KAC'S WALK ON n-SPHERE MIXES IN $n \log n$ STEPS

BY NATESH S. PILLAI[1] AND AARON SMITH[2]

Harvard University and University of Ottawa

Determining the mixing time of Kac's random walk on the sphere S^{n-1} is a long-standing open problem. We show that the total variation mixing time of Kac's walk on S^{n-1} is between $\frac{1}{2}n \log(n)$ and $200n \log(n)$ for all n sufficiently large. Our bound is thus optimal up to a constant factor, improving on the best-known upper bound of $O(n^5 \log(n)^2)$ due to Jiang [*Ann. Appl. Probab.* **22** (2012) 1712–1727]. Our main tool is a "non-Markovian" coupling recently introduced by the second author in [*Ann. Appl. Probab.* **24** (2014) 114–130] for obtaining the convergence rates of certain high dimensional Gibbs samplers in continuous state spaces.

1. Introduction. In his 1954 paper [13], Marc Kac introduced a random walk on a sphere as a model for a Boltzmann gas. Kac's walk on the $(n-1)$-sphere,

$$S^{n-1} = \left\{ X \in \mathbb{R}^n : \sum_{i=1}^{n} X[i]^2 = 1 \right\},$$

is a discrete-time Markov chain $\{X_t \in S^{n-1}\}_{t \geq 0}$ that evolves as follows. At every step t, choose two coordinates $1 \leq i_t < j_t \leq n$ and an angle $\theta_t \in [0, 2\pi)$ uniformly at random and set

$$
\begin{aligned}
\begin{pmatrix} X_{t+1}[i_t] \\ X_{t+1}[j_t] \end{pmatrix} &= \begin{bmatrix} \cos(\theta_t) & -\sin(\theta_t) \\ \sin(\theta_t) & \cos(\theta_t) \end{bmatrix} \begin{pmatrix} X_t[i_t] \\ X_t[j_t] \end{pmatrix}, \\
X_{t+1}[k] &= X_t[k], \qquad k \notin \{i_t, j_t\}.
\end{aligned}
$$

(1.1)

Let $F : \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \times [0, 2\pi) \times S^{n-1} \mapsto S^{n-1}$ be the map associated with this representation, so that $X_{t+1} = F(i_t, j_t, \theta_t, X_t)$. Physically, Kac motivated this random walk by considering the velocities of n particles in a one-dimensional box. He assumed that these particles were uniformly distributed in space, and the random walk X_t models the change in their velocities over time as collisions occur. The condition that X_t be constrained to a sphere corresponds to the principle of

conservation of energy. Understanding the mixing properties of this process is central to Kac's program in kinetic theory. The article [18] gives a useful description of this program.

To state the main result of this paper, we recall some standard definitions. For measures ν_1, ν_2 on a measure space $(\Omega, \mathcal{F})$, the *total variation distance* between ν_1, ν_2 is given by

$$\|\nu_1 - \nu_2\|_{\text{TV}} = \sup_{A \in \mathcal{F}} (\nu_1(A) - \nu_2(A)).$$

We denote the distribution of a random variable Z by $\mathcal{L}(Z)$ and write $Z \sim \nu$ as a shorthand for $\mathcal{L}(Z) = \nu$.

Let μ denote the normalized Haar measure on S^{n-1}. The Markov chain $\{X_t\}_{t \geq 0}$ has μ as its unique stationary distribution on S^{n-1}. The *mixing profile* of X_t is defined as

$$\tau(\varepsilon) = \min\left\{ t : \sup_{X_0 = x \in S^{n-1}} \|\mathcal{L}(X_t) - \mu\|_{\text{TV}} < \varepsilon \right\}$$

and the *mixing time* is given by $\tau_{\text{mix}} = \tau(0.25)$.

Our main result is the following.

THEOREM 1. *Fix $C_1 < \frac{1}{2}$ and $C_2 > 200$. If the sequence of times $\{T_1(n)\}_{n \in \mathbb{N}}$ satisfies $T_1(n) < C_1 n \log(n)$ for all n, then*

$$(1.2) \qquad \lim_{n \to \infty} \inf_{X_0 \in S^{n-1}} \|\mathcal{L}(X_{T_1(n)}) - \mu\|_{\text{TV}} = 1.$$

If the sequence of times $\{T_2(n)\}_{n \in \mathbb{N}}$ satisfies $T_2(n) > C_2 n \log(n)$ for all n, then

$$(1.3) \qquad \lim_{n \to \infty} \sup_{X_0 \in S^{n-1}} \|\mathcal{L}(X_{T_2(n)}) - \mu\|_{\text{TV}} = 0.$$

Theorem 1 implies that, for any $\delta > 0$ and $n > N_0(\delta)$ sufficiently large, the mixing time τ_{mix} of Kac's walk on S^{n-1} satisfies

$$(1 - \delta)\frac{1}{2}n \log(n) \leq \tau_{\text{mix}} \leq (1 + \delta)200 n \log(n).$$

Theorem 1 also establishes the stronger result that Kac's walk exhibits *pre-cutoff* with window $[\frac{1}{2}n \log(n), 200n \log(n)]$ (see, e.g., Chapter 18 of [15] for an introduction to cut-off). We conjecture that Kac's walk also exhibits cutoff at time $2n \log(n)$. While we have made no effort to optimize the size of the cutoff window obtained in this paper and both our upper and lower bounds for τ_{mix} could be easily improved with our methods, a proof of cut-off does not seem to follow from our methods and will likely require new ideas.

1.1. *Relationship to prior work.* There is an extensive literature on the mixing time of Kac's walk under various measures. In the sequence of papers [3, 4, 11, 17], the spectrum of Kac's walk on S^n was bounded and then computed. Although these results imply a convergence *rate* for Kac's walk in L^2, and a bound on the *distance* to stationarity in L^2 implies a bound on the total variance distance to stationarity, these bounds *do not* imply any bound at all on the total variation mixing time of Kac's walk. This is because, when $\mathcal{L}(X_0)$ is concentrated at a point, the initial L^2 distance to stationarity is not finite. Stronger entropy bounds [5] also give no bound on the total variation mixing of Kac's walk. The previous best known bound on the mixing time of Kac's walk on S^n, due to Jiang [12], is $O(n^5 \log(n)^2)$.

There are other versions of Kac's walk; see the literature review in [19] for an overview of one of the most-studied such walks. The recent works [8, 18] discuss the relationship of the study of Kac's walk to the original goals of Kac's paper [13].

2. Technical overview. We bound the mixing time of a Markov chain by applying the standard *coupling lemma* (see, e.g., Theorem 5.2 of [15]).

LEMMA 2.1. *Let K be the transition kernel of a Markov chain with unique stationary distribution ν on state space Ω. Let $\{X_t\}_{t\geq 0}$, $\{Y_t\}_{t\geq 0}$ be two Markov chains started at $X_0 = x \in \Omega$ and $Y_0 \sim \nu$. Define the coalescence time*

$$(2.1) \qquad\qquad \tau(x) = \min\{t : X_t = Y_t\}.$$

Assume that the coupling of $\{X_t\}_{t\geq 0}$, $\{Y_t\}_{t\geq 0}$ satisfies $X_t = Y_t$ for all $t \geq \tau(x)$. Then for any $t \geq 0$,

$$\|\mathcal{L}(X_t) - \nu\|_{\mathrm{TV}} \leq \mathbb{P}[\tau(x) > t].$$

The skeleton of the coupling argument in our paper is quite similar to that in [22]. As in that paper, we construct a coupling of two copies of Kac's walk over two phases. In the first phase, we construct a "proportionate coupling" (see Definition 3.1) between two copies of Kac's walk. We show that in this first phase, the two copies become close to each other in the Euclidean norm in $O(n \log(n))$ time. In the second phase, also of length $O(n \log(n))$, we construct a non-Markovian coupling that allows two copies of Kac's walk that start very close to each other to actually coalesce. We then apply Lemma 2.1, bounding the probability $\mathbb{P}[\tau(x) > T]$ that the random coalescence time $\tau(x)$ is larger than a particular deterministic time $T \approx 200n \log(n)$.

