The Annals of Probability
2018, Vol. 46, No. 4, 2345–2399
https://doi.org/10.1214/17-AOP1230

ON THE MIXING TIME OF KAC'S WALK AND OTHER HIGH-DIMENSIONAL GIBBS SAMPLERS WITH CONSTRAINTS

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

Harvard University and University of Ottawa†*

Determining the total variation mixing time of Kac's random walk on the special orthogonal group SO(n) has been a long-standing open problem. In this paper, we construct a novel non-Markovian coupling for bounding this mixing time. The analysis of our coupling entails controlling the smallest singular value of a certain random matrix with highly dependent entries. The dependence of the entries in our matrix makes it not amenable to existing techniques in random matrix theory. To circumvent this difficulty, we extend some recent bounds on the smallest singular values of matrices with independent entries to our setting. These bounds imply that the mixing time of Kac's walk on the group SO(n) is between $C_1 n^2$ and $C_2 n^4 \log(n)$ for some explicit constants $0 < C_1, C_2 < \infty$, substantially improving on the bound of $O(n^5 \log(n)^2)$ in the preprint of Jiang [Jiang (2012)]. Our methods may also be applied to other high dimensional Gibbs samplers with constraints, and thus are of independent interest. In addition to giving analytical bounds on the mixing time, our approach allows us to compute rigorous estimates of the mixing time by simulating the eigenvalues of a random matrix.

CONTENTS

1. Introduction . . . 2346
   1.1. Motivations outside of physics . . . 2348
   1.2. Previous work . . . 2348
   1.3. Our contributions . . . 2349
2. Preliminaries . . . 2350
   2.1. Coupling . . . 2351
   2.2. Random mapping representation of Kac's walk . . . 2351
   2.3. Lie groups and SO(n) . . . 2352
   2.4. Key distance scales . . . 2354
3. Proof strategy . . . 2354
4. Coupling inequality . . . 2355
   4.1. Coupling inequality . . . 2356
   4.2. Choice of $\mathcal{R}$ . . . 2357
   4.3. Choice of κ . . . 2358

Received June 2016; revised May 2017.
[1]Supported by an ONR grant.
[2]Supported by a grant from NSERC.
MSC2010 subject classifications. Primary 60J10; secondary 60J20.
Key words and phrases. Markov chain, Gibbs sampler, coupling, mixing times, Kac's random walk.

5. Technical lemmas . . . 2358
5.1. Matrix estimates . . . 2358
5.2. Probability estimates . . . 2360
6. Coupling argument . . . 2361
7. Relating mixing times to singular values . . . 2364
8. Singular values of random matrices . . . 2365
8.1. Bounds on determinant of random matrices . . . 2366
8.2. Bounding the smallest singular value of D_∞ . . . 2371
8.3. Application to Kac's walk . . . 2375
9. Proof of Theorem 1.1 . . . 2377
10. Discussion . . . 2378
10.1. Open problems . . . 2378
10.2. Applications to other Gibbs samplers . . . 2379
10.3. Applications to simulation . . . 2379
Appendix A: Proof of Theorem 4.3 . . . 2380
A.1. Density of random mapping representation . . . 2381
A.2. Coupling construction . . . 2385
A.3. Proof of Theorem 4.3 . . . 2387
Appendix B: Proof of Lemma 4.6 . . . 2388
Appendix C: Proofs of technical bounds . . . 2390
C.1. Matrix estimates . . . 2391
C.2. Probability estimates . . . 2396
Acknowledgements . . . 2396
References . . . 2396

1. Introduction. Mark Kac introduced a random walk on the sphere in his 1954 paper [39] as a model for a Boltzmann gas. In this paper, we study Kac's walk on the special orthogonal group SO(n), which was first introduced in a statistical context [27] and has been studied as a generalization of Kac's walk on the sphere since [17] (see also, e.g., [8, 32, 35, 45, 46]).

Kac's walk on SO(n) is a discrete-time Markov chain $\{X_t\}_{t\in\mathbb{N}}$ that evolves as follows. Fix an ordering of the $N \equiv \frac{n(n-1)}{2}$ planes generated by two coordinates in $\mathbb{R}^n$ and choose $X_1 \in \mathrm{SO}(n)$. For $t \in \mathbb{N}$, choose $1 \le i_t \le N$ and $\theta_t \in [0, 2\pi)$ uniformly at random and set

$$X_{t+1} = \mathrm{R}(i_t, \theta_t) X_t, \tag{1.1}$$

where $\mathrm{R}(i,\theta)$ denotes a rotation by the angle θ in the ith coordinate plane. If the ith coordinate plane is associated with the coordinate axes $1 \le k < \ell \le n$, $\mathrm{R}(i,\theta)$ is an $n \times n$ matrix with entries

$$\begin{aligned} &\mathrm{R}(i,\theta)_{jj} = \cos(\theta), \qquad j \in \{k,\ell\},\\ &\mathrm{R}(i,\theta)_{k\ell} = \sin(\theta), \qquad \mathrm{R}(i,\theta)_{\ell k} = -\sin(\theta),\\ &\mathrm{R}(i,\theta)_{jj} = 1, \qquad j \notin \{k,\ell\},\\ &\mathrm{R}(i,\theta)_{jj'} = 0, \qquad j' \notin \{j,k,\ell\}. \end{aligned} \tag{1.2}$$

If we write $X_t = [v_t^{(1)} \ v_t^{(2)} \ \dots \ v_t^{(n)}]$, the law of $\{v_t^{(1)}\}_{t\in\mathbb{N}}$ is known as *Kac's walk on the sphere* S^{n-1}. Physically, Kac motivated this random walk by considering n particles in a one-dimensional box. He assumed that these particles were uniformly distributed in space, and the vector $v_t^{(1)}$ models the change in their velocities over time as collisions occur; the condition that $v_t^{(1)}$ be constrained to the 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 (see [44] for a useful description of this program). Kac's walks on the sphere and on SO(n) have attracted great attention and estimating their mixing times has been a long standing open problem (see Sections 1.1 and 1.2). Recently, in [47], the authors of this paper obtained a matching upper bound and lower bound for the mixing time of Kac's walk in S^{n-1}, thus settling this problem up to a constant factor.

To state our main result, 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

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

We denote the distribution of a random variable X by $\mathcal{L}(X)$ and write $X \sim \nu$ as a shorthand for $\mathcal{L}(X) = \nu$. For a Markov chain $\{X_t\}_{t\in\mathbb{N}}$ with unique associated stationary distribution ν on state space Ω, we define the associated *mixing profile* by

$$\tau(\varepsilon) = \inf\Big\{t \in \mathbb{N} : \sup_{X_1 = x\in\Omega} \big\|\mathcal{L}(X_t) - \nu\big\|_{\mathrm{TV}} < \varepsilon\Big\}$$

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

Let μ denote the normalized Haar measure on SO(n). Our main result is the following bound on the mixing time of Kac's walk on SO(n).

THEOREM 1.1. *Let $\{X_t\}_{t\geq 0}$ be a copy of Kac's walk on* SO(n). *Then for all sequences $T = T(n) > 10^7 n^4 \log(n)$,*

$$\limsup_{n\to\infty} \sup_{X_1 = x\in\mathrm{SO}(n)} \big\|\mathcal{L}(X_T) - \mu\big\|_{\mathrm{TV}} = 0, \tag{1.3}$$

and for all sequences $T = T(n) < N$,

$$\liminf_{n\to\infty} \sup_{X_1 = x\in\mathrm{SO}(n)} \big\|\mathcal{L}(X_T) - \mu\big\|_{\mathrm{TV}} = 1. \tag{1.4}$$

We have not tried to optimize the constant 10^7 appearing in Theorem 1.1.

1.1. *Motivations outside of physics*. Kac's random walk has been studied in a wide range of fields including computer science, statistics and numerical analysis. To our knowledge, the Markov chain that we call Kac's walk on $\mathrm{SO}(n)$ was initially proposed in [27] as a Gibbs sampler targeting the Haar measure on $\mathrm{SO}(n)$. The problem of sampling from Haar measure on $\mathrm{SO}(n)$ was motivated by [33], but the walk itself has been suggested as a computationally efficient method for finding projections onto random small-dimensional subspaces [2]. Bounds on the mixing time of Kac's walk are required to check that this approach is, in fact, computationally efficient.

Our analysis of Kac's walk is also interesting as a worked example that belongs to several active areas of research. The Markov chains we study are a sequence of high-dimensional Gibbs samplers (see [10]). Despite three decades of extensive work in this area, there are few effective bounds on the mixing times of Gibbs samplers in high dimensions (see [16, 38] for an introduction to the large literature on this problem). Of the existing effective bounds on the mixing times of high-dimensional Gibbs samplers on continuous state spaces, almost all target distributions with support equal to a union of quadrants of $\mathbb{R}^n$ (e.g., [38]) or involve explicitly computing spectral information for the transition kernel (e.g., the analyses [32, 48, 50] of other random walks on $\mathrm{SO}(n)$). Our analysis gives one of relatively few results for Gibbs samplers on a complicated sample space for which spectral information cannot easily be used. Some closely related papers are [41, 42] on Gibbs samplers on convex sets, as well as the papers [36, 53] directly motivated by the study of Kac's walk on the sphere. There is also a large number of papers studying the mixing time of Markov chains on groups [15, 24, 32, 50, 51] using Fourier analysis and representation theory. Unlike these papers, we do not use Fourier analytic tools, and thus our methods are in principle generalizable to other Gibbs samplers in $\mathbb{R}^n$ for which spectral information is not available.

1.2. *Previous work*. The central question in the study of Kac's walk is to determine the speed at which it converges to equilibrium. This question is somewhat vague, as it does not specify the metric under which convergence is to be measured. Early work focused on proving that the spectral gap of the chain was large. In [17], the authors showed that the spectral gap of the walk on $\mathrm{SO}(n)$ was at least order of n^{-3}. Janvresse first showed in [34] that the spectral gap of the walk on the sphere was exactly on the order of n^{-1}. Janvresse also showed in [35] that the spectral gap of the walk on $\mathrm{SO}(n)$ was on the order of n^{-2}. The exact spectral gap for both walks was found in [8], and the full spectrum was computed in [43]. Some of this work was generalized in [7]. Although these bounds 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 $\mathcal{L}(X_1)$ is concentrated at a point, the initial L^2 distance to stationarity is not finite.

Later work has focused on stronger metrics for convergence or more demanding versions of the problem. In [9], a very strong convergence condition as measured by entropy was discovered. These bounds, like the bounds relating to spectral gap, only imply convergence for sufficiently smooth initial distribution $\mathcal{L}(X_1)$. In this note, we focus on convergence bounds that do not depend on the initial distribution $\mathcal{L}(X_1)$. The first bound with this property was obtained [46], in which the authors showed a convergence time of order $O(n^{2.5})$ in the L^1 Wasserstein metric, and [45] improved this bound to $O(n^2 \log(n))$ in the stronger L^2 Wasserstein metric. This latter bound is tight up to factors of $\log(n)$, and will be essential to our work. Related Wasserstein bounds have also been found in [12] for several similar models. However, a mixing bound in the Wasserstein metric does not directly imply any mixing bound in the total variation metric.

Thus the bounds obtained in [45, 46], despite their strength, do not give any information at all about the mixing time in total variation distance. The first bound on convergence in total variation was on the order of 4^{n^2} steps, obtained by Diaconis and Saloff-Coste in [17]. No progress was made on this problem until the recent unpublished work of Yunjiang Jiang [37], in which the author obtained a mixing bound of order $n^5 \log^2(n)$.

Theorem 1.1 of our paper also implies a mixing bound on the order of $n^4 \log(n)$ for Kac's walk on S^{n-1}. This improves upon all bounds prior to the present author's recent work [47], which shows matching upper and lower bounds on the order of $n \log(n)$ for this walk. The papers [28, 32, 44, 48, 50] all study variants or projections of Kac's walk on SO(n).

1.3. *Our contributions*. We have three main contributions: an order of magnitude improvement on the previous best bound for the convergence rate of Kac's walk in the strong total variation metric, new bounds on the smallest singular values of certain random matrices with dependent entries, and a general approach to bounding the mixing times of Gibbs samplers on spaces that are not "rectangular". Our method also gives a way to compute effective mixing bounds via simulation. We now give a broad overview of our approach, and its relationship to some previous work.

The upper bound of $O(n^4 \log(n))$ on the mixing time of Kac's walk is proved by using the popular *coupling* technique: we run two copies $\{X_t\}_{t\in\mathbb{N}}$, $\{Y_t\}_{t\in\mathbb{N}}$ of Kac's walk, and study the first time $\inf\{t \in \mathbb{N} : X_t = Y_t\}$ that they collide. Like many non-Markovian couplings (see, e.g., [13, 29, 47, 53]), the main idea is to construct a coupling in two passes: an initial Markovian coupling of "most" of the randomness in the chain, followed by a very general coupling of some "leftover" randomness. Our initial coupling is exactly the one constructed in [45]. Under this coupling, two copies of Kac's walk mix in Wasserstein distance after after $O(n^2 \log(n))$ steps. Our contribution is in the construction and analysis of the second stage of the coupling.

The usual approach to converting a Wasserstein mixing bound for a high-dimensional Gibbs sampler to a total variation mixing bound is via a greedy coupling: one attempts to match more and more coordinates as time progresses. Unfortunately, this approach works poorly for Kac's walk on $\mathrm{SO}(n)$. Indeed, as the authors discuss in [47], the greedy approach does not even work in the simpler case of Kac's walk on the sphere. Instead, we first imagine running two coupled copies of Kac's walk $(\tilde{X}_t, \tilde{Y}_t)$ according to a coupling from [45]. These will act as "scaffolding" for our final coupling. We then construct an N-dimensional perturbation of this scaffold by adding a small amount of additional randomness at each of N time steps. The key point here is that it turns out to be easier to analyze the coupling probability of our two chains by using the N bits of randomness all at once, rather than analyzing the coupling probability of the individual coordinate at each step as done in [47]. Our approach is somewhat reminiscent of the "sprinkling" strategy used in random graph theory [3], which also involves coupling "most" random variables in an intuitive way and then carefully analyzing a small amount of "leftover" randomness.

We expect that a non-Markovian coupling with the properties described in Section A.3 should exist for any Gibbs sampler that satisfies certain rather weak technical conditions (see Section 10.2 for discussion), though we do not give a formal statement of a result of this form. We believe that a similar analysis should also be applicable for other constrained high-dimensional Gibbs samplers, with the main technical difficulties being a bound on the singular value of a certain random matrix (see Lemma 8.10) and certain continuity conditions (see Lemma 5.6). This is in stark contrast to our previous work [47] on Kac's walk on the sphere, which used a similar approach that could not easily be extended.

Our simple results on random matrices in Section 8 are novel, giving bounds on the smallest singular values of random matrices with significant dependence between entries. These bounds are closely related to the results of [20, 22] on the smallest singular values of random matrices with independent entries, and give bounds that are qualitatively similar to [22].

In addition to giving asymptotic results on mixing times, our method allows us to numerically estimate the mixing time of Kac's walk for fixed n by simulating the singular values of a certain random matrix (see Section 10.3). This is useful for those interested in knowing the mixing time of a particular Gibbs sampler, and will generally give sharper results than our mathematical analysis. Note that estimating the mixing time of a Markov chain in this way is not trivial—*a priori* it is not obvious how to obtain *any* rigorous bounds on the mixing time of Kac's walk by a finite computation; see, e.g., [11] for further discussion on estimating mixing times via simulation, and [52] for tests of weak mixing on $\mathrm{SO}(n)$.

2. Preliminaries. We give notation that will be used throughout the paper, including a review of some important definitions and results from probability, differential geometry and the theory of Lie groups. For functions $f, g : \mathbb{N} \mapsto \mathbb{R}$, we write

$f = O(g)$ if $\limsup_{n\to\infty} \frac{|f(n)|}{|g(n)|} < \infty$. We also write $f = \Omega(g)$ if $g = O(f)$ and we write $f = \Theta(g)$ if both $f = O(g)$ and $f = \Omega(g)$. Finally, we write $f = o(g)$ if $\limsup_{n\to\infty} \frac{|f(n)|}{|g(n)|} = 0$. Unless otherwise noted, the terms inside of such "big-O" notation should always be taken with respect to the problem dimension n.

2.1. *Coupling*. Recall that a *coupling* of two probability measures π_1, π_2 on measure spaces $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ is a probability measure Π on the product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ with marginals

$$\Pi(A, \Omega_2) = \pi_1(A),$$
$$\Pi(\Omega_1, B) = \pi_2(B)$$

for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ (where $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the smallest σ-algebra containing $\mathcal{F}_1 \times \mathcal{F}_2$). Although we define couplings in terms of measures, we sometimes slightly abuse notation and refer to a coupling of a pair of random variables X, Y, possibly on different probability spaces. This should always be interpreted in the following way: we construct a measure π that is a coupling of the two measures $\mathcal{L}(X)$, $\mathcal{L}(Y)$ and replace the random variables X, Y with a new pair of random variables $(\tilde{X}, \tilde{Y}) \sim \pi$ on the same probability space.

In constructing our coupling, we sometimes have to combine two transition kernels (sometimes also called "probability kernels"). For a pair of transition kernels $\mu : \Omega_1 \mapsto \Omega_2$ and $\nu : \Omega_2 \mapsto \Omega_3$, we denote by $\mu \otimes \nu : \Omega_1 \mapsto \Omega_2 \times \Omega_3$ the "usual" combination of two probability kernels; see Lemma 1.38 of [40] and the discussion around it for a careful definition. Informally, for $x \in \Omega_1$ a sample $(X_2, X_3) \sim (\mu \otimes \nu)(x, \cdot)$ can be thought of as being obtained by first sampling $X_2 \sim \mu(x, \cdot)$ and then sampling $X_3 \sim \nu(X_2, \cdot)$. This informal description corresponds exactly to the precise definition when $\Omega_1, \Omega_2, \Omega_3$ are all finite and have the usual σ-algebras. Similarly, when μ is a a measure on Ω_1 and $\nu : \Omega_1 \mapsto \Omega_2$ is a transition kernel, we denote by $\mu \otimes \nu$ the "usual" combined measure on $\Omega_1 \times \Omega_2$. Informally, a sample $(X_1, X_2) \sim \mu \otimes \nu$ can be thought of as being obtained by first sampling $X_1 \sim \mu$ and then sampling $X_2 \sim \nu(X_1, \cdot)$.

2.2. *Random mapping representation of Kac's walk*. We recall the definition of a *random mapping representation* of a Markov chain. Let Q be the transition kernel of a time-homogenous Markov chain on measure space $(\Omega_1, \mathcal{F}_1)$, let $(\Omega_2, \mathcal{F}_2, \nu)$ be a probability space, and let $Z \sim \nu$. We call a measurable function $f : \Omega_1 \times \Omega_2 \mapsto \Omega_1$ a *random mapping representation of Q* if it satisfies the equality

$$\mathbb{P}\big[f(x, Z) \in A\big] = Q(x, A)$$

for all $x \in \Omega_1$ and all $A \in \mathcal{F}_1$. We recall that it is possible to construct a Markov chain by its random mapping representation as follows. Let $\{Z_t\}_{t\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution ν, and let $x \in \Omega_1$ be a starting point. Then the sequence $\{X_t\}_{t\in\mathbb{N}}$ defined by the recursion

$$X_1 = x,\qquad X_{t+1} = f(X_t, Z_t) \tag{2.1}$$

is a Markov chain with transition kernel Q and starting point $X_1 = x$. If a Markov chain is constructed in terms of a random mapping, we call the associated sequence of i.i.d. random variables $\{Z_t\}_{t\in\mathbb{N}}$ the *update sequence* associated with $\{X_t\}_{t\in\mathbb{N}}$. Conversely, if $\{Z_t\}_{t=1}^T$ is a sequence of update variables for a random mapping representation f, we call the sequence $\{X_t\}_{t=1}^{T+1}$ defined in equation (2.1) the Markov chain *associated with* random mapping representation f and update sequence $\{Z_t\}_{t=1}^T$.

Throughout, we denote by unif(S) the uniform probability measure on a set S when there is an obvious uniform measure on S, for example, when S is finite, or S inherits a natural measure from Haar or Lebesgue measure.

We use the following random mapping representation F of Kac's walk:

$$\begin{aligned}\Omega_1 &= \mathrm{SO}(n),\\ \Omega_2 &= \mathfrak{A} \equiv \{1, 2, \ldots, N\} \times [0, 2\pi),\\ \nu &= \mathrm{Unif}(\mathfrak{A}),\\ \mathrm{F}\big(x, (i, \theta)\big) &= \mathrm{R}(i, \theta)x,\end{aligned} \tag{2.2}$$

where Ω_1, Ω_2 are given their usual topologies, $\mathcal{F}_1$, $\mathcal{F}_2$ are the usual completions of the associated Borel σ-algebras on their respective spaces, and $N \equiv \frac{n(n-1)}{2}$ and R are as defined following equation (1.1). For $T \in \mathbb{N}$, and vector

$$V = \big(x, (i_1, \theta_1), \ldots, (i_T, \theta_T)\big) \in \mathrm{SO}(n) \times \mathfrak{A}^T,$$

we define the iterated map $\mathrm{F}_T : \mathrm{SO}(n) \times \mathfrak{A}^T \mapsto \mathrm{SO}(n)^{T+1}$ by the recursion

$$\begin{aligned}\mathrm{F}_T(V)[1] &= x,\\ \mathrm{F}_T(V)[t+1] &= \mathrm{F}\big(\mathrm{F}_T(V)[t], (i_t, \theta_t)\big), \qquad 1 \le t \le T.\end{aligned} \tag{2.3}$$

Finally, we denote by K the transition kernel of Kac's walk.

