the new density $h_{m+1}(x_1, \dots, x_{m+2})$ with respect to U_{m+1} after step j is bounded by*

$$\frac{1}{2\pi} g_m\big(x_1, \dots, (x_{m+1} + x_{m+2})^{1/2}\big)(x_{m+1}^2 + x_{m+2}^2)^{-1/2}.$$

Observe that $(x_{m+1}^2 + x_{m+2}^2)^{-1/2} \leq |\min_{1\leq i\leq n} x_i|^{-1}$.

PROOF. Denote the new density with respect to U_{m+1} by $h_{m+1}(x_1, \dots, x_{m+2})$ with a slight abuse of notation. Then we have

$$h_{m+1}\big(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \cos\theta, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \sin\theta\big)$$

is independent of θ and in particular equals

$$h_{m+1}\big(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2}, 0\big).$$

Furthermore, the total contribution of density from $(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \times \cos\theta, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \sin\theta)$ for all θ should add up to the previous density at the point $(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2})$. In other words,

$$\begin{aligned}&(x_{m+1}^2 + x_{m+2}^2)^{1/2}\\ &\quad\times \int_{\theta=0}^{2\pi} h_{m+1}\big(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \cos\theta, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \sin\theta\big)\, d\theta\\ &\quad= h_m\big(x_1, \dots, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2}\big).\end{aligned}$$

Notice that the factor $(x_{m+1}^2 + x_{m+2}^2)^{1/2}$ accounts for the measure of the circle $\{(y_1, \dots, y_{m+2}, 0, \dots, 0)$: with $y_1 = x_1, \dots, y_m = x_m$ and $y_{m+1}^2 + y_{m+2}^2 = x_{m+1}^2 + x_{m+2}^2\}$, over which we aggregate.

Thus we get

$$\begin{aligned}&h_{m+1}\big(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \cos\theta, (x_{m+1}^2 + x_{m+2}^2)^{1/2} \sin\theta\big)\\ &\quad= \frac{1}{2\pi}(x_{m+1}^2 + x_{m+2}^2)^{-1/2} h_m\big(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2}\big)\\ &\quad\le \frac{1}{2\pi}(x_{m+1}^2 + x_{m+2}^2)^{-1/2} g_m\big(x_1, \dots, (x_{m+1}^2 + x_{m+2}^2)^{1/2}\big).\end{aligned}$$ □

The Case 2 rotations will contribute the remaining factors in the bound of $g(x)$ in Claim 1. More precisely, we have the following lemma.

LEMMA 3.2. *Assume at step $j-1$, the running distribution is supported on some $S^m \subset S^{n-1}$, which is viewed as the standard sphere in $\mathbb{R}^{m+1} = \{x_1, \dots, x_{m+1}\}$, and that the density g_j with respect to U_m satisfies*

$$\begin{aligned}&g_{j-1}(x_1, \dots, x_{m+1})\\ &\quad\le C(j, m)(a_1^2 + b_1^2)^{-1/2} \cdots (a_{m-1}^2 + b_{m-1}^2)^{-1/2}\\ &\qquad\times [(-\log|x_1|)^{j-1} + \cdots + (-\log|x_{m+1}|)^{j-1}],\end{aligned} \tag{18}$$

where $C(j, m)$ is a constant that varies with j and m. Here $a_i \neq b_i$ for each i and $(a_1, b_1), \dots, (a_{m-1}, b_{m-1})$ are pairs in $\{x_1, \dots, x_{m+1}\}^2$ with the property that no two pairs are the same and each coordinate appears at most twice.

If furthermore the j*th rotation is as in Case* 2, *then the new density bound takes the form*

$$\begin{aligned} &g_j(x_1, \dots, x_{m+1}) \\ &\quad \le 512C(j,m)(j+1)!(a_1^2+b_1^2)^{-1/2}\cdots(a_{m-1}^2+b_{m-1}^2)^{-1/2} \\ &\qquad \times [(-\log|x_1|)^j + \cdots + (-\log|x_{m+1}|)^j] \end{aligned}$$

with possibly a different sequence of (a_i, b_i) *satisfying the same property as above.*

Notice that starting with a density satisfying the bound (18), a type 1 or type 3 rotation would preserve its form, with j replaced by $j+1$. Type 1 rotation does that trivially, due to the fact that the polylogarithmic factor always increases with j. Type 2 rotation introduces an additional factor of $(a_m^2+b_m^2)^{-1/2}$, but decreases the other existing factors, hence also preserves the bound with $j \to j+1$.

