J. Appl. Probab. **59,** 860–879 (2022)
*doi:*10.1017/jpr.2021.96

THE HIT-AND-RUN VERSION OF TOP-TO-RANDOM

SAMUEL BOARDMAN,* *Cornell University*
DANIEL RUDOLF,** *University of Passau*
LAURENT SALOFF-COSTE,*** *Cornell University*

Abstract

We study an example of a *hit-and-run* random walk on the symmetric group $\mathbf{S}_n$. Our starting point is the well-understood *top-to-random* shuffle. In the hit-and-run version, at each *single step*, after picking the point of insertion j uniformly at random in $\{1, \ldots, n\}$, the top card is inserted in the jth position k times in a row, where k is uniform in $\{0, 1, \ldots, j-1\}$. The question is, does this accelerate mixing significantly or not? We show that, in L^2 and sup-norm, this accelerates mixing at most by a constant factor (independent of n). Analyzing this problem in total variation is an interesting open question. We show that, in general, hit-and-run random walks on finite groups have non-negative spectrum.

Keywords: Markov chain; card shuffling; cut-off phenomenon

2020 Mathematics Subject Classification: Primary 60J10; 60B10
Secondary 60B15; 60G51

1. Introduction

Given a finite group and a generating k-tuple, consider the simple random walk on G associated with this k-tuple. At each integer time, this random walk moves from the current position X_n to $X_n g$, where g is picked uniformly at random among the k generators, independently of all previous steps. To define the *hit-and-run walk* based on the same generating k-tuple, for any group element g, call m_g the order (i.e. exponent) of g. At each step, pick one of the k generators uniformly at random, call it g, pick ℓ uniformly in $\{0, \ldots, m_g - 1\}$, and move to $X_n g^\ell$.

This defines a natural variation on simple random walks which allows for long jumps when the orders of some of the generators are relatively large. As often in the study of random walks on finite groups, it is easier to think about the problem for a family of random walks on a sequence of finite groups whose sizes increase to infinity.

Two of the most basic questions one can ask concerning a family of ergodic random walks on some finite groups whose sizes increase to infinity are: How long does the walk take to be approximately uniformly distributed? Does the cut-off phenomenon occur? That is, is there a rapid transition from being far from equilibrium to reaching approximate equilibrium?

Received 18 March 2021; revision received 20 October 2021.
* Postal address: 530 Church St, 2074 East Hall, Ann Arbor, MI 48109, USA. Email address: stb89@cornell.edu
** Postal address: Fakultät für Informatik und Mathematik, Innstraße 33, 94032 Passau, Germany. Email address: daniel.rudolf@uni-passau.de
*** Postal address: 567 Malott Hall, Department of Mathematics, Ithaca, NY 14853, USA. Email address: lps2@cornell.edu

See [1], [5], and [6] for introductions to these problems. In the context of hit-and-run random walks, the following additional question emerges: Does the hit-and-run version converge faster than the simple random walk version? That is, does the extra randomization help, and if so, how much?

We study these problems in the case of the hit-and-run walk associated with one of the classic random walks on the symmetric group, top-to-random. See Example 1.2 below. We show that if convergence is measured in L^2, the hit-and-run walk and the original top-to-random walk both take order $n \log n$ to converge. What exactly happens to the hit-and-run walk in total variation is left as an open question, but it seems plausible that, again, it takes order $n \log n$ to converge, as top-to-random does [1, 5]. We give an analysis of the Markov chain consisting in following a fixed single card. While studying this example and based on some numerical evidence, the first and last authors conjectured that the hit-and-run top-to-random walk had only non-negative eigenvalues. The second author provided a proof of this fact, and more, based on earlier works on hit-and-run algorithms [11]: for any generating tuple on any finite group, the associated hit-and-run walk has non-negative spectrum. See Theorem 1.1 and Section 4.

1.1. Random walks based on generating k-tuples

Let G be a finite group with identity element e. For any generating k-tuple $S = (g_1, \dots, g_k)$, let μ_S be the probability measure

$$\mu_S = \frac{1}{k} \sum_{i=1}^{k} \delta_{g_i}, \quad \delta_g(h) = \begin{cases} 1 & \text{if } h = g, \\ 0 & \text{otherwise.} \end{cases}$$

The random walk on the group G driven by the measure μ_S above or any probability measure μ, for that matter, is the Markov chain with state space G and Markov kernel

$$M(x, y) = \mu(x^{-1}y).$$

The uniform measure $u = u_G$ on G is always invariant for such a Markov chain and it is useful to consider the (convolution) operator

$$f \mapsto Mf(x) = \sum_y M(x, y) f(y)$$

acting on $L^2 = L^2(G, u)$. At any (discrete) time t, the iterated kernel $M^t(x, y)$ is given by the t-fold convolution $\mu^{(t)}$ of μ by itself in the form $M^t(x, y) = \mu^{(t)}(x^{-1}y)$. The adjoint M^* of M satisfies $M = M^*$ if and only if μ is symmetric in the sense that $\check{\mu}(x) = \mu(x^{-1}) = \mu(x)$.

Example 1.1. The following examples on the symmetric group $\mathbf{S}_n$ will be of particular interest to us. See [1], [2], [3], [5], [7], [8], [9], and [13].

- (Top-to-random.) $S = (\sigma_i)_1^n$, where σ_i takes the top card of the deck and places it in position i. In cycle notation, $\sigma_i = (i, i-1, \dots, 2, 1)$. The probability measure μ_S in this example is not symmetric.
- (Random-to-random or random insertions.) $S = (\sigma_{ij})_{1 \le i,j \le n}$ (ordered lexicographically), where σ_{ij} is 'take the card in position i and insert it in position j'. In cycle notation, when $i < j$, $\sigma_{ij} = (j, j-1, \dots, i)$. Note also that $\sigma_{ij} = \sigma_{ji}^{-1}$ and $\sigma_{ii} = e$. The corresponding measure μ_S gives probability $1/n$ to the identity element e and probability $1/n^2$ to

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

each σ_{ij}, $i \neq j$ with the caveat that when $|j-i|=1$, $\sigma_{ij}=\sigma_{ji}$, so that the corresponding transposition $\tau=\sigma_{ij}=\sigma_{ji}$ actually has probability $2/n^2$.

- (Random transposition.) Take $S=(\tau_{ij})_{1\leq i\leq j\leq n}$, where τ_{ij} transposes the cards in positions i and j (i.e. $\tau_{ij}=(i,j)$). This tuple S contains each true transposition (i,j), $1\leq i<j\leq n$, twice, and also includes n copies of the identity (i,i), $1\leq i\leq n$. Equivalently, we can think of (i,j) being picked uniformly independently at random from $\{1,\ldots,n\}$ so that the probability measure μ_S gives probability $1/n$ to the identity and probability $2/n^2$ to any transposition.

All these examples are ergodic in the sense that the distribution at time t of the associated Markov chain converges to the uniform distribution u on $\mathbf{S}_n$.

1.2. Hit-and-run walks based on generating tuples

We now consider the following modification of the measure μ_S associated with a fixed generating tuple $S=(s_1,\ldots,s_k)$, which we call q_S. For each $s_i\in S$, let m_i be its order in G (the smallest m such that $s_i^m=e$). Define

$$q_S=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{m_i}\sum_{j=0}^{m_i-1}\delta_{s_i^j}. \tag{1.1}$$

To describe q_S in words, q_S is the distribution of a random element in G chosen as follows: pick i uniformly in $\{1,\ldots,k\}$, pick m uniformly in $\{0,\ldots,m_i-1\}$, output $s_i^m\in G$. This is reminiscent of the so-called hit-and-run algorithms, hence the name.

The question we want to address is whether or not the random walk driven by q_S mixes faster than the random walk driven by μ_S. Does taking a uniform step in the direction of the generator s_i, i.e. along the one parameter subgroup $\{s_i^m : 0\leq m\leq m_i-1\}$ instead of just a single s_i-step, speed up convergence or not?