2.3. *Lie groups and* SO(n). For $x, y \in \mathbb{R}^m$, we write $\langle x, y\rangle$ for the Euclidean inner product and $\|x - y\|$ for the associated norm. We also use "0" as a shorthand for the vector $(0, 0, \ldots, 0) \in \mathbb{R}^m$. Denote by $M(n)$ the collection of $n \times n$ matrices with real valued entries. For $h \in M(n)$, let $h^\dagger$ denote its transpose. For a linear map $T : \mathbb{R}^k \mapsto \mathbb{R}^\ell$, define the operator norm

$$\|T\|_{\mathrm{Op}} = \sup_{\|v\|=1} \|Tv\|.$$

For elements $X_1, X_2, \dots$ belonging to $\mathrm{SO}(n)$, we use the convention

$$\prod_{s=1}^{t} X_s \equiv X_1 X_2 \cdots X_{t-1} X_t.$$

Note that the order of multiplication matters due to the noncommutativity. We write ∂S for the usual topological boundary of any set S (the relevant topology will always be clear from the context). For any smooth manifold $\mathcal{M}$ and any $x \in \mathcal{M}$, we denote by $T_x\mathcal{M}$ the tangent space of $\mathcal{M}$ at x. We will often identify $T_x\mathbb{R}^m$ with $\mathbb{R}^m$. For any pair of smooth manifolds $\mathcal{M}, \mathcal{N}$ and any smooth function $f : \mathcal{M} \mapsto \mathcal{N}$, we define for $p \in \mathcal{M}$ the usual associated derivative map $\mathrm{d}f_p : T_p\mathcal{M} \to T_{f(p)}\mathcal{N}$. We recall an explicit construction of the derivative $\mathrm{d}f_p$. Fix $v \in T_p\mathcal{M}$ and let $\gamma : [0, 1] \to \mathcal{M}$ satisfy $\gamma(0) = p$, $\gamma'(0) = v$. Then

$$\mathrm{d}f_p(v) = (f \circ \gamma)'(0).$$

The quantity $\mathrm{d}f_p(v)$ is independent of the path γ as long as $\gamma'(0) = v$. The rank of the linear map $\mathrm{d}f_p$ is denoted by $\mathrm{Rank}(\mathrm{d}f_p)$.

Let G be a Lie group G with Lie algebra $\mathcal{G}$. For $a \in G$, let $\mathrm{L}_a : G \mapsto G$ be the left multiplication map

$$\mathrm{L}_a(g) = ag.$$

The exponential map, denoted by exp, maps $\mathcal{G}$ to G. When G is a matrix group and $\mathcal{G}$ is identified with a subset of $M(n)$, the exponential map has the explicit form

$$\exp(A) = \sum_{i=0}^{\infty} \frac{A^i}{i!}, \qquad A \in \mathcal{G} \tag{2.4}$$

(see Section 1.1 of [1]). The Hilbert–Schmidt inner product on $M(n)$ is

$$\langle A, B\rangle_{\mathrm{HS}} = \mathrm{Tr}\big(A^\dagger B\big),$$

where Tr is the trace. The corresponding inner product is denoted by $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$.

Throughout the paper, we will use the convention that the addition of angles θ, θ' is always done modulo 2π, and that the distance between two angles is measured with respect to the usual metric on the torus with 0 curvature, rather than the usual metric on the line.

The Lie algebra $\mathcal{G} = \mathfrak{so}(n)$ of $\mathrm{SO}(n)$ is the set of $n \times n$ skew-symmetric matrices

$$\mathfrak{so}(n) = \big\{h \in M(n) : h = -h^\dagger\big\}.$$

We denote by D_{HS} the Riemannian metric on $\mathrm{SO}(n)$ induced by the inner product $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$ on $\mathfrak{so}(n)$. We recall two important facts about the exponential map and this metric. Since D_{HS} is a bi-invariant metric on $\mathrm{SO}(n)$, we have that for any fixed $A \in \mathfrak{so}(n)$, sets of the form $\{\exp(tA)\}_{t\in\mathbb{R}}$ are geodesic paths with respect to the metric D_{HS} (see, e.g., Corollary 1 of [18]). We also have that the exponential map is a surjective map from $\mathfrak{so}(n)$ to $\mathrm{SO}(n)$ (see, e.g., Theorem 6.9.3 of [1]).

The Haar measure μ on $\mathrm{SO}(n)$ is also induced by the inner product $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$. We denote by $\mathcal{P}:\mathrm{SO}(n)\mapsto\mathfrak{so}(n)$ the orthogonal projection operator into $\mathfrak{so}(n)$ with respect to the Hilbert–Schmidt norm, that is,

$$\mathcal{P}(g)=\arg\min_{A\in\mathfrak{so}(n)}\|g-A\|_{\mathrm{HS}}.$$

We construct an orthonormal basis for $\mathfrak{so}(n)$ as follows. For $1\le i\le N$, $\mathrm{R}_i(\theta)\equiv\mathrm{R}(i,\theta)$ [see equation (1.1)] is a map from the N-dimensional torus $\mathbb{T}=[0,2\pi)^N$ to $\mathrm{SO}(n)$. Set

$$a_i=\frac{1}{\sqrt{2}}\mathrm{dR}_i(0)\in\mathfrak{so}(n).$$

The set $\{a_i\}_{1\le i\le N}$ constitute an orthonormal basis in $\mathfrak{so}(n)$.

2.4. *Key distance scales*. Throughout the paper, the quantities ϕ_n, ε_n and ω_n will control the three distance scales that are key to our coupling proof. The quantity

$$\phi_n=4^{-5N}N^{-120N}$$

controls the scale on which a certain function "looks flat". The quantity

$$\varepsilon_n=\phi_n^{30}$$

will control the total amount of "injected randomness" available to our coupling. Finally,

$$\omega_n=\varepsilon_n^{30}$$

controls the typical distance between the two Markov chains that we are trying to couple.

3. Proof strategy. We give an overview of the proof of Theorem 1.1. The basic plan is to directly construct and analyze, for any pair of starting positions $x,y\in\mathrm{SO}(n)$ and large time $n^4\log(n)\ll T\ll n^5$, a coupling of two random variables $X_{T+1}\sim K^{T+1}(x,\cdot)$, $Y_{T+1}\sim K^{T+1}(y,\cdot)$. The goal is for this construction to satisfy

$$\mathbb{P}[X_{T+1}=Y_{T+1}]=1-o(1). \tag{3.1}$$

The most common approach to proving an inequality of the form (3.1) directly is to construct a coupling between two copies $\{X_t\}_{t\in\mathbb{N}}$, $\{Y_t\}_{t\in\mathbb{N}}$ of the Markov chain of interest and force more and more of their coordinates to agree over time. Unfortunately, because the moves of Kac's walks occur on 1-dimensional slices of the N-dimensional curved space $\mathrm{SO}(n)$, it is difficult to write down an explicit coupling that has this property. Instead of doing this, we build a two-step coupling: we begin with a standard coupling, due to Oliveira [45], that forces the two copies of Kac's walk to become close; we then give a very nonexplicit but small perturbation of this coupling, which forces two copies to actually meet.

Below, we give an informal summary of the main ingredients of the proof.

• *Step 1: Construction of perturbed map.* For $x \in \mathrm{SO}(n)$ and $\tilde{X}_{T+1} \sim K^T(x,\cdot)$, we construct a (random) family of small and smooth high-dimensional perturbations $\{f_{\mathcal{A}_x}(\cdot)\}$ of $\tilde{X}_{T+1}$ on SO(n) that preserves the distribution of $\tilde{X}_{T+1}$. More precisely, for any $\delta \in [-\varepsilon,\varepsilon]^N$, $f_{\mathcal{A}_x}(\delta) \in \mathrm{SO}(n)$ will satisfy

$$\|f_{\mathcal{A}_x}(\delta) - \tilde{X}_{T+1}\|_{\mathrm{HS}} \lesssim \varepsilon_n, \qquad \mathcal{L}\big(f_{\mathcal{A}_x}(\delta)\big) = K^T(x,\cdot). \tag{3.2}$$

See Definition 4.1 and Lemma A.1.

• *Step 2: Coupling and overlap.* Fix $x, y \in \mathrm{SO}(n)$ and let $I_x \equiv f_{\mathcal{A}_x}([-\varepsilon_n,\varepsilon_n]^N)$ and $I_y \equiv f_{\mathcal{A}_y}([-\varepsilon_n,\varepsilon_n]^N)$ be the (random) images of $f_{\mathcal{A}_x}, f_{\mathcal{A}_y}$ on SO(n). For $\varepsilon_n \ll 1$, the maps $f_{\mathcal{A}_x}, f_{\mathcal{A}_y}$ are approximately linear. Thus if δ_x, δ_y are sampled from unif($[-\varepsilon_n,\varepsilon_n]^N$), then $f_{\mathcal{A}_x}(\delta_x)$, $f_{\mathcal{A}_y}(\delta_y)$ are approximately uniformly distributed on I_x, I_y. Thus, equality (3.2) and Taylor approximation at $\delta_x, \delta_y = 0$ suggests that

$$\begin{aligned}
&\|K^T(x,\cdot) - K^T(y,\cdot)\|_{\mathrm{TV}} \\
&\quad\lesssim 1 - \frac{|I_x \cap I_y|}{\min(|I_x|,|I_y|)} \\
&\quad\approx 1 - \frac{|\{f_{\mathcal{A}_x}(0)\exp(J_x[-\varepsilon_n,\varepsilon_n]^N)\} \cap \{f_{\mathcal{A}_y}(0)\exp(J_y[-\varepsilon_n,\varepsilon_n]^N)\}|}{(2\varepsilon_n)^N \min(|\mathrm{Det}(J_x)|,|\mathrm{Det}(J_y)|)},
\end{aligned}\tag{3.3}$$

where $J_x = \mathrm{d}f_{\mathcal{A}_x}(0)$, $J_y = \mathrm{d}f_{\mathcal{A}_y}(0)$ are the Jacobians of $f_{\mathcal{A}_x}, f_{\mathcal{A}_y}$ at 0. Thus $\|K^T(x,\cdot) - K^T(y,\cdot)\|_{\mathrm{TV}} \ll 1$ as long as

$$\|J_x - J_y\| + \|f_{\mathcal{A}_x}(0) - f_{\mathcal{A}_y}(0)\|_{\mathrm{HS}} \ll \min\big(\sigma_1(J_x), \sigma_1(J_y)\big), \tag{3.4}$$

where $\sigma_1(J_x), \sigma_1(J_y)$ are the smallest singular values of J_x, J_y; see Theorems 4.3 and 1.

• *Step 3: Bounds on singular values.* Using some ideas from random matrix theory, we show that $\sigma_1(J_x), \sigma_1(J_y) \gtrsim n^{-n^2}$ with high probability; see Theorem 8.10.

• *Step 4: Using the contractive coupling.* The only remaining step is to have a coupling of $(\tilde{X}_{T+1}, \tilde{Y}_{T+1}) \sim (K^T(x,\cdot), K^T(y,\cdot))$ so that

$$\|\tilde{X}_{T+1} - \tilde{Y}_{T+1}\|_{\mathrm{HS}} \approx \|f_{\mathcal{A}_x}(0) - f_{\mathcal{A}_y}(0)\|_{\mathrm{HS}} \ll n^{-n^2} \ll \min\big(\sigma_1(J_x), \sigma_1(J_y)\big).$$

We use the contractive coupling from [45] to achieve the above.

Thus the main idea is to reduce the (highly nonlinear) problem of optimally coupling two copies of Kac's walk to the (linear) problem of analyzing the singular values of a random matrix.

4. Coupling inequality. In this section, we prove a variant of the standard coupling inequality for Markov chains. We also construct some of the functions and measures that will be used to define our coupling.

4.1. *Coupling inequality.* We state and prove a coupling inequality. The following map defines a family of small perturbations of a single chain of Kac's walk.

DEFINITION 4.1 (Perturbed map of Kac's walk). Fix $T \in \mathbb{N}$ and a set of integers $\mathcal{S} = \{s_1 < \cdots < s_N\}$ with $1 \le s_1 < s_N < T$. Fix a sequence $\mathcal{I} = \{i_1, \dots, i_T\}$ with $i_t \in \{1, 2, \dots, N\}$ and a sequence $\theta = \{\theta_1, \dots, \theta_T\}$ with $\theta_t \in [0, 2\pi)$. Fix $x \in \mathrm{SO}(n)$ and let

$$\mathcal{A} = \{x, T, \mathcal{S}, \mathcal{I}, \theta\}. \tag{4.1}$$

Define the maps $e_{\mathcal{A},t} : [0, 2\pi)^N \mapsto [0, 2\pi)$ by

$$e_{\mathcal{A},t}(\delta_1, \dots, \delta_N) = \begin{cases} \theta_t, & t \notin \mathcal{S}, \\ \theta_t + \delta_\ell, & t = s_\ell \in \mathcal{S}. \end{cases}$$

Define the map $f_{\mathcal{A}} : [0, 2\pi)^N \mapsto \mathrm{SO}(n)$ by

$$f_{\mathcal{A}}(\delta_1, \dots, \delta_N) = \mathrm{F}_T\big(x, (i_1, e_{\mathcal{A},1}), \dots, (i_T, e_{\mathcal{A},T})\big)[T+1], \tag{4.2}$$

where F_T is the iterative map defined in (2.3). For an i.i.d. sequence $\delta_1, \dots, \delta_N \sim \mathrm{unif}[-\varepsilon_n, \varepsilon_n]$, define the following probability measure on $\mathrm{SO}(n)$:

$$\mu_{\mathcal{A}} = \mathcal{L}\big(f_{\mathcal{A}}(\delta_1, \dots, \delta_N)\big). \tag{4.3}$$

REMARK 4.2. Since we perform addition modulo 2π, for $\tilde{\theta} \sim \mathrm{unif}((0, 2\pi])$, we have $\tilde{\theta} + \tilde{\delta} \sim \mathrm{unif}((0, 2\pi])$ for any independent real valued random variable $\tilde{\delta}$. This fact is key to our use of the perturbed map in Definition 4.1 in constructing an alternative random representation of Kac's walk (see Lemma A.1).

We use this map to state a coupling inequality. Fix $x, y \in \mathrm{SO}(n)$, $T \in \mathbb{N}$. Define

$$\mathfrak{B}_T = \big\{S \subset \{1, 2, \dots, T\} : |S| = N\big\} \tag{4.4}$$

to be the collection of size-N subsets of $\{1, 2, \dots, T\}$, and let $\mathcal{R} : \{1, 2, \dots, N\}^T \mapsto \mathfrak{B}_T$ be any map. Define the associated collection $\mathcal{G}_{\mathcal{R}}$ of "good" sequences by

$$\mathcal{G}_{\mathcal{R}} = \big\{\{j_t\}_{t=1}^T \in \{1, 2, \dots, N\}^T : \{j_r\}_{r \in \mathcal{R}(\{j_t\}_{t=1}^T)} = \{1, 2, \dots, N\}\big\}.$$

We will discuss our choice of $\mathcal{R}$ in Section 4.2.

For $\mathcal{I} \in \{1, 2, \dots, N\}^T$, we denote by $\nu_{\mathcal{I}}$ any measure on $[0, 2\pi)^T \times [0, 2\pi)^T$ that is a coupling of two copies of $\mathrm{unif}([0, 2\pi)^T)$; denote by $\kappa = \{\nu_{\mathcal{I}}\}_{\mathcal{I} \in \{1,2,\dots,N\}^T}$ this collection of couplings. We will discuss our choice of κ in Section 4.3. Using this notation, we have the following.

THEOREM 4.3 (Coupling inequality). *Let*

$$(\mathcal{I}, \Theta^{(x)}, \Theta^{(y)}) \sim \mathrm{unif}(\{1, 2, \dots, N\}) \otimes \kappa.$$

Let $\mathcal{S} = \mathcal{R}(\mathcal{I})$. *Let* $\mathcal{A}_x = \{x, T, \mathcal{S}, \mathcal{I}, \Theta^{(x)}\}$ *and* $\mathcal{A}_y = \{y, T, \mathcal{S}, \mathcal{I}, \Theta^{(x)}\}$ *be as in equation* (4.1), *and let* $\mu_{\mathcal{A}_x}, \mu_{\mathcal{A}_y}$ *be the associated* (*random*) *measures as given in equation* (4.3). *Then*

$$\|K^T(x,\cdot) - K^T(y,\cdot)\|_{\mathrm{TV}} \le \mathbb{E}\big[\|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} \mathbf{1}_{\mathcal{I}\in\mathcal{G}_{\mathcal{R}}}\big] + \mathbb{P}[\mathcal{I} \notin \mathcal{G}_{\mathcal{R}}], \tag{4.5}$$

where the expectation is taken over the random choice of $(\mathcal{I}, \Theta^{(x)}, \Theta^{(y)})$.

PROOF. The proof is deferred to Appendix A. □

REMARK 4.4. It is not obvious that the expression on the right-hand side of inequality (4.5) exists. In particular, it is not obvious that the integrand $\|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} \mathbf{1}_{\mathcal{I}\in\mathcal{G}_{\mathcal{R}}}$ is measurable. We do not assume that this expression exists; its existence is part of the conclusion of Theorem 4.3.

In order to use Theorem 4.3, we must fix the function $\mathcal{R} : \{1, 2, \dots, N\}^T \mapsto \mathfrak{B}_T$ and the collection of couplings $\kappa = \{\nu_{\mathcal{I}}\}_{\mathcal{I}\in\{1,2,\dots,N\}^T}$. We describe our choices in the next two sections.

4.2. *Choice of* $\mathcal{R}$. In this section, we describe the function $\mathcal{R} : \{1, 2, \dots, N\}^T \mapsto \mathfrak{B}_T$ for choosing the set $\mathcal{S}$.

DEFINITION 4.5 (Choice of marked times). Fix constant $\mathcal{Q} > 0$ and running time $T > N$. Let $\{i_t\}_{t\in\mathbb{N}}$ be a sequence with $i_t \in \{1, 2, \dots, N\}$. We inductively define the sequence $\{s_\ell\}_{\ell=1}^N$ by setting

$$s_1 = \min\{t \ge 1 : i_t = 1\},$$
$$s_{\ell+1} = \min\big\{t \ge s_\ell + \mathcal{Q}n^2 \log(n) : i_t = \ell + 1\big\}.$$

Define

$$\mathcal{R}_{\mathrm{marked}}\big(\{i_t\}_{t=1}^T\big) = \begin{cases} \{s_1, \dots, s_N\}, & \text{if } T \ge s_N + 1, \\ \{1, 2, \dots, N\}, & \text{otherwise.} \end{cases} \tag{4.6}$$

Henceforth, we use the above choice of $\mathcal{R}_{\mathrm{marked}}$, with $\mathcal{Q} \gg 1$ and $T \approx n^4 \log(n)$. Requiring $s_{\ell+1} - s_\ell \gtrsim n^2 \log(n)$ ensures that the chain mixes fairly well between times s_ℓ and $s_{\ell+1}$. This is crucial for our random matrix comparison arguments in Section 8, and in particular the bound (8.18). See Section 10.3 for a discussion of a different construction that we conjecture to give sharper bounds on the mixing time.

4.3. *Choice of* κ. We use a small modification of the contractive coupling described by Oliviera in [45] for the choice of κ. This coupling has the following properties.

LEMMA 4.6. *Fix* $T \in \mathbb{N}$ *and* $x, y \in \mathrm{SO}(n)$. *There exists a collection of measures* $\kappa = \{\nu_{\mathcal{I}}\}_{\mathcal{I} \in \{1,2,\dots,N\}^T}$ *with the following properties. For*

$$(\{i_t\}_{t=1}^T, \{\theta_t^{(x)}\}_{t=1}^T, \{\theta_t^{(y)}\}_{t=1}^T) \sim \mathrm{unif}(\{1, 2, \dots, N\}^T) \otimes \kappa,$$

we have the following:

1. *Marginally, both* $\{(i_t, \theta_t^{(x)})\}_{t=1}^T$ *and* $\{(i_t, \theta_t^{(y)})\}_{t=1}^T$ *are draws from* $\mathrm{unif}(\mathfrak{A}^T)$, *where* $\mathfrak{A}$ *is as in* (2.2).
2. *Fix* $0 < A < \infty$. *For all* $T > n^2(20A \log(n) - \log(\omega_n))$,

$$\mathbb{P}[\|X_{T+1} - Y_{T+1}\|_{\mathrm{HS}} \le n^{-5}\omega_n] \ge 1 - n^{-A} \tag{4.7}$$

holds, where $\{X_t\}_{t=1}^{T+1}, \{Y_t\}_{t=1}^{T+1}$ *are the Markov chains associated to the random mapping representation* (2.2), *initial points* x, y *and update sequences* $\{(i_t, \theta_t^{(x)})\}_{t=1}^T, \{(i_t, \theta_t^{(y)})\}_{t=1}^T$.

PROOF. This lemma essentially follows from the calculations in the proof of Theorem 1 of [45]. We give a summary of the results of [45] and how they imply this lemma in Appendix B. $\square$

5. Technical lemmas.

We give a collection of general estimates that will be used throughout the paper. The proofs are deferred to Appendix C, and can be safely skipped on a first reading of this paper.

5.1. *Matrix estimates.* We use the following result repeatedly.

LEMMA 5.1. *Fix* $k \ge 1$ *and let* $P_1, \dots, P_k$ *and* $Q_1, \dots, Q_k$ *be two sequences of elements of* $M(n)$. *Then*

$$\left\| \prod_{i=1}^k Q_i - \prod_{i=1}^k P_i \right\|_{\mathrm{HS}} \le \sum_{i=1}^k \left\| \prod_{\ell=1}^{i-1} Q_\ell \right\|_{\mathrm{Op}} \|Q_i - P_i\|_{\mathrm{HS}} \left\| \prod_{\ell=i+1}^k P_\ell \right\|_{\mathrm{Op}}. \tag{5.1}$$

We also have the elementary bound.

