Mixing Times and Kac's Walks

Aaron Smith

Department of Mathematics and Statistics, uOttawa

29 July 2017

Co-Author and Thanks

Natesh Pillai

Thanks also: Tristan Collins, Persi Diaconis, John Jiang, Federico Poloni, Terence Tao and Roman Vershynin.

Overview

- **Goal:** Mixing bounds for specific chains of historical interest - Kac's walks.
- **Problem:** Going from weak to strong mixing on interesting state spaces.
- **General(?) Technique:** Strategies for building coupling.

Main Examples: Kac's Walks

1. Kac's walk $\{X_t\}_{t \in \mathbb{N}} \in \mathbb{SO}(n)$.
2. To get X_{t+1} from X_t,
   1. Choose $1 \leq i(t) < j(t) \leq n$, $\theta(t) \in [0, 2\pi)$ uniformly.
   2. Multiply X_t by rotation $R(i(t), j(t), \theta(t))$ of $\theta(t)$ degrees in $(i(t), j(t))$-plane.
3. Kac's walk on sphere: just take first column of walk on $\mathbb{SO}(n)$.

Next: three stories about Kac's walks.

Balls in a Box

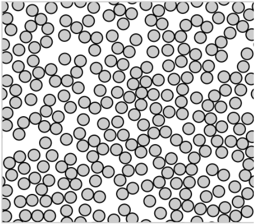


Figure: Hard Balls in a Box

How do their velocities evolve as particles collide?

Balls in a Box - a Condensed History

1. Few balls, exact solution: Newton's equations of motion (1687).
2. Many balls, "statistical" solution: Boltzmann's equation (1872).
3. Natural question (Hilbert's 6'th problem, 1900): how to "derive" Boltzmann's equation?

Kac's Program

- Kac (1956) proposes Kac's walk as simplest version of problem; gives informal derivation of the Boltzmann equation.
- **Key technical issue:** To rigorize argument, need to check that Kac's model *equilibrates quickly.*
- Finishing Kac's argument is still open, even though Kac's original conjecture was proved by Janvresse (2001).

MCMC in Statistics

- We all know: MCMC is very popular; works iff Markov chains mix quickly.
- Historical trivia: "biggest" example in original MCMC paper (Hastings, 1970) is Kac's walk on $\mathbb{SO}(n)$!
- Given big mixing time literature - would be nice to know the mixing time of first interesting example from statistics!

Walks on Groups, Manifolds, etc

- Large literature on “conjugacy-invariant” walks on groups.
- **Technical warmups for $\mathbb{SO}(n)$:** Conjugacy-invariant analogues to Kac’s walk has been studied by Matthews, Rosenthal, Porod, Jiang and Hough (1985-2017).
- **Analogies and predictions for sphere:** Behaviour “similar” to (famous, conjugacy-invariant) random transposition walk on S_n.

NEXT: Previous results.

Best Published Orders: Kac's Walk on Sphere

- Inverse spectral gap bounded by Janvresse (2001):

$$\lambda(K)^{-1} \approx n.$$

- Wasserstein mixing estimated by Oliveira (2009):

$$\tau_{\text{mix}}^{(W_2)} \lesssim n^2 \log(n).$$

- Total variation bounded by Jiang (2012):

$$n \leq \tau_{\text{mix}} \lesssim n^5 \log(n)^3$$

Best Published Orders: Kac's Walk on $\mathbb{SO}(n)$

- Inverse spectral gap estimated by Janvresse (2003):

$$\lambda(K)^{-1} \approx n^2.$$

- Wasserstein mixing estimated by Oliveira (2009):

$$\tau_{\mathrm{mix}}^{(W_2)} \lesssim n^2 \log(n).$$

- Total variation bounded by Diaconis/Saloff-Coste (2001):

$$n^2 \leq \tau_{\mathrm{mix}} \lesssim e^{n^2}.$$

(See also unpublished work of Jiang).

Main Results (Pillai/S. 2016; Pillai/S. preprint)

Mixing of Kac's Walk on the Sphere

The mixing time of Kac's walk on the sphere satisfies

$$\frac{1}{2} n \log(n) \leq \tau_{\mathrm{mix}} \leq 200 n \log(n).$$

Mixing of Kac's Walk on $\mathbb{SO}(n)$

The mixing time of Kac's walk on $\mathbb{SO}(n)$ satisfies

$$\frac{n(n-1)}{2} \leq \tau_{\mathrm{mix}} \leq 10^7 n^4 \log(n).$$

Next: Heuristics and proof approach.

Heuristics

- Expect

$$\tau_{\mathrm{mix}}^{(\mathrm{TV})} \approx \tau_{\mathrm{mix}}^{(W_2)} \approx \tau_{\mathrm{rel}}.$$

- Whole spectrum known; good bounds on Wasserstein mixing.
- How to transfer to TV mixing?
- **NEXT:** a standard approach.