Example 1.2. (*Our main example: hit-and-run for top-to-random.*) Top-to-random on $\mathbf{S}_n$ is obtained by considering the generating n-tuple

$$S=\{(k,k-1,\ldots,2,1) : k=1,\ldots,n\}=\{\sigma_k : k=1,\ldots,n\}$$

where $\sigma_k:=(k,k-1,\ldots,2,1)$. The associated simple random walk measure is (as mentioned earlier, it is not symmetric)

$$\mu_S(\sigma)=\begin{cases}\dfrac{1}{n} & \text{if } \sigma\in S,\\ 0 & \text{otherwise.}\end{cases}$$

The associated hit-and-run measure is given by

$$q(\sigma)=q_S(\sigma)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}\sum_{j=0}^{i-1}\delta_{\sigma_i^j}(\sigma). \tag{1.2}$$

This probability measure is symmetric and gives positive probability to order n^2 distinct permutations.

Let us now describe our findings (informally), and related open questions regarding the hit-and-run for top-to-random shuffle.

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

- (Facts.) In L^2, the mixing time for hit-and-run for top-to-random with n cards is of order $n \log n$, the same order as the top-to-random shuffle. See Section 3. There is a cut-off in L^2 but the cut-off time is not known. See Theorem 1.2 below. In L^1 (i.e. total variation), the mixing time is at least of order n and no more than order $n \log n$.
- (Open questions.) What is the cut-off time in L^2 for the hit-and-run version of top to random? How does it compare precisely with $n \log n$, the cut-off time for the top-to-random shuffle?
- (Open questions.) Is there a cut-off in L^1 (i.e. total variation)? What is the order of magnitude of the L^1-mixing time? Describe a simple statistic that provides a good lower bound for the mixing time in L^1.
- (Conjecture.) There is a cut-off in L^1 and the rough order of the L^1 cut-off time is $n \log n$.

Regarding general hit-and-run walks, we prove the following result.

Theorem 1.1. *Let G be a finite group and let $S = (s_1, \dots, s_k)$ be a generating tuple. The eigenvalues $-1 \le \beta_{|G|-1} \le \cdots \le \beta_1 \le \beta_0 = 1$ of the hit-and-run walk on G based on S driven by the symmetric measure q_S at (1.1) are all non-negative, i.e. $0 \le \beta_{|G|-1} \le \cdots \le \beta_1 \le \beta_0 = 1$.*

The proof of this theorem is in Section 4. Section 2 provides exact computations concerning the Markov chains obtained by following a single card. We explore the time to equilibrium for this Markov chain as a function of the starting position of the card that is followed, both in total variation and in L^2. Section 3 studies the convergence of the hit-and-run top-to-random walk on the symmetric group $\mathbf{S}_n$ in the L^2-norm $\|\cdot\|_2$, that is, we estimate

$$d_2\big(q^{(t)}, \big) = \|\big(q^{(t)}/u\big) - 1\|_2.$$

We prove the following result, which shows that the L^2-mixing time is of order $n \log n$.

Theorem 1.2. *For any n, t, we have*

$$d_2\big(q^{(t)}, u\big) \ge \sqrt{n-1}\left(1 - \frac{1}{n}\right)^t.$$

The second largest eigenvalue β_1 of q satisfies $\beta_1 \in [1 - 1/n, 1 - 1/(8n)]$ and, for any n large enough and $t(n, c) \ge 9n \log n + 12nc$, $c > 0$,

$$d_2\big(q^{(t(n,c))}, u\big) \le (2 + \mathrm{o}(1))\,\mathrm{e}^{-c} \le \sqrt{5}\,\mathrm{e}^{-c}.$$

The convergence of $q^{(t)}$ to u in the L^2 sense occurs with a cut-off.

Remark 1.1. The well-known definition of cut-off can be found in the next section. The existence of this L^2-cut-off follows from the other assertions in the theorem and [4, Theorem 3.3]. Indeed, we know that the spectral gap $\lambda = 1 - \beta_1$ for q is at least $1/(8n)$ and the time to stationarity in L^2, T, is at least $\frac{1}{2}n \log n$ so that the product λT tends to infinity with n. Referring to the vocabulary and definitions from [4], the L^2-cut-off stated in the theorem occurs at about $n \log n$ and has a window of order n. We do not know the more precise behavior of the cut-off time. The upper bound on $d_2(q^{(t)}, u)$ is stated 'for n large enough'. This comes from the proof of Theorem 1 in [2] and whatever 'large enough' means in the proof of [2, Theorem 1] works here as well. It seems that $n \ge 6$ is enough. We do not expect the constants in the definition of $t(n, c)$ to be sharp.

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

1.3. Notions of convergence

We will discuss convergence to the uniform distribution using two different distances between probability measures (or between their densities with respect to the uniform measure u). Let ν be a probability measure on a finite group G and let u be the uniform distribution on G.

Total variation (or $\frac{1}{2}$-$L^1(G, u)$-norm) is defined by

$$\begin{aligned}\|\nu - u\|_{\mathrm{TV}} &= \max_{A\subseteq G}\{\nu(A) - u(A)\}\\ &= \max_{A\subseteq G}\{|\nu(A) - u(A)|\}\\ &= \frac{1}{2}\|(\nu/u) - 1\|_1\\ &= \frac{1}{2}\sum_{g\in G}|\nu(g) - u(g)|.\end{aligned}$$

We also set $d_1(\nu, u) = \|(\nu/u) - 1\|_1 = 2\|\nu - u\|_{\mathrm{TV}}$. Convergence in $L^2(G, u)$ is measured using the distance

$$\begin{aligned}d_2(\nu, u)^2 &= \|(\nu/u) - 1\|_2^2\\ &= \sum_{g\in G}|(\nu(g)/u(g)) - 1|^2 u(g)\\ &= |G|\sum_{g\in G}|\nu(g) - u(g)|^2.\end{aligned}$$

Let $(G_n)_1^\infty$ be a sequence of finite groups such that $|G_n|$ tends to infinity with n. Let u_n be the uniform probability on G_n. We say that a sequence of probability measures μ_n on G_n, $n = 1, 2, \ldots$, has a cut-off at time t_n in L^p, $p = 1, 2$, if $t_n \to \infty$ and, for any $\epsilon > 0$,

$$\lim_{n\to\infty} d_p\big(\mu_n^{((1+\epsilon)t_n)}, u_n\big) = 0 \quad \text{and} \quad d_p\big(\mu_n^{((1-\epsilon)t_n)}, u_n\big) = l_\infty(p),$$

where $l_\infty(1) = 1$ and $l_\infty(2) = +\infty$.

Whenever the probability measure μ is symmetric, i.e. $\check{\mu} = \mu$, the associated convolution operator $f \mapsto Mf$ is diagonalizable with real eigenvalues $-1 \le \beta_{|G|-1} \le \cdots \le \beta_1 \le \beta_0 = 1$ and

$$d_2\big(\mu^{(t)}, u\big)^2 = \sum_1^{|G|-1}\beta_i^{2t}, \quad t = 1, 2, \ldots.$$

This follows from the usual spectral decomposition. See e.g. [13, Theorem 5.2]. Moreover,

$$d_\infty\big(\mu^{(2t)}, u\big) = \max_G\left\{\left|\frac{\mu}{\nu} - 1\right|\right\} = |G|\mu^{(2t)}(e) - 1 = d_2\big(\mu^{(t)}, u\big)^2.$$

The second equality (no absolute value) uses

$$\mu^{(2t)}(e) = \sum_y \mu^{(t)}(y)\mu^{(t)}(y^{-1}) = \sum_y |\mu^{(t)}(y)|^2.$$

The Cauchy–Schwarz inequality easily gives $d_\infty\big(\mu^{(2t)}, u\big) \le d_2\big(\mu^{(t)}, u\big)^2$ and it follows that this inequality must be an equality. Let us illustrate these definitions using the classical examples described above.