LEMMA 5.2. *Let* M_1, M_2 *be two* N *by* N *symmetric matrices. For a general matrix* M, *denote by* $\sigma_1(M) \le \sigma_2(M) \le \cdots \le \sigma_N(M)$ *the ordered singular values of* M. *Assume that*

$$\|M_1 - M_2\|_{\mathrm{Op}} \le \delta \sigma_1(M_1) \tag{5.2}$$

for some $0 < \delta < 1$. *Then the determinants of* M_1, M_2 *satisfy*

$$\left|\frac{\det(M_2)}{\det(M_1)} - 1\right| \le N^{\frac{N}{2}}\delta^N.$$

Our next collection of bounds will require further notation. For fixed constant $c > 0$, fixed sequences $\theta_k, \tilde{\theta}k \in [0, 2\pi)$, $\mathrm{R}_k, \tilde{\mathrm{R}}_k \in \mathrm{SO}(n)$ and $a_k \in \mathfrak{so}(n)$, define the two functions $f : [-c, c]^N \mapsto \mathrm{SO}(n)$ and $g : [-c, c]^N \mapsto \mathrm{SO}(n)$:

$$\begin{aligned} f(u_1, u_2, \dots, u_N) &= \prod_{k=1}^{N} \mathrm{R}_k \exp\bigl((\theta_k + u_k)a_k\bigr), \\ g(v_1, v_2, \dots, v_N) &= \prod_{k=1}^{N} \tilde{\mathrm{R}}_k \exp\bigl((\tilde{\theta}k + v_k)a_k\bigr). \end{aligned} \tag{5.3}$$

For $1 \le i + 1 < j \le N$, define

$$M_{i,j} = M_{i,j}(u_1, \dots, u_N) \equiv \prod_{k=i+1}^{j-1} \mathrm{R}_k e^{(\theta_k + u_k)a_k},$$

for $1 \le j + 1 < i \le N$, define

$$M_{i,j} = \prod_{k=j+1}^{i-1} \mathrm{R}_k e^{(\theta_k + u_k)a_k},$$

and for $1 \le i < N$ define $M_{i,i+1} = M_{i+1,i} = \mathrm{Id}$. The derivative map $\mathrm{d}f : T_p[-c, c]^N \mapsto T_{f(p)}\,\mathrm{SO}(n)$ in the direction $h = \sum_k h_k \frac{\partial}{\partial e_k}$ is

$$\mathrm{d}f_u(h) = \sum_{j=1}^{N} \prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k + u_k)a_k}\bigl(\mathrm{Id} + (h_k a_k - \mathrm{Id})1_{k=j}\bigr). \tag{5.4}$$

We note that the functions f, g are essentially of the same form as those given in equation (4.2), with small differences in notation that will make the expressions in this section simpler to work with [e.g., we combine the products of certain sequences of matrices of the form $R(i, \theta)$ into individual elements R_k, and explicitly write the remaining matrices of the form $R(i, \theta)$ in terms of elements of the Lie algebra $\mathfrak{so}(n)$]. We state two lemmas that bound various approximations related to f and g.

LEMMA 5.3 (Closeness of tangent maps). *Let* $f : [-c, c]^N \mapsto \mathrm{SO}(n)$ *be of the form given in equation* (5.3). *Assume that* $c < N^{-3}$ *and* $\max_k |\theta_k - \tilde{\theta}k| \le c^2$, *where* $\theta_k, \tilde{\theta}k \in [0, 2\pi)$ *are as in equation* (5.3). *Then for all* $u, v \in [-c, c]^N$ *and all* $h \in \mathbb{R}^N$ *with* $\|h\| \le 1$,

$$\begin{aligned} \|\mathrm{d}f_u(h) - \mathrm{d}f_v(h)\|_{\mathrm{HS}} &\le 4N^2 c, \\ \|\mathrm{dL}_{f(v)(f(u))^{-1}}\mathrm{d}f_u(h) - \mathrm{d}f_v(h)\|_{\mathrm{HS}} &\le 8N^2 c. \end{aligned} \tag{5.5}$$

LEMMA 5.4 (Approximation by exponential map). *Let f be of the form given in equation* (5.3). *Then for $c < N^{-3}$,*

$$\| f(u) - f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(u)\big)\|_{\mathrm{HS}}^2 \le 8N^2c^2.$$

We need the following expression for the Jacobian of f and g.

LEMMA 5.5 (Jacobian formula). *Let f be the function defined in equation* (5.3). *For $1 \le i < j \le N$, define*

$$D_{i,j} = -\operatorname{Tr}\big[a_i M_{i,j}\mathrm{R}_j a_j \mathrm{R}_j^{-1} M_{i,j}^{-1}\big]; \tag{5.6}$$

for $1 \le j < i \le N$, define

$$D_{i,j} = D_{j,i}.$$

Finally, set $D_{i,i} = 1$ for all $1 \le i \le N$ and let D be the matrix with entries $[D_{i,j}]$. Then for all $h = (h_1, \dots, h_N) \in \mathbb{R}^N \setminus \{0\}$, $h' = (h'_1, \dots, h'_N) \in \mathbb{R}^N \setminus \{0\}$, and all $u \in [-c, c]^N$,

$$\big\langle \mathrm{d}f_u(h), \mathrm{d}f_u(h')\big\rangle_{\mathrm{HS}} = \langle h, h'\rangle + \sum_{i \ne j} h_i h'_j D_{i,j} = h^\dagger D h'. \tag{5.7}$$

We next show that f is generally a diffeomorphism for c sufficiently small.

LEMMA 5.6. *Let f be of the form given in equation* (5.3) *and assume that it satisfies*

$$\inf_{h \ne 0} \frac{2\|\mathrm{d}f_0(h)\|}{\|h\|} \ge \phi_n, \qquad \sup_{h \ne 0} \frac{\|\mathrm{d}f_0(h)\|}{\|h\|} \le N. \tag{5.8}$$

Then there exist constants $0 < C_0 < 1$, $N_0 < \infty$ such that the function f is a diffeomorphism whenever $c < C_0 n^{-6}\phi_n$ for all $n > N_0$ sufficiently large.

5.2. *Probability estimates.* We note that the "marked times" in Definition 4.5 are generally not too large.

LEMMA 5.7. *Fix $k \ge 2$, $\mathcal{Q} \ge 1$ and $T > N\lceil \mathcal{Q} n^2 \log(n)\rceil + kN^2$. Let s_N be defined as in Definition* 4.5. *Then, for all $k \ge 2$,*

$$\mathbb{P}[s_N > T] \le e^{-\frac{N}{4}}. \tag{5.9}$$

6. Coupling argument. We now give a generic bound on the total variation distances appearing in equation (4.5), under the assumption that nothing "too bad" happened. This formalizes the heuristic calculation (3.3).

LEMMA 6.1 (Coupling construction). *Fix $n \in \mathbb{N}$ and a constant $0 < C < \frac{\log(\omega_n)}{\log(\varepsilon_n)} - 6$. Fix $T \in \mathbb{N}$, $\mathcal{Q} > 0$, $x, y \in \mathrm{SO}(n)$, $\mathcal{I} \in \{1, 2, \dots, N\}^T$, $\mathcal{S} \in \mathfrak{B}_T$ and two sets $\Theta^{(x)}, \Theta^{(y)} \in [0, 2\pi)^T$. For $z \in \{x, y\}$, let $\mathcal{A}_z = \mathcal{A}(z, T, \mathcal{S}, \mathcal{I}, \Theta^{(z)})$ be defined as in Definition 4.1. Finally, let $f = f_{\mathcal{A}_x}$ and $g = f_{\mathcal{A}_y}$ be as defined following equation (4.1). Assume that*

$$\|f(0) - g(0)\|_{\mathrm{HS}} \le \omega_n n^{5+C} \tag{6.1}$$

and that for $w \in \{f, g\}$, we have for all $\zeta \in [-c, c]^N$

$$\inf_{h \neq 0} \frac{2\|\mathrm{d}w_\zeta(h)\|}{\|h\|} \ge \phi_n, \qquad \sup_{h \neq 0} \frac{\|\mathrm{d}w_\zeta(h)\|}{\|h\|} \le N. \tag{6.2}$$

For all $n > N_0$ sufficiently large, it is possible to couple two sequences of i.i.d. random variables $U_1, \dots, U_N \sim \mathrm{unif}(-\varepsilon_n, \varepsilon_n)$ and $V_1, \dots, V_N \sim \mathrm{unif}(-\varepsilon_n, \varepsilon_n)$ so that

$$\mathbb{P}[f(U_1, \dots, U_N) \neq g(V_1, \dots, V_N)] = 513 N^2 \phi_n^{-1} \varepsilon_n = o(1).$$

PROOF. By Lemma 5.4,

$$\begin{aligned} &\|f(p) - f(0)\exp(\mathrm{dL}_{f(0)^{-1}} \circ \mathrm{d}f_0(p))\|_{\mathrm{HS}} \le 8N^2\varepsilon_N^2, \\ &\|g(p) - g(0)\exp(\mathrm{dL}_{g(0)^{-1}} \circ \mathrm{d}g_0(p))\|_{\mathrm{HS}} \le 8N^2\varepsilon_N^2 \end{aligned} \tag{6.3}$$

for all $p \in (-\varepsilon_n, \varepsilon_n)^N$. For $0 < r < \infty$, define

$$\mathcal{H}_f(r) = \{f(0)\exp(\mathrm{dL}_{f(0)^{-1}} \circ \mathrm{d}f_0(p)) : p \in [-r, r]^N\},$$
$$\mathcal{H}_g(r) = \{g(0)\exp(\mathrm{dL}_{g(0)^{-1}} \circ \mathrm{d}g_0(p)) : p \in [-r, r]^N\}.$$

We claim the following.

PROPOSITION 6.2. *Fix a satisfying $\frac{1}{2}\varepsilon_n \le a \le 2\varepsilon_n$ and set*

$$\begin{aligned} u_1 &= a - 32N^2\phi_n^{-1}\varepsilon_n^2, \\ u_2 &= a + 32N^2\phi_n^{-1}\varepsilon_n^2. \end{aligned} \tag{6.4}$$

Then for all $n > N_0$ sufficiently large,

$$\begin{aligned} &\mathcal{H}_f(u_1) \subset f([-a, a]^N) \subset \mathcal{H}_f(u_2), \\ &\mathcal{H}_g(u_1) \subset g([-a, a]^N) \subset \mathcal{H}_g(u_2). \end{aligned} \tag{6.5}$$

PROOF. Assume without loss of generality that $f(0) = \mathrm{Id}$. Fix $p \in \partial([-a,a]^N)$. Since the exponential map is surjective (see, e.g., Theorem 6.9.3 of [1]), we can write $f(p) = \exp(h)$ for some $h \in \mathfrak{so}(n)$. By taking a solution h to $f(p) = \exp(h)$ with small norm, we can also assume that $\|h\|_{\mathrm{HS}} \le 2N^2\phi_n^{-1}\varepsilon_n$. Since $\mathrm{d}f_0$ has full rank, we can write $h = \mathrm{d}f_0(q)$ for some q. We calculate

$$\begin{aligned}
\|p-q\|_\infty \le \|p-q\| &\le 2\phi_n^{-1}\|\mathrm{d}f_0(p) - \mathrm{d}f_0(q)\|_{\mathrm{HS}} \\
&= 2\phi_n^{-1}\|h - \mathrm{d}f_0(p)\|_{\mathrm{HS}} \\
&\le 4\phi_n^{-1}\|\exp(h) - \exp(\mathrm{d}f_0(p))\|_{\mathrm{HS}} \qquad (6.6)\\
&= 4\phi_n^{-1}\|f(p) - \exp(\mathrm{d}f_0(p))\|_{\mathrm{HS}} \\
&\le 32N^2\phi_n^{-1}\varepsilon_n^2.
\end{aligned}$$

The first line follows from inequality (6.2). The third line follows for n sufficiently large from a Taylor expansion and the fact that $\varepsilon_n \ll n^{-4}$. The last line follows from inequality (6.3). This proves that if $p \in \partial[-a,a]^N$ and $f(p) = \exp(\mathrm{d}f_0(q))$, then $q \notin [-u_1,u_1]^N$.

Thus we have shown $f(\partial[-a,a]^N) \subset \mathcal{H}_f(u_1)^c$. Since f is a diffeomorphism (by Lemma 5.6) and a map between manifolds of the same dimension N, this implies $f(\partial[-a,a]^N) = \partial f([-a,a]^N) \subset \mathcal{H}_f(u_1)^c$. Using the fact that both f and the exponential map are diffeomorphisms, the condition $\partial f([-a,a]^N) \subset \mathcal{H}_f(u_1)^c$ together with the fact that $f(0) \subset f([-a,a]^N) \cap \mathcal{H}_f(u_1) \neq \varnothing$, implies the containment condition $\mathcal{H}_f(u_1) \subset f([-a,a]^N)$. This is exactly the left-hand side of the first containment condition (6.5).

To prove the right-hand side (6.5), essentially the same calculation shows that for any $p = \exp(\mathrm{d}f_0(q)) \in f([-a,a]^N)$, we have $q \in [-u_2,u_2]^N$. This immediately implies the right-hand side of the first containment condition (6.5).

The proof of the second containment condition (6.5) is identical, so this completes the proof of the proposition. $\square$

Since the exponential map is surjective and $\mathrm{d}f_0$ has full rank, there exists some h so that $g(0) = \mathrm{L}_{f(0)} \exp(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(h))$. By the fact that the exponential map takes geodesic paths in the bi-invariant metric $\mathrm{D_{HS}}$ on $\mathrm{SO}(n)$ (see, e.g., Corollary 1 of [18]) and the assumption that $\varepsilon_n = o(n^{-5})$, followed by inequality (6.1), we have

$$\|h\|_{\mathrm{HS}} \le \mathrm{D_{HS}}(f(0), g(0)) \le \|f(0) - g(0)\|_{\mathrm{HS}} \le \omega_n n^{5+C}. \qquad (6.7)$$

We claim the following.

PROPOSITION 6.3. *Fix a, u_1, and u_2 as in Proposition 6.2, and set*

$$v_1 = u_1 - 64N^2\phi_n^{-1}\varepsilon_n^2,$$

$$v_2 = u_2 + 64N^2\phi_n^{-1}\varepsilon_n^2.$$

Then for all $n > N_0$ sufficiently large,

$$\mathcal{H}_f(v_1) \subset \mathcal{H}_g(u_1) \subset \mathcal{H}_g(u_2) \subset \mathcal{H}_f(v_2). \tag{6.8}$$

PROOF. We prove the left-most containment in (6.8) by contradiction. Assume this containment is not true. Then there exists $\ell \in \mathcal{H}_f(v_1) \setminus \mathcal{H}_g(u_1)$. Write $\ell = f(0)\exp(\mathrm{dL}_{f(0)^{-1}} \circ \mathrm{d}f_0(p))$ for some $p \in [-v_1, v_1]^N$. By the local surjectivity of the exponential map, we can write $\ell = g(0)\exp(\mathrm{dL}_{g(0)^{-1}} \circ \mathrm{d}g_0(q))$ for some q. Since $\ell \notin \mathcal{H}_g(u_1)$, we have $q \notin [-u_1, u_1]^N$. However, by essentially the same calculation as in (6.6) combined with inequality (6.7),

$$\|p - q\|_\infty \le \|p - q\| \le 32N^2\phi_n^{-1}\varepsilon_n^2 + \omega_n n^{5+C} \le 64N^2\phi_n^{-1}\varepsilon_n^2.$$

Thus

$$v_1 \ge \|p\|_\infty \ge \|q\|_\infty - 64N^2\phi_n^{-1}\varepsilon_n^2 > u_1 - 64N^2\phi_n^{-1}\varepsilon_n^2 = v_1.$$

This is a contradiction, and so no such ℓ exists. This completes the proof of the first containment relationship in (6.8). The second containment is trivial, and the third is proved in essentially the same way as the first. This completes the proof of the proposition. □

Combining Propositions 6.2 and 6.3, we have

$$f\big([-\varepsilon_n, \varepsilon_n]^N\big) \supset \mathcal{H}_g(a_1) \supset g\big([-a_2, a_2]^N\big), \tag{6.9}$$

where $a_1 = \varepsilon_n - 256N^2\phi_n^{-1}\varepsilon_n^2$ and $a_2 = a_1 - 256N^2\phi_n^{-1}\varepsilon_n^2$.

Let ρ_f and ρ_g denote the densities of $\mathcal{L}(f(U_1, \dots, U_N))$ and ρ_g of $\mathcal{L}(g(V_1, \dots, V_N))$. By Lemma 5.2,

$$\left|\frac{\rho_g(p)}{\rho_f(q)} - 1\right| \le N^{\frac{N}{2}}\left(\frac{\|\mathrm{d}f_p - \mathrm{d}g_q\|_{\mathrm{Op}}}{\sigma_1(\mathrm{d}f_p)}\right)^N$$

uniformly in p (resp., q) in the range of f (resp., g).

By Lemma 5.3 and Assumption (6.2), $\frac{\|\mathrm{d}f_p - \mathrm{d}g_q\|_{\mathrm{Op}}}{\sigma_1(df_p)} \le 32\phi_n^{-1}N^2\varepsilon_n$, and so

$$\left|\frac{\rho_g(p)}{\rho_f(q)} - 1\right| \le \big(32N^{2.5}\phi_n^{-1}\varepsilon_n\big)^N \ll N^{-N}. \tag{6.10}$$

Combining inequality (6.10) with the containment condition (6.9),

$$\begin{aligned} &\big\|\mathcal{L}\big(f(U_1, \dots, U_N)\big) - \mathcal{L}\big(g(V_1, \dots, V_N)\big)\big\|_{\mathrm{TV}} \\ &\qquad \le 1 - \frac{\mu(f([-\varepsilon_n, \varepsilon_n]^N) \cap g([-\varepsilon_n, \varepsilon_n]^N))}{\mu(g([-\varepsilon_n, \varepsilon_n]^N))}\big(1 + N^{-N}\big) \\ &\qquad \le 1 - \frac{\mu(g([-a_2, a_2]^N))}{\mu(g([-\varepsilon_n, \varepsilon_n]^N))}\big(1 + N^{-N}\big) \qquad (6.11) \end{aligned}$$

$$= 1 - \frac{(2a_2)^N}{(2\varepsilon_N)^N}(1 + N^{-N})$$
$$\le 1 - \frac{(2\varepsilon_n - 1024N^2\phi_n^{-1}\varepsilon_n^2)^N}{(2\varepsilon_n)^N}(1 + N^{-N})$$
$$\le 513N^2\phi_n^{-1}\varepsilon_n = o(1).$$

This completes the proof of the lemma. □

7. Relating mixing times to singular values. In this section, we show that a bound on the smallest singular value of the Jacobian of the map $f_{\mathcal{A}}$ given in Definition 4.1 implies a bound on the mixing time of Kac's walk.

Fix time $T \in \mathbb{N}$ and function $\mathcal{R} : \{1, 2, \ldots, N\}^T \mapsto \mathfrak{B}_T$. Let $\{(i_t, \theta_t)\}_{t=1}^T \sim \text{unif}(\mathfrak{A}^T)$, fix $x \in \text{SO}(n)$ and then set $\mathcal{A} = \{x, T, \mathcal{R}(\{i_t\}_{t=1}^T), \{i_t\}_{t=1}^T, \{\theta_t\}_{t=1}^T\}$. Finally, let $f_{\mathcal{A}}$ be as in equation A.1 and let D be the Jacobian of $f_{\mathcal{A}}$ at $(0, 0, \ldots, 0)$. Define the critical scale

$$\psi_n = \min\left((2n)^{-30}, \sup\left\{r > 0 : \mathbb{P}[\sigma_1(D) \le r] < \frac{1}{\sqrt{n}}\right\}\right), \tag{7.1}$$

where the probability is taken over the random variables $\{(i_t, \theta_t)\}_{t=1}^T$. Note that $\psi_n = \psi_n(x, T, \mathcal{R})$ is a (deterministic) function of our choice of x, T and $\mathcal{R}$. We now have the following bound on the mixing time of Kac's walk.

THEOREM 1 [Intermediate bound on the mixing time of Kac's walk on SO(n)]. *Let $\{T_n\}_{n\in\mathbb{N}}$, $\{\mathcal{R}_n\}_{n\in\mathbb{N}}$ be sequences that satisfy*:

1. *for* $\{i_t\}_{t=1}^{T_n} \sim \text{unif}(\{1, 2, \ldots, N\})$,

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{R}_n(\{i_t\}_{t=1}^{T_n}) \in \mathcal{G}_{\mathcal{R}_n}] = 1, \tag{7.2}$$

and

2. *the bounds on the smallest singular values satisfy*

$$\psi_n \ge \phi_n \tag{7.3}$$

for all $x \in \text{SO}(n)$, *and*

3. *the sequence of times satisfies* $T_n > n^2(40\log(n) - \log(\omega_n))$.

Then

$$\limsup_{n\to\infty} \sup_{X_1 = x \in \text{SO}(n)} \|\mathcal{L}(X_{T_n}) - \mu\|_{\text{TV}} = 0.$$

PROOF. This will be a consequence of Theorem 4.3, Lemma 4.6 and Lemma 6.1. We will need to confirm that the conditions (6.1) and (6.2) of Lemma 6.1 are satisfied with high probability.

Fix $y \in \mathrm{SO}(n)$ and let $(\mathcal{I}, \Theta^{(x)}, \Theta^{(y)}) \sim \mathrm{unif}(\{1, 2, \dots, N\}^T) \otimes \kappa$, where κ is as defined in Lemma 4.6. Let $f_{\mathcal{A}_x}$, $f_{\mathcal{A}_y}$ be the usual associated maps as in equation (4.2). Let $\mathcal{E}_1$ be the event that *either* inequality (6.1) holds with constant $C = 15$, *or* $\mathcal{R}_n(\{i_t\}_{t=1}^{T_n}) \notin \mathcal{G}_{\mathcal{R}_n}$. Applying Lemma 4.6,

$$\mathbb{P}[\mathcal{E}_1] = 1 - o(1) \tag{7.4}$$

as n goes to infinity.