PROOF. Without loss of generality assume $(u_j, v_j) = (1, 2)$.

The new density h' is obtained from the old density h by averaging over $\theta \in [0, 2\pi]$ of $h(R(1,2,\theta)x)$, where $R(1,2,\theta)x$ denotes the rotation of the vector $x \in S^m$ by angle θ along $x_1 \wedge x_2$. In formula, we have

$$h'(x) = \frac{1}{2\pi}\int_0^{2\pi} h(R(1,2,\theta)x)\,d\theta. \tag{19}$$

We write the bound (18) as a sum of $m+1$ terms and consider one of the terms

$$g_i(x) = C(a_1^2+b_1^2)^{-1/2}\cdots(a_{m-1}^2+b_{m-1}^2)^{-1/2}(-\log|x_i|)^{j-1}.$$

By assumption, at most two elements in $a_1, b_1, \dots, a_{m-1}, b_{m-1}$ equal x_1 and at most two other elements equal x_2.

By the circle averaging formula (19), we have

$$g_i'(x) = \frac{1}{2\pi}\int_0^{2\pi} g\big((x_1^2+x_2^2)^{1/2}\cos\theta, (x_1^2+x_2^2)^{1/2}\sin\theta, x_3, \dots, x_{m+1}\big)\,d\theta.$$

We shall break the integral into two parts, where the range of integration is over $I_{\cos} = [0, \pi/4] \cup [3\pi/4, 5\pi/4] \cup [7\pi/4, 2\pi]$ and its complement $I_{\sin}$ in $[0, 2\pi]$, respectively; that is, the ranges are where $\cos\theta$ is close to 1 and $\sin\theta$ is close to 1, respectively. By symmetry, we just have to deal with the integral over the range $\theta \in I_{\sin}$, and multiply the final bound by 1 in the end.

First we look at the case when $i \notin \{1, 2\}$, which means the rotation $(1, 2)$ does not affect the logarithmic factor $(-\log|x_i|)^j$ at the end. In this case, all the factors in $g_i(x)$ of the form $(x_2^2 + x_s^2)^{-1/2}$ that involve x_2 but not x_1 upon the rotation $R(1,2,\theta)$ become $((x_1^2+x_2^2)\sin^2\theta + x_s^2)^{-1/2}$, which can be bounded above by $\sqrt{2}(x_1^2+x_2^2+x_s^2)^{-1/2}$.

As of the factors that involve both x_1 and x_2, that is, $(x_1^2+x_2^2)^{-1/2}$, there can be at most one of such. And it remains the same under the rotation $R(1,2,\theta)$ since $(x_1^2+x_2^2)\cos^2\theta+(x_1^2+x_2^2)\sin^2\theta = x_1^2+x_2^2$.

The factors that involve x_s and x_s, $s\neq 2$, become $((x_1^2+x_2^2)\cos^2\theta+x_s^2)^{-1/2}$, which we can bound as follows:

Using the fact that $\frac{1}{\sqrt{2}}(|a|+|b|)\le (a^2+b^2)^{1/2}\le |a|+|b|$, we get

$$\big((x_1^2+x_2^2)\cos^2\theta+x_s^2\big)^{-1/2}\sim [(|x_1|+|x_2|)|\cos\theta|+|x_s|]^{-1},$$

where $a\sim b$ means $b/C\le a\le bC$ for some constant C. Here we can take C to be 2.

More difficult is the case when $i\in\{1,2\}$, when we also have to deal with a $(-\log[(x_1^2+x_2^2)^{-1/2}\cos\theta])^{j-1}$ factor that goes to infinity for $\theta\in I_{\sin}$.