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

- (Top-to-random.) Convergence in total variation occurs at time $n \log n$ in the sense that if we set $t(n, c) = n \log n + cn$,

$$\lim_{n\to\infty} \left\| \mu^{(t(n,c))} - u \right\|_{\mathrm{TV}} = \begin{cases} 1 & \text{if } c < 0, \\ 0 & \text{if } c > 0. \end{cases}$$

  See [5] and [9]. With a little work, the results in [9] easily imply

$$\lim_{n\to\infty} d_2\big(\mu^{(t(n,c))}, u\big) = \begin{cases} \infty & \text{if } c < 0, \\ 0 & \text{if } c > 0. \end{cases}$$

- (Random-to-random.) Convergence in total variation (and in L^2) occurs with a cut-off at time $(3n/4) \log n$. See [2].
- (Random transposition.) Convergence in total variation (and in L^2) occurs with a cut-off at time $(n/2) \log n$. See [5], [8], and [14].

2. Single-card Markov chain

To investigate the complex behavior of the hit-and-run top-to-random chain, it behooves us to explore the dynamics of just a single card. We do so by defining a Markov chain $(X_t)_{t=0}^{\infty}$ with state space $\{1, 2, \ldots, n\}$ that represents the position of an arbitrarily chosen card after t shuffle iterations. This is a classical example of a function of a Markov chain that produces a Markov chain.

2.1. Abstract projection

Before proceeding with the example, we review some general aspects of this situation. Abstractly, we start with a Markov kernel Q on a state space X (in our case $Q(x, y) = q_S(x^{-1}y)$ on $\mathbf{S}_n$) and a lumping (or projection) map $p : X \to \underline{X}$ which is surjective and has the property that

$$\sum_{y\in X\,:\,p(y)=\underline{y}} Q(x, y) = \underline{Q}(\underline{x}, \underline{y})$$

depends only on $p(x) = \underline{x}$. This defines a Markov kernel on $\underline{X}$. If Q has stationary measure π then its push-forward $\underline{\pi}(\underline{x}) = \pi(p^{-1}(\underline{x}))$ is stationary for $\underline{Q}$. Moreover,

$$\|Q^t(x, \cdot) - \pi\|_{\mathrm{TV}} \geq \|\underline{Q}^t(\underline{x}, \cdot) - \underline{\pi}\|_{\mathrm{TV}}.$$

This simple comparison does not work well for the L^2 and L^∞ convergence measured using d_2 and d_∞ because normalization becomes an issue.

Let β and $\underline{\phi}$ be an eigenvalue and associated eigenfunction for the chain $\underline{Q}$. Then it is plain that the function $\phi(x) = \underline{\phi} \circ p(x)$ is an eigenfunction for Q with eigenvalue β. Also, two orthogonal eigenfunctions $\underline{\phi}_1, \underline{\phi}_2$ for $\underline{Q}$ on $L^2(\underline{\pi})$ give orthogonal ϕ_1, ϕ_2 in $L^2(\pi)$ (we will not use this second fact).

2.2. Single-card chain in L^2

Let q be the measure for the hit-and-run version of top-to-random defined in (1.2). We consider the projection of $Q(x, y) = q(x^{-1}y)$ on $\{1, \ldots, n\}$ corresponding to following the position of a single card. To simplify notation, we set $\underline{Q} = K$ and notice that the stationary (and

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

reversible) measure for K is the uniform measure on $\{1, \dots, n\}$. The transition probabilities $K(i, j)$, $i, j \in \{1, \dots, n\}$ are given by

$$K(i,j) = \begin{cases} \frac{1}{n}\sum_{k \geq j}^{n} \frac{1}{k} + \frac{i-1}{n} & \text{if } i = j, \\ \frac{1}{n}\sum_{k \geq i}^{n} \frac{1}{k} & \text{if } j < i, \\ \frac{1}{n}\sum_{k \geq j}^{n} \frac{1}{k} & \text{if } j > i. \end{cases}$$

The following lemma gives the eigenvalues and eigenvectors of K. The form of the eigenvectors was guessed by extrapolation from the cases $n \leq 4$.

Lemma 2.1. *The eigenvalues and associated eigenvectors of the stochastic matrix* $(K(i, j))_{1 \leq i,j \leq n}$ *are* $\beta_0 = 1$, $\mathbf{\Psi}_0 = (1, \dots, 1)$ *and*

$$\beta_i = 1 - \frac{i}{n}, \quad \mathbf{\Psi}_i = \left(\frac{-1}{n-i}, \dots, \frac{-1}{n-i}, 1, 0, \dots, 0 \right), \quad i = 1, \dots, n-1,$$

where, in $\mathbf{\Psi}_i$, *the value* $-1/(n-i)$ *is repeated* $n - i$ *times.*

Proof. It suffices to verify the guessed formula. For β_{n-j} and $\mathbf{\Psi}_{n-j}$, $j \in \{1, \dots, n-1\}$, and $k < j + 1$, we have

$$\begin{aligned} &(K\mathbf{\Psi}_{n-j})_k \\ &= \frac{1}{n}\left[\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n}\right][k-1]\frac{(-1)}{j} \\ &\quad + \left[\frac{1}{n}\left[\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n}\right] + \frac{k-1}{n}\right]\frac{(-1)}{j} \\ &\quad + \frac{1}{n}\left[\frac{1}{k+1} + \frac{1}{k+2} + \cdots + \frac{1}{n}\right]\frac{(-1)}{j} + \cdots + \frac{1}{n}\left[\frac{1}{j} + \frac{1}{j+1} + \cdots + \frac{1}{n}\right]\frac{(-1)}{j} \\ &\quad + \frac{1}{n}\left[\frac{1}{j+1} + \frac{1}{j+2} + \cdots + \frac{1}{n}\right] \cdot 1 + 0 + \cdots + 0 \\ &= \frac{1}{n}\left[\frac{(-1)}{j}\left(\frac{j}{n} + \frac{j}{n-1} + \cdots + \frac{j}{j+1}\right) + \frac{(-1)}{j}[j-k] \cdot 1 + \frac{(-1)}{j}k + \frac{1}{j} - \frac{1}{j}\right. \\ &\quad \left. + \frac{1}{j+1} + \frac{1}{j+2} + \cdots + \frac{1}{n}\right] \\ &= -\frac{1}{n}. \end{aligned}$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

For $k = j+1$, we obtain

$$(K\boldsymbol{\Psi}_{n-j})_k = \frac{1}{n}\left(\frac{1}{j+1}+\frac{1}{j+2}+\cdots+\frac{1}{n}\right)j\frac{(-1)}{j} + \frac{1}{n}\left(\frac{1}{j+1}+\frac{1}{j+2}+\cdots+\frac{1}{n}\right)+\frac{j}{n}=\frac{j}{n}.$$

Likewise, for $k > j+1$, we find

$$(K\boldsymbol{\Psi}_{n-j})_k = \frac{1}{n}\left(\frac{1}{j+1}+\frac{1}{j+2}+\cdots+\frac{1}{n}\right)j\frac{(-1)}{j} + \frac{1}{n}\left(\frac{1}{j+1}+\frac{1}{j+2}+\cdots+\frac{1}{n}\right)=0. \qquad \square$$

These eigenvectors are not normalized and

$$\|\boldsymbol{\Psi}_i\|_2^2 = \frac{1}{n(n-i)}+\frac{1}{n}=\frac{n-i+1}{n(n-i)}, \quad i=1,\dots,n-1. \tag{2.1}$$

In the next lemma we use this knowledge (including (2.1)) to compute

$$d_2(K^t(i,\cdot),u)^2 = n\sum_{j=1}^{n}\left|K^t(i,j)-\frac{1}{n}\right|^2 = \sum_{k=1}^{n-1}\beta_k^{2t}\frac{\boldsymbol{\Psi}_k(i)^2}{\|\boldsymbol{\Psi}_k\|_2^2}.$$

For the last equality, see e.g. [13, equation (5.2)] or [12, Lemma 1.3.3].