Next, let $\mathcal{E}_2$ be the event that inequality (6.2) holds. The first line of this inequality holds with probability at least $1 - \frac{1}{\sqrt{n}}$ by the definition (7.1) of ϕ_n. We claim that the second part of this inequality holds with probability 1. To see this, note that the entries of the matrix D, as given by equation (5.6), satisfy

$$|D_{i,j}| \le 1.$$

This implies $\|D\|_{\mathrm{Op}} \le \max_i \sum_j |D_{i,j}| \le N$. Therefore,

$$\mathbb{P}[\mathcal{E}_2] = 1 - o(1). \tag{7.5}$$

Next, let $\mathcal{E}_3 = \{\mathcal{I} \notin \mathcal{G}_{\mathcal{R}_n}\}$. By assumption,

$$\mathbb{P}[\mathcal{E}_3] = o(1). \tag{7.6}$$

By Theorem 4.3 and Lemma 6.1,

$$\begin{aligned}
\sup_{X_1 = x \in \mathrm{SO}(n)} \|\mathcal{L}(X_{T_n}) - \mu\|_{\mathrm{TV}} &\le \sup_{x, y \in \mathrm{SO}(n)} \|K^{T_n}(x, \cdot) - K^{T_n}(y, \cdot)\|_{\mathrm{TV}} \\
&\le \mathbb{E}\big[\|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} \mathbf{1}_{\mathcal{I} \in \mathcal{G}_{\mathcal{R}_n}}\big] + \mathbb{P}[\mathcal{I} \notin \mathcal{G}_{\mathcal{R}_n}] \\
&\le \big(1 - \mathbb{P}[\mathcal{E}_1]\big) + \big(1 - \mathbb{P}[\mathcal{E}_2]\big) + \mathbb{P}[\mathcal{E}_3] + o(1).
\end{aligned}$$

Applying inequalities (7.4), (7.5) and (7.6) to this bound completes the proof. $\square$

REMARK 7.1. This result is correct, with proof as stated, for any pair of sequences $\{\phi_n, \omega_n\}_{n \in \mathbb{N}}$ with $0 < \omega_n < \phi_n^{30} < n^{-900}$. This is useful for the simulation discussion in Section 10.3.

8. Singular values of random matrices. The last ingredient in the proof of our upper bound on the mixing time of Kac's walk is a *lower* bound on the smallest singular value $\sigma_1(D)$ of the matrix D defined in equation (8.15). In this section, we use some ideas from random matrix theory to obtain the required bound.

The notation used in this section is essentially independent of the notation of the remainder of the paper, except where explicitly noted. This section begins by giving a generic bound on the smallest singular value of a random matrix whose entries have continuous but strongly dependant entries (see Section 8.1), then applies this bound to a simple random matrix D_∞ for which some exact calculations are possible (see Section 8.2), and finally compares the smallest singular value of

D_∞ to that of the matrix D defined in equation (8.15) (see Section 8.3). This argument gives us a lower bound on the constant ψ_n that is defined in equation (7.1), and which has a critical role in the conclusion of Theorem 1.

We believe that the bounds in this section may be of independent interest, and so we give a brief overview of the results and their relationship to the existing literature. Our main abstract result, given in Lemma 8.3, is qualitatively similar to the main bound, Theorem 1.1, of [22]. In [22], the authors show that, if the diagonal entries $\{M[i,i]\}$ of a random matrix M are independent and have distributions with densities uniformly bounded by a constant $C < \infty$, then M cannot have a "very small" determinant. Their proof is essentially an inductive argument on the k by k upper submatrices $M_k \equiv \{M[i,j]\}_{1\le i,j\le k}$: an analysis of the polynomial $\det(M_{k+1})$ shows that it cannot concentrate on any interval much smaller than $C^{-1}\det(M_k)$, which implies $\det(M_{k+1}) \gtrsim (Cn)^{-1}\det(M_k)$ for all $1 \le k \le n$ with moderately high probability.

Unfortunately, Theorem 1.1 of [22] does not apply in our situation for several reasons. Trivially, the matrices we are interested in are symmetric and have deterministic diagonals, and their off-diagonal entries are not independent. More importantly, the densities of the conditional distributions of the near-diagonal entries $M[k,k+1]$ given the "previous" entries $\{M[i,j]\}_{i<k}$ *are* sometimes concentrated on a very small interval. Our Lemma 8.3 relaxes the conditions of [22], allowing for symmetric matrices with dependant entries. Most importantly, it allows for the conditional distribution of the near-diagonal entries $M[k,k+1]$ to be concentrated on a small interval for *some* values of the previous entries, as long as it is not concentrated for *most* values of the previous entries. Like [22], our proof is a simple inductive argument. Our Lemma 8.4 gives a simple restatement of the main criterion of Lemma 8.3 when the conditional distributions of $M[k,k+1]$ have densities.

While Lemma 8.3 does give useful bounds for the random matrix D of interest, it is difficult to obtain them directly. The main problem is that the densities of the conditional distributions of the near-diagonal entries $D[k,k+1]$ are extremely complicated and sometimes very large, and we see no plausible way to obtain useful bounds on them. To avoid this calculation, we define a closely related matrix D_∞ for which we can find explicit and simple formulas for the required conditional densities (see Lemma 8.6). These calculations immediately imply that D_∞ satisfies the conditions of Lemma 8.3, and it is straightforward to check that the nearby matrix D must also satisfy the conditions of Lemma 8.3 with similar constants (see the proof of Lemma 8.10).

8.1. *Bounds on determinant of random matrices*. Let M be an $n \times n$ symmetric random matrix with associated measurable space (Ω, Σ). Let $\mathcal{F}_0$ denote the σ-algebra generated by the entries of M. For $1 \le i \le n-1$, let $\mathcal{F}_i$ be a σ-algebra under which $M[k,\ell]$ is $\mathcal{F}_i$-measurable for all (k,ℓ) satisfying either $i \le k$ and $i+2 \le \ell$, or $k = \ell \in \{i, i+1\}$. Finally, let $\zeta_i \sim \mathcal{L}(M[i,i+1]|\mathcal{F}_i)$. We make the following assumptions for the matrix M.

ASSUMPTION 8.1. 1. The random variable ζ_i satisfies the anti-concentration bound

$$\mathbb{P}\Big[\sup_{w\in\mathbb{R}}\sup_{\beta\in\mathbb{R},(\alpha,\varepsilon)\in\mathcal{R}(C)}\mathbb{P}\big[|\alpha(\zeta_i)^2+\beta\zeta_i-w|<\varepsilon|\mathcal{F}_i\big]<\frac{4C\sqrt{\varepsilon}}{\sqrt{\alpha}}+n^{-2}\Big] > 1-n^{-2} \tag{8.1}$$

for some fixed $1\le C<\infty$, where

$$\mathcal{R}(C)=\{(\alpha,\varepsilon);\alpha>0,\varepsilon\ge(4C^2n^4)^{-1}\alpha\}.$$

2. We have

$$\begin{aligned}\mathbb{P}\big[|M[n,n]|<(4Cn)^{-4}\big]&\le\frac{1}{n^2}, \qquad n \text{ odd}.\\ \mathbb{P}\big[|M[n-1,n-1]M[n,n]-M[n-1,n]^2|<(4Cn)^{-4}\big]&\le\frac{1}{n^2}, \qquad n \text{ even}.\end{aligned} \tag{8.2}$$

REMARKS 8.2. The assumption given by inequality (8.1) is often easy to verify in practice. For example, it holds if the conditional density ρ_i of ζ_i is bounded by the constant $1\le C<\infty$ with high probability (see Lemma 8.4). The second part of Assumption 8.1 is often straightforward to check by hand.

Let $|M|$ denote the determinant of the matrix M.

LEMMA 8.3. *For a matrix M satisfying the hypotheses of Assumption 8.1,*

$$\mathbb{P}\big[|M|<(4Cn)^{-4(n+1)}\big]\le\frac{3}{n}$$

for all $n\in\mathbb{N}$.

PROOF. For an $n\times n$ matrix A and indices $1\le i,j\le n$, denote by $A_{(i,j)}$ the matrix obtained by removing the ith row and jth column from A, and $A_{(i,j),(k,\ell)}$ the matrix obtained by removing the ith and kth rows and the jth and ℓth columns.

Let $m=\frac{n-1}{2}$ when n is odd, and let $m=\frac{n}{2}-1$ when n is even. Let $M^{(1)}=M$. We inductively define $\{M^{(k)}\}_{k=1}^m$ by setting

$$M^{(k+1)}=M^{(k)}_{(1,2),(2,1)}.$$

Observe that $M^{(k)}$ is just the $(n-2k+2)$ by $(n-2k+2)$ lower right-hand submatrix of $M=M^{(1)}$. We define the events

$$U_k=\{|M^{(k)}|^2>(4Cn)^{-4(m-k+2)}\},$$

$$V_k=\Big\{\sup_{w\in\mathbb{R}}\sup_{\beta\in\mathbb{R},(\alpha,\varepsilon)\in\mathcal{R}(C)}\mathbb{P}\big[|\alpha(\zeta_{2k-1})^2+\beta\zeta_{2k-1}-w|<\varepsilon|\mathcal{F}_{2k-1}\big] < \frac{4C\sqrt{\varepsilon}}{\sqrt{\alpha}}+n^{-2}\Big\}.$$

By definition, $U_k \in \mathcal{F}_{2k-2}$ and $V_k \in \mathcal{F}_{2k-1}$. We now expand the determinant of $M^{(k)}$:

$$
\begin{aligned}
|M^{(k)}| &= -M^{(k)}[1,2]|M^{(k)}_{(1,2)}| + C_1 \\
&= -M^{(k)}[1,2]M^{(k)}[2,1]|M^{(k)}_{(1,2),(2,1)}| + M^{(k)}[1,2]C_2 + C_1 \\
&= -M^{(k)}[1,2]^2|M^{(k+1)}| + M^{(k)}[1,2]C_2 + C_1,
\end{aligned} \tag{8.3}
$$

where

$$
\begin{aligned}
C_1 &= \sum_{1\le j\le n, j\ne 2} (-1)^{j+1} M^{(k)}[1,j]|M^{(k)}_{(1,j)}|, \\
C_2 &= \sum_{3\le j\le n} (-1)^{j} M^{(k)}[2,j]|M^{(k)}_{(1,2),(2,j)}|.
\end{aligned}
$$

By assumption, $C_1, C_2 \in \mathcal{F}_{2k-1}$. Thus from equation (8.3) and choosing $\varepsilon^2 = \frac{1}{16}C^{-4}n^{-8}|M^{(k+1)}|^2$ and $\alpha = |M^{(k+1)}|$ in equation (8.1), we obtain

$$
\mathbb{P}\left(|M^{(k)}|^2 < \frac{1}{16}C^{-4}n^{-8}|M^{(k+1)}|^2 \Big| V_k\right) > 1 - \frac{2}{n^2}. \tag{8.4}
$$

Using equation (8.4), we deduce that

$$
\begin{aligned}
\mathbb{P}[U_k^c \cap U_{k+1} \cap V_k] &= \mathbb{E}[1_{U_k^c} 1_{U_{k+1}} 1_{V_k}] \\
&\le \mathbb{E}[1_{U_k^c} | V_k, U_{k+1}] \\
&\le \frac{4C(4Cn)^{-2(m-k+2)}}{(4Cn)^{-2(m-k+1)}} + n^{-2} \le 2n^{-2}.
\end{aligned} \tag{8.5}
$$

Using this inequality repeatedly, and defining $V = \bigcup_i V_i$, we have

$$
\begin{aligned}
\mathbb{P}[U_1^c] &\le \mathbb{P}[U_1^c \cap V] + \mathbb{P}[V^c] \\
&\le \mathbb{P}[U_1^c \cap V] + \frac{m}{n^2} \\
&= \mathbb{P}[U_1^c \cap U_2 \cap V] + \mathbb{P}[U_1^c \cap U_2^c \cap V] + \mathbb{P}[V^c] \\
&= \cdots \\
&= \sum_{j=2}^{m} \mathbb{P}[U_1^c \cap \cdots \cap U_{j-1}^c \cap U_j \cap V] + \mathbb{P}[U_m^c] + \mathbb{P}[V^c] \\
&\le 2\frac{m-1}{n^2} + \frac{1}{n^2} + \frac{m}{n^2} \le \frac{3}{n},
\end{aligned}
$$

where the first few lines are repeated use of exact equalities and union bounds, and the three terms in the last line use, respectively, inequality (8.5), inequality (8.2) and the assumption in inequality (8.1) that $\mathbb{P}[V^c] \le n^{-2}$. This completes the proof. $\square$

We give a simple sufficient condition for the assumption given by inequality (8.1) to hold.

LEMMA 8.4 (Nonconcentration). *Define* $f : \mathbb{R} \mapsto \mathbb{R}$ *by*

$$f(w) = \alpha w^2 + \beta w + \gamma$$

and let W *be a random variable with density* ρ *satisfying* $\sup_w \rho(w) < C < \infty$. *Then for any* $w \in \mathbb{R}$ *and* $\varepsilon > 0$,

$$\mathbb{P}[|f(W) - w| < \varepsilon] \le \frac{4C\sqrt{\varepsilon}}{\sqrt{|\alpha|}}.$$

PROOF. Assume without loss of generality that $\alpha > 0$, and let η be the density of $f(W)$. Fix $r \in \mathbb{R}$ and define the quantity

$$\Gamma = \Gamma(r) \equiv \frac{r}{\alpha} + \frac{\beta^2}{4\alpha^2} - \frac{\gamma}{\alpha}.$$

We have

$$\begin{aligned}
\mathbb{P}[f(W) \le r] &= \mathbb{P}\left[\left(W - \frac{\beta}{2\alpha}\right)^2 \le \Gamma\right] \\
&= \mathbb{P}\left[W \le \frac{\beta}{2\alpha} + \sqrt{\Gamma}\right] - \mathbb{P}\left[W \le \frac{\beta}{2\alpha} - \sqrt{\Gamma}\right].
\end{aligned}$$

Thus

$$\begin{aligned}
\eta(r) &= \frac{d}{dr}\mathbb{P}[f(W) \le r] \\
&= \rho\left(\frac{\beta}{2\alpha} + \sqrt{\Gamma}\right)\frac{1}{2\alpha\sqrt{\Gamma}} + \rho\left(\frac{\beta}{2\alpha} - \sqrt{\Gamma}\right)\frac{1}{2\alpha\sqrt{\Gamma}} \\
&\le \frac{C}{\alpha}\frac{1}{\sqrt{\Gamma}}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
\mathbb{P}[|f(W) - w| \le \varepsilon] &= \int_{w-\varepsilon}^{w+\varepsilon} \eta(r)\,dr \\
&\le \frac{C}{\alpha}\int_{w-\varepsilon}^{w+\varepsilon} \frac{1}{\sqrt{\Gamma(r)}}\,dr \\
&\le \frac{C}{\alpha}\int_{-\varepsilon}^{\varepsilon} \frac{\sqrt{\alpha}}{\sqrt{|r|}}\,dr \\
&= \frac{4C\sqrt{\varepsilon}}{\sqrt{\alpha}}.
\end{aligned}$$

This completes the proof of the bound. $\square$

We next show that the assumption given by inequality (8.1) remains true under small perturbations. For two probability measures ν_1, ν_2 on $\mathbb{R}^d$, define the *Wasserstein distance*

$$W_2(\nu_1, \nu_2)^2 = \inf_{(U,V)\in\mathcal{C}, U\sim\nu_1, V\sim\nu_2} \mathbb{E}\left[\|U - V\|^2\right],$$

where $\mathcal{C}$ is the set of all couplings on $\mathbb{R}^d \times \mathbb{R}^d$ with marginal distributions ν_1 and ν_2.

Let M' be another $n \times n$ symmetric random matrix, and let $\mathcal{F}_i'$, ζ_i', etc. be defined analogously to $\mathcal{F}_i$, ζ_i. For fixed $1 \le i \le n$ and sequence $\{m_{k,\ell}\}_{k,\ell\ge i+2}$, let $\mathcal{G}_i = \mathcal{G}_i(\{m_{k,\ell}\})$ be an event in the σ-algebra generated by $\{M[k,\ell]\}_{k,\ell\ge i+2}$ on which

$$M[k,\ell] = m_{k,\ell}, \qquad k,\ell \ge i+2.$$

Let $\mathcal{G}_i' = \mathcal{G}_i'(\{m_{k,\ell}\})$ be defined analogously. We then have the following.

LEMMA 8.5. *Let the density ρ_i of ζ_i satisfies*

$$\mathbb{P}\Big[\sup_w \rho_i(w) > C\Big] \le \frac{1}{2n^5} \tag{8.6}$$

for some fixed $1 \le C < \infty$. Let M' satisfy (8.2) of Assumption 8.1 for this choice of C. Assume that

$$\begin{aligned} &W_2\big(\mathcal{L}\big(M[i,(i+1):n]|\mathcal{G}_i\big), \mathcal{L}\big(M'[i,(i+1):n]|\mathcal{G}_i'\big)\big)^2 \\ &\quad \le \delta < \frac{1}{8n^5}(4Cn)^{-4}. \end{aligned} \tag{8.7}$$

Then there exists a universal constant N_0 so that the determinant $|M'|$ of M' satisfies

$$\mathbb{P}\big[|M'| < \big(4Cn^3\big)^{-4(n+1)}\big] \le \frac{3}{n}$$

for all $n > N_0$.

PROOF. We begin with a generic bound. Let $U, V \in \mathbb{R}$ be two random variables, and define $f : \mathbb{R} \mapsto \mathbb{R}$ by $f(w) = \alpha w^2 + \beta w$ for some $\alpha > 0$, $\beta \in \mathbb{R}$. Fix $\varepsilon > 0$. Then

$$\begin{aligned} \sup_{w\in\mathbb{R}} \mathbb{P}\big[\big|f(U) - w\big| < \varepsilon\big] &= \sup_{w\in\mathbb{R}} \mathbb{P}\big[\big|\alpha U^2 + \beta U - w\big| \le \varepsilon\big] \\ &\le \sup_{w\in\mathbb{R}} \mathbb{P}\Big[U \in \Big(w - \sqrt{\frac{\varepsilon}{\alpha}}, w + \sqrt{\frac{\varepsilon}{\alpha}}\Big)\Big] \\ &\le \sup_{w\in\mathbb{R}} \mathbb{P}\Big[V \in \Big[w - \sqrt{\frac{\varepsilon}{4\alpha}}, w + \sqrt{\frac{\varepsilon}{4\alpha}}\Big]\Big] \qquad (8.8) \end{aligned}$$

$$+\mathbb{P}\left[|U-V|>\sqrt{\frac{\varepsilon}{4\alpha}}\right]$$
$$\le \sup_{w\in\mathbb{R}} \mathbb{P}\left[V\in\left(w-\sqrt{\frac{\varepsilon}{4\alpha}}, w+\sqrt{\frac{\varepsilon}{4\alpha}}\right)\right]$$
$$+\frac{4\alpha}{\varepsilon}W_2\big(\mathcal{L}(U),\mathcal{L}(V)\big)^2.$$

By equation (8.6) and Lemma 8.4, Assumption 8.1 is satisfied the by the matrix M. For ε, f as above, equation (8.8) then implies

$$\begin{aligned}
&\sup_{w\in\mathbb{R}} \mathbb{P}\big[f\big(M'[i,i+1]\big)\in[w-\varepsilon,w+\varepsilon]|\mathcal{G}_i'\big]\\
&\quad\le \sup_{w\in\mathbb{R}} \mathbb{P}\big[f\big(M[i,i+1]\big)\in[w-\varepsilon,w+\varepsilon]|\mathcal{G}_i\big]\\
&\quad\quad+\frac{4\alpha}{\varepsilon}W_2\big(\mathcal{L}\big(M\big[i,(i+1):n\big]|\mathcal{G}_i\big),\mathcal{L}\big(M'\big[i,(i+1):n\big]|\mathcal{G}_i'\big)\big)^2\\
&\quad\le \frac{1}{2n^5}+\frac{4C\sqrt{\varepsilon}}{\alpha}+\frac{4\alpha}{\varepsilon}W_2\big(\mathcal{L}\big(M\big[i,(i+1):n\big]|\mathcal{G}_i\big),\mathcal{L}\big(M'\big[i,(i+1):n\big]|\mathcal{G}_i'\big)\big)^2\\
&\quad\le \frac{1}{2n^5}+\frac{4C\sqrt{\varepsilon}}{\alpha}+\frac{4\alpha\delta}{\varepsilon},
\end{aligned}$$

where the first term in the second-last line comes from inequality (8.6) and the other terms come from an application of Lemma 8.4, and where the last line comes from inequality (8.7). This bound implies

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{w\in\mathbb{R}} \mathbb{P}\big[f\big(M'[i,i+1]\big)\in[w-\varepsilon,w+\varepsilon]|\mathcal{F}_i'\big]|\mathcal{G}_i'\Big]\\
&\quad\le \frac{1}{2n^5}+\frac{4C\sqrt{\varepsilon}}{\alpha}+\frac{4\alpha\delta}{\varepsilon}\le\frac{1}{n^5}+\frac{4C\sqrt{\varepsilon}}{\alpha}.
\end{aligned}$$

By Markov's inequality, this implies

$$\mathbb{P}\Big[\sup_{w\in\mathbb{R}} \mathbb{P}\big[f\big(M'[i,i+1]\big)\in[w-\varepsilon,w+\varepsilon]|\mathcal{F}_i'\big]>n^2\Big(n^{-5}+\frac{4C\sqrt{\varepsilon}}{\alpha}\Big)\Big]\le n^{-2}.$$

Thus M' satisfies Assumption 8.1 with constant $\tilde{C}=n^2C$. Applying Lemma 8.3 now completes the proof. □

8.2. *Bounding the smallest singular value of D_∞.* In this section, we define a specific random matrix D_∞, and show that it satisfies the requirements of Lemma 8.3. Define the n-sphere

$$\mathrm{S}^{(n-1)}=\left\{w\in\mathbb{R}^n:\sum_{i=1}^n w[i]^2=1\right\}.$$

Denote the Euclidean inner product by $\langle\cdot,\cdot\rangle$. We begin with the technical lemma.

LEMMA 8.6 (Conditional densities on spheres). *Fix $0 \le k \le n-2$. Let $W, v_1, \dots, v_{n-1} \sim \mathrm{Unif}(\mathrm{S}^{(n-1)})$ be i.i.d. For $0 \le k \le n-2$, let $\mathcal{G}_k$ be the σ-algebra generated by $\langle W, v_1\rangle, \dots, \langle W, v_k\rangle$ and $v_1, \dots, v_{n-1}$. Let*

$$Z \sim \mathcal{L}\big(\langle W, v_{n-1}\rangle | \mathcal{G}_k\big),$$

and let ρ be the density of Z. We have

$$\mathbb{P}\Big[\sup_z \rho(z) > n^{20}\Big] \le n^{-2}$$

for all $n > N_0$ sufficiently large, uniformly in $0 \le k \le n-2$.