In fact when $i=1$, the only factors that have singularities for $\theta\in I_{\sin}$ and for the coordinates bounded away from 0 take the following form:

$$\big((|x_1|+|x_2|)|\cos\theta|+|x_s|\big)^{-1}\big((|x_1|+|x_2|)|\cos\theta|+|x_t|\big)^{-1} \\ \times \big(-\log[(x_1^2+x_2^2)^{-1/2}\cos\theta]\big)^{j},$$

where $s\neq t$, or without the x_t factor. In the former case we will show in Lemma 3.3 below that the following integral:

$$\frac{1}{2\pi}\int_{\theta\in I_{\sin}}\big((|x_1|+|x_2|)|\cos\theta|+|x_s|\big)^{-1}\big((|x_1|+|x_2|)|\cos\theta|+|x_t|\big)^{-1} \\ \times \big(-\log[(x_1^2+x_2^2)^{1/2}\cos\theta]\big)^{j-1}\,d\theta$$

is bounded by

$$j!(x_1^2+x_2^2)^{-1/2}(x_s^2+x_t^2)^{-1/2} \\ \times [(-\log|x_1|)^j+(-\log|x_2|)^j+(-\log|x_s|)^j+(-\log|x_t|)^j] \tag{20}$$

whereas in the case where the x_t factor is not present, the same bound (20) multiplied by $\sqrt{2}$ applies the expression

$$\frac{1}{2\pi}\int_{\theta\in I_{\sin}}\big((|x_1|+|x_2|)|\cos\theta|+|x_s|\big)^{-1}\big(-\log[(x_1^2+x_2^2)^{1/2}\cos\theta]\big)^{j-1}\,d\theta \tag{21}$$

using the fact that for $\theta\in I_{\sin}$,

$$\big((|x_1|+|x_2|)\cos\theta+|x_t|\big)^{-1}\ge 1/\sqrt{2}.$$

When $i\neq 1$, the logarithmic singularity will not arise when integrating over $\theta\in I_{\sin}$, so it will trail off as a remaining factor of the form $(-\log|x_i|)^{j-1}\le 1+(-\log|x_i|)^j$.

Recall also that we have factors of the form

$$2(x_1^2+x_2^2+x_s^2)^{-1/2}(x_1^2+x_2^2+x_t^2)^{-1/2} \tag{22}$$

coming from the uniform bound on the factors involving x_2 but not x_1; here s, t are possibly different indices than those appearing in the singular factors. Equation (22) can be trivially bounded above by $2(x_1^2 + x_s^2)^{-1/2}(x_2^2 + x_t^2)^{-1/2}$. The remaining inverse factors in $g(R(1, 2, \theta)x)$ do not contain x_1 or x_2, so one can easily check that the inductive hypothesis is satisfied.

The best way to visualize this branching inductive argument is to consider a simple, possibly disconnected graph on $m+1$ vertices with degrees bounded above by 2. The edges between i and j represent a factor of the form $(x_i^2 + x_j^2)^{-1/2}$ in the bound on the density. A rotation in the $x_1 \wedge x_2$ plane has the effect of producing two new graphs on $m+1$ vertices, and the density bound we get will be a sum over all the resulting graphs. Without loss of generality let us describe one of those two descendant graphs, the one associated with x_1.

There will be edges $(1, 2)$, $(3, 4)$, $(1, 3)$ and $(2, 4)$ if x_3 and x_4 were incident to x_1 in the previous graph, or simply $(1, 2)$ when x_1 only has degree 1. If x_1 had degree 0, then it remains isolated in the x_1 component of the descendant graph. In the process of this rewiring, some logarithmic factors $(\log |x_s|)^j$ and factorial factors $j!$ are also introduced, namely, if $(-\log |x_3|)^{j-1}$ or $(-\log |x_4|)^{j-1}$ was a factor in the bound for the previous step running disribution, then the new bound will have $j!(-\log |x_4|)^j$. If there is originally a log factor of other coordinates, then the exponent on that factor remains the same.

