Forum of Mathematics, Sigma (2023), **Vol. 11**:e2 1–45
doi:10.1017/fms.2022.106

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Double coset Markov chains

Persi Diaconis[1,2], Arun Ram[3] and Mackenzie Simper[4]

[1]Department of Mathematics, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305-2125 USA;
E-mail: diaconis@math.stanford.edu
[2]Department of Statistics, Stanford University, 390 Jane Stanford Way, Stanford, CA 94305-4020, USA
[3]School of Mathematics & Statistics, The University of Melbourne, Parkville, VIC 3010, Australia;
E-mail: aram@unimelb.edu.au
[4]Department of Mathematics, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305-2125 USA;
E-mail: msimper@stanford.edu

Received: 12 October 2022; **Revised:** 16 November 2022; **Accepted:** 02 December 2022

2020 Mathematics Subject Classification: *Primary* – 60J10; *Secondary* – 20C08, 05E05

Abstract

Let G be a finite group. Let H, K be subgroups of G and $H\backslash G/K$ the double coset space. If Q is a probability on G which is constant on conjugacy classes ($Q(s^{-1}ts) = Q(t)$), then the random walk driven by Q on G projects to a Markov chain on $H\backslash G/K$. This allows analysis of the lumped chain using the representation theory of G. Examples include coagulation-fragmentation processes and natural Markov chains on contingency tables. Our main example projects the random transvections walk on $GL_n(q)$ onto a Markov chain on S_n via the Bruhat decomposition. The chain on S_n has a Mallows stationary distribution and interesting mixing time behavior. The projection illuminates the combinatorics of Gaussian elimination. Along the way, we give a representation of the sum of transvections in the Hecke algebra of double cosets, which describes the Markov chain as a mixture of Metropolis chains. Some extensions and examples of double coset Markov chains with G a compact group are discussed.

Contents

1. Introduction

This paper develops tools which allow projecting a random walk on a group to a Markov chain on special equivalence classes of the group. Fourier analysis on the group can then be harnessed to give sharp analysis of rates of convergence to stationarity for the Markov chain on equivalence classes. We begin with a motivating example.

Example 1.1 (Coagulation-fragmentation processes). In chemistry and physics, coagulation-fragmentation processes are models used to capture the behavior of 'blobs' that combine and break up over time. These processes are used in population genetics to model the merging and splitting of family groups. A simple mean field model considers n unlabeled particles in a partition $\lambda = (\lambda_1, \ldots, \lambda_k)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$, $\sum_{i=1}^{k} \lambda_i = n$. At each step of the process, a pair of particles is chosen uniformly at random and the partition evolves according to the rules:

1. If the particles are in distinct blocks, combine the blocks.
2. If the particles are in the same block, break the block uniformly into two blocks.
3. If the same particle is chosen twice, do nothing.

This defines a Markov chain on partitions of n. Natural questions are:

- What is the stationary distribution $\pi(\lambda)$?
- How does the process evolve?
- How long to reach stationarity?

All of these questions can be answered by considering the random transpositions process on the symmetric group S_n. The transition probabilities for this process are constant on conjugacy classes, the conjugacy classes are indexed by partitions and the conjugacy class containing the current permutation of the walk evolves as the coagulation-fragmentation process on partitions. The answers are (see section 2.3):

- The stationary distribution is $\pi(\lambda) = \prod_{i=1}^{n} 1/(i^{a_i} a_i!)$ for a partition λ with a_i parts of size i.
- Starting at $\lambda = 1^n$, the pieces evolve as the connected components of a growing Erdős-Rényi random graph process: Initially there are n vertices and no edges. Random edges are added to the graph so that the connected components correspond to the parts of a partition λ of n. This works as long as there are no repeated edges (i, j) (so for number of steps of smaller order than n). Repeated edges can be easily handled so that there is a tight connection between the growth of the random graph and the random transpositions Markov chain. In particular, random transpositions mix in order $(1/2)n \log n$, which is the same amount of time for the random graph to become connected. See [6] for more details.
- It takes order $n \log n$ steps to reach stationarity.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

The coagulation-fragmentation process is a special case of a double coset walk: Let G be a finite group, H, K subgroups of G. The equivalence relation

$$s \sim t \iff hsk = t, \quad h \in H, \quad k \in K$$

partitions the group into double cosets $H\backslash G/K$. Let $Q(s)$ be a probability measure on G. That is, $0 \le Q(s) \le 1$ for all $s \in G$ and $\sum_{s\in G} Q(s) = 1$. Further assume $Q(s)$ is a *class function*: It is constant on conjugacy classes, that is, $Q(s) = Q(t^{-1}st)$ for all $s, t \in G$. Then Q defines a random walk on G by repeated multiplication of random elements chosen according to Q. In other words, the random walk is induced by convolution, $Q^{*k}(s) = \sum_{t\in G} Q(t)Q^{*(k-1)}(st^{-1})$ and a single transition step has probability $P(x, y) = Q(yx^{-1})$.

This random walk induces a random process on the space of double cosets. While usually a function of a Markov chain is no longer a Markov chain, in this situation the image of the random walk on $H\backslash G/K$ is Markov. Section 2.2 proves the following general result. Throughout, we pick double coset representatives $x \in G$ and write x for HxK.

Theorem 1.2. *For $Q(s) = Q(t^{-1}st)$ a probability on G, the induced process on $H\backslash G/K$ is Markov with the following properties.*

1. *The transitions are*

$$P(x, y) = Q(HyKx^{-1}), \quad x, y \in H\backslash G/K.$$

2. *The stationary distribution is*

$$\pi(x) = \frac{|HxK|}{|G|}.$$

3. *If $Q(s^{-1}) = Q(s)$, then P is reversible with respect to π:*

$$\pi(x)P(x, y) = \pi(y)P(y, x).$$

Further, suppose Q is concentrated on a single conjugacy class $\mathcal{C}$ (that is, $Q(s) = \delta_{\mathcal{C}}(s)/|\mathcal{C}|$). Then the Markov chain P has the following properties.

1. *The eigenvalues of P are among the set*

$$\left\{\frac{\chi_\lambda(\mathcal{C})}{\chi_\lambda(1)}\right\}_{\lambda\in\widehat{G}},$$

   where $\widehat{G}$ is the set of all irreducible representations of G and χ_λ is the character of the irreducible representation indexed by λ.
2. *The multiplicity of $\chi_\lambda(\mathcal{C})/\chi_\lambda(1)$ is*

$$m_\lambda = \langle \chi_\lambda|_H, 1\rangle \cdot \langle \chi_\lambda|_K, 1\rangle,$$

   where $\langle \chi_\lambda|_H, 1\rangle$ is the number of times the trivial representation appears when χ_λ is restricted to H.
3. *For any time $\ell > 0$,*

$$\sum_{x\in H\backslash G/K} \pi(x)\|P_x^\ell - \pi\|_{TV}^2 \le \frac{1}{4}\sum_{\lambda\neq 1} m_\lambda \left|\frac{\chi_\lambda(\mathcal{C})}{\chi_\lambda(1)}\right|^{2\ell}.$$

This theorem shows that the properties of the induced chain are available via the character theory of G. It is proved with variations and extensions in Section 2.2. The main example is introduced next.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Example 1.3 ($GL_n(q)$ and Gaussian elimination). Fix a prime power q, and let $GL_n(q)$ be the invertible $n \times n$ matrices over $\mathbb{F}_q$. Let $H = K = \mathcal{B}$ be the Borel subgroup: upper-triangular matrices in $GL_n(q)$. A classical result is the Bruhat decomposition,

$$GL_n(q) = \bigsqcup_{\omega \in S_n} \mathcal{B}\omega\mathcal{B},$$

where ω is the permutation matrix for the permutation $\omega \in S_n$. This decomposition shows that the double cosets $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$ are indexed by permutations. As explained in Section 2.5 below, the permutation ω associated to $M \in GL_n(q)$ is the 'pivotal' permutation when M is reduced to upper-triangular form by row reduction (Gaussian elimination).

The set of *transvections* $\mathcal{T}_{n,q}$ is the conjugacy class containing the basic row operations $I + \theta E_{ij}$; here, $\theta \in \mathbb{F}_q$, and E_{ij} is the matrix with a 1 in position (i, j) and zeroes everywhere else (so $I + \theta E_{ij}$ acts by adding θ times row i to row j). Hildebrand [35] gave sharp convergence results for the Markov chain on $GL_n(q)$ generated by the class function Q which gives equal probability to all transvections. He shows that n steps are necessary and sufficient for convergence to the uniform distribution for any q. Of course, convergence of the lumped chain on $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$ might be faster. The results we found surprised us. Careful statements are included below. At a high level, we found:

- Starting from a 'typical' state $x \in S_n$, order $\log n/\log q$ steps are necessary and sufficient for convergence. This is an exponential speed-up from the original chain.
- Starting from id, order n steps are necessary and sufficient for convergence.
- Starting from ω_0, the reversal permutation, order $\log n/(2 \log q)$ steps are necessary and sufficient for convergence.

To simplify the statement of an honest theorem, let us measure convergence in the usual chi-square or L^2 distance:

$$\chi^2_x(\ell) = \sum_y \frac{\left(P^\ell(x, y) - \pi_q(y)\right)^2}{\pi_q(y)}.$$

In Section 3, the usual total variation distance is treated as well.

Theorem 1.4. *The random transvections walk on $GL_n(q)$ induces a Markov chain $P(x, y)$ on $S_n \cong \mathcal{B}\backslash GL_n(q)/\mathcal{B}$ with stationary distribution.*

$$\pi_q(\omega) = q^{I(\omega)}/[n]_q!, \quad \text{where } [n]_q! := (1 + q)(1 + q + q^2) \dots (1 + q + \dots + q^{n-1}),$$

for $\omega \in S_n$, where $I(\omega) = |\{(i, j) : i < j, \omega(i) > \omega(j)\}|$ is the number of inversions in ω.

Furthermore, if $\log q > 6/n$ then the following statements are true.

1. *(Typical start) If $\ell \geq (\log n + c)/(\log q - 6/n)$ for any $c > 0$, then*

$$\sum_{x \in S_n} \pi(x)\chi^2_x(\ell) \leq (e^{e^{-c}} - 1) + e^{-cn}.$$

Conversely, for any ℓ

$$\sum_{x \in S_n} \pi(x)\chi^2_x(\ell) \geq (n - 1)^2 q^{-4\ell}.$$

These results show order $\log_q(n)$ steps are necessary and sufficient for convergence.

2. *(Starting from id) If $\ell \geq (n \log q/2 + c)/(\log q - 6/n), c > 0$, then*

$$\chi^2_{id}(\ell) \leq (e^{e^{-2c}} - 1) + e^{-cn}.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Conversely, for any ℓ,

$$\chi^2_{id}(\ell) \geq (n-1)(q^{n-1}-1)q^{-4\ell}.$$

These results show that order n steps are necessary and sufficient for convergence starting from the identity.

3. *(Starting from ω_0) If $\ell \geq (\log n/2 + c)/(\log q - 6/n)$ for $c \geq 2\sqrt{2}$, then there is a universal constant $K > 0$ (independent of q, n) such that*

$$\chi^2_{\omega_0}(\ell) \leq -2K \log(1 - e^{-c}) + Ke^{-cn}.$$

Conversely, for any ℓ,

$$\chi^2_{\omega_0}(\ell) \geq q^{-(n-2)}(n-1)(q^{n-1}-1)q^{-4\ell}.$$

These results show that order $\log_q(n)/2$ steps are necessary and sufficient for convergence starting from ω_0.

Remark 1.5. Note that, while Hildebrand's result of order n convergence rate was independent of the parameter q, the rates in Theorem 1.4 depend on q.

The stationary distribution π_q is the Mallows measure on S_n. This measure has a large enumerative literature; see [24] Section 3 for a review or [64]. It is natural to ask what the induced chain 'looks like' on S_n. After all, the chain induced by random transpositions on partitions has a simple description and is of general interest. Is there a similarly simple description of the chain in S_n? This question is treated in Section 5 using the language of Hecke algebras.