PROOF. We first prove the claim for $k = n-2$, the maximum value. Let $\mathcal{F} = \mathcal{G}_{n-2}$. We can write down the dependence of $\sup_z \rho(z)$ on $\mathcal{F}$ explicitly. Let $H = \mathrm{span}(v_1, \dots, v_{n-2})$ be the hyperplane spanned by $v_1, \dots, v_{n-2}$. Let $\mathcal{P}_H : \mathbb{R}^n \mapsto H$ be the operator associated with orthogonal projection onto H, and define

$$\begin{aligned} v^0 &= \mathcal{P}_H(v_{n-1}), \\ v^+ &= v_{n-1} - \mathcal{P}_H(v_{n-1}). \end{aligned}$$

We note that $\langle W, v^0\rangle$ and v^+ are both $\mathcal{F}$-measurable. Let $W_H = \|\mathcal{P}_H W\|^2$. The random variable W_H is also $\mathcal{F}$-measurable.

Let $Z^+ = \langle W, v^+\rangle$, so that

$$\mathcal{L}(Z) = \mathcal{L}\big(Z^+ + \langle W, v^0\rangle | \mathcal{G}_k\big). \tag{8.9}$$

Let $W' = \|v^+\| \sqrt{1 - W_H} \|S\|$, where $S \sim \mathrm{Unif}(\mathrm{S}^{(1)})$. Then $Z^+ \stackrel{D}{=} W'[1]$. Thus the density ρ^+ of Z^+ satisfies

$$\begin{aligned} \rho^+(z) &= \frac{2}{\pi} \frac{\sqrt{\|v^+\|^2(1 - W_H) - z^2}}{\|v^+\|^2(1 - W_H)} \\ &\le \frac{2}{\pi} \frac{1}{\sqrt{\|v^+\|^2(1 - W_H)}}. \end{aligned}$$

By equation (8.9) and the fact that $\langle W, v^0\rangle \in \mathcal{F}$, it follows that

$$\sup_z \rho(z) = \sup_z \rho^+(z) \le \frac{2}{\pi} \frac{1}{\sqrt{\|v^+\|^2(1 - W_H)}}. \tag{8.10}$$

Thus, to complete the proof, it suffices to bound $\frac{1}{\sqrt{\|v^+\|^2(1-W_H)}}$ with high probability.

First, we bound $1 - W_H$. Since $W_H = \|\mathcal{P}_H\|^2$ where $W \sim \mathrm{Unif}(\mathrm{S}^{(n-1)})$ and H is an independently and randomly chosen hyperplane of dimension $n-2$, we can assume without loss of generality that $H = \{w \in \mathbb{R}^n : w[n-1] = w[n] = 0\}$. Thus

$$\begin{aligned}
\mathbb{P}[1 - W_H \le n^{-20}] &= \mathbb{P}[1 - W[1]^2 - \cdots - W[n-2]^2 \le n^{-20}] \\
&= \mathbb{P}[W[n-1]^2 + W[n]^2 \le n^{-20}] \\
&= O(n^{-15}).
\end{aligned} \tag{8.11}$$

By exactly the same reasoning, we deduce that

$$\begin{aligned}
\mathbb{P}[\|v^+\|^2 \le n^{-5}] &= \mathbb{P}[W[n]^2 \le n^{-20}] \\
&= O(n^{-15}).
\end{aligned} \tag{8.12}$$

Combining inequalities (8.11) and (8.12) with inequality (8.10), we conclude that

$$\mathbb{P}\Big[\sup_z \rho(z) > n^{-20}\Big] = O(n^{-15}).$$

This completes the proof of the lemma. □

Let $P_1, \dots, P_N$ be i.i.d. draws from the Haar measure on SO(N). Define a symmetric $N \times N$ matrix D_∞ by

$$\begin{aligned}
&D_\infty[i,i] = 1, \\
&D_\infty[i,j] = -\operatorname{Tr}\left[a_i \left(\prod_{\ell=i+1}^{j} P_\ell\right) a_j \left(\prod_{\ell=i+1}^{j} P_\ell\right)^{-1}\right], \qquad i < j, \\
&D_\infty[i,j] = D_\infty[j,i], \qquad i > j.
\end{aligned} \tag{8.13}$$

For $i < j$, let $P_{i,j} = \prod_{\ell=i+1}^{j} P_\ell$. We note that, for $1 \le i < j \le n$,

$$D_\infty[i,j] = -\operatorname{Tr}[a_i P_{i,j} a_j (P_{i,j})^{-1}].$$

REMARK 8.7. The matrix D_∞ has two useful properties. First, the appearance of the Haar measure in the definition of D_∞ means that it is easier to make exact calculations involving the entries of D_∞ than those of D. Second, D_∞ is "close" to D, and so bounds on D_∞ can easily be transferred to bounds on D. More precisely, Theorem 1 of [45] implies that the entries of D converge to those of D_∞ as $\mathcal{Q}$ goes to infinity. The $\Theta(n^2 \log(n))$ scaling of the lower bound $s_{i+1} - s_i \ge \mathcal{Q} n^2 \log(n)$ in Definition 4.5 was chosen so that we could guarantee that D and D_∞ must be close in distribution.

For $1 \le i < N$, we define $\mathcal{F}_i$ to be the σ-algebra generated by the matrices $\{P_j\}$ and the inner products $\{-\operatorname{Tr}[a_i P_{i,j} a_j (P_{i,j})^{-1}]\}$ for $j > i + 1$.

LEMMA 8.8. *The matrix D_∞ and σ-algebras $\{\mathcal{F}_i\}_{1\le i<n}$ given above satisfy Assumption 8.1 with*

$$C = N^{20}$$

for all $N > N_0$ sufficiently large. Thus for all $N > N_0$ sufficiently large,

$$\mathbb{P}\bigl[|D_\infty| < (4N^{21})^{-4(N+1)}\bigr] \le \frac{3}{N}.$$

PROOF. We make some initial observations. Let

$$\mathfrak{S}_n = \{a \in \mathfrak{so}(n) : \|a\|_{\mathrm{HS}} = 1\},$$

and define the bijection $\mathcal{M} : \mathfrak{S}_n \mapsto \mathrm{S}^{(N-1)}$ by

$$\mathcal{M}(a) = \bigl(\langle a, a_1\rangle_{\mathrm{HS}}, \ldots, \langle a, a_N\rangle_{\mathrm{HS}}\bigr).$$

We note that, if $P \sim \mu$ [the Haar measure on $\mathrm{SO}(n)$] and $a \in \mathfrak{S}_n$, then

$$\mathcal{M}(PaP^{-1}) \sim \mathrm{unif}\bigl(\mathrm{S}^{(N-1)}\bigr). \tag{8.14}$$

We first prove inequality (8.2). When N is odd, this is trivial. When N is even, we let $W \sim \mathrm{unif}(\mathrm{S}^{(n-1)})$. By equation (8.14), we have that $D_\infty[N-1, N] \stackrel{D}{=} W[1]$, so

$$\begin{aligned}
&\mathbb{P}\bigl[\bigl|D_\infty[N,N]D_\infty[N-1,N-1] - D_\infty[N-1,N]^2\bigr| < (4N^6)^{-4}\bigr]\\
&\quad= \mathbb{P}\bigl[\bigl|1 - W[1]^2\bigr| < (4N^6)^{-4}\bigr]\\
&\quad= O(N^{-3}).
\end{aligned}$$

We next prove inequality (8.1). Observe that:

1. P_{i+1} is independent of $\{P_j\}_{j>i+1}$.
2. By equality (8.14), the vectors $\{\mathcal{M}((P_{i+2}\cdots P_j)a_j(P_{i+2}\cdots P_j)^{-1})\}_{j>i+1}$ are i.i.d. choices from the sphere $S^{(N-1)}$, and the entries $\{D_\infty[i,j]\}_{j>i+1}$ are inner products of these vectors with $\mathcal{M}(a_i)$:

$$\begin{aligned}
D_\infty[i, j+1] &= -\mathrm{Tr}\Biggl[a_i\Biggl(\prod_{\ell=i+1}^{j} P_\ell\Biggr)a_j\Biggl(\prod_{\ell=i+1}^{j} P_\ell\Biggr)^{-1}\Biggr]\\
&= -\bigl\langle\mathcal{M}(a_i), \mathcal{M}\bigl((P_{i+2}\cdots P_j)a_j(P_{i+2}\cdots P_j)^{-1}\bigr)^{\dagger}\bigr\rangle.
\end{aligned}$$

3. Combining the previous two observations, the distribution of

$$Z \sim \mathcal{L}\bigl(-\bigl\langle\mathcal{M}(a_i), \mathcal{M}\bigl(P_{i+1}a_{i+1}P_{i+1}^{-1}\bigr)\bigr\rangle|\mathcal{F}_i\bigr)$$

satisfies the hypotheses of Lemma 8.6.

Thus, by Lemmas 8.4 and Lemma 8.6, the matrix D_∞ satisfies inequality (8.1) of Assumption 8.1 with constant $C = N^{20}$. The conclusion follows immediately from Lemma 8.3. This completes the proof. □

We apply our results to obtain a bound on the smallest singular value of D_∞.

LEMMA 8.9 (Smallest singular values of D_∞). *Let D_∞ be the matrix defined as in* (8.13), *and let $\sigma_1(D_\infty) \le \sigma_2(D_\infty) \le \cdots \le \sigma_N(D_\infty)$ be its singular values. Then*

$$\mathbb{P}\big[\sigma_1(D_\infty) \le N^{-N}\big(4N^{21}\big)^{-4(N+1)}\big] = o(1).$$

PROOF. Using the trivial bound $\sigma_i(D_\infty) \le N \max_{k,\ell} |D_\infty[k,\ell]| \le N$, we have

$$\begin{aligned}\sigma_1(D_\infty) &= |D_\infty| \prod_{i=2}^{n} \sigma_i(D_\infty)^{-1} \\ &\ge |D_\infty| N^{-N}.\end{aligned}$$

Thus, for any $0 < r < \infty$.

$$\mathbb{P}\big[\sigma_1(D_\infty) \le r\big] \le \mathbb{P}\big[|D_\infty| \le r N^N\big].$$

Choosing $r = N^{-N}(4N^{21})^{-4(N+1)}$ and applying Lemmas 8.3 and 8.8, we have

$$\begin{aligned}\mathbb{P}\big[\sigma_1(D_\infty) \le N^{-N}\big(4N^{21}\big)^{-4(N+1)}\big] &\le \mathbb{P}\big[|D_\infty| \le N^{-N}\big(4N^{21}\big)^{-4(N+1)} N^N\big] \\ &= \mathbb{P}\big[|D_\infty| \le \big(4N^{21}\big)^{-4(N+1)}\big] \\ &\le \frac{3}{N}.\end{aligned}$$

This completes the proof. □

8.3. *Application to Kac's walk.* We show that the sequence $\{\psi_n^{-1}\}_{n\ge1}$ defined in equation (7.1) does not grow too quickly, and thus complete our proof of Theorem 1.1. We do this by comparing the matrix D of interest, defined as in equation (5.6), with the matrix D_∞ studied in Section 8.2.

LEMMA 8.10. *Let $\mathcal{Q} > 100$, $T_b = 0$, $T > \lceil \mathcal{Q}N^2 \log(N) + 2N^2 \rceil$. Then the sequence $\{\psi_n^{-1}\}_{n\ge1}$ defined in equation* (7.1) *and associated with the coupling $\mathcal{R}_{\text{marked}}$ defined in equation* (4.6) *satisfies*

$$\phi_n \ge \psi_n$$

for all $n > N_0$ sufficiently large.

REMARK 8.11. We believe that Lemma 8.10 may hold with $\psi_n^{-1} = O(n^k)$ for some $k < \infty$.

PROOF OF LEMMA 8.10. Define the matrix D as in equation (5.6):

$$D[i,j] = -\operatorname{Tr}\left[a_i\left(\prod_{\ell=i+1}^{j} M_\ell\right)a_j\left(\prod_{\ell=i+1}^{j} M_\ell\right)^{-1}\right], \qquad i<j, \tag{8.15}$$

where

$$M_\ell = \prod_{t=s_\ell+1}^{s_{\ell+1}} R(i_s,\theta_s).$$

Note that, by an application of Lemma 5.5, this matrix is the Jacobian matrix D that appears in equation (7.1). We will now relate D to D_∞. Recall that the off-diagonal entries of D_∞ are written

$$D_\infty[i,j] = -\operatorname{Tr}\left[a_i\left(\prod_{\ell=i+1}^{j} M'_\ell\right)a_j\left(\prod_{\ell=i+1}^{j} M'_\ell\right)^{-1}\right], \qquad i<j,$$

where $\{M'_\ell\}_{\ell=1}^N$ are an i.i.d. sequence from the Haar measure on $\mathrm{SO}(n)$. Applying Lemma 5.1, we have for $1 \le i < j \le N$,

$$\begin{aligned}\bigl|D[i,j]-D_\infty[i,j]\bigr| &\le \sqrt{N}\left\|a_i\left(\prod_{\ell=i+1}^{j} M_\ell\right)a_j\left(\prod_{\ell=i+1}^{j} M_\ell\right)^{-1}\right.\\ &\qquad\left. - a_i\left(\prod_{\ell=i+1}^{j} M'_\ell\right)a_j\left(\prod_{\ell=i+1}^{j} M'_\ell\right)^{-1}\right\|_{\mathrm{HS}}\\ &\le 8\sqrt{N}\sum_{\ell=i+1}^{j}\|M_\ell - M'_\ell\|_{\mathrm{HS}}.\end{aligned} \tag{8.16}$$

Let $\{M_\ell\}_{\ell=1}^N, \{M'_\ell\}_{\ell=1}^N \in \mathrm{SO}(n)$ be the matrices associated to D, D_∞ as above. By Theorem 1 of [45] and the Markov property, for fixed $\mathcal{Q} \ge 4$ it is possible to couple these two sequences so that

$$\mathbb{E}\bigl[\|M_\ell - M'_\ell\|_{\mathrm{HS}} | \{M_k, M'_k\}_{k>\ell}\bigr] \le 4n^{-\mathcal{Q}+3}. \tag{8.17}$$

Note that we are approximating an i.i.d. sequence $\{M'_\ell\}$ by the dependent sequence $\{M_\ell\}$, which is essentially a very sparse subsequence of a run of Kac's walk. The fact that a very sparse sequence of elements from a Markov chain can approximate an i.i.d. sequence from that Markov's chain stationary distribution should not be completely surprising – indeed we should expect that any sufficiently sparse subsequence should approximate such an i.i.d. sequence as long as the Markov chain has sufficiently strong mixing properties.

For fixed $1 \le i < N$ and sequence $\{m_\ell\}_{\ell \ge i+2} \in \mathrm{SO}(n)$, let $\mathcal{G}_i$ (resp., $\mathcal{G}_i'$) be the event $\{M_\ell = m_\ell, \ell \ge i+2\}$ (resp., $\{M_\ell' = m_\ell, \ell \ge i+2\}$). By inequalities (8.16) and (8.17),

$$W_2\big(\mathcal{L}\big(D[i,(i+1):n]|\mathcal{G}_i\big), \mathcal{L}\big(D_\infty[i,(i+1):n]|\mathcal{G}_i'\big)\big)^2 \le 32n^{-\mathcal{Q}+3.5} \tag{8.18}$$

for all $\mathcal{Q} \ge 4$.

With this initial calculation complete, we can now prove Lemma 8.10. Fix notation as in Lemma 8.18. We will apply Lemma 8.5, with D_∞ playing the role of M and D playing the role of M'. By Lemma 8.8, D_∞ satisfies the conditions of Lemma 8.5 with constant $C = N^{20}$. By inequality 8.18, D_∞, D satisfy Condition (8.7) of Lemma 8.5 for all $\mathcal{Q} > 95$. Thus, for fixed $\mathcal{Q} > 95$,

$$\mathbb{P}\big[|D| > \big(4N^{23}\big)^{5N}\big] \le \frac{3}{n}$$

for all $n > N_0$ sufficiently large. By a calculation identical to that in Lemma 8.9, we conclude that

$$\mathbb{P}\big[\sigma_1(D) > \big(4N^{24}\big)^{5N}\big] \le \frac{3}{n},$$

completing the proof. $\square$

9. Proof of Theorem 1.1. We first prove inequality (1.3), the upper bound in Theorem 1.1. Fix constant $\mathcal{Q} = 101$ and sequence $T = T_n \equiv \lceil n^2(40\log(n) - \log(\omega_n))\rceil$. Let the sequence $\mathcal{R}_n$ be the associated sequence as in Definition 4.5, and let the sequence $\kappa = \kappa_n$ be as in Lemma 4.6. Inequality (1.3) will follow immediately from verifying that Conditions (7.2) and (7.3) in Theorem 1 for this choice.

Condition (7.2) follows immediately from Lemma 5.7. Condition (7.3) follows immediately from Lemma 8.10. This completes the proof of inequality (1.3).

Next, we prove inequality (1.4), the lower bound in Theorem 1.1. The proof is essentially a matter of counting dimensions. Fix $T < N$ and $x \in \mathrm{SO}(n)$. For fixed sequence $\mathcal{I} = \{i_t\}_{t=1}^T \in \{1, 2, \dots, N\}^T$, define $\Pi_{\mathcal{I},T} : [0, 2\pi)^T \mapsto \mathrm{SO}(n)$ by

$$\Pi_{\mathcal{I},T}\big(\{\theta_t\}_{t=1}^T\big) = \prod_{t=T}^{1} \mathrm{R}(i_t, \theta_t)x.$$

From this definition, $\mathrm{Rank}((\mathrm{d}\Pi_{\mathcal{I},T})_u) \le T < N$ for all $u \in [0, 2\pi)^T$. Let $A_{\mathcal{I},T} = \Pi_{\mathcal{I},T}([0, 2\pi)^T)$. By Sard's theorem (see page 205 of [26]), this implies

$$\pi(A_{\mathcal{I},T}) = 0. \tag{9.1}$$

Next, define

$$\mathrm{Max}_T = \bigcup_{\mathcal{I} \in \{1,2,\dots,N\}^T} A_{\mathcal{I},T}.$$

Since this is a union of only $\binom{n}{2}^T$ elements, equation (9.1) implies

$$\pi(\mathrm{Max}_T) = 0.$$

Let $\{X_t\}_{t\in\mathbb{N}}$ be a copy of Kac's walk with initial point $x \in \mathrm{SO}(n)$. We have, deterministically, the inclusion $X_T \in \mathrm{Max}_T$. Thus

$$\|\mathcal{L}(X_T) - \pi\|_{\mathrm{TV}} \geq |\mathbb{P}[X_T \in \mathrm{Max}_T] - \pi(\mathrm{Max}_T)| = 1,$$

completing the proof of the lower bound. Thus the proof of Theorem 1.1 is completed.

REMARK 9.1 (Dimension counting and curved spaces). A natural approach for obtaining a lower bound is to count the rank of the *tangent map* associated with the function f_t at 0 rather than bounding the dimension of the image of f_t itself. This approach suggests that the chain will not have mixed until the first time T that the span of $\{a_{i_1}, a_{i_2}, \ldots, a_{i_T}\}$ has dimension N with high probability; by the usual coupon collector argument, this requires $T \gtrsim n^2 \log(n)$.

While this approach works for Gibbs samplers where all moves are straight lines in Euclidean space (for which the rank of the tangent map of f_t at 0 is an upper bound on the dimension of the image of f_t), and it works for Kac's walk on the sphere for a slightly different reason (see [47]), it does *not* work for Kac's walk on $\mathrm{SO}(n)$. In particular, it is possible for f_t to have full dimension N, despite $\mathrm{Dimension}(\mathrm{span}(\{a_{i_1}, a_{i_2}, \ldots, a_{i_t}\})) < N$. See the famous "Euler angle" decomposition of SO(3) for an illustration of this fact [25].

10. Discussion. We mention some consequences of our approach, as well as some open questions.

10.1. *Open problems*. Our work leaves open the question as to whether the mixing time of Kac's walk is indeed $\Theta(n^2 \log(n))$ as conjectured, and whether it exhibits the *cutoff phenomenon*.

The main difficulty in applying our method is obtaining a bound on ψ_n, which measures the smallest singular value $\sigma_1(D)$ of the matrix D. In this paper, we were only able to bound $\sigma_1(D)$ by comparing D to a simpler limiting matrix D_∞ for which exact calculations were available. To improve this bound further, we believe that it is necessary to analyze $\sigma_1(D)$ directly. To obtain an $O(n^2 \log(n))$ bound on the mixing time of Kac's walk using our method, it would be enough to obtain any polynomial bound $\psi_n = O(n^k)$ for some $0 < k < \infty$. The main obstacles to proving such a bound are:

1. Our weak random matrix bound in Lemma 8.3. Our argument for Lemma 8.3, like that in [22], only takes advantage of the randomness of entries within distance 1 of the diagonal of an n by n random matrix M. Unfortunately, no argument that only analyzes these entries can give any bound that is stronger than $|M|^{-2} = 2^{O(n)}$. Since the matrix D of interest has "many more" than $3n$ "pieces" of randomness, we can hope to take advantage of them and obtain a stronger bound, as in [4, 20].

2. We bound the determinant $|D|$ of the random matrix D and use this to obtain a bound on the smallest singular value $\sigma_1(D)$; we give up a factor of N^{N-1} in the process (see Lemma 8.9). To avoid this loss, we must either obtain stronger bounds on the joint distribution of the remaining singular values $\sigma_2(D) \le \sigma_3(D) \le \cdots \le \sigma_N(D) \gtrsim \log(n)$, or to bound $\sigma_1(D)$ directly as in [4, 20].

3. Our bound on $\sigma_1(D)$ is obtained through a comparison of D and the related matrix D_∞ (see Theorem 8.10). While D, D_∞ are nearby under the choice of $\mathcal{R}$ in Definition 4.5, they are very far for more reasonable choice of $\mathcal{R}$ (see e.g., Definition 10.1 below). Thus, to get a better mixing bound, the matrix D should be studied directly.

It seems possible to extend our arguments to resolve any two of these three obstacles together; however, we see no route to resolving all three simultaneously without substantial changes.

10.2. *Applications to other Gibbs samplers*. We suspect that a result that is qualitatively similar to Theorem 4.3 can hold for many other Gibbs samplers. Roughly speaking, we expect that the following three ingredients should be enough to obtain such a result:

1. a random mapping representation of the form (2.3) must exist, and
2. the representation must be sufficiently "nice" that the measurability results given in Lemmas A.1–A.6 hold with the obvious modifications, and
3. there must be a candidate contractive coupling κ.