It remains to prove the bound (20), and notice that we only need to prove it for $\theta \in [\pi/4, \pi/2]$ and then multiply the resulting bound by 4. This is given by the following technical lemma. $\square$

LEMMA 3.3. *For* $0 \le x_t, x_s, 0 \le x_1, x_2$,

$$\begin{aligned}\int_0^{\pi/4} &((x_1 + x_2)\sin\theta + x_s)^{-1}((x_1 + x_2)\sin\theta + x_t)^{-1} \\ &\quad \times (-\log[(x_1 + x_2)\sin\theta])^{j-1}\, d\theta \\ &\le 4(j+1)!(x_s + x_t)^{-1}(x_1 + x_2)^{-1} \\ &\quad \times [(-\log x_1)^j + (-\log x_2)^j + (-\log x_s)^j + (-\log x_t)^j].\end{aligned}$$

REMARK. Note this is equivalent to the bound (20), by replacing sin with cos and changing the range of integration to $[\pi/4, \pi/2]$.

PROOF. Without loss of generality, we can assume $x_t \le x_s$. Furthermore, we can replace $\sin\theta$ by its linearization at 0, and multiply the resulting bound by 2 in the end, since for $\theta \in [0, \pi/4]$, we have $\theta/2 \le \sin\theta \le 2\theta$. So instead we just need to prove

$$\int_0^1 \big((x_1+x_2)\theta+x_s\big)^{-1}\big((x_1+x_2)\theta+x_t\big)^{-1}\big(-\log[(x_1+x_2)\theta]\big)^{j-1}\,d\theta$$
$$\le 4\pi(j+1)!(x_s+x_t)^{-1}(x_1+x_2)^{-1}$$
$$\times [(-\log x_1)^j+(-\log x_2)^j+(-\log x_s)^j+(-\log x_t)^j].$$

First of all, the factor $((x_1+x_2)\theta+x_s)^{-1}$ can be bounded above by $2(x_s+x_t)^{-1}$ for $\theta \in [0,1]$. So it remains to bound the integral of the remaining factors:

$$\int_0^1 \big((x_1+x_2)\theta+x_t\big)^{-1}\big(-\log[(x_1+x_2)\theta]\big)^{j-1}\,d\theta$$
$$\le x_t^{-1}\int_0^{\varepsilon}\big(-\log[(x_1+x_2)\theta]\big)^{j-1}\,d\theta$$
$$+\big(-\log[(x_1+x_2)\varepsilon]\big)^{j-1}\int_{\varepsilon}^1\big((x_1+x_2)\theta+x_t\big)^{-1}\,d\theta$$
$$\le x_t^{-1}j!\varepsilon\big(-\log[(x_1+x_2)\varepsilon]\big)^{j}$$
$$+\big(-\log[(x_1+x_2)\varepsilon]\big)^{j-1}(x_1+x_2)^{-1}\log\left[\frac{x_1+x_2+x_t}{(x_1+x_2)\varepsilon+x_t}\right]$$
$$\le x_t^{-1}j!\varepsilon\big(-\log[(x_1+x_2)\varepsilon]\big)^{j}$$
$$+\big(-\log[(x_1+x_2)\varepsilon]\big)^{j-1}(x_1+x_2)^{-1}\log[(x_1+x_2)\varepsilon]$$
$$= x_t^{-1}j!\varepsilon\big(-\log[(x_1+x_2)\varepsilon]\big)^{j}+\big(-\log[(x_1+x_2)\varepsilon]\big)^{j}(x_1+x_2)^{-1}.$$

In the second equality we used the fact that

$$\int_0^{\varepsilon}(-\log\theta)^j\,d\theta=\int_{-\log\varepsilon}^{\infty}y^je^{-y}\,dy\le j!$$

and in the third inequality we used $\frac{\varepsilon(x_1+x_2)+x_t}{x_1+x_2+x_t} > \varepsilon(x_1+x_2)$ for $\varepsilon < 1$.