Theorem 1.6. *Let $H_n(q)$ be the Hecke algebra corresponding to the $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$ double cosets and $D = \sum_{T \in \mathcal{T}_{n,q}} T \in H_n(q)$ be the sum of all transvections. Then,*

$$D = (n-1)q^{n-1} - [n-1]_q + (q-1) \sum_{1 \leq i < j \leq n} q^{n-1-(j-i)} T_{ij},$$

with T_{ij} in the Hecke algebra.

This gives a probabilistic description of the induced chain on S_n. Roughly stated, from $\omega \in S_n$ pick $(i, j), i < j$, with probability proportional to $q^{-(j-i)}$ and transpose i and j in ω using the Metropolis algorithm (reviewed in Section 2.6). This description is explained in Section 5; see also [18]. The probabilistic description is crucial in obtaining good total variation lower bounds for Theorem 1.4.

Outline

Section 2 develops and surveys background material on double cosets, Markov chains (proving Theorem 1.2), transpositions and coagulation-fragmentation processes, transvections and Gaussian elimination. Theorem 1.4 is proved in Section 3. Theorem 1.6 is proved in Section 5 using a row reduction. Section 6 contains another Markov chain from a lumping of the transvections chain, and Section 7 surveys further examples—contingency tables and extensions of the GL_n results to finite groups of Lie type, for which the Bruhat decomposition holds and there are natural analogs of transvections. Of course, there are an infinite variety of groups G, H, K, and we also indicate extensions to compact Lie groups.

We have posted a more leisurely, expository version of the present paper on the arXiv [19]. This contains more examples and proof details.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Notation

Throughout, q will be a prime power. For a positive integer n, define the quantities

$$[n]_q := \frac{q^n - 1}{q - 1} = q^{n-1} + q^{n-2} + \ldots + q + 1, \quad [n]_q! := [n]_q [n-1]_q \ldots [1]_q.$$

2. Background

This section gives the basic definitions and tools needed to prove our main results. Section 2.1 gives background on double cosets. In Section 2.2, Markov chains are reviewed and Theorem 1.2 is proved, along with extensions. Section 2.3 reviews the coagulation-fragmentation literature along with the random transpositions literature. Section 2.4 develops what we need about transvections, and Section 2.5 connects the Bruhat decomposition to Gaussian elimination. Finally, Section 2.6 reviews the Metropolis algorithm.

2.1. Double cosets

Let H, K be subgroups of a finite group G. The double coset decomposition is a standard tool of elementary group theory. The original proof of Sylow's first theorem uses double cosets, as does Mackey's basic theorems on decomposing restrictions of induced representations. The Hecke algebra $\mathrm{End}_G(G/H)$—the linear maps of the right H-invariant functions that commute with the action of G—has a basis induced by H-H double cosets. Hecke algebras are basic objects of study in modern number theory. For a detailed survey, see [24] or [14]. Double cosets can have very different sizes and [24], [48] develop a probabilistic and enumerative theory. For present applications, an explicit description of the double cosets is needed.

Example 2.1. Let S_λ, S_μ be parabolic subgroups of the symmetric group S_n. Here, $\lambda = (\lambda_1, \ldots, \lambda_I)$ and $\mu = (\mu_1, \ldots, \mu_J)$ are partitions of n. The subgroup S_λ consists of all permutations in which the first λ_1 elements may only be permuted amongst each other, the next $\lambda_1 + 1, \ldots, \lambda_1 + \lambda_2$ elements may only be permuted amongst each other and so on. It is a classical fact that the double cosets $S_\lambda \backslash S_n / S_\mu$ are in bijection with 'contingency tables'—arrays of nonnegative integers with row sums λ and column sums μ. See [38], Section 1.3. For proofs and much discussion of the connections between the group theory and applications and statistics, see [24], Section 5. Random transpositions on S_n induces a natural Markov chain on these tables, see [60] and Chapter 3 of [59]. Contingency tables also label the double cosets of parabolic subgroups in $GL(n, Q)$. See [39] and Section 7.1 below.

Example 2.2. Let M be a finite group, $G = M \times M$, and $H = K = M$ embedded diagonally as subgroups of G (that is, $\{(m, m) : m \in M\}$). The conjugacy classes in G are products of conjugacy classes in each coordinate of M. In the double coset equivalency classes, note that

$$(s, t) \sim (id, s^{-1}t) \sim (id, k^{-1}s^{-1}tk),$$

and so double cosets can be indexed by conjugacy classes of M. If Q_1 is a conjugacy invariant probability on M, then $Q = Q_1 \times \delta_{id}$ is conjugacy invariant on G. The random walk on G induced by Q maps to the random walk on M induced by Q_1. In this way, the double coset walks extend conjugacy invariant walks on M. Example 1.1 in the introduction is a special case.

Of course, the conjugacy classes in M (and so the double cosets) can be difficult to describe. Describing the conjugacy classes of $U_n(q)$—the unit upper-triangular matrices in $GL_n(q)$—is a well-known 'wild' problem. See [2] or [17] for background and details. Describing the double cosets for the Sylow p-subgroup in S_n seems difficult.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Example 2.3. Let G be a finite group of Lie type, defined over $\mathbb{F}_q$, with Weyl group W. Let $\mathcal{B}$ be the Borel subgroup (maximal solvable subgroup). Take $H = K = \mathcal{B}$. The Bruhat decomposition gives

$$G = \bigsqcup_{\omega \in W} \mathcal{B}\omega\mathcal{B},$$

so the double cosets are indexed by W. See [11], Chapter 8, for a clear development in the language of groups with a $(\mathcal{B}, N)$ pair.

Conjugacy invariant walks on G have been carefully studied in a series of papers by David Gluck, Bob Guralnick, Michael Larsen, Martin Liebeck, Aner Shalev, Pham Tiep and others. These authors develop good bounds on the character ratios needed. See [30] for a recent paper with careful reference to earlier work. Of course, Example 2 with $G = GL_n(q), W = S_n$ is a special case. The present paper shows what additional work is needed to transfer character ratio results from G to W.

The double cosets form a basis for the algebra of H-K bi-invariant functions $L(H\backslash G/K)$ with product

$$f \star g(s) = \sum_t f(t)g(st^{-1}).$$

This is usually developed for $H = K$ [14], [12], [15], but the extra flexibility is useful. We add a caveat: When $H = K$, the algebra of bi-invariant functions (into $\mathbb{C}$) is semisimple and with a unit. This need not be the case for general H and K. David Craven tried many pairs of subgroups of S_4 and found examples which were not semisimple. For $G = S_4$, Marty Issacs produced the example H the cyclic subgroup generated by (1234) and K the cyclic subgroup generated by (1243). The algebra doesn't have a unit and so cannot be semisimple. This occurs even for some pairs of distinct parabolic subgroups of S_n. There are also distinct pairs of parabolics where the algebra is semisimple. Determining when this occurs is an open question.

Further examples are in Section 7. Since the theory is developed for general H, K, G there is a large set of possibilities. What is needed are examples where the double cosets are indexed by familiar combinatorial objects and the walks induced on $H\backslash G/K$ are of independent interest.

2.2. Markov chain theory

Let H, K be subgroups of a finite group G, and Q a probability on G. See [43] for an introduction to Markov chains; see [15] or [53] for random walks on groups.

Proposition 2.4. *Let Q be a probability on G which is H-conjugacy invariant ($Q(s) = Q(h^{-1}sh)$ for $h \in H, s \in G$). The image of the random walk driven by Q on G maps to a Markov chain on $H\backslash G/K$ with transition kernel*

$$P(x, y) = Q(HyKx^{-1}).$$

The stationary distribution of P is $\pi(x) = |HxK|/|G|$. If $Q(s) = Q(s^{-1})$, then (P, π) is reversible.

Proof. The kernel P is well defined; that is, it is independent of the choice of double coset representatives for x, y. Dynkin's criteria ([40] Chapter 6, [49]) says that the image of a Markov chain in a partitioning of the state space is Markov if and only if for any set in the partition and any point in a second set, the chance of the original chain moving from the point to the first set is constant for points in the second set.

Fixing x, y, observe

$$Q(HyK(hxk)^{-1}) = Q(HyKx^{-1}h^{-1}) = Q(HyKx^{-1}).$$

Since the uniform distribution on G is stationary for the walk generated by Q, the stationary distribution of the lumped chain is $\pi(x) = |HxK|/|G|$. Finally, any function of a reversible chain is reversible and $Q(s) = Q(s^{-1})$ gives reversibility of the walk on G. □

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Remark 2.5. A different sufficient condition for Proposition 2.4 is $Q(sh) = Q(s)$ for all $s \in G, h \in H$.

Remark 2.6. Usually, a function of a Markov chain is *not* Markov. For relevant discussion of similar 'orbit chains', see [9].

In all of our examples, the measure Q is a class function ($Q(s) = Q(t^{-1}st)$ for all $s, t \in G$), which is a stronger requirement than that in Proposition 2.4. The eigenvalues of the walk on G can be given in terms of the irreducible complex characters of G. Let $\widehat{G}$ be an index set for these characters. We write $\lambda \in \widehat{G}$ and $\chi_\lambda(\mathcal{C})$ for the character value at the conjugacy class $\mathcal{C}$. Let

$$\beta_\lambda = \frac{1}{\chi_\lambda(1)} \sum_{s \in G} Q(s)\chi_\lambda(s).$$

If Q is simply concentrated on a single conjugacy class $\mathcal{C}$, then β_λ is the character ratio

$$\beta_\lambda = \frac{\chi_\lambda(\mathcal{C})}{\chi_\lambda(1)}.$$

For a review of a large relevant literature on character ratios and their applications, see [30].

The restriction of χ_λ to H is written $\chi_\lambda|_H$ and $\langle \chi_\lambda|_H, 1\rangle$ is the number of times the trivial representation of H appears in $\chi_\lambda|_H$. By reciprocity, this is $\left\langle \chi_\lambda, \mathrm{Ind}_H^G(1)\right\rangle$, where Ind_H^G is the induced representation from H to G.

Proposition 2.7. *Let Q be a class function on G. The induced chain $P(x, y)$ of Proposition 2.4 has eigenvalues*

$$\beta_\lambda = \frac{1}{\chi_\lambda(1)} \sum_{s \in G} Q(s)\chi_\lambda(s), \tag{2.1}$$

with multiplicity

$$m_\lambda = \langle \chi_\lambda|_H, 1\rangle \cdot \langle \chi_\lambda|_K, 1\rangle. \tag{2.2}$$

The average square total variation distance to stationarity satisfies

$$\sum_x \pi(x)\|P_x^\ell - \pi\|_{TV}^2 \le \frac{1}{4} \sum_{\lambda \in \widehat{G}, \lambda \neq 1} m_\lambda \beta_\lambda^{2\ell}. \tag{2.3}$$

Proof. The eigenvalues of a lumped chain are always some subset of the eigenvalues of the original chain. To determine the multiplicity of the eigenvalue β_λ in the lumped chain, fix $\lambda : G \to GL_{d_\lambda}$ an irreducible representation of G. Let M^λ be the $d_\lambda \times d_\lambda$ matrix representation of λ. That is, each entry $M_{ij}^\lambda : G \to \mathbb{C}$ is a function of G. These functions are linearly independent and can be chosen to be orthogonal with respect to

$$\langle f_1, f_2\rangle := \frac{1}{|G|} \sum_{g \in G} f_1(g)\overline{f_2}(g)$$

(see Chapter 3 of [58]). Let V_λ be the space of all linear combinations of the functions M_{ij}^λ. If $f \in V_\lambda$, then

$$Pf(x) = \sum_{y \in G} P(x, y)f(y) = \beta_\lambda f(x)$$

That is, V_λ is the eigenspace for the eigenvalue β_λ and it has dimension $d_\lambda^2 = \chi_\lambda(1)^2$.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

In the lumped chain on $H\backslash G/K$, a basis for the eigenspace for eigenvalue β_λ are the $H\times K$ invariant functions in V_λ [9]. To determine the dimension of this subspace, note that $G\times G$ can act on V_λ by $f^{g_1,g_2}(x)=f(g_1^{-1}xg_2)$. This gives a representation of $G\times G$ on V_λ. The matrix of this representation is isomorphic to $M\otimes M$, since $M_{ij}(s^{-1}tu)=M_{ij}(s^{-1})M_{ij}(t)M_{ij}(u)$.