Informally, these requirements do not appear to be onerous. Condition (1) is the usual way to write down a Gibbs sampler, condition (2) consists of technical measurability issues that we do not expect to provide difficulties for "natural" Gibbs samplers, while any optimal 1-step coupling is a reasonable candidate for (3).

The key ingredients in extending our bounds to new examples are the analysis of an underlying "scaffolding" coupling ν, the analysis of the smallest singular value of the Jacobian of the "perturbation map" $f_{\mathcal{A}}$, and soft bounds on the smoothness of $f_{\mathcal{A}}$.

10.3. *Applications to simulation*. Although we have improved on the best mixing time bound for Kac's walk, we strongly suspect that our upper bound is far from optimal. In this section, we mention that our bounds can be combined with computer simulation to obtain improved bounds on the mixing time.

Recall that in Section A.3, we defined our coupling in terms of a function $\mathcal{R} : \{1, 2, \dots, N\}^T \mapsto \mathfrak{B}_T$. In that section, we chose the function $\mathcal{R}$ according to Definition 4.5. We then analyzed this choice of coupling for the remainder of the paper. We now consider the following alternative choice of function.

DEFINITION 10.1 (Greedy subset choice). Fix $T \in \mathbb{N}$. Let $\{i_t\}_{t=1}^{T}$ be a sequence with $i_t \in \{1, 2, \dots, N\}$. For $1 \le t \le T$, define $V_t = \operatorname{span}(\{a_{i(1)}, a_{i(2)}, \dots, a_{i_t}\})$. Then define $s_1 = 1$ and inductively define

$$s_{\ell+1} = \inf\{s > s_\ell : V_s \neq V_{s_\ell}\}.$$

Then define $\mathcal{R}_{\text{greedy}}$ by

$$\mathcal{R}_{\text{greedy}}(\{i_t\}_{t=1}^{T}) = \begin{cases} \{s_1, \dots, s_N\}, & \text{if } T \ge s_N + 1, \\ \{1, 2, \dots, N\}, & \text{otherwise.} \end{cases}$$

We give the following analogue to Lemma 5.7.

LEMMA 10.2. *Fix $c > 0$ and $T = \lceil N \log(N) + cN \rceil$. Let s_N be defined as in Definition 10.1. Let $\{i_t\}_{t=1}^{T} \sim \operatorname{Unif}(\{1, 2, \dots, N\}^T)$. Then*

$$\lim_{n \to \infty} \mathbb{P}[s_N < T] = e^{-e^c}. \tag{10.1}$$

PROOF. As with Lemma 5.7, the proof is deferred to Appendix C.2. □

Recall the definition of ψ_n from equation (7.1), with choice $T = n^3$ and $\mathcal{R} = \mathcal{R}_{\text{greedy}}$. By a minor modification of Theorem 1 with these choices, combined with Remark 7.1, we find that

$$\tau_{\text{mix}} = O\big(n^2\big(-\log\big(\min(n^{-1}, \psi_n)\big)\big)\big)$$

as long as $\psi_n = O(n^{-n})$. We do not know how to prove that $\psi_n = O(n^{-n})$. However, we can directly simulate the random matrix D that appears in equation (7.1) easily on a computer. Thus, the quantiles of the distribution of $\sigma_1(D)$ can be estimated by simulation, which allows us to calculate upper bounds on the constant ψ_n with high confidence.

We point out that, a priori, it is not clear how to obtain, via simulation, *any* reasonable estimate on the mixing time of a Gibbs sampler on a continuous state space. For this reason, we hope that this bound and its extension to other Gibbs samplers may be of independent interest.

APPENDIX A: PROOF OF THEOREM 4.3

We prove Theorem 4.3, our main coupling inequality. This largely entails checking that all of the random variables have the claimed distributions, and that the "usual" maximal coupling of measures can be extended in a measurable way to our setting.

We begin by showing that the map $f_{\mathcal{A}}$ constructed in Definition (4.1) can be used to give an alternative random mapping representation of Kac's walk.

LEMMA A.1. *Fix $T \in \mathbb{N}$ and $x \in \mathrm{SO}(n)$. Let $\{i_t, \theta_t\}_{t=1}^T \sim \mathrm{Unif}(\mathfrak{A}^T)$, and let $\{\delta_i\}_{i=1}^N$ be any sequence of independent random variables that is independent of the sequence $\{i_t, \theta_t\}_{t=1}^T$. Finally, let $\mathcal{R} : \{1, 2, \dots, N\}^T \mapsto \mathfrak{B}_T$ be arbitrary and set $\mathcal{S} = \mathcal{R}(\{i_t\}_{t=1}^T)$.*

Then for $\mathcal{A}$ and $f_{\mathcal{A}}$ as given in Definition 4.1,

$$\mathbb{P}\big[f_{\mathcal{A}}(\delta_1, \dots, \delta_N) \in B\big] = K^T(x, B)$$

for all measurable $B \subset \mathrm{SO}(n)$.

REMARK A.2. In the hypothesis of Lemma A.1, we do not require $\{\delta_i\}_{i=1}^N$ to be identically distributed.

PROOF OF LEMMA A.1. We follow the notation of Definition 4.1 and define the sequence $\{X_t\}_{t=1}^{T+1}$ by the recursion:

$$\begin{aligned} X_1 &= x, \\ X_{t+1} &= \mathrm{F}\big(X_t, \big(i_t, e_{\mathcal{A},t}(\delta_1, \dots, \delta_N)\big)\big), \qquad 1 \le t \le T, \end{aligned} \tag{A.1}$$

where F is the random mapping representation of Kac's walk given in equation (2.2). Note that, under this definition, $X_{T+1} = f_{\mathcal{A}}(\delta_1, \dots, \delta_N)$. Therefore, to check that $X_{T+1} \sim K^T(x, \cdot)$, it is sufficient to check that the update sequence $\{(i_t, e_{\mathcal{A},t}(\delta_1, \dots, \delta_N))\}_{t=1}^T$ used in equation (A.1) has the correct distribution. The sequence $\{i_t\}_{t=1}^T$ is defined to be an i.i.d. sequence of unif$(\{1, 2, \dots, N\})$ random variables, so the marginal distribution of this sequence is correct.

To check that $\{e_{\mathcal{A},t}(\delta_1, \dots, \delta_N)\}_{t=1}^T$ is an i.i.d. sequence of unif$[0, 2\pi)$ random variables conditional on $\{i_t\}_{t=1}^T$, we note that conditional on $\{\delta_i^{(x)}\}_{i=1}^N$, $\{i_t\}_{t=1}^T$ and $\mathcal{S} = \mathcal{R}(\{i_t\}_{t=1}^T)$, the sequence $\{\theta_t\}_{t=1}^T$ is an i.i.d. sequence of unif$[0, 2\pi)$ random variables. Since the sum (modulo 2π) of any random variable and an independent unif$[0, 2\pi)$ random variable is itself uniformly distributed on $[0, 2\pi)$, this implies $\{e_{\mathcal{A},t}(\delta_1, \dots, \delta_N)\}_{t=1}^T$ is, conditional on $\{i_t\}_{t=1}^T$, an i.i.d. sequence of unif$[0, 2\pi)$ random variables. This completes the proof. $\square$

A.1. Density of random mapping representation. We now check that the random mapping representation described in Lemma A.1 has a useful density formula. Fix $T \in \mathbb{N}$ and $x \in \mathrm{SO}(n)$. Fix $\{i_t\}_{t=1}^T \in \{1, 2, \dots, N\}^T$ and $\mathcal{S} \in \mathfrak{B}_T$ that satisfy

$$\bigcup_{s \in \mathcal{S}} \{i_s\} = \{1, 2, \dots, N\}. \tag{A.2}$$

For fixed $\{\theta_t\}_{t=1}^T \in [0, 2\pi)^T$, set

$$\mathcal{A} = \big\{x, T, \mathcal{S}, \{i_t\}_{t=1}^T, \{\theta_t\}_{t=1}^T\big\},$$

and let $f_{\mathcal{A}}$ be as in Definition 4.1. Let $\delta \equiv \{\delta_t\}_{t=1}^{N} \sim \mathrm{Unif}([-\varepsilon_n, \varepsilon_n]^N)$ be a sequence of i.i.d. random variables, and define the set of "nonsingular" angles by

$$\Theta_{\mathrm{ns}} = \{\{\theta_t\}_{t=1}^{T} \in [0, 2\pi)^T : \mathbb{P}[\det(\mathrm{d} f_{\mathcal{A}}(\delta)) = 0] = 0\}, \tag{A.3}$$

where in equation (A.3), x, $\{i_t\}_{t=1}^{T} \in \{1, 2, \dots, N\}^T$ and $\mathcal{S} \in \mathfrak{B}_T$ are viewed as fixed parameters and the only random variables are $(\delta_1, \dots, \delta_N)$. In particular, in equation (A.3), $f_{\mathcal{A}}$ is viewed as a deterministic function of the arguments $\{(i_t, \theta_t)\}_{t=1}^{T} \in \mathfrak{A}^T$.

For $v \in \mathrm{SO}(n)$, define

$$\mathfrak{P}_{\mathcal{A}}(v) = \{u \in [-\varepsilon_n, \varepsilon_n]^N : f_{\mathcal{A}}(u) = v, \det(\mathrm{d} f_{\mathcal{A}}(u)) \neq 0\}$$

to be the set of pre-images of v that are not critical points of $f_{\mathcal{A}}$.

THEOREM 2 (Density of random mapping representation). *We have the following*:

1. *The Lebesgue measure of the set* $[0, 2\pi)^T \setminus \Theta_{\mathrm{ns}}$ *is zero*.
2. *For every* $v \in \mathrm{SO}(n)$, *the set* $\mathfrak{P}_{\mathcal{A}}(v) \subset [0, 2\pi)^T$ *has only countably many elements*.
3. *For any fixed* $\{\theta_t\}_{t=1}^{T} \in \Theta_{\mathrm{ns}}$, *the random variable* $f_{\mathcal{A}}(\delta_1, \dots, \delta_N)$ *has a density* $\rho_{\mathcal{A}}$ *with respect to the Haar measure* μ *on* $\mathrm{SO}(n)$

$$\rho_{\mathcal{A}}(v) \propto \sum_{u \in \mathfrak{P}_{\mathcal{A}}(v)} \frac{1}{\det(\mathrm{d} f_{\mathcal{A}}(u))} \tag{A.4}$$

for μ*-almost every* $v \in \mathrm{SO}(n)$.

The remainder of this section will be dedicated to the proof of Theorem 2. We note that we use several calculations and facts that are stated in Section 5. Although Section 5 appears after the statement of Theorem 4.3 in our paper, the results in Section 5 do not make use of Theorem 4.3.

We begin by recalling the following change of variables formula.

PROPOSITION A.3. *Fix constants* $0 < b < \infty$ *and* $k \in \mathbb{N}$ *and a diffeomorphism* $f : [-b, b]^k \mapsto \mathrm{SO}(n)$. *Let* $U_1, \dots, U_k \in \mathrm{unif}([-b, b])$ *be i.i.d. Then the random variable* $f(U_1, \dots, U_N)$ *has density*

$$\rho(v) \propto \frac{1}{\det(\mathrm{d} f)(f^{-1}(v))}$$

with respect to the Haar measure μ *on* $\mathrm{SO}(n)$.

We will use the following extension of Proposition A.3. We suspect that this extension is standard, but are not aware of any references, and so provide a proof.

PROPOSITION A.4 (Formula for densities). *Fix constants* $0 < b < \infty$ *and a smooth function* $f : [-b, b]^N \mapsto \mathrm{SO}(n)$ *for which the complement of the set*

$$\Delta_f = \{u \in [-b, b]^N : \det(\mathrm{d}f(u)) \neq 0\}$$

has measure 0. *Let* $U_1, \dots, U_N \in \mathrm{unif}([-b, b])$ *be i.i.d. Then*:

1. *For every* $v \in \mathrm{SO}(n)$, *the set* $\{u \in \Delta_f : f(u) = v\}$ *has only countably many elements*.
2. *The random variable* $f(U_1, \dots, U_N)$ *has density*

$$\rho(v) \propto \sum_{u \in \Delta_f : f(u) = v} \frac{1}{\det(\mathrm{d}f(u))}$$

with respect to the Haar measure on $\mathrm{SO}(n)$.

PROOF. We set some notation. For a set $S \subset [-b, b]^N$, we denote by $f|_S$ the *restriction of* f *to* S, that is, the function with domain S that agrees with f at every point. Also, for a point x in Euclidean space and radius $r > 0$, we denote by $B_r(x) = \{y : \|x - y\| \leq r\}$ the ball of radius r around x.

Next, for $u \in \Delta_f$, let

$$r(u) = \frac{1}{2} \sup\{r' : f|_{B_{r'}(u)} \text{ is a diffeomorphism}\}.$$

By the inverse function theorem and the definition of f, we have $0 < r(u) < 3Nb$ for all $u \in \Delta_f$. It is clear that

$$\Delta_f \subset \bigcup_{u \in \Delta_f} B_{r(u)}(u),$$

and so by the Besicovitch covering theorem (see, e.g., [23]) there exists a countable set $\{u_i\}_{i \in \mathbb{N}}$ with the properties that $r(u_i) > 0$ for all $i \in \mathbb{N}$, and also

$$\Delta_f \subset \bigcup_{i \in \mathbb{N}} B_{r(u_i)}(u_i), \tag{A.5}$$

with

$$\sup_{u \in [-b,b]^N} \left|\{i : u \in B_{r(u_i)}(u_i)\}\right| \leq C < \infty$$

for some constant $0 < C < \infty$. Thus (A.5) implies that the set Δ_f has a countable subcover, for which f is a diffeomorphism on each piece. Because $f|_{S_i}$ is a diffeomorphism for every $i \in \mathbb{N}$, we know that for every $v \in \mathrm{SO}(n)$, the set $\Delta_f \cap S_i \cap f^{-1}(v)$ has at most one element. This implies that the set $\Delta_f \cap f^{-1}(v)$ is countable, proving the first assertion.

Define the sequence of disjoint sets

$$S_1 = B_{r(u_1)}(u_1),$$
$$S_i = B_{r(u_i)}(u_i) \setminus \bigcup_{1 \le j < i} B_{r(u_j)}(u_j), \qquad i > 1.$$

Since $\{S_i\}_{i \in \mathbb{N}}$ cover Δ_f, the set E defined by

$$E \equiv [-b, b]^N \setminus \bigcup_{i \in \mathbb{N}} S_i = [-b, b]^N \setminus \bigcup_{i \in \mathbb{N}} B_{r(u_i)}(u_i)$$

has measure 0 by assumption. Denoting $f_i = f|_{S_i}$, for all measurable $A \subset \mathrm{SO}(n)$ we have that

$$\begin{aligned}
\mathbb{P}\bigl[f(U_1, \dots, U_N) \in A\bigr] &= \sum_{i \in \mathbb{N}} \mathbb{P}\bigl[f_i(U_1, \dots, U_N) \in A \;\&\; (U_1, \dots, U_N) \in S_i\bigr] \\
&= \mathcal{Z} \sum_{i \in \mathbb{N}} \int_{A \cap f(S_i)} \frac{1}{\det(\mathrm{d} f_i(f_i^{-1}(v)))} \, dv \\
&= \mathcal{Z} \int_A \sum_{i \in \mathbb{N}} \frac{1}{\det(\mathrm{d} f_i(f_i^{-1}(v)))} \mathbf{1}_{v \in f(S_i)} \, dv \\
&= \mathcal{Z} \int_A \sum_{u \in \Delta_f : f(u) = v} \frac{1}{\det(\mathrm{d}(f)(u))} \, dv,
\end{aligned}$$

where the second equality follows from Proposition A.3, the constant $\mathcal{Z}$ is the normalizing constant that appears in the statement of Proposition A.3, and the third equality is an application of Fubini's theorem. This proves the second assertion and the proof is completed. □

We now prove Theorem 2.

PROOF OF THEOREM 2. Denote by $\Theta_s = [0, 2\pi)^T \setminus \Theta_{ns}$ the set of singular angles. Viewing $\{\theta_t\}_{t=1}^T \in [0, 2\pi)^T$ as parameters, inspection of the determinant formula in equation (5.6) yields three important facts about the determinant

$$D_{\mathcal{A}}(\theta_1, \dots, \theta_T, \delta_1, \dots, \delta_N) \equiv \det(\mathrm{d} f_{\mathcal{A}})(\delta_1, \dots, \delta_N)$$

of $\mathrm{d} f_{\mathcal{A}}$ at a point $\{\delta_t\}_{t=1}^N \in [0, 2\pi)^T$:

1. The determinant $D_{\mathcal{A}}(\theta_1, \dots, \theta_T, \delta_1, \dots, \delta_N)$ is a multivariate polynomial in the $2(T + N)$ variables

$$(z_1, \dots, z_{2(N+T)}) \equiv \bigl(\cos(\theta_1), \sin(\theta_1), \dots, \sin(\theta_T), \cos(\delta_1), \dots, \sin(\delta_N)\bigr).$$

2. The polynomial $D_{\mathcal{A}} = D_{\mathcal{A}}(z_1, \dots, z_{2(T+N)})$ can be written (via the usual formula for determinants and product-sum formula for trigonometric functions) as the sum of polynomials in the same variables with degree at most $(4T)^N$.

3. Finally, due to our requirement of (A.2), $\mathrm{d}f_{\mathcal{A}}(0)$ is the identity matrix, and thus

$$D_{\mathcal{A}}(0, \dots, 0) = 1.$$

Fix $\varepsilon > 0$, and denote by λ the Lebesgue measure on $[0, 2\pi)^{T+N}$. By a multivariate generalization of the Turan theorem (see Theorem 2 of [21]), these three facts imply the very weak bound

$$\lambda\big(\big\{\Gamma \in [0, 2\pi)^{T+N} : \big|D_{\mathcal{A}}(\Gamma)\big| \le \varepsilon\big\}\big) \le 14(T + N)\varepsilon^{(4T)^{-N}}.$$

In particular, this immediately implies

$$\lambda\big(\Gamma \in [0, 2\pi)^{T+N} : \big|D_{\mathcal{A}}(\Gamma)\big| = 0\}\big) = 0, \tag{A.6}$$

which in turn implies that Θ_{s} is measure 0. This completes the proof of the first claim in Theorem 2. Applying Proposition A.4 completes the proof of the main claim of Theorem 2. $\square$

A.2. Coupling construction. Recall the measures $\mu_{\mathcal{A}}$ from Definition 4.1. Now, we will use the density functions of the measures $\mu_{\mathcal{A}}$ from Theorem 2 to construct the standard "maximal coupling" of two such measures. Roughly speaking, the following coupling proceeds by trying to force the two random variables to be equal with maximal probability. This is a standard construction; see for instance, Theorem 2.12 of [14].

Fix $T \in \mathbb{N}$, $x, y \in \mathrm{SO}(n)$. Fix three sequences $\mathcal{I} = \{i_1, \dots, i_T\}$ and $\Theta^{(z)} = \{\theta_1^{(z)}, \dots, \theta_T^{(z)}\}$, $z \in \{x, y\}$, with $i_t \in \{1, 2, \dots, N\}$ and $\theta_t^{(z)} \in [0, 2\pi)$. Fix $\mathcal{R} : \{1, 2, \dots, N\} \mapsto \mathfrak{B}_T$ and set $\mathcal{S} = \mathcal{R}(\mathcal{I})$. Finally, let $\mathcal{A}_z = \{z, T, \mathcal{S}, \mathcal{I}, \Theta^{(z)}\}$ with $z \in \{x, y\}$ and write $\mathcal{A}_{x,y} = (\mathcal{A}_x, \mathcal{A}_y)$.

DEFINITION A.5 (Coupling of induced maps). Let Θ_{ns} be as in equation (A.3). We now define a coupling $\mu_{\mathcal{A}_{x,y}}$ on $\mathrm{SO}(n) \times \mathrm{SO}(n)$ of the pair of measures $\mu_{\mathcal{A}_x}$ and $\mu_{\mathcal{A}_y}$ following the construction in Theorem 2.12 of [14]. We do this in two cases.

1. *Nonsingular case*: The set $\mathcal{S}$ satisfies $\{i_{s_\ell}\}_{\ell=1}^N = \{1, 2, \dots, N\}$ and $\{\theta_t^{(z)}\}_{t=1}^T \notin \Theta_{\mathrm{ns}}$ for both of $z \in \{x, y\}$. By Theorem 2, for $z \in \{x, y\}$, the measure $\mu_{\mathcal{A}_z}$ has a density, with respect to the Haar measure μ on $\mathrm{SO}(n)$, that is given by the formula

$$\rho^{(z)}(v) = \sum_{u \in \mathfrak{P}_{\mathcal{A}}} \frac{1}{\det(\mathrm{d}f_{\mathcal{A}_z}(u))}$$

v-almost everywhere. Define the "minorization" measure $\tilde{\mu}_{\min}$ on $\mathrm{SO}(n)$ by

$$Z_{\min} = \int_{\mathrm{SO}(n)} \min\big(\rho^{(x)}(v), \rho^{(y)}(v)\big)\mu(dv),$$

$$\tilde{\mu}_{\min}(A) = \begin{cases} \dfrac{1}{Z_{\min}} \displaystyle\int_A \min\big(\rho^{(x)}(v), \rho^{(y)}(v)\big)\mu(dv), & Z_{\min} > 0, \\ \mu(A), & Z_{\min} = 0 \end{cases}$$

for measurable $A \subset \mathrm{SO}(n)$. Let $\mathrm{diag} : \mathrm{SO}(n) \mapsto \mathrm{SO}(n) \times \mathrm{SO}(n)$ be given by $\mathrm{diag}(x) = (x, x)$, and define

$$\mu_{\min} = \tilde{\mu}_{\min} \circ \mathrm{diag}^{-1},$$

a measure that concentrates on the diagonal of $\mathrm{SO}(n) \times \mathrm{SO}(n)$. Define the "remainder" densities on $\mathrm{SO}(n)$ as

$$\rho_{\mathrm{rem}}^{(z)}(v) = \begin{cases} \dfrac{1}{1 - Z_{\min}} \big(\rho^{(z)}(v) - \min\big(\rho^{(x)}(v), \rho^{(y)}(v)\big)\big), & Z_{\min} < 1, \\ \rho^{(z)}(v), & Z_{\min} = 1 \end{cases}$$

and on $\mathrm{SO}(n) \times \mathrm{SO}(n)$ as

$$\rho_{\mathrm{rem}}(v_1, v_2) = \rho_{\mathrm{rem}}^{(x)}(v_1) \rho_{\mathrm{rem}}^{(y)}(v_2).$$

Finally, define for $A \subset \mathrm{SO}(n) \times \mathrm{SO}(n)$,

$$\mu_{\mathcal{A}_{x,y}}(A) = Z_{\min} \mu_{\min}(A) + (1 - Z_{\min}) \int_A \rho_{\mathrm{rem}}(v) \mu(dv).$$

2. *Singular case*: The set $\mathcal{S}$ does not satisfy $\{i_{s_\ell}\}_{\ell=1}^N = \{1, 2, \dots, N\}$, or $\{\theta_t^{(z)}\}_{t=1}^T \in \Theta_{\mathrm{ns}}$ for one of $z \in \{x, y\}$. In this case, let $\mu_{\mathcal{A}_{x,y}}$ be product measure of $\mu_{\mathcal{A}_x}$ and $\mu_{\mathcal{A}_y}$.