Lemma 2.2. *The quantity $d_2(K^t(i,\cdot),u)^2$ equals*

$$\begin{cases} \displaystyle\sum_{k=1}^{n-2}\left(1-\frac{k}{n}\right)^{2t}\frac{n}{(n-k)(n-k+1)}+\left(\frac{1}{n}\right)^{2t}\frac{n}{2} & \text{if } i=1,\\ \displaystyle\sum_{k=1}^{n-i}\left(1-\frac{k}{n}\right)^{2t}\frac{n}{(n-k)(n-k+1)}+\left(\frac{i-1}{n}\right)^{2t}\frac{n(i-1)}{i} & \text{if } 1<i<n,\\ \displaystyle\left(1-\frac{1}{n}\right)^{2t}(n-1) & \text{if } i=n. \end{cases}$$

Proof. This follows from Lemma 2.1, equation (2.1), and inspection. $\square$

We need to understand what these formulas mean. The term

$$\sum_{k=1}^{n-i}\left(1-\frac{k}{n}\right)^{2t}\frac{n}{(n-k)(n-k+1)}$$

can be bounded above by

$$\frac{1}{n}\sum_{k=1}^{n-i}\left(1-\frac{k}{n}\right)^{2t-2}$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

and bounded below by one-half of this quantity. Set

$$B(n,t,i)\left(1-\frac{1}{n}\right)^{2t-1}=\frac{1}{n}\sum_{k=1}^{n-i}\left(1-\frac{k}{n}\right)^{2t-2}.$$

Lemma 2.3. *For $n\geq 4$, $t\geq 1$, the quantity $B(n,t)$ is bounded as follows.*

- *If $2\leq i\leq an$, $a\leq 1/2$,*

$$\left(\frac{1}{n-1}+\frac{1}{4(2t-1)}\right)\leq B(n,t,i)\leq\left(\frac{1}{n-1}+\frac{1}{2t-1}\right).$$

- *If $i\leq an$, $a<1$, and $n\geq 2/(1-a)$, then there exists $c_a>0$ such that*

$$\left(\frac{1}{n-1}+\frac{c_a}{2t-1}\right)\leq B(n,t,i)\leq\left(\frac{1}{n-1}+\frac{1}{2t-1}\right).$$

- *If $n-i_0\leq i\leq n-2$,*

$$\frac{1}{n-1}\leq B(n,t,i)\leq\frac{i_0}{n-1}.$$

Proof. Observe that

$$B(n,t,i)=\frac{1}{n}\left(1-\frac{1}{n}\right)^{-1}\sum_{k=1}^{n-i}\left(\frac{1-k/n}{1-1/n}\right)^{2t-2}=\frac{1}{n-1}+\frac{1}{n-1}\sum_{k=2}^{n-i}\left(1-\frac{k-1}{n-1}\right)^{2t-2}.$$

The stated results easily follow, for example, by comparing Riemann sums with integrals in the first and second cases. □

Proposition 2.1.

(a) *For each fixed $i=1,2,\ldots,$ set $t_i(n,c)=(2n/i)(\log n+c)$. Then*

$$\lim_{n\to\infty} d_2\big(K^{t_i(n,c)}(n-i,\cdot),u\big)=\begin{cases}+\infty & \text{if } c<0,\\ 0 & \text{if } c>0.\end{cases}$$

That is, the position of the card starting in position $n-i$ becomes random in the L^2 sense with a cut-off at time $(2n/i)\log n$.

(b) *For each fixed $i=1,2,\ldots$ and any t_n tending to infinity,*

$$\lim_{n\to\infty} d_2(K^{t_n}(i,\cdot),u)=0.$$

Moreover, there exists a constant $c_i>0$ such that, for any $\epsilon\in(2/n,1)$,

$$d_2(K^t(i,\cdot),u)=\epsilon\Rightarrow t(n)\in(c_i/\epsilon^2,10/\epsilon^2).$$

(c) *For each fixed $a\in(0,1)$, set*

$$t_a(n,c)=\frac{1}{2\log(1/a)}(\log n+c).$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

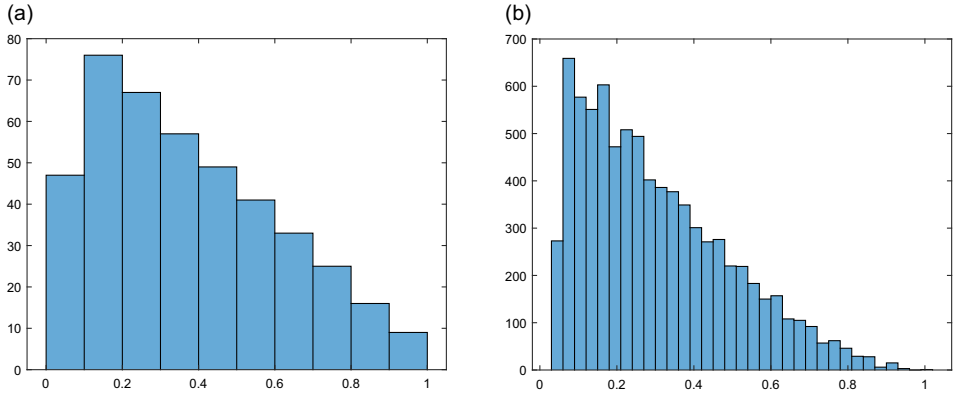

FIGURE 1. 21 cards. (a) The eigenvalue distribution for the two-card chains. (b) The eigenvalue distribution for the three-card chains. Note that all eigenvalues are positive.

Then

$$\lim_{n\to\infty} d_2\big(K^{t_a(n,c)}([an],\cdot),u\big) = \begin{cases} +\infty & \text{if } c<0,\\ 0 & \text{if } c>0. \end{cases}$$

That is, the position of the card starting in position [an] becomes random in the L^2 sense with a cut-off at time $\log n/(2\log(1/a))$.

Remark 2.1. The first and second statements are for fixed i, that is, they cannot be applied with i depending on n. Because of this, statement (a) is about starting somewhere near the bottom and statement (b) about starting somewhere near the top. Similarly, in (c), the real $a \in (0, 1)$ is fixed, which means this statement is about starting from somewhere in the middle. In statements (a) and (c), where exactly one starts is important as it appears in the definition of the cut-off time.

Proof. Avoiding the two special cases of the first and last starting positions (which are easily treated), for a starting position $j \in \{2, \dots, n-1\}$, we have

$$\begin{aligned} &\frac{1}{2}\left(1-\frac{1}{n}\right)^{2t-1} B(n,j,t) + \left(\frac{j-1}{n}\right)^{2t} \frac{n(j-1)}{i}\\ &\quad \le d_2(K^t(j,\cdot),u)^2\\ &\quad \le \left(1-\frac{1}{n}\right)^{2t-1} B(n,j,t) + \left(\frac{j-1}{n}\right)^{2t} \frac{n(j-1)}{j}. \end{aligned}$$

The stated results follow from Lemma 2.3 and careful inspection. □

We end this section with two eigenvalue data plots, shown in Figure 1, which concern different projections, namely, those corresponding to following a given pair or a triplet of cards instead of just one. These chains become more complex and we have not computed all their eigenvalues and eigenfunctions. Instead, these plots are based on computer-assisted computations of the eigenvalues of these chains. The first plot is for the two-card chain on 21 cards and the second is for the three-card chains on 21 cards.

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

2.3. Single-card chain in L^1

The relatively simple form of the eigenvalues and eigenvectors of the single-card chain K also allows us to determine the L^1-distance of $K^t(i,\cdot)$ from its stationary measure u. Namely, the diagonalization of K shows that the ith row of K^t, $K^t(i,\cdot)$, $1 \le i \le n-1$, consists of the repeated entry

$$\frac{1}{n^t}\left(\frac{-(i-1)^t}{i} + \frac{i^{t-1}}{i+1} + \cdots + \frac{(n-1)^{t-1}}{n}\right) + \frac{1}{n}$$

in columns $j = 1$ through $i-1$,

$$\frac{1}{n^t}\left(\frac{(i-1)^{t+1}}{i} + \frac{i^{t-1}}{i+1} + \cdots + \frac{(n-1)^{t-1}}{n}\right) + \frac{1}{n}$$

in column i,

$$\frac{1}{n^t}\left(\frac{-(i+k-1)^t}{i+k} + \frac{(i+k)^{t-1}}{i+k+1} + \cdots + \frac{(n-1)^{t-1}}{n}\right) + \frac{1}{n}$$

in column $j = i+k$, $i+1 \le i+k < n$, and

$$\frac{1}{n^t}\left(\frac{-(n-1)^t}{n}\right) + \frac{1}{n}$$

in column n. The last row, $i = n$, consists of the entries

$$\frac{1}{n^t}\left(\frac{-(n-1)^t}{n}\right) + \frac{1}{n}$$

in columns 1 through $n-1$ and

$$\frac{1}{n^t}\left(\frac{(n-1)^{t+1}}{n}\right) + \frac{1}{n}$$

in column n.