Taking $\varepsilon = x_t$, we obtain the result. □

4. Proof of Claim 2. We prove the claim by a contradiction argument. Here we use the result from [9] that after $k = n^2\log n\log\varepsilon$ steps the $\mathcal{L}^2$ transportation distance between the running distribution of the Kac random walk on S^{n-1} and the uniform distribution U_{n-1} is less than ε. So by Holder's inequality, the $\mathcal{L}^1$ transportation distance is also less than ε. We know that the uniform measure $U_{n-1}(H_\varepsilon)$ varies linearly with ε; in fact using the marginal density formula (9) for a single coordinate on the unit sphere, together with the fact that $H = \bigcup H_\varepsilon^i$ where $H_\varepsilon^i := \{x \in S^{n-1} : |x_i| \le \varepsilon\}$, one sees that it is bounded above by $n^{3/2}\varepsilon$, and similarly $U_{n-1}(H_{\varepsilon^\alpha+\varepsilon}) \le 2n^{3/2}(\varepsilon^\alpha+\varepsilon)$. Next let α, β be two real numbers between 0 and 1 satisfying

$$\alpha+\beta<1$$

and

$$\alpha - \beta > 1/2.$$

Then with $\varepsilon \le n^{-3}$, one verifies easily that

$$(\varepsilon^{\beta} - (\varepsilon^{\alpha} + \varepsilon)n^{3/2})\varepsilon^{\alpha} > \varepsilon. \tag{23}$$

So if $\mu_k(H_\varepsilon) > \varepsilon^{\beta}$, with $\varepsilon \le n^{-3}$, then in order to transport the mass of H_ε under μ_k in excess of $H_{\varepsilon+\varepsilon^{\alpha}}$ under U_{n-1}, the left-hand side of (23) gives a lower bound on the transportation cost for that alone, because each particle of mass originally in H_ε must traverse at least a distance of ε^{α} to go outside of $H_{\varepsilon+\varepsilon^{\alpha}}$. Since the *total* transport cost cannot exceed ε after k steps, this is a contradiction. Hence we must have $\mu_k(H_\varepsilon) < \varepsilon^{\beta}$. One set of choices for α and β is $\alpha = 3/4$ and $\beta = 1/8$, which is the content of Claim 2.

Acknowledgments. I would like to thank my advisor Persi Diaconis for introducing me to this problem and pointing me to the relevant literature. I would also like to thank the referee for suggesting numerous changes.

REFERENCES

[1] CAPUTO, P. (2008). On the spectral gap of the Kac walk and other binary collision processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 205–222. MR2429910

[2] CARLEN, E. A., CARVALHO, M. C., LE ROUX, J., LOSS, M. and VILLANI, C. (2010). Entropy and chaos in the Kac model. *Kinet. Relat. Models* **3** 85–122. MR2580955

[3] CARLEN, E. A., CARVALHO, M. C. and LOSS, M. (2003). Determination of the spectral gap for Kac's master equation and related stochastic evolutions. *Acta Math.* **191** 1–54. MR2020418

[4] CARLEN, E. A., GERONIMO, J. S. and LOSS, M. (2008). Determination of the spectral gap in the Kac model for physical momentum and energy-conserving collisions. *SIAM J. Math. Anal.* **40** 327–364. MR2403324

[5] DIACONIS, P. and SALOFF-COSTE, L. (2000). Bounds for Kac's master equation. *Comm. Math. Phys.* **209** 729–755. MR1743614

[6] JANVRESSE, E. (2001). Spectral gap for Kac's model of Boltzmann equation. *Ann. Probab.* **29** 288–304. MR1825150

[7] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[8] MASLEN, D. K. (2003). The eigenvalues of Kac's master equation. *Math. Z.* **243** 291–331. MR1961868

[9] OLIVEIRA, R. I. (2009). On the convergence to equilibrium of Kac's random walk on matrices. *Ann. Appl. Probab.* **19** 1200–1231. MR2537204

[10] SIDENKO, S. (2008). Kac's random walk and coupon collector's process on poset. Ph.D. thesis, Massachusetts Inst. Technology. MR2717533

[11] SMITH, A. (2011). A Gibbs sampler on the n-simplex. Available at arXiv:1107.5829.

126 BLACKWELDER CT., APT 605A
STANFORD, CALIFORNIA 94305
USA
E-MAIL: jyj@math.stanford.edu