We next verify that this coupling can be used to describe a transition kernel, and that it has the desired optimal-coupling property.

LEMMA A.6. *Fix $x, y \in \mathrm{SO}(n)$ and $T \in \mathbb{N}$. Fix a measurable function $\mathcal{R} : \{1, 2, \dots, N\}^T \mapsto \mathfrak{B}_T$. For any sequence $\{i_t, \theta_t^{(x)}, \theta_t^{(y)}\}_{t=1}^T$ and $\mathcal{S} = \mathcal{R}(\{i_t\}_{t=1}^T)$, let*

$$\mathcal{A}_{x,y} = \mathcal{A}_{x,y}\big(\{i_t, \theta_t^{(x)}, \theta_t^{(y)}\}_{t=1}^T\big) \equiv (\mathcal{A}_x, \mathcal{A}_y),$$

where $\mathcal{A}_x, \mathcal{A}_y$ are as in (4.1). Then collection of measures $\{\mu_{\mathcal{A}_{x,y}}\}$ from Definition A.5 is a transition kernel from $\{1, 2, \dots, N\}^T \times [0, 2\pi)^{2T}$ to $\mathrm{SO}(n) \times \mathrm{SO}(n)$. Furthermore, for $(X, Y) \sim \mu_{\mathcal{A}_{x,y}}$,

$$\mathbb{P}[X \neq Y] = \|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} \tag{A.7}$$

whenever $\{i_s\}_{s \in \mathcal{S}} = \{1, 2, \dots, N\}$ and $\{\theta_t^{(x)}\}_{t=1}^T, \{\theta_t^{(y)}\}_{t=1}^T \notin \Theta_{\mathrm{ns}}$.

PROOF. It is clear from the construction that $\mu_{\mathcal{A}_{x,y}}$ is a measure for each fixed $\{i_t\}_{t=1}^T \in \{1, 2, \dots, N\}^T, \{\theta_t^{(x)}\}_{t=1}^T, \{\theta_t^{(y)}\}_{t=1}^T \in [0, 2\pi)^T$. Thus, to check that $\{\mu_{\mathcal{A}_{x,y}}\}$ is a transition kernel, it is enough to check that the map $M_B : (\{1, 2, \dots, N\} \times [0, 2\pi) \times [0, 2\pi))^T \mapsto [0, 1]$ given by

$$M_B\big(\{i_t, \theta_t^{(x)}, \theta_t^{(y)}\}_{t=1}^T\big) \equiv \mu_{\mathcal{A}_{x,y}}(B)$$

is measurable for each measurable $B \subset \mathrm{SO}(n) \times \mathrm{SO}(n)$. Since the variables $\{i_t\}_{t=1}^T$ are discrete, it is enough to check that the dependence on $\{\theta_t^{(x)}, \theta_t^{(y)}\}_{t=1}^T$ is measurable for each fixed sequence $\{i_t\}_{t=1}^T$.

This follows immediately from the fact that Definition A.5 gives an explicit (if rather complicated) formula for the distribution $\mu_{\mathcal{A}_x,\mathcal{A}_y}$ density in terms of a collection of measurable functions of the parameters $\{\theta_t^{(x)}, \theta_t^{(y)}\}_{t=1}^T$ and the arguments $\{\delta_t^{(x)}, \delta_t^{(y)}\}_{t=1}^N$. In particular, in Definition A.5, all terms that appear in the joint distribution are obtained by starting with the smooth functions $f_{\mathcal{A}_x}$, $f_{\mathcal{A}_y}$ and then making a finite number of measurable transformations. These transformations are limited to taking derivatives of smooth functions, replacing a measurable function f with an indicator function of the form $\mathbf{1}_{f(x)=c}(x)$ or $\mathbf{1}_{f(x)\le c}(x)$, taking products, differences, minima and inverses of measurable functions, and finally taking integrals of the results. Since $M_B(\{i_t, \theta_t^{(x)}, \theta_t^{(y)}\}_{t=1}^T)$ is defined in terms of this measurable formula for $\mu_{\mathcal{A}_{x,y}}$, it is also measurable. This completes our proof that $\{\mu_{\mathcal{A}_{x,y}}\}$ is a transition kernel.

Finally, equality (A.7) follows immediately from Theorem 2.12 of [14], since the coupling is constructed exactly as in the proof of that theorem. This completes the proof of the lemma. □

A.3. Proof of Theorem 4.3. We now complete the proof of Theorem 4.3. Since unif($\{1, 2, \ldots, N\}$) is discrete, the measure unif($\{1, 2, \ldots, N\}$) $\otimes\, \kappa$ exists, where we recall that the symbol "$\otimes$" denotes the usual combination of two kernels (see again the notation of Theorem 5.17 of [40]). Let

$$\kappa_{\min} = \{\mu_{\mathcal{A}_{x,y}}\}$$

be the collection of measures as given in Definition A.5. By Lemma A.6, this collection of measures is in fact a transition kernel. Thus, by the Ionescu–Tulcea theorem (see Theorem 5.17 of [40]), the combined measure unif($\{1, 2, \ldots, N\}$) $\otimes$ $\kappa \otimes \kappa_{\min}$ also exists. Let

$$\{i_t\}_{t=1}^T, \{\theta_t^{(x)}\}_{t=1}^T, \{\theta_t^{(y)}\}_{t=1}^T, X_{T+1}, Y_{T+1} \sim \mathrm{unif}(\{1, 2, \ldots, N\}) \otimes \nu \otimes \kappa_{\min}.$$

For any $\mathcal{I} \in \{1, 2, \ldots, N\}^T$, we have by this construction

$$\mathcal{L}(\{\theta_t^{(x)}\}_{t=1}^T, \{\theta_t^{(y)}\}_{t=1}^T | \{i_t\}_{t=1}^T = \mathcal{I}) = \nu_{\mathcal{I}}.$$

Next, by Lemma A.1, the random variables X_{T+1}, Y_{T+1} have marginal distributions

$$X_{T+1} \sim K^T(x, \cdot), \qquad Y_{T+1} \sim K^T(y, \cdot). \tag{A.8}$$

Let $\mathcal{A}$ be as defined in Definition 4.1.

By equality (A.7) in Lemma A.6, combined with the fact from Theorem 2 that $[-\varepsilon_n, \varepsilon_n]^T \setminus \Theta_{\mathrm{ns}}$ has Lebesgue measure 0, we also have

$$\begin{aligned}\mathbb{P}[X_{T+1} \neq Y_{T+1}] &\leq \sum_{\mathcal{J} \in \mathcal{G}_{\mathcal{R}}} \mathbb{E}\big[\|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} | \{i_t\}_{t=1}^T = \mathcal{J}\big] \mathbb{P}\big[\{i_t\}_{t=1}^T = \mathcal{J}\big] \\ &\quad + \sum_{\mathcal{J} \notin \mathcal{G}_{\mathcal{R}}} \mathbb{P}\big[\{i_t\}_{t=1}^T = \mathcal{J}\big] \\ &\leq \mathbb{E}\big[\|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} \mathbf{1}_{\{i_t\}_{t=1}^T \in \mathcal{G}_{\mathcal{R}}}\big] + \mathbb{P}\big[\{i_t\}_{t=1}^T \notin \mathcal{G}_{\mathcal{R}}\big],\end{aligned} \tag{A.9}$$

where $\{i_t\}_{t=1}^T$, $\{\theta_t^{(x)}\}_{t=1}^T$, and $\{\theta_t^{(y)}\}_{t=1}^T$ are viewed as random variables in the above expression, and all other parameters are viewed as fixed.

By equality (A.8) and inequality (A.9),

$$\begin{aligned}\|K^T(x, \cdot) - K^T(y, \cdot)\|_{\mathrm{TV}} &\leq \mathbb{P}[X_{T+1} \neq Y_{T+1}] \\ &\leq \mathbb{E}\big[\|\mu_{\mathcal{A}_x} - \mu_{\mathcal{A}_y}\|_{\mathrm{TV}} \mathbf{1}_{\mathcal{I} \in \mathcal{G}_{\mathcal{R}}}\big] + \mathbb{P}[\mathcal{I} \notin \mathcal{G}_{\mathcal{R}}].\end{aligned}$$

This completes the proof the proof of Theorem 4.3.

APPENDIX B: PROOF OF LEMMA 4.6

Denote by D_{HS} the Riemannian distance on $\mathrm{SO}(n)$ and by W_{D} the Wasserstein distance associated with this metric. In Lemma 1 of [45], the author shows that

$$W_D\big(K(x, \cdot), K(y, \cdot)\big) \leq \sqrt{1 - \frac{2}{n(n-1)}} \mathrm{D}_{\mathrm{HS}}(x, y)$$

for all $x, y \in \mathrm{SO}(n)$. This implies (via the standard disintegration theorem—see Theorem 5.4 of [40]) that, for any $\gamma > 0$ and $x, y \in \mathrm{SO}(n)$, it is possible to couple $(i^{(x)}, \theta^{(x)}) \sim \mathrm{unif}(\mathfrak{A})$, $(i^{(y)}, \theta^{(y)}) \sim \mathrm{unif}(\mathfrak{A})$ so that

$$\begin{aligned}&\mathbb{E}\big[\mathrm{D}_{\mathrm{HS}}\big(\mathrm{F}\big(x, \big(i^{(x)}, \theta^{(x)}\big)\big), \mathrm{F}\big(y, \big(i^{(y)}, \theta^{(y)}\big)\big)\big)\big] \\ &\quad \leq \sqrt{1 - \frac{2}{n(n-1)}} \mathrm{D}_{\mathrm{HS}}(x, y) + \gamma.\end{aligned} \tag{B.1}$$

We note that this bound is not sufficient for our coupling, since Lemma 1 of [45] allows for the possibility that $i^{(x)} \neq i^{(y)}$ in the associated coupling.

However, below we give an argument to show that inequality (B.1) can still hold if we insist that $i^{(x)} = i^{(y)}$. Fix $x, y \in \mathrm{SO}(n)$. This argument is based on Lemma 1 of [45] and an explicit version of the usual path coupling construction (see [6] for the first version of this construction in the Markov chain literature).

In the proof of Lemma 1 of [45], the author constructs a coupling $\tilde{\nu}_\gamma$ of $i \sim \mathrm{unif}(\{1, 2, \dots, N\})$ and $\theta^{(x)}, \theta^{(y)} \sim \mathrm{unif}([0, 2\pi))$ with the property

$$\begin{aligned}&\mathbb{E}\big[\mathrm{D}_{\mathrm{HS}}\big(\mathrm{F}\big(x, \big(i, \theta^{(x)}\big)\big), \mathrm{F}\big(y, \big(i, \theta^{(y)}\big)\big)\big)^2\big] \\ &\quad \leq \left(1 - \frac{2}{n(n-1)}\right) \mathrm{D}_{\mathrm{HS}}(x, y)^2 + C \mathrm{D}_{\mathrm{HS}}(x, y)^3,\end{aligned} \tag{B.2}$$

where the constant C is bounded by a polynomial in n and is independent of x, y. Fix $0 < \gamma < 0.1/C$, and let $x = x_1, x_2, \dots, x_k = y$ be a sequence that satisfies

$$\begin{aligned}
&\mathrm{D_{HS}}(x_j, x_{j+1}) = \gamma, \qquad 1 \le j < k-1,\\
&\mathrm{D_{HS}}(x_{k-1}, x_k) \le \gamma,\\
&\sum_{j=1}^{k-1} \mathrm{D_{HS}}(x_j, x_{j+1}) = \mathrm{D_{HS}}(x, y).
\end{aligned}$$

The existence of such a sequence is an immediate consequence of the fact that there exists a distance minimizing geodesic between any pair of points on SO(n) [see the Hopf–Rinow theorem [31] for the proof that distance-minimizing geodesics exist on SO(n)]. By the Ionescu–Tulcea theorem (see Theorem 5.17 of [40]), the coupling $\tilde{\nu}_\gamma$ that appears before inequality (B.2) can be extended to a coupling $\hat{\nu}_\gamma$ of $i \sim \mathrm{unif}(\{1, 2, \dots, N\})$ and $\theta^{(1)}, \dots, \theta^{(k)} \sim \mathrm{unif}([0, 2\pi))$ with the property

$$\begin{aligned}
&\mathbb{E}\bigl[\mathrm{D_{HS}}\bigl(\mathrm{F}\bigl(x_j, \bigl(i, \theta^{(j)}\bigr)\bigr), \mathrm{F}\bigl(x_{j+1}, \bigl(i, \theta^{(j+1)}\bigr)\bigr)\bigr)^2\bigr]\\
&\quad\le \left(1 - \frac{2}{n(n-1)}\right)\mathrm{D_{HS}}(x_j, x_{j+1})^2 + C\mathrm{D_{HS}}(x_j, x_{j+1})^3 \qquad \text{(B.3)}\\
&\quad\le \left(1 - \frac{2}{n(n-1)}\right)\mathrm{D_{HS}}(x_j, x_{j+1})^2 + C\gamma^3.
\end{aligned}$$

Summing over j and applying Jensen's inequality, we have

$$\begin{aligned}
&\mathbb{E}\bigl[\mathrm{D_{HS}}\bigl(\mathrm{F}\bigl(x, \bigl(i, \theta^{(1)}\bigr)\bigr), \mathrm{F}\bigl(y, \bigl(i, \theta^{(k)}\bigr)\bigr)\bigr)\bigr]\\
&\quad= \sum_{j=1}^{k-1} \mathbb{E}\bigl[\mathrm{D_{HS}}\bigl(\mathrm{F}\bigl(x_j, \bigl(i, \theta^{(j)}\bigr)\bigr), \mathrm{F}\bigl(x_{j+1}, \bigl(i, \theta^{(j+1)}\bigr)\bigr)\bigr)\bigr]\\
&\quad\le \sum_{j=1}^{k-1} \sqrt{\mathbb{E}\bigl[\mathrm{D_{HS}}\bigl(\mathrm{F}\bigl(x_j, \bigl(i, \theta^{(j)}\bigr)\bigr), \mathrm{F}\bigl(x_{j+1}, \bigl(i, \theta^{(j+1)}\bigr)\bigr)\bigr)^2\bigr]}\\
&\quad\le \sum_{j=1}^{k-1} \sqrt{\left(1 - \frac{2}{n(n-1)}\right)\mathrm{D_{HS}}(x_j, x_{j+1})^2 + C\gamma^3}\\
&\quad\le (k-1)\sqrt{\left(1 - \frac{2}{n(n-1)}\right)\gamma^2 + C\gamma^3}\\
&\quad\le (k-1)\sqrt{1 - \frac{2}{n(n-1)}}\gamma + (k-1)\sqrt{C}\gamma^{1.5}\\
&\quad\le \sqrt{1 - \frac{2}{n(n-1)}}\mathrm{D_{HS}}(x, y) + (1 + \sqrt{C})\gamma + \sqrt{C\gamma}\mathrm{D_{HS}}(x, y).
\end{aligned}$$

We note that $0 < \gamma < 0.1/C$ above is arbitrary. Thus set $0 < \gamma' < 0.01\omega_n n^{-n^n} \times \min(\frac{1}{C}, \frac{1}{\mathrm{diam}(\mathrm{SO}(n))})$. The above inequality shows that it is possible to couple $i \sim \mathrm{unif}(\{1, 2, \dots, N\})$ and $\theta^{(x)} = \theta^{(1)}, \theta^{(y)} = \theta^{(k)} \sim \mathrm{unif}([0, 2\pi))$ so that

$$\mathbb{E}\big[\mathrm{D_{HS}}\big(\mathrm{F}\big(x, \big(i, \theta^{(x)}\big)\big), \mathrm{F}\big(y, \big(i, \theta^{(y)}\big)\big)\big)\big] \le \sqrt{1 - \frac{2}{n(n-1)}}\mathrm{D_{HS}}(x, y) + \gamma'. \tag{B.4}$$

Fix $x, y \in \mathrm{SO}(n)$, sample $\{i_t\}_{t=1}^{T} \sim \mathrm{unif}(\{1, 2, \dots, N\})$, $\{\theta_t^{(x)}\}_{t=1}^{T} \sim \mathrm{unif}([0, 2\pi)^T)$ and $\{\theta_t^{(y)}\}_{t=1}^{T} \sim \mathrm{unif}([0, 2\pi)^T)$, and let $\{X_t\}_{t=1}^{T}$ and $\{Y_t\}_{t=1}^{T}$ be the Markov chains associated with these starting points and update sequences. For $1 \le t \le T$, let $\mathcal{F}_t$ be the σ-algebra generated by $\{X_t, Y_t\}$. By the Ionescu–Tulcea theorem, the single-step coupling of $\{i, \theta^{(x)}, \theta^{(y)}\}$ that satisfies equation (B.4) can be extended to a coupling $\nu_{\gamma'}$ of $\{i_t\}_{t=1}^{T} \sim \mathrm{unif}(\{1, 2, \dots, N\})$ and $\{\theta_t^{(x)}\}_{t=1}^{T}, \{\theta_t^{(y)}\}_{t=1}^{T} \sim \mathrm{unif}([0, 2\pi)^T)$ with the property that, for all $1 \le t \le T$,

$$\begin{aligned}\mathbb{E}\big[\mathrm{D_{HS}}\big(\mathrm{F}\big(X_{t+1}, \big(i_t, \theta_t^{(x)}\big)\big), \mathrm{F}\big(Y_{t+1}, \big(i_t, \theta_t^{(y)}\big)\big)\big) | \mathcal{F}_t\big] \\ \le \sqrt{1 - \frac{2}{n(n-1)}}\mathrm{D_{HS}}(X_t, Y_t) + \gamma'.\end{aligned}$$

Thus, for all $1 \le t \le T$, we have under this coupling

$$\begin{aligned}\mathbb{E}\big[\mathrm{D_{HS}}(X_{t+1}, Y_{t+1})\big] &\le \mathbb{E}\big[\mathbb{E}\big[\mathrm{D_{HS}}(X_{t+1}, Y_{t+1}) | \mathcal{F}_t\big]\big] \\ &\le \mathbb{E}\bigg[\sqrt{1 - \frac{2}{n(n-1)}}\mathrm{D_{HS}}(X_t, Y_t) + \gamma'\bigg] \\ &\le \cdots \\ &\le \bigg(\sqrt{1 - \frac{2}{n(n-1)}}\bigg)^t \mathrm{D_{HS}}(x, y) + \frac{n(n-1)}{2}\gamma'.\end{aligned}$$

Under the same coupling and this choice of γ', we have for all $t \ge n^2(20A\log(n) - \log(\omega_n))$ that

$$\mathbb{E}\big[\mathrm{D_{HS}}(X_t, Y_t)\big] \le \omega_n n^{-A-5}.$$

Thus, by Markov's inequality, we have for $t \ge n^2(20A\log(n) - \log(\omega_n))$ that

$$\mathbb{P}\big[\|X_t - Y_t\|_{\mathrm{HS}} \le n^{-5}\omega_n\big] \ge 1 - n^{-A} \tag{B.5}$$

under the coupling $\nu_{\gamma'}$. By the standard disintegration theorem (Theorem 5.4 of [40]), there exists a transition kernel κ so that $\nu_{\gamma'} = \mathrm{unif}(\{1, 2, \dots, N\}) \otimes \kappa$. By inequality (B.5), this kernel κ satisfies the requirements of the lemma, completing our proof.

APPENDIX C: PROOFS OF TECHNICAL BOUNDS

We prove the bounds in Section 5.