This representation restricts to a representation $M_H\otimes M_K$ of $H\times K$, and the dimension of the $H\times K$ invariant functions in V_λ is the multiplicity of the trivial representation on $M_H\otimes M_K$. This is

$$\langle \chi_\lambda|_H\otimes\chi_\lambda|_K,1\rangle=\langle\chi_\lambda|_H,1\rangle\cdot\langle\chi_\lambda|_K,1\rangle.$$

To note the total variation inequality, let $1=\beta_1\geq\beta_2\geq\cdots\geq\beta_n\geq\cdots\geq\beta_{|S_n|}\geq-1$ be the eigenvalues with eigenfunctions f_j (chosen to be orthonormal with respect to π), and we have

$$\chi_x^2(\ell):=\|\frac{P_x^\ell}{\pi}-1\|_{2,\pi}^2=\sum_{j\neq1}f_j(x)^2\beta_j^{2\ell}$$

where $\|\cdot\|_{2,\pi}$ denotes the ℓ^2 norm with respect to the distribution π. Multiplying by $\pi(x)$ and summing over all x in the state space gives

$$\sum_x\pi(x)\left\|\frac{P_x^\ell}{\pi}-1\right\|_{2,\pi}^2=\sum_x\pi(x)\sum_{j\neq1}f_j(x)^2\beta_j^{2\ell}=\sum_{\lambda\neq1}m_\lambda\beta_\lambda^{2\ell},\tag{2.4}$$

using orthonormality of f_j. The total variation bound arises since $4\|P_x^\ell-\pi\|_{TV}^2\leq\|P_x^\ell/\pi-1\|_{2,\pi}^2$. □

2.3. Random transpositions and coagulation-fragmentation processes

Let $G=S_n$ be the symmetric group. The random transpositions Markov chain, studied in [22], is generated by the measure

$$Q(\omega)=\begin{cases}1/n, & \omega=id,\\ 2/n^2, & \omega=(ij),\\ 0, & \text{otherwise.}\end{cases}$$

This was the first Markov chain where a sharp cutoff for convergence to stationarity was observed. A sharp, explicit rate is obtained in [54]. They show

$$\|Q^\ell-u\|_{TV}\leq 2e^{-c},\quad\text{with }\ \ell=\frac{1}{2}n(\log n+c).$$

The asymptotic 'profile' (the limit of $\|Q^\ell-u\|_{TV}$ as a function of c for n large) is determined in [63]. Schramm [56] found a sharp parallel between random transpositions and the growth of an Erdős-Rényi random graph: Given vertices $1,2,\ldots,n$, for each transposition (i,j) chosen, add an edge from vertices i to j to generate a random graph. See [6] for extensions and a comprehensive review. The results, translated by the coagulation-fragmentation description of the cycles, give a full and useful picture for the simple mean field model described in the introduction.

It must be emphasized that this mean field model is a very special case of coagulation-fragmentation models studied in the chemistry-physics-probability literature. These models study the dynamics of particles diffusing in an ambient space, and allow general collision kernels (e.g., particles close in space may be more likely to join). The books by Bertoin [7] and Pitman [50] along with the survey paper of Aldous [3] are recommended for a view of the richness of this subject. On the other hand, the sharp rates of convergence results available for the mean field model are not available in any generality.

There is a healthy applied mathematics literature on coagulation-fragmentation. A useful overview which treats discrete problems such as the ones treated here is [5]. A much more probabilistic

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

development of the celebrated Becker–Doring version of the problem is in [36]. This develops rates of convergence using coupling. See also [27] for more models with various stationary distributions on partitions.

Other lumpings of random transpositions include classical urn models—the Bernoulli–Laplace model [23], [28], and random walks on phylogenetic trees [26]. The sharp analysis of random transpositions transfers, via comparison theory, to give good rates of convergence for quite general random walks on the symmetric group [20], [34]. For an expository survey, see [16].

2.4. Transvections

Fix n, a prime p and $q = p^a$ for some positive integer a. A *transvection* is an invertible linear transformation of $\mathbb{F}_q^n$ which fixes a hyperplane, is not the identity and has all eigenvalues equal to 1. Transvections are convenient generators for the group $SL_n(q)$ because they generalize the basic row operations of linear algebra. These properties are carefully developed in [62] Chapter 1, 9; [4] Chapter 4.

Using coordinates, let $\mathbf{a}, \mathbf{v} \in \mathbb{F}_q^n$ be two nonzero vectors with $\mathbf{a}^\top \mathbf{v} = 0$. A transvection, denoted $T_{\mathbf{a},\mathbf{v}} \in GL_n(q)$ is the linear map given by

$$T_{\mathbf{a},\mathbf{v}}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{a}^\top \mathbf{x}), \quad \text{for } \mathbf{x} \in \mathbb{F}_q^n.$$

It adds a multiple of $\mathbf{v}$ to $\mathbf{x}$, the amount depending on the 'angle' between $\mathbf{a}$ and $\mathbf{x}$. As a matrix, $T_{\mathbf{a},\mathbf{v}} = I + [a_1\mathbf{v}a_2\mathbf{v}\dots a_n\mathbf{v}] = I + \mathbf{v}\mathbf{a}^\top$. Multiplying $\mathbf{a}$ by a nonzero constant and dividing $\mathbf{v}$ by the same constant doesn't change $T_{\mathbf{a},\mathbf{v}}$. Let us agree to normalize $\mathbf{v}$ by making its last nonzero coordinate equal to 1. Let $\mathcal{T}_{n,q} \subset SL_n(q)$ be the set of all transvections.

An elementary count shows

$$|\mathcal{T}_{n,q}| = \frac{(q^n - 1)(q^{n-1} - 1)}{q - 1}. \tag{2.5}$$

It is easy to generate $T \in \mathcal{T}_{n,q}$ uniformly: Pick $\mathbf{v} \in \mathbb{F}_q^n$ uniformly, discarding the zero vector. Normalize $\mathbf{v}$ so the last nonzero coordinate, say index j, is equal to 1. Pick $a_1, a_2, \dots, a_{j-1}, a_{j+1}, \dots a_n$ uniformly in $\mathbb{F}_q^{n-1} - \{0\}$, and set a_j so that $\mathbf{a}^\top \mathbf{v} = 0$. The transvection $T_{\mathbf{a},\mathbf{v}}$ fixes the hyperplane $\{\mathbf{x} : \mathbf{a}^\top \mathbf{x} = 0\}$.

Example 2.8. Taking $\mathbf{v} = \mathbf{e}_1, \mathbf{a} = \mathbf{e}_2$ gives the transvection with matrix

$$T_{\mathbf{e}_1,\mathbf{e}_2} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

This acts on $\mathbf{x}$ by adding the second coordinate to the first. Similarly, the basic row operation of adding θ times the ith coordinate to the jth is given by $T_{\mathbf{e}_j,\theta\mathbf{e}_i}$.

Lemma 2.9. *The set of transvections $\mathcal{T}_{n,q}$ is a conjugacy class in $GL_n(q)$.*

Proof. Let $M \in GL_n(q)$, so $MT_{\mathbf{e}_2,\mathbf{e}_1}M^{-1}$ is conjugate to $T_{\mathbf{e}_2,\mathbf{e}_1}$. Then,

$$MT_{\mathbf{e}_2,\mathbf{e}_1}M^{-1}(\mathbf{x}) = M(M^{-1}\mathbf{x} + (M^{-1}\mathbf{x})_2\mathbf{e}_1) = \mathbf{x} + (M^{-1}\mathbf{x})_2 M\mathbf{e}_1.$$

Let $\mathbf{a}$ be the second column of $(M^{-1})T_{\mathbf{e}_2,\mathbf{e}_1}$ and $\mathbf{v}$ the first column of M, and check this last is $T_{\mathbf{a},\mathbf{v}}$ (and $\mathbf{a}^\top \mathbf{v} = 0$). Thus, transvections form a conjugacy class. □

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

2.5. Gaussian elimination and the Bruhat decomposition

The reduction of a matrix $M \in GL_n(q)$ to standard form by row operations is a classical topic in introductory linear algebra courses. It gives efficient, numerically stable ways to solve linear equations, compute inverses and calculate determinants. There are many variations.

Example 2.10. Consider the sequence of row operations

$$M = \begin{pmatrix} 0 & 3 & 2 \\ 1 & 2 & 0 \\ 3 & 0 & 5 \end{pmatrix} \to \begin{pmatrix} 0 & 3 & 2 \\ 1 & 2 & 0 \\ 0 & -6 & 5 \end{pmatrix} \to \begin{pmatrix} 0 & 3 & 2 \\ 1 & 2 & 0 \\ 0 & 0 & 9 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 2 \\ 0 & 0 & 9 \end{pmatrix} = U.$$

The first step subtracts 3 times row 2 from row 3, multiplication by

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix}.$$

The second step adds 2 times row 1 to row 3, multiplication by

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

The third (pivot) step brings the matrix to upper-triangular form by switching rows 1 and 2, which corresponds to multiplication by the matrix by

$$\omega_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This gives $\omega_1 L_2 L_1 M = U \implies M = L_2^{-1} L_1^{-1} \omega_1^{-1} U = L\omega U$ with $L = L_2^{-1} L_1^{-1}, \omega = \omega_1^{-1} = \omega_1$.

If $\mathcal{L}, \mathcal{B}$ are the subgroups of lower and upper-triangular matrices in $GL_n(q)$, this gives

$$GL_n(q) = \bigsqcup_{\omega \in S_n} \mathcal{L}\omega\mathcal{B}. \tag{2.6}$$

Any linear algebra book treats these topics. A particularly clear version which uses Gaussian elimination as a gateway to Lie theory is in Howe [37]. Articles by Lusztig [45] and Strang [61] have further historical, mathematical and practical discussion.

Observe that carrying out the final pivoting step costs $d_c(\omega, \mathrm{id})$ operations, where $d_c(\omega, \mathrm{id})$, the Cayley distance of ω to the identity, is the minimum number of transpositions required to sort ω (with arbitrary transpositions (i, j) allowed). Cayley proved $d_c(\omega, \mathrm{id}) = n - \#\text{cycles in } \omega$ (see [16]). In the example above $n = 3$, $\omega = 213$ has two cycles and $3 - 2 = 1$—one transposition sorts ω.

How many pivot steps are needed 'on average'? This becomes the question of the number of cycles in a pick from Mallows measure π_q. Surprisingly, this is a difficult question. Following partial answers by Gladkich and Peled [29], this problem was recently solved by Jimmy He, Tobias Möller and Teun Verstraaten in [33]. They show that, when $q > 1$, the limiting behavior of the number of even cycles under π_q has an approximate normal distribution with mean and variance proportional to n, and that the number of odd cycles has bounded mean and variance.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

The Bruhat decomposition
In algebraic group theory, one uses

$$GL_n(q) = \bigsqcup_{\omega \in S_n} \mathcal{B}\omega\mathcal{B}. \tag{2.7}$$

This holds for any semisimple group over any field with $\mathcal{B}$ replaced by the Borel group (the largest solvable subgroup) and S_n replaced by the Weyl group.

Let $\omega_0 = \left(\begin{smallmatrix} 1 & 2 & \cdots & n \\ n & n-1 & \cdots & 1 \end{smallmatrix}\right)$ be the reversal permutation in S_n. Since $\mathcal{L} = \omega_0\mathcal{B}\omega_0$, equation (2.7) is equivalent to the LU decomposition (2.6). Given $M \in GL_n(q)$, Gaussian elimination on $\omega_0 M$ can be used to find $\omega_0 M \in L\omega' U$ and thus $M \in B\omega B$ with $\omega = \omega_0\omega'$.