This "two-step" approach, with an initial "contracting" phase followed by a second "coalescing" phase, is a popular approach for proving the convergence of Metropolis–Hastings chains on continuous state spaces (see, e.g., [16, 21]). However, there are some difficulties in extending this approach to the study of high-dimensional Gibbs samplers such as Kac's walk. For most Metropolis–Hastings

chains, two nearby Markov chains can coalesce in a single step, and so the co-
alescence phase can be of length 1. Gibbs samplers do not share this property:
in n dimensions, arbitrarily nearby Markov chains generally have 0 probability
of coalescing in less than n steps. This means that the coalescence phase is quite
lengthy, and it becomes necessary to ensure that the two chains do not drift too far
apart during this period. We also mention that the presence of constraints on the
state space (in this case, the constraint that X_t must be an element of S^{n-1} and
cannot be an arbitrary element of $\mathbb{R}^n$) presents additional complications that are
not present for most Metropolis–Hastings chains. In our case, this constraint guar-
antees that any Markovian coupling scheme in the coalescence phase will have
$\mathbb{E}[\tau(x)] > Cn^2$ for some $C > 0$. The same phenomenon occurs in [22] and many
well-studied discrete chains (see, e.g., Lemma 8 of [2]). See [9, 10] for earlier
examples of non-Markovian couplings applied to the study of Markov chains on
discrete state spaces and [6, 14] for more general discussion of when a "good"
Markovian coupling exists.

2.1. *Structure of the paper.*   Section 3 describes phase 1 of the coupling, Sec-
tion 4 describes phase 2, and the proof of Theorem 1 is given in Section 5.

2.2. *Informal discussion of "greedy" couplings.*   Before giving a formal de-
scription of the second phase of our coupling in Section 4.1, we briefly explain
where it comes from.

The obvious "greedy" approach to coupling a pair of Gibbs samplers $\{W_t,$
$Z_t\}_{t\geq 0}$ in $\mathbb{R}^n$ is to run a Markovian coupling that matches as many coordinates
as possible at each step. That is, we try to grow the set $E_t = \{1 \leq i \leq n : W_t[i] =$
$Z_t[i]\}$ by as much as possible at every step. Such a greedy coupling always exists,
but it is easiest to analyze if $E_t \subset E_{t+1}$ for all t, that is, if coordinates that agree
once will agree forever. To phrase this second condition in a slightly unfamiliar
way, this means that the sequence of hyperplanes

$$(2.2) \qquad H_t = \{v \in \mathbb{R}^n : \forall i \in E_t, v[i] = 0\}$$

should satisfy

$$(2.3) \qquad \begin{aligned} H_{t+1} &\subset H_t, \\ (W_t - Z_t) &\in H_t \end{aligned}$$

for all $t \geq 0$. When this occurs, the coalescence time $\tau(x)$ satisfies

$$(2.4) \qquad \tau(x) \leq \inf\{t \geq 0 : H_t = \{0\}\}.$$

We claim that no Markovian coupling of the square $\{X_t^2, Y_t^2\}_{t\geq 0}$ of Kac's walk
can satisfy inequalities (2.3). Indeed, for two copies of Kac's walk X_t, Y_t, the
dynamics often *force* $X_{t+1}^2 - Y_{t+1}^2$ to "pop out" out of a candidate H_t. However,
our notation suggests a much larger family of "greedy" couplings: *any* sequence

$\{H_t\}_{t \geq 0}$ that satisfies inequalities (2.3) and (2.4) gives rise to a "greedy" coupling of $\{Z_t, W_t\}_{t \geq 0}$, even if the hyperplanes are not of the form (2.2) or the coupling is not Markovian.

There are many candidates for the sequence $\{H_t\}_{t \geq 0}$. Roughly speaking, in this paper we construct the simplest possible sequence $\{H_t\}_{t \geq 0}$ that is not of the form (2.2): rather than starting with $H_0 = \mathbb{R}^n$ and greedily *reducing* the dimension of our candidate H_t as quickly as possible until it reaches $\{0\}$, which gives hyperplanes of the form (2.2), we start with $H_T = \{0\}$ at some fixed time $T > 0$, and then go back in time while greedily *increasing* the dimension of our candidate H_t as quickly as possible until it reaches $\mathbb{R}^n$. This sequence of candidate hyperplanes, which are described carefully in Section 4.1 [Equations (4.1) and (4.2) define the sequence of hyperplanes used in this paper] turn out to satisfy (2.3) and (2.4) with high probability for Kac's walk. Because of this, the resulting non-Markovian coupling is almost as straightforward to analyze as the standard greedy Markovian coupling.

Our construction for $\{H_t\}_{t \geq 0}$ will not satisfy (2.3) with high probability for all Markov chains, just as the usual construction in (2.2) did not. Nonetheless, we hope that this general approach may yield other sophisticated couplings that are straightforward to analyze. We also point out that, when coupling Gibbs samplers, it is very common to force the two Markov chains to always select the same sequence of coordinates at every step,[4] and we make this choice in the present paper. In this situation, the above coupling constructions would be applied to the Markov chains that are given after *conditioning* on the sequence of coordinates to be updated. This modification makes no difference to the remainder of the discussion.

3. Contraction estimates. In this section, we couple two copies of the Kac's walk on S^{n-1} and obtain a one-step contraction for the distance between these two copies in a suitable psuedo-metric. This is achieved via the following coupling.

DEFINITION 3.1 (Proportional coupling). We define a coupling of two copies $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$ of Kac's walk for a single step. Fix $X_0, Y_0 \in S^{n-1}$. Let (i_0, j_0, θ_0) be the *update variables* used by the random walk X_1 in representation (1.1), so that $X_1 = F(i_0, j_0, \theta_0, X_0)$. Choose $\varphi \in [0, 2\pi)$ uniformly at random among all angles that satisfy[5]

$$X_1[i_0] = \sqrt{X_0[i_0]^2 + X_0[j_0]^2} \cos(\varphi),$$

$$X_1[j_0] = \sqrt{X_0[i_0]^2 + X_0[j_0]^2} \sin(\varphi).$$

---

[4]In the notation representation (1.1), this corresponds to choosing the same sequence $\{i_t, j_t\}_{t \geq 0}$.

[5]If $X_0[i_0] = X_0[j_0] = 0$, all angles satisfy these equations. Otherwise, they have a unique solution, and the value of $\varphi - \theta_0$ modulo 2π does not depend on θ_0.

Then choose $\theta_0' \in [0, 2\pi)$ uniformly among the angles that satisfy

$$F(i_0, j_0, \theta_0', Y_0)[i_0] = \sqrt{Y_0[i_0]^2 + Y_0[j_0]^2} \cos(\varphi),$$

$$F(i_0, j_0, \theta_0', Y_0)[j_0] = \sqrt{Y_0[i_0]^2 + Y_0[j_0]^2} \sin(\varphi)$$

and set $Y_1 = F(i_0, j_0, \theta_0', Y_0)$.

REMARK 3.2. This coupling forces Y_1 to be as close as possible to X_1 in the Euclidean distance. For example, in dimension $n = 2$, we always have $X_1 = Y_1$ under this coupling. For $X_0, Y_0 \in S^{n-1}$, this coupling forces the three points $(0, 0)$, $(X_1[i_0], X_1[j_0])$ and $(Y_1[i_0], Y_1[j_0])$ to be collinear; in particular, it forces $X_1[i_0]Y_1[i_0], X_1[j_0]Y_1[j_0] \geq 0$.