In the case $i = n$ (single card starting at the bottom of the deck), we find that

$$\|K^t(n,\cdot) - u\|_{\mathrm{TV}} = \left(1 - \frac{1}{n}\right)^{t+1}$$

(indeed, by definition of our shuffling, this card position becomes uniform as soon as it is touched). For $1 \le i \le n-1$,

$$\begin{aligned} 2\|K^t(i,\cdot) - u\|_{\mathrm{TV}} &= \frac{1}{n^t}\left|-\frac{(n-1)^t}{n}\right| + \frac{i-1}{n^t}\left|\frac{-(i-1)^t}{i} + \sum_{\ell=i}^{n-1}\frac{\ell^{t-1}}{\ell+1}\right| \\ &\quad + \frac{1}{n^t}\left(\frac{(i-1)^{t+1}}{i} + \sum_{\ell=i}^{n-1}\frac{\ell^{t-1}}{\ell+1}\right) \\ &\quad + \frac{1}{n^t}\sum_{k=1}^{n-i-1}\left|\frac{-(i+k-1)^t}{i+k} + \sum_{\ell=1}^{n-(i+k)}\frac{(i+k+\ell-1)^{t-1}}{i+k+\ell}\right| \\ &= J_1 + J_2 + J_3 + J_4. \end{aligned} \tag{2.2}$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

Looking at J_2 and J_4 for large t, i.e. $t \geq (n-1)\log n + (n-1)/(n-2)$, we have

$$\frac{-(i-1)^t}{i} + \sum_{\ell=i}^{n-1} \frac{\ell^{t-1}}{\ell+1} \geq 0$$

and

$$\frac{-(i+k-1)^t}{i+k} + \sum_{\ell=1}^{n-(i+k)} \frac{(i+k+\ell-1)^{t-1}}{i+k+\ell} \geq 0 \quad \text{for } k \in \{1, \dots, n-i-1\}.$$

Because the sum of all the same terms in $J_1 + J_2 + J_3 + J_4$ but without any absolute value is

$$\sum_{\ell=1}^{n} (K^t(i,\ell) - u(\ell)) = 0,$$

it follows that for $t \geq (n-1)\log n + (n-1)/(n-2)$, we have $2\|K^t(i,\cdot) - u\|_{\mathrm{TV}} = 2|J_1|$, that is,

$$\|K^t(i,\cdot) - u\|_{\mathrm{TV}} = \frac{1}{n}\left(1 - \frac{1}{n}\right)^t, \quad i \in \{1, \dots, n-1\}.$$

This, of course, occurs only after much approximate convergence has taken place. It only describes the long-term asymptotic behavior of $\|K^t(i,\cdot) - u\|_{\mathrm{TV}}$, $i < n$. To describe the shorter-term behavior, we consider three cases: bottom starting positions of the type $n-i$ for fixed $i = 1, 2, \dots$, top starting positions of the type $i = 1, 2, \dots$, and middle of the pack starting positions of the type $[an]$, $a \in (0,1)$ (see Remark 2.1 above). The key difficulty in these estimates is to identify which of the terms J_1, J_2, J_3, J_4 plays the key role. In what follows, we omit the details regarding upper bounds. The details given for the lower bounds indicate, in each case, which term is dominant and hence what the target should be for upper bounds. Verifying that all terms are upper-bounded appropriately follows from simple careful arguments including basic Riemann sum estimates.

For starting position $n-i$, i fixed, we get a reasonable lower bound by focusing on the first and third terms in (2.2). Write

$$\begin{aligned} 2\|K^t(n-i,\cdot) - u\|_{\mathrm{TV}} &\geq \frac{1}{n}\left(1 - \frac{1}{n}\right)^t \\ &\quad + \frac{1}{n^t}\left(\frac{(n-i-1)^{t+1}}{n-i} + \sum_{\ell=n-i+1}^{n} \frac{(\ell-1)^{t-1}}{\ell}\right) \\ &\geq \frac{1}{n}\left(1 - \frac{1}{n}\right)^t + \left(1 - \frac{i+1}{n}\right)^{t+1}. \end{aligned}$$

An upper bound of the type

$$\|K^t(n-i,\cdot) - u\|_{\mathrm{TV}} \leq C_i\left(\frac{1}{n}\left(1 - \frac{1}{n}\right)^t + \left(1 - \frac{i+1}{n}\right)^t\right)$$

holds as well. This proves convergence in time of order $n/(i+1)$ with no cut-off for the bottom cards.

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

For starting position i, i fixed (starting position towards the top), we have

$$2\|K^t(i,\cdot)-u\|_{\mathrm{TV}} \geq \left(\frac{1}{n}+\frac{c_i}{t}\right)\left(1-\frac{1}{n}\right)^t.$$

The term $(c_i/t)(1-1/n)^t$ is contributed by the last summand, J_4, in (2.2). In this last summand, namely

$$\frac{1}{n^t}\sum_{k=1}^{n-i-1}\left|\frac{-(i+k-1)^t}{i+k}+\sum_{\ell=1}^{n-(i+k)}\frac{(i+k+\ell-1)^{t-1}}{i+k+\ell}\right|,$$

restrict the first summation to those k less than, say, $n/4$. In this range of k values, the positive term in the absolute value dominates the negative term and we obtain a lower bound of the type (we assume $t \geq 4$)

$$\begin{aligned}\frac{c_i}{n^t}\sum_{k=1}^{n/4}\sum_{\ell=n/2}^{n-1}\ell^{t-2} &\geq c_i'\left(1-\frac{1}{n}\right)^{t-1}\left(\frac{1}{n-1}\sum_{n/2}^{n-1}\left(\frac{\ell}{n-1}\right)^{t-2}\right)\\ &\geq \frac{c_i''}{t}\left(1-\frac{1}{n}\right)^t,\end{aligned}$$

where we used an integral to lower-bound the Riemann sum in parentheses. A matching upper bound

$$\|K^t(i,\cdot)-u\|_{\mathrm{TV}} \leq C_i\left(\frac{1}{n}+\frac{1}{t}\right)\left(1-\frac{1}{n}\right)^t$$

is easily obtained for all four terms in (2.2). The key rate of convergence is thus in $1/t$ for the top starting positions.

For a starting position in the middle of the pack, $i=[an]$, $a\in(0,1)$ fixed, a similar analysis shows that $\|K^t([an],\cdot)-u\|_{\mathrm{TV}}$ is also of order

$$\left(\frac{1}{n}+\frac{1}{t}\right)\left(1-\frac{1}{n}\right)^t.$$

This time, for the lower bound, we use the second term

$$\frac{i-1}{n^t}\left|\frac{-(i-1)^t}{i}+\sum_{\ell=i}^{n-1}\frac{\ell^{t-1}}{\ell+1}\right|.$$

It provides a lower bound of the type $c_a(1-1/n)^t/t$. Indeed, for $i=[an]$ and n large enough,

$$\begin{aligned}\sum_{\ell=i}^{n-1}\frac{\ell^{t-1}}{\ell+1} &\geq \frac{(n-1)^{t-1}}{2}\sum_{\ell=[an]}^{n-1}\left(\frac{\ell}{n-1}\right)^{t-2}\frac{1}{n-1}\\ &\geq \frac{(n-1)^{t-1}}{2}\int_{(a+1)/2}^{1}x^{t-2}\,\mathrm{d}x\\ &\geq c_a\frac{(n-1)^{t-1}}{t-1}.\end{aligned}$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

For $t \geq t_a$, this is larger than twice $(i-1)^{t-1}$, $i = [an]$. It follows that

$$\frac{i-1}{n^t}\left|\frac{-(i-1)^t}{i} + \sum_{\ell=i}^{n-1} \frac{\ell^{t-1}}{\ell+1}\right| \geq \frac{c_a}{2}\frac{an}{n^t}\frac{(n-1)^{t-1}}{t-1} \geq \frac{c'_a}{t}\left(1-\frac{1}{n}\right)^t.$$

Again, using similar techniques, all terms in (2.2) can be seen to be bounded above appropriately.