The subgroup $\mathcal{B}$ gives rise to the quotient $GL_n(q)/\mathcal{B}$. This may be pictured as the space of 'flags'. Here, a flag F consists of an increasing sequence of subspaces $F = F_1 \subset F_2 \subset \ldots \subset F_n$ with $\dim(F_i) = i$. Indeed, $GL_n(q)$ operates transitively on flags and the subgroup fixing the *standard flag* $E = \langle \mathbf{e}_1 \rangle \subset \langle \mathbf{e}_1, \mathbf{e}_2 \rangle \ldots \subset \langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$ is exactly $\mathcal{B}$. This perspective will be further explained and used in Section 6 to study a function of the double coset Markov chain on $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$.

Remark 2.11. The double cosets of $GL_n(q)$ define equivalence classes for any subgroup of $GL_n(q)$. For the matrices $SL_n(q)$ with determinant 1, these double cosets again induce the Mallows distribution on permutations. More precisely, for $x \in SL_n(q)$, let $[x]_{SL_n(q)} = \{x' \in SL_n(q) : x' \in \mathcal{B}x\mathcal{B}\}$ be the equivalence class created by the double coset relation $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$, within $SL_n(q)$. Note that two matrices $x, x' \in SL_n(q)$ could be in the same double coset with $x' = b_1 x b_2$, but $b_1, b_2 \notin SL_n(q)$ (necessarily, $\det(b_1) = \det(b_2)^{-1}$).

Then, $|[\omega]_{SL_n(q)}|/|SL_n(q)| = p_q(\omega)$. This follows since $|GL_n(q)| = (q-1) \cdot |SL_n(q)|$, and $|\mathcal{B}\omega\mathcal{B}| = (q-1) \cdot |[\omega]_{SL_n(q)}|$. If $M \in GL_n(q)$ and $M \in B\omega B$, then $M/\det(M) \in [\omega]_{SL_n(q)}$. Conversely, for each $M \in SL_n(q)$ there are $(q-1)$ unique matrices in $GL_n(q)$ created by multiplying M by $1, 2, \ldots, q-1$.

2.6. The Metropolis algorithm

The Metropolis algorithm is a basic algorithm of scientific computing which arises in describing the random walk induced by transvections on the double cosets $\mathcal{B}_n\backslash GL_n(q)/\mathcal{B}_n$ (Section 3.2). This section gives background.

Given a probability distribution π on a space $\mathcal{X}$, the Metropolis algorithm gives a way of changing the output of a Markov chain with transition matrix $K(x, y)$ to have stationary distribution π on $\mathcal{X}$. For simplicity, suppose the original chain is symmetric, $K(x, y) = K(y, x)$ (as in our examples). This implies that $K(x, y)$ has a uniform stationary distribution. Define the Metropolis Markov chain with the transition matrix:

$$M(x,y) = \begin{cases} K(x,y), & \text{if } \pi(y) \geq \pi(x),\ x \neq y, \\ K(x,y) \cdot \frac{\pi(y)}{\pi(x)}, & \text{if } \pi(y) < \pi(x),\ x \neq y, \\ K(x,x) + \sum_{z:\pi(z)<\pi(x)} K(x,z)\left(1 - \frac{\pi(z)}{\pi(x)}\right), & \text{if } x = y. \end{cases}$$

These transition probabilities have a simple implementation: From x, pick y according to $K(x, y)$. If $\pi(y) \geq \pi(x)$, move to y. If $\pi(y) < \pi(x)$, flip a coin with heads probability $\pi(y)/\pi(x)$. If the coin is heads, move to y. If the coin is tails, stay at x. Elementary calculations show that $\pi(x)M(x, y) = \pi(y)M(y, x)$, that is, M has π as stationary distribution. Note that the normalizing constant of π is not needed.

For background, applications, and theoretical properties of the Metropolis algorithm, see the textbook of Liu [44] or the survey [21]. Sharp analysis of rates of convergence of the Metropolis algorithm is still an open research problem. The special cases developed in Section 3.2 show it can lead to fascinating mathematics.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

3. Double coset walks on $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$

Throughout this section, $\mathcal{B}$ is the group of upper-triangular matrices in $GL_n(q)$, $\mathcal{T}_{n,q}$ is the conjugacy class of transvections in $GL_n(q)$. This gives the probability measure on $GL_n(q)$ defined by

$$Q(M) = \begin{cases} \frac{1}{|\mathcal{T}_{n,q}|}, & \text{if } M \in \mathcal{T}_{n,q}. \\ 0, & \text{else.} \end{cases}$$

Note the random transvections measure Q is supported on $SL_n(q)$, a subgroup of $GL_n(q)$. This means that the random walk on $GL_n(q)$ driven by Q is not ergodic (there is zero probability of moving x to y if x, y are matrices with different determinants). However, Q is a class function on $GL_n(q)$ since transvections form a conjugacy class. The image of the uniform distribution on $GL_n(q)$ mapped to $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$ is the Mallows measure $\pi_q(\omega) = q^{I(\omega)}/[n]_q!$.

Section 3.1 introduces the definition of the Markov chain as multiplication in the Hecke algebra, which is further explained in Section 5. Section 3.3 gives the combinatorial expressions for the eigenvalues and their multiplicities, which are needed to apply Theorem 1.2 for this case. Section 4.2 shows that for the induced Markov chain on S_n, starting from $id \in S_n$, order n steps are necessary and sufficient for convergence. Section 4.3 studies the chain starting from the reversal permutation, for which only order $\log n/2 \log q$ steps are required. Finally, Section 4.4 considers starting from a 'typical' element, according to the stationary distribution, for which $\log n/\log q$ steps are necessary and sufficient.

These results can be compared to Hildebrand's Theorem 1.1 [35] which shows that the walk driven by Q on $GL_n(q)$ converges in $n + c$ steps (uniformly in q). Our results thus contribute to the program of understanding how functions of a Markov chain behave and how the mixing time depends on the starting state. In this case, changing the starting state gives an exponential speed up.

3.1. Hecke algebras and the Metropolis algorithm

The set of $\mathcal{B}$-$\mathcal{B}$ double cosets of $GL_n(q)$ has remarkable structure. For $\omega \in S_n$, let $T_\omega = \mathcal{B}\omega\mathcal{B}$. Linear combinations of double cosets form an algebra (over $\mathbb{C}$, for example).

Definition 3.1. The *Iwahori–Hecke algebra* $H_n(q)$ is spanned by the symbols $\{T_\omega\}_{\omega \in S_n}$ and generated by $T_i = T_{s_i}$ for $s_i = (i, i+1), 1 \le i \le n-1$, with the relations

$$\begin{cases} T_{s_i} T_\omega = T_{s_i \omega} & \text{if } I(s_i\omega) = I(\omega) + 1, \\ T_{s_i} T_\omega = q T_{s_i \omega} + (q-1) T_\omega & \text{if } I(s_i\omega) = I(\omega) - 1, \end{cases} \tag{3.1}$$

where $I(\omega)$ is the usual length function on S_n ($I(s_i\omega) = I(\omega) \pm 1$).

Consider the flag space $\mathcal{F} = G/\mathcal{B}$. The group $GL_n(q)$ acts on the left on $\mathcal{F}$. One can see $H_n(q)$ acting on the right of $\mathcal{F}$ and in fact

$$H_n(q) = \mathrm{End}_{GL_n(q)}(GL_n(q)/\mathcal{B}).$$

The Hecke algebra is the full commuting algebra of $GL_n(q)$ acting on $GL_n(q)/\mathcal{B}$. Because transvections form a conjugacy class, the sum of transvections is in the center of the group algebra $\mathbb{C}[GL_n(q)]$, and so it may be regarded as an element of $H_n(q)$. This will be explicitly delineated and the character theory of $H_n(q)$ used to do computations.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

3.2. The Metropolis connection

The relations (3.1) can be interpreted probabilistically. Consider what equation (3.1) says as linear algebra: Left multiplication by T_{s_i} can take ω to ω or $s_i\omega$. The matrix of this map (in the basis $\{T_\omega\}_{\omega\in S_n}$) has ω, ω' entry

$$\begin{cases} 1, & \text{if} \quad I(s_i\omega) = I(\omega) + 1, \text{ and } \omega' = s_i\omega, \\ q, & \text{if} \quad I(s_i\omega) = I(\omega) - 1, \text{ and } \omega' = \omega, \\ q-1, & \text{if} \quad I(s_i\omega) = I(\omega) - 1, \text{ and } \omega' = s_i\omega, \\ 0, & \text{otherwise.} \end{cases}$$

For example, on $GL_3(q)$ using the ordered basis $T_{id}, T_{s_1}, T_{s_2}, T_{s_1s_2}, T_{s_2s_1}, T_{s_1s_2s_1}$, the matrix of left multiplication by s_1 is

$$T_{s_1} = \begin{pmatrix} 0 & q & 0 & 0 & 0 & 0 \\ 1 & q-1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q & 0 & 0 \\ 0 & 0 & 1 & q-1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q \\ 0 & 0 & 0 & 0 & 1 & q-1 \end{pmatrix}.$$

The first column has a 1 in row s_1 because $I(s_1) > I(id)$. The second column has entries q and $q-1$ in the first two rows because $I(s_1^2) = I(id) < I(s_1)$.

We can also write the matrices for multiplication defined by T_{s_2} and $T_{s_1s_2s_1}$ as

$$T_{s_2} = \begin{pmatrix} 0 & 0 & q & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q & 0 \\ 1 & 0 & q-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q \\ 0 & 1 & 0 & 0 & q-1 & 0 \\ 0 & 0 & 0 & 1 & 0 & q-1 \end{pmatrix}$$

$$T_{s_1s_2s_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & q^3 \\ 0 & 0 & 0 & q^2 & 0 & q^2(q-1) \\ 0 & 0 & 0 & 0 & q^2 & q^2(q-1) \\ 0 & q & 0 & q(q-1) & q(q-1) & q(q-1)^2 \\ 0 & 0 & q & q(q-1) & q(q-1) & q(q-1)^2 \\ 1 & q-1 & q-1 & (q-1)^2 & (q-1)^2 & (q-1)^3 + q(q-1) \end{pmatrix}.$$

Observe that all three matrices above have constant row sums (q, q and q^3, respectively). Dividing by these row sums gives three Markov transition matrices: M_1, M_2 and $M_{121} = M_1M_2M_1$.

These matrices have a simple probabilistic interpretation: Consider, for $\overline{q} = 1/q$, the matrix defined

$$M_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \overline{q} & 1-\overline{q} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \overline{q} & 1-\overline{q} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \overline{q} & 1-\overline{q} \end{pmatrix}.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

The description of this Markov matrix is: From ω, propose $s_1\omega$:

- If $I(s_1\omega) > I(\omega)$, go to $s_1\omega$.
- If $I(s_1\omega) < I(\omega)$, go to $s_1\omega$ with probability $1/q$, else stay at ω.

This is exactly the Metropolis algorithm on S_n for sampling from $\pi_q(\omega)$ with the proposal given by the deterministic chain 'multiply by s_1'. The matrices $M_i, 1 \le i \le n-1$, have a similar interpretation and satisfy

$$\pi_q(\omega)M_i(\omega,\omega') = \pi_q(\omega')M_i(\omega',\omega).$$

The Metropolis algorithm always results in a reversible Markov chain. See Section 2.6 or [21] for background. It follows that any product of $\{M_i\}$ and any convex combination of such products yields a π_q reversible chain. Note also that the Markov chain on S_n is automatically reversible since it is induced by a reversible chain on GL_n.