We set some notation for the remainder of the paper. Whenever we consider a pair of copies of Kac's walk $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$, we define

$$(3.1) \qquad A_t[i] = X_t[i]^2, \qquad B_t[i] = Y_t[i]^2$$

for all $t \geq 0$ and all $1 \leq i \leq n$. For $x \in \mathbb{R}^n$, we define

$$\|x\|_1 = \sum_i |x[i]|, \qquad \|x\|_2 = \left(\sum_i x[i]^2\right)^{1/2}.$$

Finally, we define $[n] = \{1, 2, \ldots, n\}$.

LEMMA 3.3. Let $X_0, Y_0 \in S^{n-1}$. For $t \geq 0$, couple (X_{t+1}, Y_{t+1}) conditional on (X_t, Y_t) according to the coupling in Definition 3.1. Then for any $t \geq 0$, Kac's walk on S^{n-1} satisfies

$$(3.2) \qquad \mathbb{E}\left[\sum_{i=1}^{n}(A_t[i] - B_t[i])^2\right] \leq 2\left(1 - \frac{1}{2n}\right)^t.$$

PROOF. Fix $X_0, Y_0 \in S^{n-1}$. For $t \geq 0$, couple (X_{t+1}, Y_{t+1}) conditional on (X_t, Y_t) according to the coupling in Definition 3.1, with update variables $i_t, j_t, \theta_t, \theta_t'$ and additional variable φ as in that definition. We calculate

$$\mathbb{E}\left[\sum_{k=1}^{n}(A_1[k] - B_1[k])^2\right]$$

$$= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{E}\left[\sum_{k=1}^{n}(A_1[k] - B_1[k])^2 | (i_0, j_0) = (i, j)\right]$$

$$= \frac{2}{n(n-1)} \frac{(n-1)(n-2)}{2} \sum_{k=1}^{n}(A_0[k] - B_0[k])^2$$

$$+ \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}\left[\left((A_0[i] + A_0[j]) \cos(\varphi)^2 - (B_0[i] + B_0[j]) \cos(\varphi)^2\right)^2\right]$$

$$+ \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}\left[\left((A_0[i] + A_0[j]) \sin(\varphi)^2 - (B_0[i] + B_0[j]) \sin(\varphi)^2\right)^2\right]$$

$$= \frac{n-2}{n} \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{4}{n(n-1)} \mathbb{E}[\cos(\varphi)^4] \sum_{i<j} \left((A_0[i] + A_0[j]) - (B_0[i] + B_0[j])\right)^2$$

$$= \left(1 - \frac{2}{n}\right) \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{3}{2n(n-1)} \sum_{i<j} \left((A_0[i] + A_0[j]) - (B_0[i] + B_0[j])\right)^2$$

$$= \left(1 - \frac{2}{n}\right) \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{3}{2n(n-1)} \sum_{i<j} \left((A_0[i] - B_0[i])^2\right)$$

$$+ (A_0[j] - B_0[j])^2 + 2(A_0[i] - B_0[i])(A_0[j] - B_0[j])\right)$$

$$= \left(1 - \frac{2}{n}\right) \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{3(n-1)}{2n(n-1)} \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{3}{n(n-1)} \sum_{i<j} (A_0[i] - B_0[i])(A_0[j] - B_0[j])$$

$$= \left(1 - \frac{1}{2n}\right) \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{3}{n(n-1)} \sum_{i<j} (A_0[i] - B_0[i])(A_0[j] - B_0[j])$$

$$= \left(1 - \frac{1}{2n}\right) \sum_{k=1}^{n} (A_0[k] - B_0[k])^2$$

$$+ \frac{3}{2n(n-1)} \left(\left(\sum_{k=1}^{n} (A_0[k] - B_0[k]) \right)^2 - \sum_{k=1}^{n} (A_0[k] - B_0[k])^2 \right)$$

$$= \left(1 - \frac{1}{2n} - \frac{3}{2n(n-1)} \right) \sum_{k=1}^{n} (A_0[k] - B_0[k])^2.$$

Thus, we have

$$(3.3) \qquad \mathbb{E} \left[\sum_{k=1}^{n} (A_1[k] - B_1[k])^2 \right] \leq \left(1 - \frac{1}{2n} \right).$$

For $t \geq 0$, let $\mathcal{F}_t = \sigma(\bigcup_{0 \leq s \leq t}(X_s, Y_s))$ be the σ-algebra generated by the random variables $X_0, \ldots, X_t$ and $Y_0, \ldots, Y_t$. Repeatedly applying equation (3.3), we have for all $t \geq 0$ that

$$\mathbb{E} \left[\sum_{k=1}^{n} (A_t[k] - B_t[k])^2 \right] = \mathbb{E} \left[\mathbb{E} \left[\sum_{k=1}^{n} (A_t[k] - B_t[k])^2 | \mathcal{F}_{t-1} \right] \right]$$

$$\leq \left(1 - \frac{1}{2n} \right) \mathbb{E} \left[\sum_{k=1}^{n} (A_{t-1}[k] - B_{t-1}[k])^2 \right]$$

$$\leq \left(1 - \frac{1}{2n} \right)^t \sum_{k=1}^{n} \mathbb{E}[(A_0[k] - B_0[k])^2] \leq 2 \left(1 - \frac{1}{2n} \right)^t.$$

This completes the proof. $\quad\square$

REMARK 3.4. We note that X_t can be recovered exactly from $(A_t, \{\frac{X_t[i]}{|X_t[i]|}\}_{i=1}^n) \in [0,1]^n \times \{-1, 1\}^n \equiv \mathcal{X}$ by a map that we temporarily denote by $G : \mathcal{X} \mapsto S^{n-1}$, whose inverse exists except on a set of measure 0. Define the metric $d_{\mathcal{X}}$ on $\mathcal{X}$ by $d_{\mathcal{X}}((A, L), (A', L')) = \|A - A'\|_2 + \sum_{i=1}^n \mathbf{1}_{L[i] \neq L'[i]}$. Inequality (3.2) can be combined with standard bounds on the contraction of simple random walk on the hypercube under the Hamming metric to show that, under the proportional coupling,

$$\mathbb{E}[d_{\mathcal{X}}(G^{-1}(X_{t+1}), G^{-1}(Y_{t+1})) | X_t, Y_t] \leq \left(1 - \frac{c}{n} \right) d_{\mathcal{X}}(G^{-1}(X_t), G^{-1}(Y_t))$$

for some $0 < c < \infty$. This contraction estimate, combined with Proposition 30 of [20], implies that the relaxation time of Kac's walk is $O(n)$. As shown in [3, 4, 11, 17], this is the correct order for the relaxation time, and our argument provides a short alternative proof of the main result in [11]. However, we cannot use this approach to calculate the exact spectral gap, which was derived in [3, 4, 17].

We close this section with two elementary bounds. For $S \subset \{1, 2, \ldots, n\}$, let $\tau_S = \min\{t : S \subset \bigcup_{0 \leq s \leq t}\{i_s, j_s\}\}$. By the standard "coupon collector" bound,

$$(3.4) \qquad \mathbb{P}[\tau_{[n]} > t] \leq ne^{-\frac{t}{n}}$$

for all $t \geq 0$. Let $S = \{1 \leq i \leq n : X_0[i]Y_0[i] < 0\}$. By Definition 3.1, we have $\min_{1 \leq i \leq n} X_t[i]Y_t[i] \geq 0$ for all $t \geq \tau_S$.