Continuity and Mixing 1

Many authors

Let transition kernel K satisfy "continuity condition"

$$\{\|x-y\| < \epsilon\} \implies \{\|K^{\ell}(x,\cdot) - K^{\ell}(y,\cdot)\|_{\mathrm{TV}} < 0.1\}$$

for some $\epsilon > 0$, $\ell \in \mathbb{N}$. Then

$$\tau_{\mathrm{mix}}^{\mathrm{TV}} \lesssim \ell + \tau_{\mathrm{mix}}^{(W_2)}(\epsilon).$$

Proof: "one-shot" Coupling.

Continuity and Mixing 2

Kac's walk, and other Gibbs samplers, *fail* the continuity assumption for ℓ moderately large:

Trivial Observation

For $\ell < \frac{n(n-1)}{2}$ and all $\epsilon > 0$, there exist $x, y \in \mathbb{SO}(n)$ so that

$$\|x - y\| < 2\epsilon, \qquad \|K^{\ell}(x, \cdot) - K^{\ell}(y, \cdot)\|_{\mathrm{TV}} = 1.$$

Main technical problem in talk: how to compare τ_{mix}, $\tau_{\mathrm{mix}}^{(W_2)}$ without obvious continuity?
Approach: More complicated coupling.

Coupling Notation

GOAL: Force two chains $\{X_t\}_{t\geq 0}$, $\{Y_t\}_{t\geq 0}$ to collide.

- Our chains defined in terms of i.i.d. sequences of *update variables* $i(t), j(t), \theta(t)$.
- In this setting, coupling Markov chains is equivalent to coupling sequences of update variables.
- In sequel, we use superscripts $i(t)^{(x)}, i(t)^{(y)}$ to denote coupled update sequences.
- **Next:** coupling arguments for Kac's walk on sphere.

Naive Coupling for Kac's Walk on the Sphere 1

- **General wish for Gibbs samplers:** try to update so that updated variables agree - *i.e.*

$$X_{t+1}[i(t)] = Y_{t+1}[i(t)]$$
$$X_{t+1}[j(t)] = Y_{t+1}[j(t)].$$

- **Immediate Problem:** even if X_t, Y_t are arbitrarily close, can only choose *one* of these equations to satisfy.
- **Question:** can *any* step-by-step "greedy" coupling work?

Naive Coupling for Kac's Walk on the Sphere 2

A coupling is *Markovian* if the joint process $\{X_t, Y_t\}_{t \in \mathbb{N}}$ is also a Markov chain.

Inefficiency of Markovian Couplings

For *any* Markovian coupling of Kac's walk,

$$\mathbb{P}[\tau_{\text{coup}} < t] \leq \frac{2t}{n(n-1)}.$$

Inefficiency Proof

$$\begin{aligned}
\mathbb{P}[\tau_{\mathrm{coup}} = t+1] &= \mathbb{E}[\mathbb{P}[\tau_{\mathrm{coup}} = t+1 | \{X_s^2, Y_s^2\}_{s \leq t}]] \\
&= \mathbb{E}[\mathbb{P}[\tau_{\mathrm{coup}} = t+1 | X_t^2, Y_t^2]] \\
&\leq \mathbb{E}[\max_{x^2 \neq y^2} \mathbb{P}[\tau_{\mathrm{coup}} = t+1 | X_t^2 = x^2, Y_t^2 = y^2]] \\
&= \frac{2}{n(n-1)}.
\end{aligned}$$

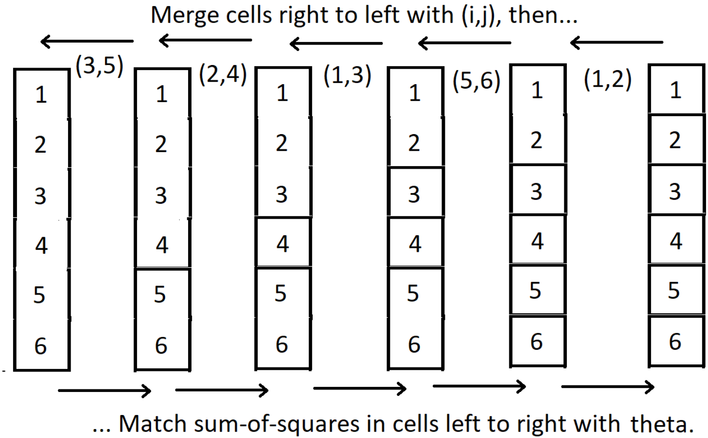
Thus for all times t and all $X_0^2 \neq Y_0^2$ fixed,

$$\mathbb{P}[\tau_{\mathrm{coup}} \leq t] = \sum_{s=1}^{t} \mathbb{P}[\tau_{\mathrm{coup}} = s] \leq \frac{2t}{n(n-1)}.$$

Building a Good Coupling

- **Plan:** As always, force two chains to "agree" more and more.
- **Problem:** Markovian greedy couplings can't work.
- **Revised plan:** Construct greedy coupling that "looks into the future."