3. Hit-and-run for top-to-random in L^2

In this section we prove Theorem 1.2, which concerns the convergence to stationarity measured in the L^2 sense, that is, using $d_2(q^{(t)}, u)$ for the hit-and-run version of top-to-random driven by the measure q in (1.2).

Proof of the lower bound in Theorem 1.2. In the section concerning following a single card, we learned that $(1-1/n)$ is an eigenvalue of that chain and consequently also an eigenvalue of convolution by q on $\mathbf{S}_n$. Now, on $\mathbf{S}_n$, each eigenvalue has multiplicity at least equal to the dimension of any irreducible representation at which it occurs. The group $\mathbf{S}_n$ has two representations of dimension 1: the trivial representation and the sign representation. All other irreducible representations have dimension at least $n-1$. So it suffices to verify that $(1-1/n)$ does not occur only at the sign representation. This can be seen from the form of the associated eigenvector we have constructed. Alternatively, one easily computes the eigenvalue for the sign representation to be $1/2$ if n is even and $(n+1)/2n$ if n is odd ($(n+1)/2$ is the number of odd integers in $\{1, \ldots, n\}$). In any case, this gives the lower bound $d_2(q^{(t)}, u) \geq \sqrt{n-1}(1-1/n)^t$ as stated. □

To prove the stated upper bound for $d_2(q^{(t)}, u)$, we use the comparison technique from [7]. Anticipating the next section, we use the fact that hit-and-run walks have non-negative spectrum. It turns out that the most efficient comparison is with the random-to-random walk of Example 1.1 driven by the measure

$$\mu(\sigma) = \begin{cases} 1/n & \text{if } \sigma = \mathrm{id}, \\ 2/n^2 & \text{if } \sigma = \sigma_{i(i+1)} = \sigma_{(i+1)i}, \\ 1/n^2 & \text{if } \sigma = \sigma_{ij},\ 1 \leq i \neq j \leq n,\ |j-i| > 1, \end{cases}$$

where $\sigma_{ij} = (j, j-1, \ldots, i)$, $\sigma_{ji} = \sigma_{ij}^{-1}$, $1 \leq i < j \leq n$. Recall that the Dirichlet form associated with a symmetric probability measure ν on a finite group G is

$$\mathcal{E}_\nu(v, w) = \frac{1}{2|G|}\sum_{x,y\in G}(v(xy)-v(x))(w(xy)-w(x))\nu(y), \qquad v, w \in L^2(G).$$

Lemma 3.1. *The Dirichlet form $\mathcal{E}_\mu$ associated with the random-to-random measure μ and the Dirichlet form $\mathcal{E}_q$ associated with hit-and-run version of top-to-random satisfy*

$$\mathcal{E}_\mu(v, v) \leq 8\mathcal{E}_q(v, v) \quad \textit{for all } v \in L^2(G).$$

Proof. Recall that $\sigma_k = \sigma_{1k} = (k, k-1, \ldots, 1)$. For each σ_{ij}, $1 \leq i \neq j \leq n$, we find a product of σ_k^ℓ, $1 \leq \ell < k \leq n$, which equals σ_{ij}. There are many ways to do this, but the following is efficient. Observe that for $i < j$, $\sigma_{ij} = \sigma_j^i \sigma_{j-1}^{j-i}$. That is, to move the card in position i down to

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

position j, insert the first i cards at position j, then insert the first $j-i$ cards now on top at position $j-1$. After the first move, the card originally in position j is at position $j-i$, so the second move places it in position $j-1$. The other $i-1$ cards moved down to position $j-1$ are returned to their original spot in the second move (barring the card originally in position i) by sliding past them all the cards they originally slid past, which were on the top after the first move. For $i>j$,

$$\sigma_{ij}=\sigma_{ji}^{-1}=\sigma_{i-1}^{-(i-j)}\sigma_i^{-j}=\sigma_{i-1}^{j-1}\sigma_i^{i-j}.$$

Now we use [7, Theorem 1] with $\widetilde{\mathcal{E}}=\mathcal{E}_\mu$, $\mathcal{E}=\mathcal{E}_q$, which gives

$$\mathcal{E}_\mu\le A\mathcal{E}_q,\quad A=\max_{\sigma\,:\,q(\sigma)>0}\left\{\frac{1}{q(\sigma)}\sum_{1\le i\ne j\le n}|\sigma_{ij}|N(\sigma,\sigma_{ij})\mu(\sigma_{ij})\right\}.$$

In the formula giving A, $|\sigma_{ij}|$ is the length of the product expressing σ_{ij}, which, in our case, is always equal to 2; $N(\sigma,\sigma_{ij})$ is the number of times σ appears in the product for σ_{ij}. So, if $\sigma=\sigma_k^\ell$ for some $2\le k\le n$ and $1\le\ell\le k-1$, $1\le i<j\le n$,

$$N(\sigma,\sigma_{ij})=\begin{cases}0 & \text{if } (k,\ell)\notin\{(j,i),(j-1,j-i)\},\\ 1 & \text{if } (k,\ell)\in\{(j,i),(j-1,j-i)\}.\end{cases}$$

When $1\le j<i\le n$, we similarly have

$$N(\sigma,\sigma_{ij})=\begin{cases}0 & \text{if } (k,\ell)\notin\{(i-1,j-1),(i,i-j)\},\\ 1 & \text{if } (k,\ell)\in\{(i-1,j-1),(i,i-j)\}.\end{cases}$$

For $1\le\ell<k<n$, this gives

$$\left\{\frac{1}{q(\sigma_k^\ell)}\sum_{1\le i\ne j\le n}|\sigma_{ij}|N(\sigma,\sigma_{ij})\mu(\sigma_{ij})\right\}=8k/n,$$

whereas for $(k,\ell)=(n,\ell)$ the result is 4. See Figure 2 for a comparison of the spectral distributions. □

Proof of the upper bound in Theorem 1.2. Given the comparison inequality $\mathcal{E}_\mu\le 8\mathcal{E}_q$ between quadratic forms, Lemma 6 of [7] (see also [13, Theorem 10.2]) provides a comparison inequality. Here we used the same idea in a slightly tighter way. Let $0\le\alpha_{|G|-1}\le\cdots\le\alpha_1<\alpha_0=1$ be the eigenvalues for random-to-random. By [7, Lemma 4], the inequality $\mathcal{E}_\mu\le 8\mathcal{E}_q$ gives

$$1-\beta_i\ge\frac{1}{8}(1-\alpha_i).$$

This is enough because there are no negative eigenvalues (to see what happens when there are negative eigenvalues, see [7]; the absence of negative eigenvalues is only a small simplification). Split the sum (the inequality here uses the fact that the β_i are all non-negative)

$$d_2\big(q^{(t)},u\big)^2=\sum_{i=1}^{|G|-1}\beta_i^{2t}\le\sum_{i=1}^{|G|-1}\mathrm{e}^{-2t(1-\beta_i)}$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