LEMMA 3.5. *Let $Y \sim \mu$. Then for all $1 < c < \infty$ and any $1 \leq i \leq n$,*

$$\mathbb{P}[Y[i] \leq n^{-3c}] \leq 2n^{-c+1}.$$

PROOF. Let $\zeta_1, \ldots, \zeta_n$ be n i.i.d. random variables with $\mathcal{N}(0, 1)$ distribution. Recall that the law $\mathcal{L}(Y[i])$ of $Y[i]$ satisfies $\mathcal{L}(Y[i]) = \mathcal{L}(\frac{\zeta_1^2}{\sum_{k=1}^n \zeta_k^2})$. By Markov's inequality and some direct computation,

$$\mathbb{P}[Y[i] < n^{-3c}] = \mathbb{P}\left[\frac{\zeta_1^2}{\sum_{k=1}^n \zeta_k^2} < n^{-3c}\right]$$

(3.5)
$$\leq \mathbb{P}[\zeta_1^2 \leq n^{-2c}] + \mathbb{P}\left[\sum_{k=1}^n \zeta_k^2 \geq n^c\right]$$

$$\leq n^{-c} + n^{-c+1} \leq 2n^{-c+1}$$

and the proof is complete. □

4. Non-Markovian coupling and coalesence estimates. In this section, we construct a more complicated non-Markovian coupling of two chains $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$ started at a pair of near by points. In the proof of Theorem 1, we will show that this coupling leads to a collision of X_t, Y_t with high probability for sufficiently large t. In order to couple the random walks $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$, it is enough to couple the *update sequences* $\{(i_t(x), j_t(x), \theta_t(x))\}_{t \geq 0}$, $\{(i_t(y), j_t(y), \theta_t(y))\}_{t \geq 0}$ used to update them in representation (1.1). Our coupling is as follows.

4.1. *Non-Markovian coupling.* Fix $T_0 \leq T \in \mathbb{N}$. We construct the coupled update sequences $\{(i_t(x), j_t(x), \theta_t(x))\}$ and $\{(i_t(y), j_t(y), \theta_t(y))\}$ for $T_0 \leq t < T$ via the following procedure:

1. For each $T_0 \leq t < T$, choose $1 \leq i_t(x) < j_t(x) \leq n$ uniformly at random and set $i_t(y) = i_t(x)$, $j_t(y) = j_t(x)$.
2. Define a sequence of partitions $\{\mathcal{P}_t\}_{t=T_0}^T$ of $\{1, 2, \ldots, n\}$ inductively by the process:

   - Set the partition $\mathcal{P}_T = \{\{1\}, \{2\}, \ldots, \{n\}\}$.
   - Write $\mathcal{P}_{t+1} = \{P_1(t+1), \ldots, P_{\ell_{t+1}}(t+1)\}$ with $P_k(t+1) \subset \{1, 2, \ldots, n\}$. Let $1 \leq u(t), v(t) \leq \ell_{t+1}$ be the indices satisfying $i_t(x) \in P_{u(t)}(t+1)$, $j_t(x) \in P_{v(t)}(t+1)$. If $u(t) = v(t)$, set $\mathcal{P}_t = \mathcal{P}_{t+1}$. If $u(t) \neq v(t)$, construct $\mathcal{P}_t$ by merging the sets $P_{u(t)}(t+1)$, $P_{v(t)}(t+1)$. Thus, if $u(t) < v(t)$, set $\mathcal{P}_t = \{P_1(t+1), \ldots, P_{u(t)-1}(t+1), P_{u(t)+1}(t+1), \ldots, P_{v(t)-1}(t+1), P_{v(t)+1}(t+1), \ldots, P_{\ell_{t+1}}(t+1), P_{u(t)}(t+1) \cup P_{v(t)}(t+1)\}$.

3. If $\mathcal{P}_{T_0} = \{1, 2, \ldots, n\}$, continue the construction of the coupling by:

- Define the ordered set

$$\mathcal{S} = \{s_1, \ldots, s_{n-1}\} \equiv \{T_0 \le t < T : \mathcal{P}_t \ne \mathcal{P}_{t+1}\}.$$

- For $t \in \mathcal{S}^c \cap \{T_0, T_0 + 1, \ldots, T - 1\}$, choose the update variables $(\theta_t(x), \theta_t(y))$ according to the proportional coupling in Definition 3.1.
- Let Π denote the set of all joint distributions on $[0, 2\pi) \times [0, 2\pi)$ with both marginal distributions uniformly distributed on $[0, 2\pi)$. For $t = s_k \in \mathcal{S}$, choose[6] $(\theta_{s_k}(x), \theta_{s_k}(y)) \sim \pi \in \Pi$ so as to maximize the probability of the following events:

$$(4.1) \qquad \sum_{i \in P_r(t+1)} X_{t+1}[i]^2 = \sum_{i \in P_r(t+1)} Y_{t+1}[i]^2, \qquad 1 \le r \le \ell_{t+1},$$

$$(4.2) \qquad X_{t+1}[k] Y_{t+1}[k] \ge 0, \qquad k \in \{i_t(x), j_t(x)\}.$$

4. If $\mathcal{P}_{T_0} \ne \{1, 2, \ldots, n\}$, continue the construction of the coupling by setting $\theta_t(y) = \theta_t(x)$ for all $T_0 \le t < T$.

REMARK 4.1.   If (4.1) holds at time T, (4.2) holds for all $T_0 \le t \le T$, and $\mathcal{P}_{T_0} = \{1, 2, \ldots, n\}$, then $X_T = Y_T$. Recall that forcing X_T and Y_T to coalesce is the goal of our coupling construction.

REMARK 4.2.   By construction, we have that $\{i_t(x), j_t(x)\}_{T_0 \le t < T}$, $\{i_t(y), j_t(y)\}_{T_0 \le t < T}$ have the correct distributions. Going through the construction, it is easy to verify that the distribution of $\theta_s(x)$ [respectively $\theta_s(y)$] conditional on $\{i_t(x), j_t(x)\}_{T_0 \le t < T}$ and $\{\theta_t(x)\}_{T_0 \le t < s}$ [resp., $\{i_t(y), j_t(y)\}_{T_0 \le t < T}$ and $\{\theta_t(y)\}_{T_0 \le t < s}$] is uniform on $[0, 2\pi)$. Thus, Definition 4.1 gives a valid coupling of two copies of Kac's walk.

4.2. *Bounds for coupling.*   The remainder of this section consists of some bounds that will allow us to prove that two chains $\{X_t\}_{T_0 \le t < T}$, $\{Y_t\}_{T_0 \le t < T}$ constructed according to the coupling described in Section 4.1 satisfy the conditions in equations (4.1) and (4.2) for all $T_0 \le t < T$ with high probability, as long as $T - T_0$ is sufficiently large and $\|X_{T_0} - Y_{T_0}\|$ is sufficiently small.

We first bound the probability that the condition in step 3 of our coupling, $\mathcal{P}_{T_0} = \{1, 2, \ldots, n\}$, fails to hold.

---

[6]Note that there is generally not a unique coupling with this "maximal" property. In our analysis, we do not care which coupling with this property is used. A reader concerned about this choice can instead use a coupling of the form implicitly constructed in the proof of Lemma 4.6 at this step, without changing any of the bounds in the rest of the paper.

LEMMA 4.3 (Splitting of partitions).   *Fix $\varepsilon > 0$, $T_0 \in \mathbb{N}$ and $T > T_0 + (\frac{1}{2} + 2\varepsilon)n \log(n)$. Then for $\mathcal{P}_{T_0}$ as in Section* 4.1 *and all $n > N_0(\varepsilon)$ sufficiently large,*

$$\mathbb{P}[\mathcal{P}_{T_0} = \{1, 2, \ldots, n\}] \geq 1 - 2n^{-\varepsilon}.$$

PROOF.   This follows immediately from Proposition 7.3 of [1]. Fix $M \geq 0$ and define the Erdős–Renyi graph G_M on n vertices with exactly M (possibly repeated) edges

$$E(G_m) = \bigcup_{s=T-M}^{T-1} (i_s, j_s).$$

Then

$$\mathbb{P}[\mathcal{P}_{T_0} = \{1, 2, \ldots, n\}] = \mathbb{P}[G_{T-T_0} \text{ is connected}] \geq 1 - 2n^{-\varepsilon},$$

where the last inequality follows from Proposition 7.3 of [1].   □

For $T_0 \leq s < T$, define the event $\mathcal{A}(s)$ by

(4.3) $\mathcal{A}(s) = \{\text{Equations (4.1) and (4.2) are satisfied for all } T_0 \leq t \leq s\}.$

For $S \subset \{1, 2, \ldots, n\}$ and $X \in \mathbb{R}^n$, define

$$\|X\|_{1,S} = \sum_{i \in S} |X[i]|.$$

Also recall the definition of A_t, B_t in equation (3.1). The following bound implies that good couplings will never drift too far apart.