Simplest Forward-Looking Coupling



Facts About Greedy Forward-Looking Coupling

1. There exists a “greedy” attempted coupling, and it is unique.
2. Distance $\max_{0 \leq t \leq T} \|X_t - Y_t\|_2 \leq 2n^2 \|X_0 - Y_0\|_2$ w.h.p. (non-obvious!).
3. Everything works fine.
4. **Problem:** This doesn't generalize well to more complicated manifolds, including $\mathbb{SO}(n)$.

Notation: Random Mappings and Perturbations

- For $T \in \mathbb{N}$, set random mapping

$$G_T(X_0, \{i(t), j(t), \theta(t)\}_{t=0}^{T}) \equiv X_T.$$

- Consider small perturbation

$$\tilde{\theta}(t) = \theta(t) + \delta(t)$$
$$F_T(\delta(0), \ldots, \delta(T)) \equiv G_T(X_0, \{i(t), j(t), \tilde{\theta}(t)\}_{t=0}^{T}).$$

- For $\delta(t) \sim \mathrm{Unif}[-\epsilon \mathbf{1}_{t \in S}, -\epsilon \mathbf{1}_{t \in S}]$ small, have linear approximation

$$F_T(\delta(0), \ldots, \delta(T)) \approx F_T(0)\, e^{J_T\,(\delta_0, \ldots, \delta_T)},$$

where J_T is the Jacobian of F_T at 0.

Greedy Couplings for Perturbation

- Since $\delta^{(z)}(t)$ are uniform, get high coupling probability if

$$\frac{|F_T^{(x)}([-\epsilon,\epsilon]^T) \cap F_T^{(y)}([-\epsilon,\epsilon]^T)|}{|F_T^{(x)}([-\epsilon,\epsilon]^T)|} \approx 1$$

  and Jacobians of $F_T^{(x)}$, $F_T^{(y)}$ roughly constant.
- From heuristic, occurs if singular values satisfy

$$\sigma_1(J_T^{(x)}) \approx \sigma_1(J_T^{(y)}) \gg \|F_T^{(x)}(0) - F_T^{(y)}(0)\|.$$

- With appropriate technical conditions, **this gives generic continuity lemma.**

Conclusions:

- Reduced problem to estimating the smallest singular value of the random matrix $J_T^{(x)}$ (plus some easy estimates).
- Good news: there is a large literature on bounding the smallest singular value of random matrices.
- Bad news: all of it assumes matrices with far more independence than $J_T^{(x)}$.
- Current state: obtain bound that is far worse than conjecture, much better and more general than "immediate" bound via Turan's inequality.
- **Time permitting:** a bit about random matrices.

Representative Results

Farrell/Vershynin 2015

Let M be an n by n random matrix with independent entries, with density less than 1. Then

$$\mathbb{P}[\sigma_1(M) \leq \frac{1}{16n^2}] \leq \frac{1}{2}.$$

Friedland/Giladi 2013

With only independent diagonals,

$$\mathbb{P}[\sigma_1(M) \leq (5n)^{-n}] \leq \frac{1}{2}.$$

Ours: high-dimensional dependence; unbounded density; hard constraints...

Ugly Formula for $J_T^{(x)}$

For $1 \leq i < j \leq \frac{n(n-1)}{2}$,

$$J_T^{(x)}[i,j] = \mathrm{Tr}[a_i M_{i,j} R_j a_j R_j^{-1} M_{i,j}^{-1}]$$

where

$$R_k = \prod_{s=S(k)+1}^{S(k+1)-1} R(s)^{(x)}, \; M_{i,j} = \prod_{k=j+1}^{i-1} (R_k \, e^{\tilde{\theta}(S(k))^{(x)} \, a_k})$$

and $\{a_k\}_{k=1}^{\frac{n(n-1)}{2}}$ is basis of $\mathbb{T}_0 \, \mathbb{SO}(n)$.
In end, obtain conclusion similar to Friedland/Giladi.

Right Answers: Computing, Cutoff, Conjectures

- Since state space is continuous, it is not obvious a priori how to obtain any sensible simulated bound on mixing time.

Right Answers: Computing, Cutoff, Conjectures

- Since state space is continuous, it is not obvious a priori how to obtain any sensible simulated bound on mixing time.
- Possible to sample distribution of $\sigma_1(J_T^{(x)})$ by computer; coupling proofs relate this to mixing times.
- Simulation gives empirical evidence for conjecture that mixing time on $\mathbb{SO}(n)$ is $O^*(n^2)$.
- See Ph.D. thesis of Amir Sepehri for additional confirmation, including conjectured cutoff windows for both walks.