Corollary 3.2. *The random transvections chain on $GL_n(q)$ lumped to $\mathcal{B}$-$\mathcal{B}$ cosets gives a π_q reversible Markov chain on S_n.*

Proof. Up to normalization, the matrix D in Theorem 1.6 is a positive linear combination of Markov chains corresponding to multiplication by

$$T_{(i,j)} = T_{s_i}T_{s_{i+1}}\cdots T_{s_{j-2}}T_{s_{j-1}}T_{s_{j-2}}\cdots T_{s_{i+1}}T_{s_i}.$$

This yields a combination of the reversible chains $M_iM_{i+1}\ldots M_{j-1}\ldots M_i$. □

Example 3.3. The transition matrix of the transvections chain on $GL_3(q)$ lumped to S_3 is $\frac{1}{|\mathcal{T}_{n,q}|}D$, with

$$D = (2q^2-q-1)\mathbf{I} + (q-1)\begin{pmatrix} 0 & q^2 & q^2 & 0 & 0 & q^3 \\ q & q(q-1) & 0 & q^2 & q^2 & q^2(q-1) \\ q & 0 & q(q-1) & q^2 & q^2 & q^2(q-1) \\ 0 & q & q & 2q(q-1) & q(q-1) & q^2+q(q-1)^2 \\ 0 & q & q & q(q-1) & 2q(q-1) & q^2+q(q-1)^2 \\ 1 & q-1 & q-1 & q+(q-1)^2 & q+(q-1)^2 & (q-1)^3+3q(q-1) \end{pmatrix}.$$

When $q = 2$, the lumped chain has transition matrix

$$\frac{1}{21}\begin{pmatrix} 5 & 4 & 4 & 0 & 0 & 8 \\ 2 & 7 & 0 & 4 & 4 & 4 \\ 2 & 0 & 7 & 4 & 4 & 4 \\ 0 & 2 & 2 & 9 & 2 & 6 \\ 0 & 2 & 2 & 2 & 9 & 6 \\ 1 & 1 & 1 & 3 & 3 & 12 \end{pmatrix}.$$

We report that this example has been verified by several different routes including simply running the transvections chain, computing the double coset representative at each step and estimating the transition rates from a long run of the chain.

Remark 3.4. The random transvections Markov chain on S_n is the 'q-deformation' of random transpositions on S_n. That is, as q tends to 1, the transition matrix tends to the transition matrix of random

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

transpositions. To see this, recall $|\mathcal{T}_{n,q}| = (q^n - 1)(q^{n-1} - 1)/(q - 1)$ and use L'Hopitals rule to note, for any integer k,

$$\lim_{q \to 1^+} \frac{(q-1)^2 q^k (q-1)^\ell}{(q^n - 1)(q^{n-1} - 1)} = \begin{cases} 0, & \text{if } \ell > 0, \\ \frac{1}{n(n-1)}, & \text{if } \ell = 0. \end{cases}$$

The interpretation of multiplication on the Hecke algebra as various 'systematic scan' Markov chains is developed in [18], [10]. It works for other types in several variations. We are surprised to see it come up naturally in the present work.

The following corollary provides the connection to [52, (3.16),(3.18),(3.20)] and [18, Proposition 4.9].

Corollary 3.5. *Let $J_1 = 1$, and let $J_k = T_{s_{k-1}} \cdots T_{s_2} T_{s_1} T_{s_1} T_{s_2} \cdots T_{s_{k-1}}$, for $k \in \{2, \ldots, n\}$. Then*

$$D = \sum_{k=1}^{n} q^{n-k} (J_k - 1).$$

Proof. Using that $J_k = T_{s_{k-1}} J_{k-1} T_{s_{k-1}}$, check, by induction, that

$$J_k = q^{k-1} + (q-1) \sum_{j=1}^{k-1} q^{j-1} T_{(j,k)}.$$

Thus,

$$\begin{aligned} \sum_{k=1}^{n} q^{n-k} (J_k - 1) &= 0 + \sum_{k=2}^{n} \left(q^{n-k+(k-1)} - q^{n-k} + (q-1) \sum_{j=1}^{k-1} q^{n-k+j-1} T_{(j,k)} \right) \\ &= (n-1) q^{n-1} - [n-1]_q + (q-1) D_{(21^{n-2})} = D, \end{aligned}$$

where $D_{(21^{n-2})} := \sum_{i<j} q^{(n-1)-(j-i)} T_{(i,j)}$. □

3.3. Eigenvalues and multiplicities

Hildebrand [35] determined the eigenvalues of the random walk driven by Q on $GL_n(q)$. His arguments use Macdonald's version of J.A. Green's formulas for the characters of $GL_n(q)$ along with sophisticated use of properties of Hall–Littlewood polynomials. Using the realization of the walk on the Hecke algebra, developed below in Section 5, and previous work of Ram and Halverson [31], we can find cleaner formulas and proofs. Throughout, we have tried to keep track of how things depend on both q and n (the formulas are easier when $q = 2$).

Theorem 3.6.

1. *The eigenvalues β_λ of the Markov chain $P(x, y)$ driven by the random transvections measure Q on $\mathcal{B} \backslash GL_n(q) / \mathcal{B}$ are indexed by partitions $\lambda \vdash n$. These are*

$$\beta_\lambda = \frac{1}{|\mathcal{T}_{n,q}|} \left(q^{n-1} \sum_{b \in \lambda} q^{ct(b)} - \frac{q^n - 1}{q - 1} \right), \tag{3.2}$$

with b ranging over the boxes of the partition λ. For the box in row i and column j, the content is defined as $ct(b) = j - i$, as in [46].

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

2. *The multiplicity of β_λ for the induced Markov chain on $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$ is f_λ^2, where*

$$f_\lambda = \frac{n!}{\prod_{b\in\lambda} h(b)}. \tag{3.3}$$

Here, $h(b)$ is the hook length of box b [46].

3. *The multiplicity of β_λ for the Markov chain induced by Q on $GL_n(q)/\mathcal{B}$ is*

$$d_\lambda = f_\lambda \cdot \frac{[n]_q!}{\prod_{b\in\lambda}[h(b)]_q}. \tag{3.4}$$

The argument uses the representation of the Markov chain as multiplication on the Hecke algebra. This is developed further in Section 5.

Proof. **(a):** Let D be the sum of transvections as in Theorem 1.6. By [31] (3.20), the irreducible representation H_n^λ of H_n indexed by λ has a 'seminormal basis' $\{v_T \mid \widehat{S}_n^\lambda\}$ such that $J_k v_T = q^{k-1}q^{c(T(k))}v_T$, where $T(k)$ is the box containing k in T. Thus,

$$Dv_T = \sum_{k=1}^{n} q^{n-k}(q^{k-1}q^{c(T(k))} - 1)v_T = \left(-\left(\frac{q^n-1}{q-1}\right) + \sum_{k=1}^{n} q^{n-1}q^{c(T(k))}\right)v_T.$$

(b): The dimension of the irreducible representation of H_n indexed by λ is the same as the dimension of the irreducible representation of S_n indexed by λ, which is well known as the hook-length formula [46, Ch. I(7.6)(ii), §6 Ex.2(a)]:

$$\dim(H_n^\lambda) = \mathrm{Card}(S_n^\lambda) = \frac{n!}{\prod_{b\in\lambda} h(b)}.$$

(c): With $G = GL_n(q)$, $H = H_n$, the result follows since [46, Ch. IV(6.7)],

$$\dim(G^\lambda) = \frac{[n]_q!}{\prod_{b\in\lambda}[h(b)]_q} \quad \text{and} \quad \mathbf{1}_{\mathcal{B}}^G = \sum_{\lambda\in\widehat{S_n}} G^\lambda \otimes H^\lambda.$$

□

Example 3.7. Equation 3.2 for some specific partitions gives

$$\beta_{(n)} = \frac{1}{|\mathcal{T}_{n,q}|}\left(q^{n-1}\cdot\left(\frac{q^n-1}{q-1}\right) - \frac{q^n-1}{q-1}\right) = 1$$

$$\beta_{(n-1,1)} = \frac{1}{|\mathcal{T}_{n,q}|}\left(q^{n-1}\left(q^{-1} + \frac{q^{n-1}-1}{q-1}\right) - \frac{q^n-1}{q-1}\right) = \frac{q^{n-2}-1}{q^{n-1}-1}$$

$$\beta_{(n-2,1^2)} = \frac{1}{|\mathcal{T}_{n,q}|}\left(q^{n-1}\left(q^{-1}+q^{-2} + \frac{q^{n-2}-1}{q-1}\right) - \frac{q^n-1}{q-1}\right) = \frac{q^{n-3}-1}{q^{n-1}-1}$$

$$\beta_{(n-2,2)} = \frac{1}{|\mathcal{T}_{n,q}|}\left(q^{n-1}\left(1+q^{-1} + \frac{q^{n-2}-1}{q-1}\right) - \frac{q^n-1}{q-1}\right) = \frac{q^{n-2}-1}{q^n-1}$$

$$\beta_{(2,1^{n-2})} = \frac{1}{|\mathcal{T}_{n,q}|}\left(q^{n-1}\left(1+q+q^{-1}+q^{-2}+\ldots+q^{-(n-2)}\right) - \frac{q^n-1}{q-1}\right) = \frac{q-1}{q^{n-1}-1}$$

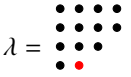
$$\beta_{(1^n)} = \frac{1}{|\mathcal{T}_{n,q}|}\left(q^{n-1}(q^0+q^{-1}+\ldots+q^{-(n-1)}) - \frac{q^n-1}{q-1}\right) = 0.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

The following simple lemma will be used several times in subsequent sections. It uses the usual majorization partial order (moving up boxes) on partitions of n, [46]. For example, when $n = 4$ the ordering is $1111 \prec 211 \prec 22 \prec 31 \prec 4$.

Lemma 3.8. *The eigenvalues β_λ of Theorem 3.6 are monotone increasing in the majorization order.*

Proof. Moving a single box (at the corner of the diagram of λ) in position (i, j) to position (i', j') necessitates $i' < i$ and $j' > j$, and so $q^{j-i} < q^{j'-i'}$. Since any $\lambda \prec \lambda'$ can be obtained by successively moving up boxes, the proof is complete. For example,

$\lambda =$  $\lambda' =$ 

□

The partition $1^n = (1, 1, \ldots, 1)$ is the minimal element in the partial order, and since $\beta_{(1^n)} = 0$, we have the following:

Corollary 3.9. *If $\lambda \vdash n$, then*

$$\beta_\lambda \geq 0.$$

Corollary 3.10. *If $\lambda = (\lambda_1, \ldots, \lambda_k), \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$ is a partition of n with $\lambda_1 = n - j$ and $j \leq n/2$, then*

$$\begin{aligned}\beta_\lambda \leq \beta_{(n-j,j)} &= \frac{q^j(q^{n-j-1}-1)(q^{n-j}-1) + (q^{n-2}-1)(q^j-1)}{(q^{n-1}-1)(q^n-1)} \\ &\leq q^{-j}(1+q^{-(n-2j+1)})(1+q^{-(n-1)})(1+q^{-(n-2)}).\end{aligned}$$

Proof. The first inequality follows from Lemma 3.8. The formula for $\beta_{(n-j,j)}$ is a simple calculation from equation (3.2). Recall the elementary inequalities:

$$\frac{1}{q^r} < \frac{1}{q^r-1} < \frac{1}{q^{r-1}}, \quad \frac{1}{1-q^{-r}} = 1 + \frac{1}{q^r-1} < 1 + q^{-(r-1)}.$$

These give the inequality for $\beta_{(n-j,j)}$:

$$\begin{aligned}\beta_{(n-j,j)} &= \frac{q^j(q^{n-j-1}-1)(q^{n-j}-1) + (q^{n-2}-1)(q^j-1)}{(q^{n-1}-1)(q^n-1)} \\ &\leq \left(\frac{q^j q^{n-j-1} q^{n-j} + q^{n-2}q^j}{q^{n-1}q^n}\right)\frac{1}{(1-q^{-(n-1)})(1-q^{-n})} \\ &\leq (q^{-j} + q^{-(n-j+1)}))(1+q^{-(n-2)})(1+q^{-(n-1)}).\end{aligned}$$

□

In the following sections, we will use a further bound from Corollary 3.10.