C.1. Matrix estimates. PROOF OF LEMMA 5.1. We note the "telescoping sum" identity

$$\prod_{i=1}^{k} Q_i - \prod_{i=1}^{k} P_i = \sum_{i=1}^{k} \left(\prod_{\ell=1}^{i-1} Q_\ell \right) (Q_i - P_i) \left(\prod_{\ell=i+1}^{k} P_\ell \right).$$

The result then follows immediately from application of the triangle inequality and the inequality $\|ABC\|_{\mathrm{HS}} \le \|A\|_{\mathrm{Op}} \|B\|_{\mathrm{HS}} \|C\|_{\mathrm{Op}}$ for any $A, B, C \in M(n)$. □

PROOF OF LEMMA 5.2. By Fact 4 in Chapter 15 of [30],

$$\sum_{i=1}^{n} \big(\sigma_i(M_1) - \sigma_i(M_2)\big)^2 \le \sum_{i=1}^{n} \sigma_i(M_1 - M_2)^2.$$

By Assumption (5.2), this implies

$$\sum_{i=1}^{n} \big(\sigma_i(M_1) - \sigma_i(M_2)\big)^2 \le N\delta^2 \sigma_1(M_1)^2.$$

In particular,

$$\max_{1 \le i \le N} \big|\sigma_i(M_1) - \sigma_i(M_2)\big| \le \sqrt{N} \delta \sigma_1(M_1).$$

For a symmetric matrix M, let $\lambda_1(M), \dots, \lambda_N(M)$ denote the eigenvalues, ordered so that $|\lambda_i(M)| = \sigma_i(M)$. We have

$$\begin{aligned}
\left| \frac{\det(M_2)}{\det(M_1)} - 1 \right| &= \left| \frac{\prod_{i=1}^{N} \lambda_i(M_2)}{\prod_{i=1}^{N} \lambda_i(M_1)} - 1 \right| \\
&= \left| \prod_{i=1}^{N} \frac{\lambda_i(M_1) + (\lambda_i(M_2) - \lambda_i(M_1))}{\lambda_i(M_1)} - 1 \right| \\
&= \prod_{i=1}^{N} \left| \frac{(\lambda_i(M_2) - \lambda_i(M_1))}{\lambda_i(M_1)} \right| \\
&\le \prod_{i=1}^{N} \left| \frac{\sqrt{N} \delta \sigma_1(M_1)}{\sigma_1(M_1)} \right| \\
&= N^{\frac{N}{2}} \delta^N
\end{aligned}$$

and the proof is complete. □

PROOF OF LEMMA 5.3. Let $u = (u_1, \dots, u_N)$, $v = (v_1, \dots, v_N)$. By equation (5.4) it follows that

$$\begin{aligned}
\|\mathrm{d} f_u(h) - \mathrm{d} f_v(h)\|_{\mathrm{HS}} &= \Bigg\| \sum_{j=1}^N \prod_{k=1}^N \mathrm{R}_k e^{(\theta_k + u_k) a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=j}\big) \\
&\qquad - \sum_{j=1}^N \prod_{k=1}^N \tilde{\mathrm{R}}_k e^{(\tilde{\theta} k + v_k) a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=j}\big) \Bigg\|_{\mathrm{HS}} \\
&\le \sum_{j=1}^N \Bigg\| \prod_{k=1}^N \mathrm{R}_k e^{(\theta_k + u_k) a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=j}\big) \\
&\qquad - \prod_{k=1}^N \tilde{\mathrm{R}}_k e^{(\tilde{\theta} k + v_k) a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=j}\big) \Bigg\|_{\mathrm{HS}} \\
&\le \sum_{j,k=1}^N \Big\| \mathrm{R}_k e^{(\theta_k + u_k) a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=j}\big) \\
&\qquad - \tilde{\mathrm{R}}_k e^{(\tilde{\theta} k + v_k) a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=j}\big) \Big\|_{\mathrm{HS}} \\
&\le 2N^2 \max\big(\|\mathrm{R}_k - \tilde{\mathrm{R}}_k\|_{\mathrm{HS}}, \|e^{(\theta_k + u_k) a_k} - e^{(\tilde{\theta} k + v_k) a_k}\|_{\mathrm{HS}}\big) \\
&\le 4N^2 \max_{1 \le j \le N} \max\big(|u_k|, |v_k|, |\theta_k - \tilde{\theta} k|\big) \le 4N^2 c,
\end{aligned}$$

where the third and fourth lines are both applications of Lemma 5.1. Applying Lemma 5.1 once more, we have

$$\begin{aligned}
\|\mathrm{dL}_{f(v)(f(u))^{-1}} \mathrm{d} f_u(h) - d f_v(h)\|_{\mathrm{HS}} &\le \|f(v) - f(u)\|_{\mathrm{HS}} + \|\mathrm{d} f_u(h) - \mathrm{d} f_v(h)\|_{\mathrm{HS}} \\
&\le 8N^2 c,
\end{aligned}$$

the second part of inequality (5.5). This completes the proof. □

PROOF OF LEMMA 5.4. Let $u = (u_1, \dots, u_N) \in [-c, c]^N$. We calculate

$$\begin{aligned}
&\|f(u) - f(0) \exp\big(\mathrm{dL}_{f(0)^{-1}} \mathrm{d} f_0(u)\big)\|_{\mathrm{HS}} \\
&\quad = \Bigg\| \prod_{k=1}^N \mathrm{R}_k e^{(\theta_k + u_k) a_k} \\
&\qquad - \prod_{k=1}^N \mathrm{R}_k e^{\theta_k a_k} \exp\Bigg(\Bigg(\prod_{k=1}^N \mathrm{R}_k e^{\theta_k a_k}\Bigg)^{-1}
\end{aligned}$$

$$
\begin{aligned}
&\quad \times \sum_{i=1}^{N}\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k}\bigl(\mathrm{Id}+(h_k a_k-\mathrm{Id})1_{k=i}\bigr)\Bigg)\Bigg\|_{\mathrm{HS}} \\
&= \Bigg\| \prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k} \\
&\quad - \prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k} \sum_{m=0}^{\infty} \frac{1}{m!}\Bigg(\Bigg(\prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k}\Bigg)^{-1} \\
&\quad \times \sum_{i=1}^{N}\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k}\bigl(\mathrm{Id}+(h_k a_k-\mathrm{Id})1_{k=i}\bigr)\Bigg)^m \Bigg\|_{\mathrm{HS}} \\
&= \Bigg\| \prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k} \sum_{m=0}^{\infty} \frac{(u_k a_k)^m}{m!} \\
&\quad - \prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k} \sum_{m=0}^{\infty} \frac{1}{m!}\Bigg(\Bigg(\prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k}\Bigg)^{-1} \\
&\quad \times \sum_{i=1}^{N}\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k}\bigl(\mathrm{Id}+(h_k a_k-\mathrm{Id})1_{k=i}\bigr)\Bigg)^m \Bigg\|_{\mathrm{HS}} \\
&\le 8N^2 \max_{1\le k\le N} |u_k|^2 \\
&\quad + \Bigg\| \sum_{i=1}^{N}\prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k}\bigl(\mathrm{Id}+(h_k a_k-\mathrm{Id})1_{k=i}\bigr) \\
&\quad - \sum_{i=1}^{N}\prod_{k=1}^{N} \mathrm{R}_k e^{\theta_k a_k}\bigl(\mathrm{Id}+(h_k a_k-\mathrm{Id})1_{k=i}\bigr)\Bigg\|_{\mathrm{HS}} \\
&= 8N^2 \max_{1\le k\le N} |u_k|^2 \le 8N^2 c^2,
\end{aligned}
$$

where the second-last line relies on the triangle inequality and repeated application of Lemma 5.1 to remove all terms that are of second or higher order in $\{a_k\}_{1\le k\le N}$. This completes the proof of the lemma. $\square$

PROOF OF LEMMA 5.5. By equation (5.4),

$$
\begin{aligned}
\langle \mathrm{d}f_u(h), \mathrm{d}f_u(h')\rangle_{\mathrm{HS}} &= \mathrm{Tr}\bigl[\mathrm{d}f_u(h)\mathrm{d}f_u(h')^{\dagger}\bigr] \\
&= \mathrm{Tr}\Bigg[\Bigg(\sum_{i=1}^{N}\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k}\bigl(\mathrm{Id}+(h_k a_k-\mathrm{Id})1_{k=i}\bigr)\Bigg)
\end{aligned}
$$

$$\times \left(\sum_{i=1}^{N} \prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k} \big(\mathrm{Id} + (h'_k a_k - \mathrm{Id}) 1_{k=i}\big) \right)^{\dagger} \Bigg]$$

$$\begin{aligned} &= \sum_{i=1}^{N} \mathrm{Tr}\Bigg[\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=i}\big) \\ &\qquad \times \prod_{k=N}^{1} \big(\mathrm{Id} - (h'_k a_k - \mathrm{Id}) 1_{k=i}\big) e^{-(\theta_k+u_k)a_k} \mathrm{R}_k^{-1} \Bigg] \\ &\qquad + \sum_{1 \le i \ne j \le N} \mathrm{Tr}\Bigg[\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=i}\big) \\ &\qquad \times \prod_{k=N}^{1} \big(\mathrm{Id} - (h'_k a_k - \mathrm{Id}) 1_{k=j}\big) e^{-(\theta_k+u_k)a_k} \mathrm{R}_k^{-1} \Bigg] \\ &\equiv \sum_{i=1}^{N} S_i + \sum_{1 \le i \ne j \le N} S_{ij}. \end{aligned} \tag{C.1}$$

We calculate terms of the form S_j and S_{ij} separately. For any S_i, applying the cyclic permutation property of the trace operator yields

$$\begin{aligned} S_i &= \mathrm{Tr}\Bigg[\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=i}\big) \\ &\qquad \times \prod_{k=N}^{1} \big(\mathrm{Id} - (h'_k a_k - \mathrm{Id}) 1_{k=i}\big) e^{-(\theta_k+u_k)a_k} \mathrm{R}_k^{-1} \Bigg] \\ &= \mathrm{Tr}\big[\mathrm{R}_i (h_i a_i)(-h'_i a_i) \mathrm{R}_i^{-1}\big] \\ &= -h_i h'_i \,\mathrm{Tr}\big[\mathrm{R}_i a_i^2 \mathrm{R}_i^{-1}\big] = -h_i h'_i \,\mathrm{Tr}\big[a_i^2\big] = h_i h'_i. \end{aligned} \tag{C.2}$$

For any S_{ij} with $i < j$,

$$\begin{aligned} S_{ij} &= \mathrm{Tr}\Bigg[\prod_{k=1}^{N} \mathrm{R}_k e^{(\theta_k+u_k)a_k} \big(\mathrm{Id} + (h_k a_k - \mathrm{Id}) 1_{k=i}\big) \\ &\qquad \times \prod_{k=N}^{1} \big(\mathrm{Id} - (h'_k a_k - \mathrm{Id}) 1_{k=j}\big) e^{-(\theta_k+u_k)a_k} \mathrm{R}_k^{-1} \Bigg] \\ &= -h_i h'_j \,\mathrm{Tr}\big[a_i M_{ij} \mathrm{R}_j e^{(\theta_j+u_j)a_j} a_j e^{-(\theta_j+u_j)a_j} \mathrm{R}_j^{-1} M_{i,j}^{-1}\big] \\ &= -h_i h'_j \,\mathrm{Tr}\big[a_i M_{i,j} \mathrm{R}_j a_j \mathrm{R}_j^{-1} M_{i,j}^{-1}\big] = h_i h'_j D_{ij}. \end{aligned} \tag{C.3}$$

For $j < i$, a similar calculation gives

$$S_{ij} = h_i h'_j D_{i,j}.$$

Combining equalities (C.1), (C.2) and (C.3) completes the proof. □

PROOF OF LEMMA 5.6. We note that f is clearly smooth, and so we must only check that it is bijective. This result will follow almost immediately from Lemma 10 of [49] and our bounds in Lemmas 5.4 and 5.5.

We begin to set up notation. For a point x in a metric space (Ω, d) and constant $\delta > 0$, let $\mathcal{B}_\delta(x) = \{y \in \Omega : d(x, y) \le \delta\}$ be the ball of radius δ around x. We then define a map $\mathcal{F}$ from $\mathcal{B}_{n^{-6}\phi_n}(\mathrm{Id}) \subset \mathrm{SO}(n)$ to $\mathbb{R}^N$ as follows. Let $x \in \mathcal{B}_{n^{-6}\phi_n}(\mathrm{Id})$. Since the exp map is surjective and sends lines to geodesic curves, we can write $x = \exp(h)$ for some $h \in \mathfrak{so}(n)$ with $\|h\|_{\mathrm{HS}} \le 2n^{-6}\phi_n$. Furthermore, we can write $h = \sqrt{2}\sum_{i=1}^N h_i a_i$. We then define $\mathcal{F}(x) = (h_1, h_2, \dots, h_N)$. Finally, we define the map $g = \mathcal{F} \circ f : [-c, c]^N \mapsto \mathbb{R}^N$.

We point out that $\mathcal{F}$ has small distortion: for $x, y \in \mathrm{SO}(n)$ with $\mathcal{F}(x) = h_x$, $\mathcal{F}(y) = h_y$,

$$\|x - y\|_{\mathrm{HS}} = \|\exp(h_x) - \exp(h_y)\|_{\mathrm{HS}} = \|h_x - h_y\| + O(N^2\|h_x - h_y\|^2). \tag{C.4}$$

We now obtain the estimates required to use Lemma 10 of [49]. Following the notation of that paper, we set $\rho = \frac{1}{256}n^{-6}\phi_n$, $\delta = \frac{\phi_n}{8}$ and $\rho_* = \frac{\delta}{8}\rho$. For $x, z \in [-c, c]^N$ with $\|x - z\| = \rho$,

$$\begin{aligned}
\|g(x) - g(z)\|_{\mathrm{HS}} &= \big\|g(x) - \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)\big) \\
&\quad + \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)\big) \\
&\quad + \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\big)\big) \\
&\quad - \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\big)\big) - g(z)\big\|_{\mathrm{HS}} \\
&\ge \big\|\mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)\big) \\
&\quad - \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\big)\big)\big\|_{\mathrm{HS}} \\
&\quad - \big\|\mathcal{F}\big(f(x)\big) - \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)\big)\big\|_{\mathrm{HS}} \\
&\quad - \big\|\mathcal{F}\big(f(z)\big) - \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\big)\big)\big\|_{\mathrm{HS}} \qquad \text{(C.5)} \\
&\ge \big\|\mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)\big) \\
&\quad - \mathcal{F}\big(f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\big)\big)\big\|_{\mathrm{HS}} \\
&\quad - 2\big\|f(x) - f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)\big\|_{\mathrm{HS}} \\
&\quad - 2\big\|f(z) - f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\big)\big\|_{\mathrm{HS}} \\
&\ge \big\|f(0)\exp\big(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\big)
\end{aligned}$$

$$
\begin{aligned}
&- f(0)\exp\bigl(\mathrm{d}L_{f(0)^{-1}}\mathrm{d}f_0(z)\bigr)\bigr\|_{\mathrm{HS}} \\
&- 32N^2c^2,
\end{aligned}
$$

where the second-last inequality follows from inequality (C.4) and Lemma 5.4, and the last inequality is due to Lemma 5.4. By inequalities (5.8) and (C.4),

$$
\begin{aligned}
&\bigl\|\mathcal{F}\bigl(f(0)\exp\bigl(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(x)\bigr)\bigr) - \mathcal{F}\bigl(f(0)\exp\bigl(\mathrm{dL}_{f(0)^{-1}}\mathrm{d}f_0(z)\bigr)\bigr)\bigr\|_{\mathrm{HS}} \\
&\quad \geq \frac{\phi_n}{4}\|x-z\| - O\bigl(N^3\|x-z\|^2\bigr).
\end{aligned}
\tag{C.6}
$$

Combining inequalities (C.5) and (C.6), we conclude

$$
\bigl\|g(x) - g(z)\bigr\|_{\mathrm{HS}} \geq \frac{\phi_n}{8}\|x-z\|.
$$

This proves the first condition of Lemma 10 of [49]: $\{\|x-z\|_{\mathrm{HS}} > \rho\}$ implies that $\{\|g(x)-g(z)\| > \delta\rho\}$. The second condition of Lemma 10 of [49] follows immediately from the second part of inequality (5.8). Thus g satisfies the requirements of Lemma 10 of [49] with ρ, ρ_*, δ as above, and so g is an injective map. But this implies that f is injective as well, completing the proof. □

C.2. Probability estimates. PROOF OF LEMMA 5.7. By observation, the random variable $s_N - N\lceil Qn^2\log(n)\rceil$ is stochastically dominated by a negative binomial distribution with parameters (N, N^{-1}). The desired inequality then follows immediately from the standard tail bound for the negative binomial distribution (see, e.g., the calculation in [5]). □

PROOF OF LEMMA 10.2. Note that s_N is exactly the time it takes to collect all N coupons in the standard "coupon collector problem" with N coupons, as studied in [19]. Thus the desired bound follows immediately from equation (2) of [19]. □

Acknowledgements. We thank Tristan Collins, Persi Diaconis, John Jiang, Federico Poloni, Terence Tao and Roman Vershynin for helpful conversations and suggestions. We also thank the referees for making detailed suggestions, leading to a substantially revised final paper. Among other changes, this included a new emphasis on a "perturbative" as opposed to "pathwise" coupling in Theorem 4.3, as well as the new inclusion of most of the details in Appendices A and B.

REFERENCES

[1] ABBASPOUR, H. and MOSKOWITZ, M. (2007). *Basic Lie Theory*. World Scientific, Hackensack, NJ. MR2364699

[2] AILON, N. and CHAZELLE, B. (2006). Approximate nearest neighbors and the fast Johnson–Lindenstrauss transform. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing* 557–563. ACM, New York. MR2277181

[3] AJTAI, M., KOMLOS, J. and SZEMERÉDI, E. (1982). Largest random component of a k-cube. *Combinatorica* **2** 1–7.

[4] BASAK, A. and RUDELSON, M. (2017). Invertibility of sparse non-Hermitian matrices. *Adv. Math*. **310** 426–483.

[5] BROWN, D. How I wasted too long finding a concentration inequality for sums of geometric variables. Available at https://cs.uwaterloo.ca/~browndg/negbin.pdf.

[6] BUBLEY, R. and DYER, M. (1997). Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science* 223–231. IEEE, Los Alamitos, CA.

[7] 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.

[8] 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

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

[10] CASELLA, G. and GEORGE, E. I. (1992). Explaining the Gibbs sampler. *Amer. Statist*. **46** 167–174. MR1183069

[11] CONGER, M. and VISWANATH, D. (2006). Shuffling cards for blackjack, bridge, and other card games. Preprint. Available at arXiv:math/0606031.

[12] CORTEZ, R. and FONTBONA, J. (2016). Quantitative propagation of chaos for generalized Kac particle systems. *Ann. Appl. Probab*. **26** 892–916. MR3476628

[13] CZUMAJ, A. and VÖCKING, B. (2014). Thorp shuffling, butterflies, and non-Markovian couplings. In *Automata, Languages, and Programming* 344–355. Springer, Berlin.

[14] DEN HOLLANDER, F. (2012). Probability theory: The coupling method. Unppublished manuscript. Available at http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf.

[15] DIACONIS, P. (1988). *Group Representations in Probability and Statistics. IMS Lecture Notes Monogr. Ser.* **11**. Inst. Math Statist., Hayward, CA. MR0964069

[16] DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci*. **23** 151–178.

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

[18] DONALDSON, S. K. (2011). Lectures on Lie groups and geometry. Lecture notes. Available at http://wwwf.imperial.ac.uk/~skdona/LIEGROUPSCONSOL.PDF.

[19] ERDÖS, P. and RÉNYI, A. (1961). On a classical problem of probability theory. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl*. **6** 215–220.

[20] FARRELL, B. and VERSHYNIN, R. (2016). Smoothed analysis of symmetric random matrices with continuous distributions. *Proc. Amer. Math. Soc*. **144** 2257–2261. MR3460183

[21] FONTES-MERZ, N. (2006). A multidimensional version of Turán's lemma. *J. Approx. Theory* **140** 27–30. MR2226674

[22] FRIEDLAND, O. and GILADI, O. (2013). A simple observation on random matrices with continuous diagonal entries. *Electron. Commun. Probab*. **18** 1–7. MR3078016

[23] FÜREDI, Z. and LOEB, P. A. (1994). On the best constant for the Besicovitch covering theorem. *Proc. Amer. Math. Soc*. **121** 1063–1073.

[24] FURMAN, A. (2002). Random walks on groups and random transformations. In *Handbook of Dynamical Systems, Vol. 1A* 931–1014. North-Holland, Amsterdam. MR1928529

[25] GOLDMAN, H., POOLE, C. and SAFKO, J. (2001). *Classical Mechanics*, 3rd ed. Addison-Wesley, Reading, MA.

[26] GUILLEMAN, V. and POLLACK, A. (1974). *Differential Topology*. Prentice-Hall, Englewood Cliffs, NJ.

[27] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437

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

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

[30] HOGBEN, L., ed. (2014). *Handbook of Linear Algebra*, 2nd ed. CRC Press, Boca Raton, FL. MR3013937

[31] HOPF, H. and RINOW, W. (1931). Ueber den Begriff der vollständigen differentialgeometrischen Fläche. *Comment. Math. Helv.* **3** 209–225. MR1509435

[32] HOUGH, B. and JIANG, Y. (2017). Cut-off phenomenon in the uniform plane Kac walk. *Ann. Probab.* **45** 2248–2308. MR3693962

[33] JAMES, A. T. (1955). A generating function for averages over the orthogonal group. *Proc. R. Soc. Lond. Ser. A* **229** 367–375. MR0070908

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

[35] JANVRESSE, É. (2003). Bounds on semigroups of random rotations on $\mathrm{SO}(n)$. *Theory Probab. Appl.* **47** 526–532.

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

[37] JIANG, Y. (2017). Kac's random walk on the special orthogonal group mixes in polynomial time. *Proc. Amer. Math. Soc.* **145** 4533–4541. MR3690635

[38] JONES, G. L. and HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447

[39] 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, CA. MR0084985

[40] KALLENBERG, O. (2006). *Foundations of Modern Probability*. Springer, New York.

[41] LOVÁSZ, L. (1999). Hit-and-run mixes fast. *Math. Program.* **86** 443–461. MR1733749

[42] LOVÁSZ, L. and VEMPALA, S. (2003). Hit-and-run is fast and fun. Technical report. Available at https://www.microsoft.com/en-us/research/publication/hit-and-run-is-fast-and-fun/.

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

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

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

[46] PAK, I. and SIDENKO, S. (2007). Convergence of Kac's random walk. Preprint. Available at http://www-math.mit.edu/~pak/research.html.

[47] PILLAI, N. S. and SMITH, A. (2017). Kac's walk on n-sphere mixes in $n \log n$ steps. *Ann. Appl. Probab.* **27** 631–650. MR3619797

[48] POROD, U. (1996). The cut-off phenomenon for random reflections. *Ann. Probab.* **24** 74–96.

[49] RAYMOND, J. S. (2002). Local inversion for differentiable functions and the Darboux property. *Mathematika* **49** 141–158.

[50] ROSENTHAL, J. S. (1994). Random rotations: Characters and random walks on $\mathrm{SO}(n)$. *Ann. Probab.* **22** 398–423. MR1258882

[51] SALOFF-COSTE, L. (2004). Random walks on finite groups. In *Probability on Discrete Structures* 263–346. Springer, Berlin. MR2023654

[52] SEPEHRI, A. (2016). New tests of uniformity on the compact classical groups as diagnostics for weak-star mixing of Markov chains. Preprint. Available at arXiv:1612.03233.

[53] 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, ONTARIO K1N 7N5
CANADA
E-MAIL: asmi28@uottawa.ca