LEMMA 4.4 (Closeness of good couplings).   *Fix $T_0 < T$, and couple two chains $\{X_t\}_{T_0 \leq t < T}$, $\{Y_t\}_{T_0 \leq t < T}$ according to the non-Markovian coupling defined in Section* 4.1. *Fix $T_0 \leq s \leq T$. Then, on the event $\mathcal{A}(s) \cap \{\mathcal{P}_{T_0} = \{1, 2, \ldots, n\}\}$, we have*

(4.4) $\|A_t - B_t\|_{1,S} \leq \|A_{T_0} - B_{T_0}\|_1$

for all $T_0 \leq t \leq s$ and all $S \in \mathcal{P}_t$. Furthermore, for all $T_0 \leq t \leq s$,

(4.5) $\|A_t - B_t\|_1 \leq n\|A_{T_0} - B_{T_0}\|_1.$

REMARK 4.5.   The final bound of Lemma 4.4, inequality (4.5), is quite weak: it allows the distance $\|A_t - B_t\|_1$ to increase by as much as a factor of n. However, this additional factor of n has a very small impact on our final bound in Theorem 1; indeed, changing this *multiplicative* factor of n to a factor of n^k would change our bound on the mixing time only by an *additive* factor of $2kn \log(n)$.

We also point out that, a priori, it is not obvious that *any* polynomial bound holds. Indeed, the distance between A_t and B_t can double over a single step. Thus, naively, one obtains only a much larger bound, on the order of n^n.

PROOF OF LEMMA 4.4.   We prove (4.4) by induction on t. It is trivial at time $t = T_0$. Now assume that it holds at time t. We will show it also holds at time $t + 1$ by considering two cases separately:

1. $t \notin \mathcal{S}$: In this case, $\ell_t = \ell_{t+1}$ and $P_k(t) = P_k(t + 1)$ for all $1 \leq k \leq \ell_t$. For all $k \neq u(t)$, it is immediate that $A_{t+1}[m] = A_t[m]$ and $B_{t+1}[m] = B_t[m]$ for all $m \in P_k(t)$, so by the induction hypothesis

$$\|A_{t+1} - B_{t+1}\|_{1, P_k(t+1)} = \|A_t - B_t\|_{1, P_k(t)} \leq \|A_{T_0} - B_{T_0}\|_1.$$

For the remaining index $k = u(t)$, we calculate

$$
\begin{aligned}
&\|A_{t+1} - B_{t+1}\|_{1, P_{u(t)}(t+1)} \\
&= \sum_{m \in P_{u(t)}(t+1)} |X_{t+1}[m]^2 - Y_{t+1}[m]^2| \\
&= \sum_{m \in P_{u(t)}(t) \setminus \{i_t, j_t\}} |X_t[m]^2 - Y_t[m]^2| + |X_{t+1}[i_t]^2 - Y_{t+1}[i_t]^2| \\
&\quad + |X_{t+1}[j_t]^2 - Y_{t+1}[j_t]^2| \\
&= \sum_{m \in P_{u(t)}(t) \setminus \{i_t, j_t\}} |X_t[m]^2 - Y_t[m]^2| \\
&\quad + |(X_t[i_t]^2 + X_t[j_t]^2) \cos(\varphi)^2 - (Y_t[i_t]^2 + Y_t[j_t]^2) \cos(\varphi)^2| \\
&\quad + |(X_t[i_t]^2 + X_t[j_t]^2) \sin(\varphi)^2 - (Y_t[i_t]^2 + Y_t[j_t]^2) \sin(\varphi)^2| \\
&= \sum_{m \in P_{u(t)}(t) \setminus \{i_t, j_t\}} |X_t[m]^2 - Y_t[m]^2| \\
&\quad + |(X_t[i_t]^2 + X_t[j_t]^2) - (Y_t[i_t]^2 + Y_t[j_t]^2)| \\
&\leq \sum_{m \in P_{u(t)}(t) \setminus \{i_t, j_t\}} |X_t[m]^2 - Y_t[m]^2| + |X_t[i_t]^2 - Y_t[i_t]^2| \\
&\quad + |X_t[j_t]^2 - Y_t[j_t]^2| \\
&= \|A_t - B_t\|_{1, P_{u(t)}(t)} \leq \|A_{T_0} - B_{T_0}\|_1,
\end{aligned}
$$

where the last step is by the induction hypothesis. This completes the proof for the case $t \notin \mathcal{S}$.

2. $t = s_k \in \mathcal{S}$: By the same argument as in the previous case, we need only check that $\|A_{t+1} - B_{t+1}\|_{1, P_{u(t)}(t+1)} \leq \|A_t - B_t\|_{1, P_{u(t)}(t+1) \cup P_{v(t)}(t+1)}$ and $\|A_{t+1} - B_{t+1}\|_{1, P_{v(t)}(t+1)} \leq \|A_t - B_t\|_{1, P_{u(t)}(t+1) \cup P_{v(t)}(t+1)}$; the other $k - 1$ expressions in equality (4.1) are satisfied automatically. By the symmetry of the problem, it

is sufficient to check the first of these two inequalities. We compute

$$\|A_{t+1} - B_{t+1}\|_{1, P_{u(t)}(t+1)}$$

$$= \sum_{m \in P_{u(t)}(t+1)} |A_{t+1}[m] - B_{t+1}[m]|$$

$$= \sum_{m \in P_{u(t)}(t+1) \backslash \{i_t\}} |A_{t+1}[m] - B_{t+1}[m]| + |A_{t+1}[i_t] - B_{t+1}[i_t]|$$

$$= \sum_{m \in P_{u(t)}(t+1) \backslash \{i_t\}} |A_t[m] - B_t[m]|$$

$$+ |A_t[i_t] + A_t[j_t] - B_{t+1}[i_t] - B_{t+1}[j_t]| \cos(\varphi)^2$$

$$\leq \sum_{m \in P_{u(t)}(t+1) \cup \{j_t\}} |A_t[m] - B_t[m]| \leq \|A_t - B_t\|_{1, P_{u(t)}(t+1) \cup P_{v(t)}(t+1)}$$

$$\leq \|A_{T_0} - B_{T_0}\|_1,$$

where the last step is by the induction hypothesis.

Thus, in either case, equation (4.4) holds at time $t + 1$, proving our first claim.

Equation (4.5) follows by calculating

$$\|A_t - B_t\|_1 = \sum_{k=1}^{\ell_t} \|A_t - B_t\|_{1, P_k(t)}$$

$$\leq \sum_{k=1}^{\ell_t} \|A_{T_0} - B_{T_0}\|_1$$

$$= \ell_t \|A_{T_0} - B_{T_0}\|_1 \leq n \|A_{T_0} - B_{T_0}\|_1,$$

where the first inequality follows from equation (4.4), and we are done. □

The following lemma will be used to show that it is possible to couple X_{t+1}, Y_{t+1} so as to satisfy equations (4.1) and (4.2), as long as X_t, Y_t satisfy those equations and have certain other good properties.