Corollary 3.11. *Define $\kappa_{n,q,j} := (1+q^{-(n-2j+1)})(1+q^{-(n-1)})(1+q^{-(n-2)})$ and*

$$\alpha_{n,q} := \max_{1 \leq j \leq n/2} \frac{\log(\kappa_{n,q,j})}{j}.$$

Then for all n, q,

$$\alpha_{n,q} \leq \frac{6}{n}.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Proof. Using $1 + x \le e^x$, there is the initial bound

$$\kappa_{n,q,j} \le \exp\left(q^{-(n-2j+1)} + q^{-(n-1)} + q^{-(n-2)}\right) \le \exp\left(3q^{-(n-2j)}\right),$$

which uses that $j \ge 1$ and so $(n-2) \ge (n-2j)$. Then,

$$\frac{\log(\kappa_{n,q,j})}{j} \le \frac{3q^{-(n-2j)}}{j}.$$

With $f(x) = 3q^{-(n-2x)}/x$, we have

$$f'(x) = \frac{3q^{-(n-2x)}(2x\log q - 1)}{x^2}.$$

Since $2\log(2) > 1$, we see that $f(x)$ is increasing for $x \ge 1$ and any $q \ge 2$. Thus, for $1 \le j \le n/2$, $f(j)$ is maximized for $j = n/2$, which gives

$$\alpha_{n,q} \le \max_{1 \le j \le n/2} \frac{3q^{-(n-2j)}}{j} \le \frac{6}{n}.$$

□

4. Mixing time analysis

In this section, the eigenvalues from Section 3.3 are used to give bounds on the distance to stationarity for the random transvections Markov chain on S_n. Section 4.1 reviews the tools which are needed for the bounds from specific starting states. Section 4.2 proves results for the chain started from the identity element, Section 4.3 proves results for the chain started from the reversal permutation, and Section 4.4 contains bounds for the average over all starting states.

4.1. Eigenvalue bounds

The following result from [18] will be the main tool for achieving bounds on the chi-square distance of the chain from different starting states.

Proposition 4.1 (Proposition 4.8 in [18]). *Let H be the Iwahori–Hecke algebra corresponding to a finite real reflection group W. Let K be a reversible Markov chain on W with stationary distribution π determined by left multiplication by an element of H (also denoted by K). The following identities are true:*

1. $\chi_x^2(\ell) = q^{-2l(x)} \sum_{\lambda \ne \mathbf{1}} t_\lambda \chi_H^\lambda(T_{x^{-1}} K^{2\ell} T_x)$, $x \in W$,
2. $\sum_{x \in W} \pi(x)\chi_x^2(\ell) = \sum_{\lambda \ne \mathbf{1}} f_\lambda \chi_H^\lambda(K^{2\ell})$,

where χ_H^λ are the irreducible characters, t_λ the generic degrees and f_λ the dimensions of the irreducible representations of W.

In general, the right-hand side of (a) could be difficult to calculate, but it simplifies for the special cases $x = id, x = \omega_0$. These calculations, and the analysis of the sum, are contained in the following sections.

The right-hand side of the equations in Proposition 4.1 involves the following quantities, defined for $\lambda \vdash n$:

- $n_\lambda = \sum_{i=1}^{|\lambda|}(i-1)\lambda_i$,
- $c_\lambda = \sum_{b \in \lambda} ct(b)$, where $ct(b) = j - i$ if box b is in column j and row i,

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Table 1. *The quantities involved in the eigenvalue and multiplicity calculations for* $n = 4$.

λ	f_λ	n_λ	t_λ	c_λ	β_λ
(4)	1	0	1	6	1
(3, 1)	3	1	$\frac{q(q^3-1)}{q-1}$	2	$\frac{q^2-1}{q^3-1}$
(2, 2)	2	2	$\frac{q^2(q^4-1)}{q^2-1}$	0	$\frac{q^2-1}{q^4-1}$
(2, 1, 1)	3	3	$\frac{q^3(q^3-1)}{q-1}$	-2	$\frac{q-1}{q^3-1}$
(1, 1, 1, 1)	1	6	q^6	-6	0

- $t_\lambda = q^{n_\lambda} \cdot r_\lambda$, where $r_\lambda = \frac{[n]_q!}{\prod_{b\in\lambda}[h_b]_q}$, $[k]_q = (q^k - 1)/(q - 1)$ and $[k]_q! = [k]_q \cdot [k - 1]_q \ldots [2]_q$,
- f_λ, which is the number of standard Young tableau of shape λ. The formula is

$$f_\lambda = \frac{n!}{\prod_{b\in\lambda} h(b)}.$$

Table 1 shows these values for $n = 4$ and general q. From this example, we can observe that c_λ is increasing with respect to the partial order on partitions, while n_λ is decreasing.

Let us record that

$$\beta_{(n-1,1)} = \frac{q^{n-2} - 1}{q^{n-1} - 1}, \qquad t_{(n-1,1)} = q \cdot \frac{q^{n-1} - 1}{q - 1} \qquad \text{and} \qquad f_{(n-1,1)} = n - 1. \tag{4.1}$$

Since $n! \le n^n = e^{n \log n}$ then

$$\frac{n!}{e^{n \log n}} \le 1. \tag{4.2}$$

We will use the following bounds from [18, Lemma 7.2]:

$$t_\lambda \le q^{-\binom{\lambda_1}{2}+\binom{n}{2}} f_\lambda, \quad \text{and} \quad \sum_{\lambda\vdash n} f_\lambda^2 = n!. \tag{4.3}$$

In addition, we need the following bounds for sums of f_λ. Part (b) of the proposition below is Lemma 7.2(b) in [18]; the proof there is incomplete, so we give the simple proof below.

Proposition 4.2.

1. *There is a universal constant* $K > 0$ *such that, for all* $1 \le j \le n$,

$$\sum_{\lambda:\lambda_1=n-j} f_\lambda \le \frac{n^j}{\sqrt{j!}} \cdot \frac{K}{j} e^{2\sqrt{2j}}.$$

2. *For* $1 \le j \le n$,

$$\sum_{\lambda:\lambda_1=n-j} f_\lambda^2 \le \frac{n^{2j}}{j!}.$$

Proof. Recall that f_λ is the number of standard tableau of shape λ, that is, the elements $1, 2, \ldots, n$ are arranged in the shape λ so that rows are increasing left to right and columns are increasing top to bottom. If $\lambda_1 = n - j$, then there are $\binom{n}{j}$ ways to choose the elements not in the first row of the tableau. For a fixed partition λ, the number of ways of arranging the remaining j elements is at most the number

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

of Young tableau corresponding to the partition of j created from the remaining rows of λ. This number is at most $\sqrt{j!}$ (Lemma 3 in [22]). Thus,

$$\sum_{\lambda:\lambda_1=n-j} f_\lambda \le \binom{n}{j} \cdot \sqrt{j!} \cdot p(j),$$

where $p(j)$ is equal to the number of partitions of j. It is well known that $\log(p(n)) \sim B \cdot \sqrt{n}$ for a constant B. More precisely, from (2.11) in [32], there is a universal constant $K > 0$ such that for all $n \ge 1$,

$$p(n) < \frac{K}{n} e^{2\sqrt{2n}}.$$

This gives (a).

For part (b), we again use the inequality $f_\lambda \le \binom{n}{j} f_{\lambda^*}$, where $\lambda^* = (\lambda_2, \dots, \lambda_k)$ is the partition of j determined by the rest of λ after the first row. Then,

$$\sum_{\lambda:\lambda_1=n-j} f_\lambda^2 \le \binom{n}{j}^2 \sum_{\lambda^* \vdash j} f_{\lambda^*}^2 = \binom{n}{j}^2 \cdot j! = \left(\frac{n!}{(n-j)!j!}\right)^2 \cdot j! \le \frac{n^{2j}}{j!}.$$

□

Proposition 4.3. *The function $s(\lambda) := q^{c_\lambda} t_\lambda$ is monotone with respect to the partial order on partitions. For any $\lambda \vdash n$,*

$$s(\lambda) \le q^{\binom{n}{2}}.$$

Proof. Suppose that $\lambda \lessdot \widetilde{\lambda}$ and $\widetilde{\lambda}$ is obtained from λ by 'moving up' one box. Suppose the box at position (i, j) is moved to (i', j'), with $i' < i, j' > j$.

Let $g(\lambda) = c_\lambda + n_\lambda$. Then,

$$\begin{aligned} c_{\widetilde{\lambda}} &= c_\lambda + (j' - j) + (i - i') \\ n_{\widetilde{\lambda}} &= n_\lambda - (i - 1) + (i' - 1) = n_\lambda + (i' - i). \end{aligned}$$

This implies that

$$g(\widetilde{\lambda}) = g(\lambda) + (j' - j) + (i - i') - (i - 1) + (i' - 1) = g(\lambda) + (j' - j).$$

Now, consider the change in r_λ. The hook lengths of $\widetilde{\lambda}$ are:

$$\begin{aligned} &\widetilde{h}(i, j) = h(i, j) - 1 = 0, \quad \widetilde{h}(i', j') = h(i, j) + 1 = 1 \\ &\widetilde{h}(k, j) = h(k, j) - 1, \quad k < i, k \ne i' \\ &\widetilde{h}(i, l) = h(i, l) - 1, \quad l < j \\ &\widetilde{h}(k, j') = h(k, j') + 1, \quad k < i' \\ &\widetilde{h}(i', l) = h(i', l) + 1, \quad l < j', l \ne j \end{aligned}$$

and $\widetilde{h}(k, l) = h(k, l)$ for all other boxes $b = (k, l)$. Thus,

$$\begin{aligned} \sum_b \widetilde{h}(b) - \sum_b h(b) &= (j' - 1) + (i' - 1) - (i - 1) - (j - 1) \\ &= (j' - j) + (i' - i). \end{aligned}$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Using the inequalities $q^{r-1} < (q^r - 1) < q^r$, we have

$$\begin{aligned} r(\widetilde{\lambda}) &= \frac{[n]_q!(q-1)^n}{\prod_b (q^{h(b)}-1)} \\ &\le \frac{[n]_q!(q-1)^n}{q^{\sum_b \widetilde{h}(b)-n}} = \frac{1}{q^{(j'-j)+(i'-i)-n}} \frac{[n]_q!(q-1)^n}{q^{\sum_b h(b)}} \\ &\le \frac{1}{q^{(j'-j)+(i'-i)-n}} \frac{[n]_q!(q-1)^n}{\prod_b (q^{h(b)}-1)} = \frac{1}{q^{(j'-j)+(i'-i)-n}} r(\lambda). \end{aligned}$$

Combining this with the result for $g(\lambda)$:

$$s(\widetilde{\lambda}) = q^{g(\widetilde{\lambda})} r_\lambda = q^{g(\lambda)+(j'-j)} \cdot \frac{1}{q^{(j'-j)+(i'-i)-n}} r(\lambda) = s(\lambda) q^{i-i'+n} > s(\lambda)$$

since $i > i'$.