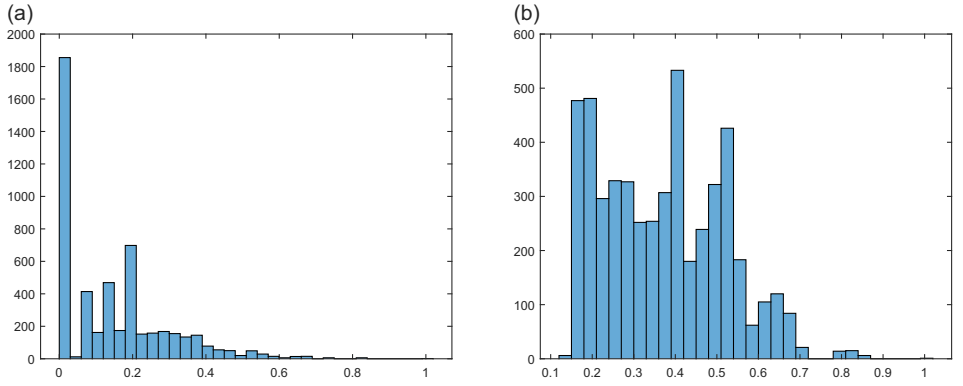

FIGURE 2. Comparison of the spectrum of random-to-random (a) and hit-and-run for top-to-random (b). The key difference is the higher multiplicity of very small eigenvalues in the random-to-random shuffle (most of those are actually equal to 0). Note the different scales on the y-axes of the two graphics. This is for a deck of seven cards so there are 7! eigenvalues.

into two sums: the sum over those indices i such that $\alpha_i \le 1/2$ and the sum over the indices for which $\alpha_i > 1/2$. For the first sum, write

$$\sum_{i\,:\,\alpha_i \le 1/2} \mathrm{e}^{-2t(1-\beta_i)} \le (n!)\,\mathrm{e}^{-t/8}.$$

For the second sum, note that $\mathrm{e}^{-3(1-x)/2} \le x$ when $x \in [.5, 1]$, and write

$$\sum_{i\,:\,\alpha_i > 1/2} \mathrm{e}^{-2t(1-\beta_i)} \le \sum_{i\,:\,\alpha_i > 1/2} \mathrm{e}^{-t(1-\alpha_i)/4} \le \sum_{i\,:\,\alpha_i > 1/2} \alpha_i^{t/6} \le d_2\big(\mu^{([t/12])}, u\big)^2.$$

This gives

$$d_2\big(q^{(t)}, u\big)^2 \le (n!)\,\mathrm{e}^{-t/8} + d_2\big(\mu^{([t/12])}, u\big)^2. \tag{3.1}$$

In [2, equation (25)], it is proved that the spectral gap for μ is asymptotically equal to $1/n$ (see the beginning of Section 3 in [2]; the exact value is $(n + 2/n^2$ and it occurs with multiplicity at least $(n-1)$) and that

$$d_2\big(\mu^{(s)}, u\big)^2 \le 4\,\mathrm{e}^{-2c} \quad \text{for any } s \ge \frac{3}{4} n \log n + cn,\ c > 0,$$

as long as n is sufficiently large (let us note that this is a rather difficult result). Using this in (3.1) yields the upper bound stated in Theorem 1.2. □

4. Positivity of the spectrum

In this final section we prove Theorem 1.1. Given a general hit-and-run random walk driven by the measure q_S at (1.1) on a finite group G, we set

$$Q(x, y) = q_S(x^{-1}y) = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{m_i} \sum_{j=0}^{m_i - 1} \delta_{s_i^j}(x^{-1}y), \qquad x, y \in G.$$

This defines a self-adjoint operator $f \mapsto Qf = \sum_y Q(\cdot, y) f(y)$ acting on the space $H = L^2(G, u)$ equipped with the inner product

$$\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{x \in G} f_1(x) f_2(x).$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

Let

$$\beta_{|G|-1} \leq \cdots \leq \beta_1 \leq \beta_0 = 1$$

be the $|G|$ eigenvalues of this operator. Because Q is Markov, these eigenvalues are contained in the interval $[-1, 1]$. The theorem we want to prove, Theorem 1.1, asserts that they are in fact in the interval $[0, 1]$, that is, Q is non-negative in the sense that $\langle Qf, f\rangle \geq 0$. The proof follows the main idea of [11], which consists in writing Q in the product form

$$Q = P^*RP$$

using auxiliary operators P, R, P^*, where $R = R^2$ is self-adjoint acting on the extended Hilbert space H_{aux}, the space of functions on $G \times \{1, \ldots, k\}$ equipped with its natural inner product $\langle \cdot, \cdot \rangle_{\text{aux}}$. Because $R = R^2$, $R^* = R$, and P^* is the formal adjoint of P, such a decomposition establishes that

$$\langle Qf, f\rangle = \langle P^*RPf, f\rangle = \langle RPf, RPf\rangle_{\text{aux}} \geq 0.$$

To use such a decomposition is a key insight from [11], but it also appears in [15, Remark 4.4] and [10, Lemma 3.1].

Define the auxiliary Hilbert space $H_{\text{aux}} = \mathbb{R}^{G\times\{1,\ldots,k\}}$ equipped with inner product

$$\langle g_1, g_2\rangle_{\text{aux}} := \frac{1}{k|G|} \sum_{x\in G} \sum_{i=1}^{k} g_1(x, i)g_2(x, i),$$

where $g_1, g_2 \in H_{\text{aux}}$. Let $P : H \to H_{\text{aux}}$ denote the bounded linear operator given by

$$Pf(x, i) = f(x), \quad (x, i) \in G \times \{1, \ldots, k\}.$$

Note that the adjoint operator, $P^* : H_{\text{aux}} \to H$ of P is given by

$$P^*g(x) = \frac{1}{k} \sum_{i=1}^{k} g(x, i).$$

That P^* is the adjoint of P means here that $\langle P^*g, f\rangle = \langle g, Pf\rangle_{\text{aux}}$ for any $f \in H$ and $g \in H_{\text{aux}}$, which can be verified by a simple calculation. The matrices (or kernels) of these operators are

$$P((x, i), y) = \delta_x(y),$$
$$P^*(x, (y, i)) = \frac{\delta_x(y)}{k},$$

for any $x, y \in G$ and $i \in \{1, \ldots, k\}$. For any pair $(x, i) \in G \times \{1, \ldots, k\}$, call

$$\mathcal{Z}(x, i) := \left\{x_0, \ldots, x_{m_i-1} \mid x_i := xs_i^j, j = 0, \ldots, m_i - 1\right\}$$

the orbit of x in G under the cyclic subgroup $\langle s_i\rangle = \left\{s_i^j : j = 0, \ldots, m_i - 1\right\}$ generated by s_i in G. Define a Markov transition kernel

$$R(\cdot, \cdot) \text{ on } (G \times \{1, \ldots, k\})^2$$

by setting

$$R((x, i_1), (y, i_2)) := \delta_{i_1}(i_2)\frac{\delta_{\mathcal{Z}(x,i_1)}(y)}{m_{i_1}}.$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

It induces a Markov operator, $R : H_{\text{aux}} \to H_{\text{aux}}$, given by

$$Rg(x, i) = \frac{1}{m_i} \sum_{z \in \mathcal{Z}(x,i)} g(z, i), \quad g \in H_{\text{aux}}.$$

Because

$$x \in \mathcal{Z}(y, i) \quad \text{if and only if} \quad y \in \mathcal{Z}(x, i)$$

the operator R is symmetric, that is,

$$R((x, i_1), (y, i_2)) = \delta_{i_1}(i_2)\frac{\delta_{\mathcal{Z}(x,i_1)}(y)}{m_{i_1}} = \delta_{i_2}(i_1)\frac{\delta_{\mathcal{Z}(y,i_2)}(x)}{m_{i_2}} = R((y, i_2), (x, i_1)).$$

Thus the corresponding operator $R : H_{\text{aux}} \to H_{\text{aux}}$ is self-adjoint.