LEMMA 4.6 (Coupling success). *Fix positive reals $1 < p < q' < \frac{q}{2}$. Let $\theta, \theta' \sim \mathrm{Unif}[0, 2\pi)$ and let $S = A + B \cos(\theta)^2$ and $S' = C + D \cos(\theta)^2$ for some $0 \leq A, B, C, D \leq 1$ that satisfy $|A - C|, |B - D| \leq n^{-q}$ and $B, D \geq n^{-p}$. Then for $n > N_0(p, q, q')$ sufficiently large, there exists a coupling of θ, θ' so that*

$$(4.6) \qquad\qquad \mathbb{P}[S = S'] \geq 1 - 6 \times 10^3 n^{-c}$$

and

$$(4.7) \qquad\qquad \cos(\theta) \cos(\theta') \geq 0, \qquad \sin(\theta) \sin(\theta') \geq 0,$$

where $c = \min(\frac{q'}{2}, q - 2q') > 0$.

REMARK 4.7.   In the notation of the construction in Section 4.1, we will use
Lemma 4.6 with $A = (\sum_{k \in P_t \setminus \{i_t\}} X_t[k]^2)$ and $B = (X_t[i_t]^2 + X_t[j_t]^2)$. Roughly
speaking, the upper bounds on $|A - C|, |B - D|$ in the statement of the lemma
represent upper bounds on the distances between the medians and densities of two
distributions that we wish to couple, and the lower bounds on B, D represent up-
per bounds on how quickly the densities change. By inspection of inequality (4.8)
below, we expect to be able to couple two random variables with high probabil-
ity if the distance between their medians and densities is small compared to the
derivative of the density.

Note that we need both bounds to obtain the desired coupling inequality; as
illustrated by the densities $\{\frac{n^{10}}{2\sqrt{2\pi}} e^{-n^{10}(x-n^{-1})^2}\}_{n \in \mathbb{N}}$, $\{\frac{n^{10}}{2\sqrt{2\pi}} e^{-n^{10}(x+n^{-1})^2}\}_{n \in \mathbb{N}}$, it is
not enough to just have a bound on the distance between the medians.

PROOF OF LEMMA 4.6.   We begin by showing that it is possible to satisfy
inequality (4.6). Recall the standard inequality that, for any distributions ν_1, ν_2
with densities f_1, f_2 on $\mathbb{R}$,

$$(4.8) \qquad \|\nu_1 - \nu_2\|_{\text{TV}} \le 1 - \int \min(f_1(x), f_2(x)) \, dx.$$

The random variables S, S' have densities

$$f_S(x) = \frac{1}{2\pi} \frac{1}{\sqrt{B^2 - 4(x - A - B/2)^2}}, \qquad x \in (A, A + B),$$

$$g_{S'}(x) = \frac{1}{2\pi} \frac{1}{\sqrt{D^2 - 4(x - C - D/2)^2}}, \qquad x \in (C, C + D).$$

Thus, applying inequality (4.8) to these two densities with $1 < p < q' < \frac{q}{2}$ and
$n > N_0(p, q, q')$ sufficiently large,

$$\|S - S'\|_{\text{TV}}$$
$$= 1 - \int_{(A,A+B) \cap (C,C+D)} \min(f_S(x), g_{S'}(x)) \, dx$$
$$\le 1 - \frac{1}{2\pi} \int_{(A+n^{-q'}, A+B-2n^{-q'})} \frac{1}{\sqrt{B^2 - 4(x - A - B/2)^2}}$$
$$\times \min\left(1, \frac{\sqrt{(D - n^{-q})^2 - 4(|x - C - D/2| + 2n^{-q})^2}}{\sqrt{D^2 - 4(x - C - D/2)^2}}\right) dx$$
$$\le 1 - \frac{1}{2\pi} \int_{(A+n^{-q'}, A+B-2n^{-q'})} \frac{1}{\sqrt{B^2 - 4(x - A - B/2)^2}}$$

$$\times \left(1 - 100 \frac{n^{-q}}{D^2 - 4(x - C - D/2)^2}\right)^{\frac{1}{2}} dx$$

$$\leq 1 - \left(1 - 8 \times 10^3 \frac{n^{-q}}{n^{-2q'}}\right)^{\frac{1}{2}}$$

$$\times \frac{1}{2\pi} \int_{(A+n^{-q'}, A+B-2n^{-q'})} \frac{1}{\sqrt{B^2 - 4(x - A - B/2)^2}} dx$$

$$\leq 1 - \left(1 - 8 \times 10^3 \frac{n^{-q}}{n^{-2q'}}\right)^{\frac{1}{2}}$$

$$\times \left(1 - \frac{1}{2\pi} \int_{(A, A+2n^{-q'}) \cup (A+B-2n^{-q'}, A+B)} \frac{1}{\sqrt{B^2 - 4(x - A - B/2)^2}} dx\right)$$

$$\leq 1 - \left(1 - 8 \times 10^3 \frac{n^{-q}}{n^{-2q'}}\right)^{\frac{1}{2}} \left(1 - \frac{2}{\pi} n^{-\frac{1}{2}q'}\right)$$

$$\leq 5 \times 10^3 n^{2q'-q} + \frac{2}{\pi} n^{-\frac{1}{2}q'}.$$

This implicitly defines a coupling of S, S' that satisfies $\mathbb{P}[S = S'] \geq 1 - 5 \times 10^3 n^{2q'-q} + \frac{2}{\pi} n^{-\frac{1}{2}q'} \geq 1 - 6 \times 10^3 \times n^{-c}$ for all sufficiently large n.

Once we have a coupling of S, S' satisfying (4.6), we can extend it to a coupling of θ, θ' that satisfies (4.7) via a suitable rotation. To be more precise, define the map $G : [0, 2\pi) \mapsto [0, 1] \times \{-1, 1\} \times \{-1, 1\}$ by

$$G(\psi) \equiv (G_1(\psi), G_2(\psi), G_3(\psi)) = \left(\cos^2(\phi), \frac{\cos(\phi)}{|\cos(\phi)|}, \frac{\sin(\phi)}{|\sin(\phi)|}\right).$$

If $\theta \sim \text{Unif}[0, 2\pi)$, then $G_1(\theta)$, $G_2(\theta)$ and $G_3(\theta)$ are independent, with $\mathbb{P}[G_2(\theta) = 1] = \mathbb{P}[G_3(\theta) = 1] = \frac{1}{2}$. Furthermore, S depends only on $G_1(\theta)$ and S' depends only on $G_1(\theta')$. Thus, when extending a coupling of S, S' to a coupling of θ, θ' we are free to choose any coupling of $G_2(\theta), G_2(\theta')$ and any coupling of $G_3(\theta), G_3(\theta')$. We choose the coupling for which

$$G_i(\theta) = G_i(\theta')$$

for $i \in \{2, 3\}$. Thus, our coupling of θ, θ' automatically satisfies both inequality (4.6) and (4.7) whenever S, S' satisfy inequality (4.6). $\square$

5. Proof of Theorem 1. We begin by proving the lower bound of Theorem 1, equation (1.2). Fix $C_1 < 1/2$ and let $\{T_1(n)\}_{n \geq 2}$ satisfy $T_1(n) < C_1 n \log(n)$. Fix $X_0 = x \in S^{n-1}$, let $\{(i_t, j_t, \theta_t)\}_{t \geq 0}$ be the update sequence associated with $\{X_t\}_{t \geq 0}$ and define

$$\zeta = \min\left\{t > 0 : \bigcup_{0 \leq s < t} \{i_s, j_s\} = \{1, 2, \ldots, n\}\right\}.$$

For $t < \zeta$, $X_t \in \mathcal{A} \equiv \bigcup_{i=1}^{n} \{X \in \mathrm{S}^{n-1} : X[i] = x[i]\}$, and $\mu(\mathcal{A}) = 0$. Thus,

(5.1)
$$\|\mathcal{L}(X_{T_1(n)}) - \mu\|_{\mathrm{TV}} \geq \mathbb{P}[X_{T(n)} \in \mathcal{A}] - \mu(\mathcal{A})$$
$$\geq \mathbb{P}[\zeta > T_1(n)] - 0.$$

By the standard coupon collector problem (see, e.g., [7]),

(5.2)
$$\lim_{n \to \infty} \mathbb{P}[\zeta > T_1(n)] = 1.$$

Combining inequality (5.1) with (5.2) completes the proof of equation (1.2).