Assuming the monotonicity, then if $\lambda_1 \ge n/2$ we have $s(\lambda) \le s((\lambda_1, n-\lambda_1))$. To calculate this quantity:

$$\begin{aligned} g(\lambda) \le g((\lambda_1, n-\lambda_1)) &= c_{(\lambda_1, n-\lambda_1)} + n_{(\lambda_1, n-\lambda_1)} \\ &= \left(\frac{\lambda(\lambda_1 - 1)}{2} + \frac{(n-\lambda_1-1)(n-\lambda_1-2)}{2} - 1 \right) + (n-\lambda_1) \\ &= \binom{\lambda_1}{2} + \binom{n-\lambda_1}{2} = \frac{n(n-1)}{2} - \lambda_1(n-\lambda_1). \end{aligned}$$

For r_λ, note that the hook lengths of $(\lambda_1, n-\lambda_1)$ are

$$\begin{aligned} &\lambda_1 + 1, \lambda_1, \ldots, 2\lambda_1 - n + 2 \\ &2\lambda_1 - n, 2\lambda_1 - n - 1, \ldots, 3, 2, 1 \\ &n - \lambda_1, n - \lambda_1 - 1, \ldots, 3, 2, 1. \end{aligned}$$

If $\lambda_1 = n - j$, we see the terms that cancel:

$$\begin{aligned} r_{(\lambda_1, n-\lambda_1)} = \frac{[n]_q!}{\prod_b [h(b)]_q} &= \frac{(q^n-1)(q^{n-1}-1)\ldots(q-1)}{\prod_b (q^{h(b)}-1)} \\ &= \frac{(q^n-1)(q^{n-1}-1)\ldots(q^{\lambda_1+2}-1)\cdot(q^{2\lambda_1-n+1}-1)}{(q^{n-\lambda_1}-1)(q^{n-\lambda-1}-1)\ldots(q-1)} \\ &< \frac{q^{\sum_{k=0}^{j-2}(n-k)+(n-2j+1)}}{q^{\sum_{k=1}^{j} k - j}} \\ &= q^{\frac{1}{2}(j-1)(2(n+1)-j)+(n-2j+1)-\frac{1}{2}j(j+1)+j} \\ &= q^{j(n-j-1)+j} = q^{j(n-j)} = q^{\lambda_1(n-\lambda_1)}. \end{aligned}$$

This uses the inequality $q^{r-1} < q^r - 1 < q^r$. Thus, if $\lambda_1 \ge n/2$, we have shown

$$s(\lambda) = q^{g(\lambda)} r_\lambda \le q^{\binom{n}{2} - \lambda_1(n-\lambda_1)} \cdot q^{\lambda_1(n-\lambda_1)} = q^{\binom{n}{2}}.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Now, suppose $\lambda_1 \le n/2$, so $s(\lambda) \le s((n/2, n/2))$ (assume n is even). To calculate this,

$$\begin{aligned} g(\lambda) &\le g((n/2,n/2)) = c_{(n/2,n/2)} + n_{(n/2,n/2)} \\ &= \frac{(n/2-1)n/2}{2} + \frac{(n/2-2)(n/2-1)}{2} - 1 + \frac{n}{2} = \frac{n^2}{4} - \frac{n}{2}. \end{aligned}$$

To bound r_λ, use the same calculation as before to get

$$r_{(n/2,n/2)} \le q^{n^2/4},$$

and so in total $s((n/2,n/2)) \le q^{n^2/2-n/2} = q^{\binom{n}{2}}$. □

4.2. Starting from id

Theorem 4.4. *Let P be the Markov chain on S_n induced by random transvections on $GL_n(q)$.*

1. *For t_λ, β_λ defined in Theorem 3.6, we have*

$$4\|P^\ell_{id} - \pi_q\|^2_{TV} \le \chi^2_{id}(\ell) = \sum_{\lambda \vdash n, \lambda \ne (n)} t_\lambda f_\lambda \beta_\lambda^{2\ell}.$$

2. *Let $\alpha_{n,q}$ be as in Corollary 3.11 and n, q such that $\log q > 6/n$. Then if $\ell \ge \frac{n \log q/2 + \log n + c}{(\log q - \alpha_{n,q})}$, $c > 0$, we have*

$$\chi^2_{id}(\ell) \le (e^{e^{-2c}} - 1) + e^{-cn}.$$

3. *For any $\ell \ge 1$,*

$$\chi^2_{id}(\ell) \ge (q^{n-1} - 1)(n-1)q^{-4\ell}.$$

4. *If $\ell \le n/8$, then for fixed q and n large,*

$$\|P^\ell_{id} - \pi_q\|_{TV} \ge 1 - o(1). \tag{4.4}$$

Theorem 4.4 shows that restricting the random transvections walk from $GL_n(q)$ to the double coset space only speeds things up by a factor of 2 when started from the identity. Hildebrand [35] shows that the total variation distance on all of $GL_n(q)$ is only small after $n+c$ steps. Note this is independent of q.

Proof. **(a):** The inequality follows from Proposition 4.1 (b):

$$\begin{aligned} \chi^2_{id}(\ell) &= q^{-2I(id)} \sum_{\lambda \ne (n)} t_\lambda \chi^\lambda_H (T_{id} K^{2\ell} T_{id}) \\ &= \sum_{\lambda \ne (n)} t_\lambda \chi^\lambda_H (K^{2\ell}) = \sum_{\lambda \ne (n)} t_\lambda f_\lambda \beta_\lambda^{2\ell}. \end{aligned}$$

(b): From Corollary 3.10 if $\lambda_1 = n - j$, then $\beta_\lambda \le \kappa_{n,q,j} q^{-j}$, where

$$\kappa_{n,q,j} = (1 + q^{-(n-2j+1)})(1 + q^{-(n-1)})(1 + q^{-(n-2)}).$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Using the bound on t_λ from equation (4.3), for $1 \le j \le \lfloor n/2 \rfloor$, we have

$$\begin{aligned}\sum_{\lambda:\lambda_1=n-j} t_\lambda f_\lambda \beta_\lambda^{2\ell} &\le (\kappa_{n,q,j} q^{-j})^{2\ell} \sum_{\lambda:\lambda_1=n-j} q^{-\binom{n-j}{2}+\binom{n}{2}} f_\lambda^2 \\ &\le (\kappa_{n,q,j} q^{-j})^{2\ell} q^{nj-j(j+1)/2} \cdot \frac{n^{2j}}{j!} \\ &\le \frac{1}{j!} \exp\left(-2\ell j\left(\log q - \frac{\log(\kappa_{n,q,j})}{j}\right) + \log q(nj - j(j+1)/2) + 2j \log n\right) \\ &\le \frac{1}{j!} \exp\big(-2\ell j\big(\log q - \alpha_{n,q}\big) + \log q(nj - j(j+1)/2) + 2j \log n\big). \end{aligned} \tag{4.5}$$

Recall the final inequality follows since $\alpha_{n,q} := \max_{1\le j\le n/2} \log(\kappa_{n,q,j})/j$. If $\ell = \frac{n \log q/2 + \log n + c}{(\log q - \alpha_{n,q})}$, then the exponent in equation (4.5) is

$$\begin{aligned}-2j(n \log q/2 + \log n + c) + \log q(nj - j(j+1)/2) + 2j \log n \\ = -j(2c + \log q(j+1)/2).\end{aligned}$$

This gives

$$\sum_{j=1}^{\lfloor n/2 \rfloor} \sum_{\lambda:\lambda_1=n-j} t_\lambda f_\lambda \beta_\lambda^{2\ell} \le \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{e^{-2jc}}{j!} \le e^{e^{-2c}} - 1.$$

Next, we need to consider the partitions λ with $\lambda_1 \le n/2$. For these partitions,

$$\beta_\lambda \le \beta_{(n/2,n/2)} \le \kappa_{n,q,n/2} q^{-n/2}.$$

Then we have

$$\begin{aligned}\sum_{j=n/2}^{n-1} \sum_{\lambda:\lambda_1=n-j} t_\lambda f_\lambda \beta_\lambda^{2\ell} &\le \beta_{(n/2,n/2)}^{2\ell} \sum_{j=n/2}^{n-1} \sum_{\lambda:\lambda_1=n-j} t_\lambda f_\lambda \\ &\le \beta_{(n/2,n/2)}^{2\ell} \sum_{j=n/2}^{n-1} \sum_{\lambda:\lambda_1=n-j} q^{nj-j(j+1)/2} f_\lambda^2 \\ &\le n! \cdot \exp\left(-2\ell(n/2)\left(\log q + \frac{\kappa_{n,q,n/2}}{n/2}\right) + n^2/2\right) \\ &\le \exp\left(-n(n \log q/2 + \log n + c) + n^2/2\right) \sum_{\lambda \vdash n} f_\lambda^2 \\ &\le n! \exp\left(-n(n \log q/2 + \log n + c) + n^2/2\right), \end{aligned}$$

using $\sum_\lambda f_\lambda^2 = n!$ and that if $n/2 \le j \le n$, then

$$nj - j(j+1)/2 \le n^2 - n^2/4 < n^2 - n(n+1)/2 \le n^2/2$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

since the function is increasing in j. Note also that if $q \geq 3$, then $\log q > 1$ and $n! \leq n^n = e^{\log n}$. To finish the bound,

$$\sum_{j=n/2}^{n-1} \sum_{\lambda:\lambda_1=n-j} t_\lambda f_\lambda \beta_\lambda^{2\ell} \leq \exp\Big(n\log n - n(n\log q/2 + \log n + c) + n^2/2\Big)$$
$$= \exp\Big(-(n^2/2)(\log q - 1) - cn\Big) \leq e^{-cn}.$$

(c): The lower bound comes from considering the $\lambda = (n-1, 1)$ term from the sum in (a). Using the quantities (4.1), this gives

$$\chi^2_{id}(\ell) = \sum_{\lambda \neq (n)} t_\lambda f_\lambda \beta_\lambda^{2\ell} \geq t_{(n-1,1)} f_{(n-1,1)} \beta_{(n-1,1)}^{2\ell}$$
$$= q \cdot \frac{q^{n-1}-1}{q-1}(n-1)\left(\frac{q^{n-2}-1}{q^{n-1}-1}\right)^{2\ell} \geq (q^{n-1}-1)(n-1)q^{-4\ell}.$$

This uses that $(q^{n-2}-1)/(q^{n-1}-1) \geq q^{-2}$.

(d): From the alternative version of the walk on the Hecke algebra, involving $D/|\mathcal{T}_{n,q}|$ with D from 1.6, the walk proceeds by picking a transposition $(i, j), i < j$ with probability proportional to

$$q^{-(j-i)}$$

and multiplying by T_{ij}. As described in Section 3.2, multiplication by T_{ij} corresponds to proposing the transposition (i, j) and proceeding via the Metropolis algorithm. Thus, multiplication by T_{ij} induces at most $2(j-i)$ inversions, always less than $2n$. From [22] Theorem 5.1, under π_q, a typical permutation has $\binom{n}{2} - \frac{n-q}{q+1} + O(\sqrt{n})$ inversions (and the fluctuations are Gaussian about this mean). If $\ell = n/8$, the measure $P^\ell_{id}(\cdot)$ is concentrated on permutations with at most $n^2/4$ inversions and π_q is concentrated on permutations with order $n^2/2 - (n-1)/(q+1) + O(\sqrt{n})$ inversions. □

4.3. Starting from ω_0

Theorem 4.5. *Let P be the Markov chain on S_n induced by random transvections on $GL_n(q)$, and let $\omega_0 \in S_n$ be the reversal permutation in S_n.*

1. *With $t_\lambda, c_\lambda, \beta_\lambda$ defined in Section 3.3,*

$$4\|P^\ell_{\omega_0} - \pi_q\|^2_{TV} \leq \chi^2_{\omega_0}(\ell) = q^{-\binom{n}{2}} \sum_{\lambda \neq (n)} q^{c_\lambda} t_\lambda f_\lambda \beta_\lambda^{2\ell}.$$

2. *Let $\alpha_{n,q}$ be as in Corollary 3.11 and n, q such that $\log q > 6/n$. If $\ell \geq (\log n/2 + c)/(\log q - \alpha_{n,q})$ for $c > 0$ with $c \geq 2\sqrt{2}$ then*

$$\chi^2_{\omega_0}(\ell) \leq -2K\log(1 - e^{-c}) + \sqrt{K}e^{-cn},$$

for a universal constant $K > 0$ (independent of q, n).

3. *For any $\ell \geq 1$,*

$$\chi^2_{\omega_0}(\ell) \geq q^{-(n-2)}(n-1)(q^{n-1}-1)q^{-4\ell}.$$

Remark 4.6. Theorem 4.5 shows that the Markov chain has a cutoff in its approach to stationarity in the chi-square metric. It shows the same exponential speed up as the walk started at a typical position

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

(Theorem 4.7 below) and indeed is faster by a factor of 2. This is presumably because it starts at the permutation ω_0, at which the stationary distribution π_q is concentrated, instead of 'close to ω_0'.