Lemma 4.1. *The operator R satisfies $R^2 = R$ and $Q = P^*RP$.*

Proof of $R^2 = R$. For arbitrary $g \in H_{\text{aux}}$, we have

$$R^2 g(x, i) = \frac{1}{m_i} \sum_{y \in \mathcal{Z}(x,i)} Rg(y, i) = \frac{1}{m_i} \sum_{y \in \mathcal{Z}(x,i)} \sum_{z \in \mathcal{Z}(y,i)} \frac{g(z, i)}{m_i} = Rg(x, i).$$

Here we used the facts that for $y \in \mathcal{Z}(x, i)$ we have $\mathcal{Z}(x, i) = \mathcal{Z}(y, i)$ and $|\mathcal{Z}(x, i)| = m_i$. □

Proof of $Q = P^*RP$. For any $x, y \in G$ we have

$$\begin{aligned} P^*RP(x, y) &= \frac{1}{k} \sum_{i=1}^{k} RP((x, i), y) \\ &= \frac{1}{k} \sum_{i=1}^{k} \sum_{z \in \mathcal{Z}(x,i)} \frac{P((z, i), y)}{m_i} \\ &= \frac{1}{k} \sum_{i=1}^{k} \frac{1}{m_i} \sum_{z \in \mathcal{Z}(x,i)} \delta_y(z) \\ &= \frac{1}{k} \sum_{i=1}^{k} \frac{1}{m_i} \sum_{j=0}^{m_i - 1} \delta_y\left(xs_i^j\right) \\ &= Q(x, y). \end{aligned}$$

Here we used the definition of $\mathcal{Z}(x, i)$ and, in the last equality, that $y = xs_i^j$ if and only if $s_i^j = x^{-1}y$. □

5. Final remarks

Example 5.1. (*Example where hit-and-run is faster.*) Assume that $G = (\mathbb{Z}/n\mathbb{Z})^d$ and $S = (0, e_1, -e_1, \ldots, e_d, -e_d)$, where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 in position j, $1 \le j \le d$. In L^2 and L^1, the walk driven by μ_S mixes in time

$$\frac{d \log d}{2(1 - \cos 2\pi/n)}$$

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

(as d (and possibly) n tends to infinity). The measure q_S is given by

$$q_S(0) = \frac{n+2d}{n(1+2d)} \sim \frac{1}{2d} + \frac{1}{n}$$

and, for $m \in \{1, \ldots, n-1\}$ and $j \in \{1, \ldots, d\}$,

$$q_S(me_j) = \frac{2}{(1+2d)n} \sim \frac{1}{dn}.$$

This is very close to the walk that simply takes a random step in a random coordinate and thus behaves similarly. The mixing times for q_S are different in L^1 and in L^2. In L^1 (or total variation), the mixing time is $d \log d$ (based on the coupon collector problem). In L^2, the mixing time is $d \log(dn)$. In both cases there is a gain over μ_S of order n^2. See [13, page 323] and [7, page 2154].

Example 5.2. (*Example when hit-and-run is a little slower.*) Let us consider briefly the example of random-transposition on $\mathbf{S}_n$. Because all generators have order 2, the measure q_S gives probability

$$\frac{1}{n} + \frac{n-1}{2n} = \frac{1}{2}\left(1 + \frac{1}{n}\right)$$

to the identity and probability $1/n^2$ to any transposition. It follows that the hit-and-run random walk based on random transposition has a cut-off in total variation and L^2 at time $n \log n$, a slow-down by a factor of $1/2$ compared to its simple random walk counterpart.

Remarks regarding hit-and-run for top-to-random. Because the hit-and-run shuffle based on top-to-random is the focus of this paper, it is worth pointing out that it can be described naturally without reference to the general hit-and-run construction. Namely, the measure q at (1.2) can be alternatively obtained as follows. Pick a position i uniformly at random in $\{1, \ldots, n\}$ and then pick a packet size j uniformly at random in $\{1, \ldots, i\}$. Pick up the packet of the top j cards and place it below the card originally at position i. This is clearly different from the top-m-to-random shuffles studied in [9]. There are two shuffle mechanisms described in [7] which bear some close similarities to the hit-and-run top-to-random shuffle described above. They are as follows.

- The *crude overhand shuffle* [7, page 2148]. The top, middle, and bottom packets are identified using two random positions $1 \le a \le b \le n$, and the order of the packets is changed as follows: top goes to the bottom, middle remains in the middle, bottom goes to the top. The pair of positions $a \le b$ is chosen by picking a uniformly in $\{1, \ldots, n\}$ and b uniformly in $\{a, \ldots, n\}$. Note that this gives weight $1/n$ to the identity which is obtained for $a = b = n$. An L^2 mixing time upper bound of order $n \log n$ is proved in [7] and an L^1 mixing time lower bound based on a coupon collector argument is also stated in [7]. However, although the coupon collector argument described in [7] makes heuristic sense, it seems that its detailed implementation is unclear because the probability that a pair of adjacent cards is broken up depends on the position of the cards. This is worth mentioning here because the exact same difficulty appears for the hit-and-run version of top-to-random, which is the focus of the present article.
- The *Borel shuffle* [7, page 2150] (which is taken from a book on the game of Bridge by Borel and Chéron from 1940). In this shuffle, a middle packet is removed from the

https://doi.org/10.1017/jpr.2021.96 Published online by Cambridge University Press

deck and placed on top. If (a, b), $1 \le a \le b \le n$, describes the position of the top and bottom card of the packet removed, (a, b) is picked with probability $1/\binom{n+1}{2}$, and this gives probability $2/(n+1)$ to the identity which is obtained for any of the choices $(1, b)$, $1 \le b \le n$. An L^2 mixing time upper bound of order $n \log n$ is proved in [7] as well as an L^1 mixing time lower bound based on a coupon collector argument (for this shuffle the coupon collector argument works fine).

Acknowledgements

We wish to thank the anonymous reviewers whose comments helped us improve the readability of the paper.

Funding information

The research of Samuel Boardman was partially supported by the NSF RTG-grant DMS-1645643. The research of Daniel Rudolf was partially supported by DFG project 389483880. The research of Laurent Saloff-Coste was partially supported by NSF grants DMS-1707589 and DMS-2054593.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ALDOUS, A. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability XVII* (Lecture Notes Math. **986**), pp. 243–297. Springer, Berlin.
[2] BERNSTEIN, M. AND NESTORIDI, E. (2019). Cutoff for random to random card shuffle. *Ann. Prob.* **47**, 3303–3320.
[3] BROWN, K. S. AND DIACONIS, P. (1998). Random walks and hyperplane arrangements. *Ann. Prob.* **26**, 1813–1854.
[4] CHEN, G.-Y. AND SALOFF-COSTE, L. (2008). The cutoff phenomenon for ergodic Markov processes. *Electron. J. Prob.* **13**, 26–78.
[5] DIACONIS, P. (1988). *Group Representations in Probability and Statistics* (Institute of Mathematical Statistics Lecture Notes: Monograph Series **11**). Institute of Mathematical Statistics, Hayward, CA.
[6] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. USA* **93**, 1659–1664.
[7] DIACONIS, P. AND SALOFF-COSTE, L. (1993). Comparison techniques for random walk on finite groups. *Ann. Prob.* **21**, 2131–2156.
[8] DIACONIS, P. AND SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitsth.* **57**, 159–179.
[9] DIACONIS, P., FILL, J. A. AND PITMAN, J. (1992). Analysis of top to random shuffles. *Combinatorics Prob. Comput.* **1**, 135–155.
[10] DYER, M., GREENHILL, C. AND ULLRICH, M. (2014). Structure and eigenvalues of heat-bath Markov chains. *Linear Algebra Appl.* **454**, 57–71.
[11] RUDOLF, D. AND ULLRICH, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Prob.* **18**, 8.
[12] SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics* (Lecture Notes Math. **1665**), ed. P. Bernard, pp. 301–413. Springer, Berlin.
[13] SALOFF-COSTE, L. (2004). Random walks on finite groups. In *Probability on Discrete Structures* (Encyclopaedia Math. Sci. **110**), pp. 263–346. Springer, Berlin.
[14] SALOFF-COSTE, L. AND ZÚÑIGA, J. (2008). Refined estimates for some basic random walks on the symmetric and alternating groups. *ALEA Lat. Am. J. Prob. Math. Statist.* **4**, 359–392.
[15] ULLRICH, M. (2014). Rapid mixing of Swendsen–Wang dynamics in two dimensions. *Dissertationes Math.* **502**, 1–64.