Let $a = 47$, $b = 18.1$. To prove inequality (1.3), we fix sequences $T_2(n)$, $T_2'(n)$ and $T_2''(n)$ that satisfy

$$200n \log(n) < T_2(n) < n^2,$$

and $T_2'(n) \geq (4a + 5)n \log(n)$, $T_2''(n) \geq n \log(n)$ with

$$T_2'(n) + T_2''(n) = T_2(n)$$

for all n sufficiently large. We then construct a coupling of two copies $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$ of Kac's walk on the sphere with starting points $X_0 = x \in \mathrm{S}^{n-1}$ and $\mathcal{L}(Y_0) = \mu$. The coupling is as follows:

1. Couple $\{X_t\}_{0 \leq t \leq T_2'(n)}$, $\{Y_t\}_{0 \leq t \leq T_2'(n)}$ by using, at each step, the *proportional coupling* as in Definition 3.1.
2. Conditional on $\{X_t\}_{0 \leq t \leq T_2'(n)}$, $\{Y_t\}_{0 \leq t \leq T_2'(n)}$, couple $\{X_t\}_{T_2'(n) \leq t \leq T_2(n)}$, $\{Y_t\}_{T_2'(n) \leq t \leq T_2(n)}$ according to the non-Markovian coupling constructed in Section 4.1. Thus in the notation of Section 4.1, we have $T_0 = T_2'(n)$ and $T = T_2(n)$.
3. Conditional on $\{X_t\}_{0 \leq t \leq T_2(n)}$, $\{Y_t\}_{0 \leq t \leq T_2(n)}$, we couple the remaining steps of the two chains as follows. First, run $\{X_t\}_{t > T_2(n)}$ conditional on $\{X_t\}_{0 \leq t \leq T_2(n)}$ according to its distribution. If $X_{T_2(n)} = Y_{T_2(n)}$, set $Y_t = X_t$ for all $t > T_2(n)$. If $X_{T_2(n)} \neq Y_{T_2(n)}$, run $\{Y_t\}_{t > T_2(n)}$ conditional on $\{Y_t\}_{0 \leq t \leq T_2(n)}$ according to its distribution and independently of $\{X_t\}_{t > T_2(n)}$.[7]

Define the events

$$\mathcal{E}_1 = \left\{ \|A_{T_2'(n)} - B_{T_2'(n)}\|_1 > n^{-a} \right\} \cup \left\{ \min_{1 \leq i \leq n} X_{T_2'(n)}[i] Y_{T_2'(n)}[i] < 0 \right\},$$

$$\mathcal{E}_2 = \{X_{T_2(n)} \neq Y_{T_2(n)}\},$$

$$\mathcal{E}_3 = \{\mathcal{P}_{T_2'(n)} \neq \{1, 2, \ldots, n\}\}.$$

---

[7] The details of this third step of the coupling are irrelevant to the following analysis. They are provided only to give a concrete coupling that satisfies the condition $\mathbb{P}[\forall t > \tau(x), X_t = Y_t] = 1$ in the statement of Lemma 2.1.

By Lemma 2.1,

$$\sup_{X_0 = x \in \mathbb{S}^{n-1}} \| \mathcal{L}(X_{T_2(n)}) - \mu \|_{\mathrm{TV}}$$

(5.3)
$$\leq \sup_{X_0 = x \in \mathbb{S}^{n-1}} \mathbb{P}[\mathcal{E}_2]$$

$$\leq \sup_{X_0 = x \in \mathbb{S}^{n-1}} \left(\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_3] + \mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c] \right).$$

The remainder of the proof of inequality (1.3) will consist of bounding these three terms, in order.

Applying inequality (3.4) and then Lemma 3.3 and Markov's inequality,

$$\lim_{n \to \infty} \sup_{X_0 = x \in \mathbb{S}^{n-1}} \mathbb{P}[\mathcal{E}_1]$$

$$\leq \lim_{n \to \infty} \sup_{X_0 = x \in \mathbb{S}^{n-1}} \left(\mathbb{P}\left[\| A_{T_2'(n)} - B_{T_2'(n)} \|_1 \geq n^{-a} \right] \right.$$

(5.4)
$$+ \mathbb{P}\left[\min_{1 \leq i \leq n} X_{T_2'(n)}[i] Y_{T_2'(n)}[i] < 0 \right] \right)$$

$$\leq \lim_{n \to \infty} \sup_{X_0 = x \in \mathbb{S}^{n-1}} \left(\mathbb{P}\left[\| A_{T_2'(n)} - B_{T_2'(n)} \|_2 \geq n^{-a-1} \right] + n e^{-\frac{T_2'(n)}{n}} \right)$$

$$\leq \lim_{n \to \infty} \left(n^{2a+2} \left(1 - \frac{1}{2n} \right)^{T_2'(n)} + n^{-1} \right) = 0.$$

By Lemma 4.3,

(5.5)
$$\lim_{n \to \infty} \sup_{X_0 = x \in \mathbb{S}^{n-1}} \mathbb{P}[\mathcal{E}_3] = 0.$$

Recall the event $\mathcal{A}(s)$ from equation (4.3):

$$\mathcal{A}(s) = \{ \text{equations (4.1) and (4.2) are satisfied for all } T_2'(n) \leq t \leq s \}.$$

Our next goal is, roughly, to show that *conditional on the first phase of the coupling having gone according to plan*, the set $\mathcal{A}(s+1) \setminus \mathcal{A}(s)$ has small probability; this is stated formally in inequality (5.10).

For $T_2'(n) \leq s < T_2(n)$, define

$$\mathcal{B}(s) = \left\{ \min_{T_2'(n) \leq t < s} \min_{1 \leq i \leq n} Y_t[i]^2 \geq \frac{1}{2} n^{-b} \right\}.$$

We will bound $\mathbb{P}[\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c]$ by showing that $\mathbb{P}[\mathcal{A}(t+1)^c \cap \mathcal{A}(t) \cap \mathcal{B}(t) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c]$ is small for all $T_2'(n) \leq t \leq T_2(n)$.

To this end, first note that

$$\mathbb{P}[\mathcal{A}(T_2'(n))^c \cap \mathcal{E}_3^c] = 0.$$

N. S. PILLAI AND A. SMITH

Recall the set $\mathcal{S}$ from the construction of non-Markovian coupling from Section 4.1. We now consider two cases: $t \in \mathcal{S}$ and $t \in \{T_2'(n), \ldots, T_2(n)\} \setminus \mathcal{S}$. In the case $t \in \{T_2'(n), \ldots, T_2(n)\} \setminus \mathcal{S}$, we have

$$\sum_{l \in P_r(t)} A_t[l] = \sum_{l \in P_r(t+1)} A_{t+1}[l], \qquad P_r(t+1) = P_r(t) \in \mathcal{P}_t, \qquad 1 \le r \le \ell_t.$$

This implies that

$$(5.6) \quad \mathbb{P}\big[\mathcal{A}(t+1)^c \cap \mathcal{A}(t) \cap \mathcal{B}(t) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] \le \mathbb{P}\big[\mathcal{A}(t+1)^c \cap \mathcal{A}(t) \cap \mathcal{E}_3^c\big] = 0.$$

Next, consider $t = s_k \in \mathcal{S}$. We will apply Lemma 4.6 with the random variables:

$$(5.7) \quad \begin{aligned} S &= \sum_{r \in P_{u(s_k)}(s_k+1) \setminus \{i_{s_k}\}} X_{s_k}[r]^2 + \big(X_{s_k}[i_{s_k}]^2 + X_{s_k}[j_{s_k}]^2\big) \cos(\theta_{s_k})^2, \\ S' &= \sum_{r \in P_{u(s_k)}(s_k+1) \setminus \{i_{s_k}\}} Y_{s_k}[r]^2 + \big(Y_{s_k}[i_{s_k}]^2 + Y_{s_k}[j_{s_k}]^2\big) \cos(\theta_{s_k})^2. \end{aligned}$$

On the event of $\mathcal{A}(s_k) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c$, equation (4.5) of Lemma 4.4 gives that

$$(5.8) \qquad \|A_{s_k} - B_{s_k}\|_1 \le n \|A_{T_2'(n)} - B_{T_2'(n)}\|_1 \le n^{1-a},$$

and so on the event $\mathcal{A}(s_k) \cap \mathcal{B}(s_k) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c$ we also have

$$\min_{T_2'(n) \le t < s} \min_{1 \le i \le n} \min\big(X_t[i]^2, Y_t[i]^2\big) \ge n^{-b}.$$

Thus, on the event $\mathcal{A}(s_k) \cap \mathcal{B}(s_k) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c$, the random variables S, S' given in equation (5.7) satisfy the assumptions of Lemma 4.6, with $p = b$, $q = a - 1$ and $q' = \frac{2(a-1)}{5} \in (b, \frac{a-1}{2})$. Thus, by Lemma 4.6,

$$\mathbb{P}\Bigg[\Bigg\{\sum_{i \in P_{u(s_k)}(s_k+1)} X_{s_k+1}[i]^2 \ne \sum_{i \in P_{u(s_k)}(s_k+1)} Y_{s_k+1}[i]^2\Bigg\}$$

$$(5.9) \qquad \cap \, \mathcal{A}(s_k) \cap \mathcal{B}(s_k) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\Bigg]$$

$$\le 6 \times 10^3 n^{-\frac{2(a-1)}{5}}.$$

By the constraint $\sum_{i=1}^n A_{t+1}[i] = \sum_{i=1}^n B_{t+1}[i] = 1$, we have

$$\Bigg\{\sum_{i \in P_{u(s_k)}(s_k+1)} X_{s_k+1}[i]^2 \ne \sum_{i \in P_{u(s_k)}(s_k+1)} Y_{s_k+1}[i]^2\Bigg\} \cap \mathcal{A}(s_k) = \mathcal{A}(s_k+1)^c \cap \mathcal{A}(s_k).$$

Combining this with inequality (5.9) gives

$$(5.10) \qquad \mathbb{P}\big[\mathcal{A}(s_k+1)^c \cap \mathcal{A}(s_k) \cap \mathcal{B}(s_k) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] \le 6 \times 10^3 n^{-\frac{2(a-1)}{5}}.$$

By Lemma 3.5 and a union bound over times $T_2'(n) \leq t \leq T_2(n)$ and indices $1 \leq i \leq n$, we have for all sufficiently large n,

$$(5.11) \qquad \mathbb{P}\big[\mathcal{B}(T_2(n))^c\big] \leq 3n^{4-\frac{b}{3}}.$$

Combining inequalities (5.10) and (5.6) and then (5.11), we have

$$
\begin{aligned}
&\mathbb{P}\big[\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] \\
&\quad = \mathbb{P}\big[\mathcal{A}(T_2(n))^c \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] \\
&\quad \leq \sum_{s=T_2(n)-T_2'(n)}^{T_2(n)} \mathbb{P}\big[\mathcal{A}(T_2(n)-s+1)^c \cap \mathcal{A}(T_2(n)-s) \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] \\
&\qquad + \mathbb{P}\big[\mathcal{A}(T_2'(n))^c \cap \mathcal{E}_1^c\big] \\
&\quad \leq \sum_{s=T_2(n)-T_2'(n)}^{T_2(n)} \big(\mathbb{P}\big[\mathcal{A}(T_2(n)-s+1)^c \cap \mathcal{A}(T_2(n)-s) \cap \mathcal{B}(T_2(n)-s) \\
&\qquad\quad \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] \\
&\qquad + \mathbb{P}\big[\mathcal{B}(T_2(n)-s)^c\big]\big) + \mathbb{P}\big[\mathcal{A}(T_2'(n))^c \cap \mathcal{E}_1^c\big] \\
&\quad \leq 6 \times 10^3 T_2(n) n^{-\frac{2(a-1)}{5}} + T_2(n)n^{4-\frac{b}{3}} + 0 \\
&\quad \leq 6 \times 10^3 n^{2-\frac{2(a-1)}{5}} + n^{6-\frac{b}{3}}.
\end{aligned}
$$

Since $b > 18$ and $a > 6$, this implies

$$(5.12) \qquad \lim_{n \to \infty} \sup_{X_0=x \in \mathrm{S}^{n-1}} \mathbb{P}\big[\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_3^c\big] = 0.$$

Inequality (1.3) follows immediately from inequalities (5.3), (5.4), (5.5) and (5.12). For sequences with $T_2(n) \leq n^2$ for all n sufficiently large, this completes the proof of inequality (1.3). For sequences with $T_2(n) > n^2$ infinitely often, inequality (1.3) follows from the case $T(n) = n^2$ by the fact that the total variation distance to stationarity of a Markov chain is monotonely decreasing in time. This completes the proof of Theorem 1.

REFERENCES

[1] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[2] BORMASHENKO, O. (2011). A coupling argument for the random transposition walk. Preprint. Available at arXiv:1109.3915.

[3] 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

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

[5] CARLEN, E. A., CARVALHO, M. C., ROUX, J. L., LOSS, M. and VILLANI, C. (2008). Entropy and chaos in the Kac model. Preprint. Available at arXiv:0808.3192.

[6] CONNOR, S. and JACKA, S. (2008). Optimal co-adapted coupling for the symmetric random walk on the hypercube. *J. Appl. Probab.* **45** 703–713. MR2455179

[7] ERDŐS, P. and RÉNYI, A. (1961). On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6** 215–220. MR0150807

[8] HAURAY, M. and MISCHLER, S. (2014). On Kac's chaos and related problems. *J. Funct. Anal.* **266** 6055–6157. MR3188710

[9] HAYES, T. P. and VIGODA, E. (2003). A non-Markovian coupling for randomly sampling colorings. In *Foundations of Computer Science*, 2003. *Proceedings*. 44*th Annual IEEE Symposium on* 618–627. IEEE, Cambridge, MA.

[10] HAYES, T. P. and VIGODA, E. (2007). Variable length path coupling. *Random Structures & Algorithms* **31** 251–272. MR2352176

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

[12] JIANG, Y. (2012). Total variation bound for Kac's random walk. *Ann. Appl. Probab.* **22** 1712–1727. MR2985175

[13] KAC, M. (1956). Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. III* 171–197. Univ. California Press, Berkeley and Los Angeles. MR0084985

[14] KENDALL, W. S. (2015). Coupling, local times, immersions. *Bernoulli* **21** 1014–1046. MR3338655

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

[16] MADRAS, N. and SEZER, D. (2010). Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli* **16** 882–908. MR2730652

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

[18] MISCHLER, S. and MOUHOT, C. (2013). Kac's program in kinetic theory. *Invent. Math.* **193** 1–147. MR3069113

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

[20] OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937

[21] ROBERTS, G. O. and ROSENTHAL, J. S. (2002). One-shot coupling for certain stochastic recursive sequences. *Stochastic Process. Appl.* **99** 195–208. MR1901153

[22] SMITH, A. (2014). A Gibbs sampler on the *n*-simplex. *Ann. Appl. Probab.* **24** 114–130. MR3161643

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1, OXFORD ST.
CAMBRIDGE, MASSACHUSETTS 01880
USA
E-MAIL: pillai@fas.harvard.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
585 KING EDWARD AVENUE
OTTAWA, ON K1N 7N5
CANADA
E-MAIL: asmi28@uottawa.ca