Proof of Theorem 4.5. **(a):** By Proposition 4.9 in [18], if ω_0 is the longest element of W, then

$$\rho^\lambda(T_{\omega_0}^2) = q^{I(\omega_0)+c_\lambda}\mathrm{Id},$$

where ρ^λ is the irreducible representation indexed by λ. Using this and 4.1 (a),

$$\begin{aligned}
\chi_{\omega_0}^2(\ell) &= q^{-2I(\omega_0)} \sum_{\lambda\neq\mathbf{1}} t_\lambda \chi_H^\lambda(T_{\omega_0^{-1}} K^{2\ell} T_{\omega_0}) \\
&= q^{-2I(\omega_0)} \sum_{\lambda\neq\mathbf{1}} t_\lambda \chi_H^\lambda(K^{2\ell} T_{\omega_0} T_{\omega_0^{-1}}) \\
&= q^{-2I(\omega_0)} \sum_{\lambda\neq\mathbf{1}} t_\lambda \chi_H^\lambda(K^{2\ell} T_{\omega_0}^2) \\
&= q^{-2I(\omega_0)} \sum_{\lambda\neq\mathbf{1}} t_\lambda \chi_H^\lambda(K^{2\ell} q^{c_\lambda + I(\omega_0)}) \\
&= q^{-I(\omega_0)} \sum_{\lambda\neq\mathbf{1}} q^{c_\lambda} t_\lambda \chi_H^\lambda(K^{2\ell}) = q^{-I(\omega_0)} \sum_{\lambda\neq\mathbf{1}} q^{c_\lambda} t_\lambda f_\lambda \beta_\lambda^{2\ell}
\end{aligned}$$

since $K \in Z(H)$, that is, K commutes with all elements of the Hecke algebra.

(b): Suppose $\lambda_1 = n - j$ for $1 \le j \le n/2$. Recall the definition $s(\lambda) = q^{c_\lambda} t_\lambda$. From Proposition 4.3, $s(\lambda) \le q^{\binom{n}{2}}$. Then,

$$\begin{aligned}
q^{-\binom{n}{2}} \sum_{\lambda:\lambda_1=n-j} q^{c_\lambda} t_\lambda f_\lambda \beta_\lambda^{2\ell} &= q^{-\binom{n}{2}} \sum_{\lambda:\lambda_1=n-j} s(\lambda) f_\lambda \beta_\lambda^{2\ell} \\
&\le (\kappa_{n,q,j} q^{-j})^{2\ell} q^{-\binom{n}{2}} \sum_{\lambda:\lambda_1=n-j} q^{\binom{n}{2}} f_\lambda \\
&\le (\kappa_{n,q,j} q^{-j})^{2\ell} \sum_{\lambda:\lambda_1=n-j} f_\lambda \\
&\le (\kappa_{n,q,j} q^{-j})^{2\ell} \cdot \frac{n^j}{\sqrt{j!}} \cdot \frac{K}{j} e^{2\sqrt{2j}} \\
&\le \frac{K}{\sqrt{j!}} \exp\left(-2\ell j\left(\log q - \frac{\log(\kappa_{n,q,j})}{j}\right) + (j-1)\log n + 2\sqrt{2j}\right).
\end{aligned}$$

The third inequality uses Proposition 4.2 for $\sum_{\lambda:\lambda_1=n-j} f_\lambda$. Recall $\alpha_{n,q} := \max_{1\le j\le n/2} \log(\kappa_{n,q,j})/j$. If $\ell = (\log n/2 + c)/(\log q - \alpha_{n,q})$, then the bound becomes

$$\begin{aligned}
\sum_{j=1}^{n/2} q^{-\binom{n}{2}} \sum_{\lambda:\lambda_1=n-j} q^{c_\lambda} t_\lambda f_\lambda \beta_\lambda^{2\ell} &\le \sum_{j=1}^{n/2} \frac{K}{\sqrt{j!}} \exp\left(-j\log n - 2jc + (j-1)\log n + 2\sqrt{2j}\right) \\
&\le 2K \sum_{j=1}^{n/2} \frac{e^{-2jc+2\sqrt{2j}}}{j},
\end{aligned}$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

using the loose bound $\sqrt{j!} > j/2$ for all $j \geq 1$. With the assumption that $c \geq 2\sqrt{2}$, we have $-2jc+2\sqrt{2j} \leq -jc$ for all $j \geq 1$. Finally,

$$\begin{aligned} 2K \sum_{j=1}^{n/2} \frac{e^{-jc}}{j} &\leq 2K \sum_{j=1}^{\infty} \frac{e^{-jc}}{j} \\ &= -2K \log(1 - e^{-c}). \end{aligned}$$

Now, for the λ with $\lambda_1 \geq n/2$, we have

$$\begin{aligned} \sum_{j=n/2}^{n-1} \sum_{\lambda : \lambda_1 = n-j} q^{-\binom{n}{2}} q^{c_\lambda} t_\lambda f_\lambda \beta_\lambda^{2\ell} &\leq \beta_{(n/2,n/2)}^{2\ell} \sum_{j=n/2}^{n-1} \sum_{\lambda : \lambda_1 = n-j} f_\lambda \\ &\leq \beta_{(n/2,n/2)}^{2\ell} \sum_{\lambda \vdash n} f_\lambda \\ &\leq \beta_{(n/2,n/2)}^{2\ell} \left(\sum_{\lambda \vdash n} f_\lambda^2 \right)^{1/2} \cdot p(n)^{1/2}, \end{aligned}$$

where $p(n)$ is equal to the number of partitions of n (the inequality is Cauchy–Schwarz). Since $p(n) \leq \frac{K}{n} e^{2\sqrt{2n}}$ for a constant $K > 0$ ([32]) and $\sum_{\lambda \vdash n} f_\lambda^2 = n!$, this gives

$$\begin{aligned} &\leq \sqrt{n!} \cdot \sqrt{\frac{K}{n}} \exp\left(\sqrt{2n} - 2\ell(n/2) \left(\log q - \frac{\kappa_{n,q,n/2}}{n/2} \right) \right) \\ &\leq \sqrt{n!} \cdot \sqrt{\frac{K}{n}} \exp\left(\sqrt{2n} - n \log n/2 - 2nc \right) \\ &\leq \sqrt{K} \exp\left(-nc - n(c - \sqrt{2} n^{-1/2}) \right) \end{aligned}$$

since $\sqrt{n!} \leq e^{n \log n/2}$. Since $c > 2\sqrt{2}$, then the bound is $\leq \sqrt{K} e^{-cn}$ for any $n \geq 1$.

(c): A lower bound comes from using equation (4.1) for the lead term on the right-hand side of (a):

$$\begin{aligned} \chi^2_{\omega_0}(\ell) &\geq q^{-\binom{n}{2}} q^{c_{(n-1,1)}} t_{(n-1,1)} f_{(n-1,1)} \beta_{(n-1,1)}^{2\ell} \\ &= q^{-\binom{n}{2}} q^{\binom{n-1}{2}+1} \cdot q \frac{q^{n-1}-1}{q-1} \cdot (n-1) \left(\frac{q^{n-2}-1}{q^{n-1}-1} \right)^{2\ell} \\ &\geq q^{-(n-2)} (q^{n-1}-1)(n-1) q^{-4\ell}. \end{aligned}$$

□

4.4. Starting from a typical site

In analyzing algorithms used repeatedly for simulations, as the algorithm is used, it approaches stationarity. This means the quantity

$$\sum_{x \in S_n} \pi_q(x) \| P_x^\ell - \pi_q \|_{TV}$$

is of interest. For the problem under study, π_q is concentrated near ω_0 so we expect rates similar to those in Theorem 4.5.

Theorem 4.7. *Let P be the Markov chain on S_n induced by random transvections on $GL_n(q)$.*

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

1. *With f_λ, β_λ defined in Section 3.3,*

$$\left(\sum_{x\in S_n} \pi_q(x)\|P_x^\ell - \pi_q\|_{TV}\right)^2 \le \frac{1}{4}\sum_{x\in S_n}\pi_q(x)\chi_x^2(\ell) = \frac{1}{4}\sum_{\lambda\neq(n)} f_\lambda^2\beta_\lambda^{2\ell}.$$

2. *Let $\alpha_{n,q}$ be as in Corollary 3.11 and n, q such that $\log q > 6/n$. If $\ell \ge (\log n + c)/(\log q - \alpha_{n,q})$, $c > 0$, then*

$$\sum_{x\in S_n}\pi_q(x)\chi_x^2(\ell) \le (e^{e^{-c}} - 1) + e^{-cn}.$$

3. *For any $\ell \ge 1$,*

$$\sum_{x\in S_n}\pi_q(x)\chi_x^2(\ell) \ge (n-1)^2 q^{-4\ell}.$$

Proof. **(a):** This is simply a restatement of 4.1 part (b).

(b): We will divide the sum depending on the first entry of the partition. By Proposition 4.2, we have the bound (true for any $1 \le j \le n$)

$$\sum_{\lambda:\lambda_1=n-j} f_\lambda^2 \le \frac{n^{2j}}{j!}.$$

Combining this with Corollary 3.10, for $j \le \lfloor n/2\rfloor$,

$$\begin{aligned}\sum_{\lambda:\lambda_1=n-j}\beta_\lambda^{2\ell} f_\lambda^2 &\le \sum_{\lambda:\lambda_1=n-j}\beta_{(n-j,j)}^{2\ell} f_\lambda^2 \le \beta_{(n-j,j)}^{2\ell}\cdot\frac{n^{2j}}{j!}\\ &\le \kappa_{n,q,j}^{2\ell} q^{-2\ell j}\cdot\frac{n^{2j}}{j!} = \frac{1}{j!}\exp\big(2\ell(\log(\kappa_{n,q,j}) - j\log q) + 2j\log n\big)\\ &= \frac{1}{j!}\exp\left(2\ell j\left(\frac{\log(\kappa_{n,q,j})}{j} - \log q\right) + 2j\log n\right).\end{aligned}$$

Define

$$\alpha_{n,q} := \max_{1\le j\le n/2}\frac{\log(\kappa_{n,q,j})}{j},$$

so then if $\ell = (\log n + c)/(\log q - \alpha_{n,q})$, the bound is

$$\begin{aligned}&\le \frac{1}{j!}\exp\big(2\ell j(\alpha_{n,q} - \log q) + 2j\log n\big)\\ &= \frac{1}{j!}\exp(-2j\log n - 2jc + 2j\log n) = \frac{e^{-2jc}}{j!}.\end{aligned}$$

Then summing over all possible j gives

$$\sum_{j=1}^{\lfloor n/2\rfloor}\sum_{\lambda:\lambda_1=n-j} f_\lambda^2\beta_\lambda^{2\ell} \le \sum_{j=1}^{\lfloor n/2\rfloor}\frac{e^{-2jc}}{j!} \le \left(e^{e^{-c}} - 1\right).$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Now, we have to bound the contribution from partitions λ with $\lambda_1 \le n/2$. Because β_λ is monotone with respect to the order on partitions, we have for all λ such that $\lambda_1 \le n/2$,

$$\beta_\lambda \le \beta_{(n/2,n/2)} \le \kappa_{n,q,n/2} q^{-n/2}$$

since $\lambda \le (n/2, n/2)$ (assuming without essential loss that n is even). Then,

$$\begin{aligned}\sum_{j=n/2}^{n-1} \sum_{\lambda:\lambda_1=n-j} f_\lambda^2 \beta_\lambda^{2\ell} &\le \beta_{(n/2,n/2)}^{2\ell} \sum_\lambda f_\lambda^2 \\ &\le n! \cdot \exp\left(2\ell(n/2)(\log q + \frac{\kappa_{n,q,n/2}}{n/2})\right) \\ &\le n! \cdot \exp(-n(\log n + c)) \\ &\le e^{-cn},\end{aligned}$$

using $\sum_\lambda f_\lambda^2 = n! \le n^n$.

(c): The sum is bounded below by the term for $\lambda = (n-1, 1)$. This is

$$f_{(n-1,1)}^2 \cdot \beta_{(n-1,1)}^{2\ell} = (n-1)^2 \left(\frac{q^{n-2}-1}{q^{n-1}-1}\right)^{2\ell} \ge (n-1)^2 q^{-4\ell}.$$

□

5. Hecke algebra computations

This section proves Theorem 1.6 which describes the transvections Markov chain on S_n as multiplication in the Hecke algebra from Definition 3.1. This is accomplished by careful and elementary row reduction. Our first proof used Hall–Littlewood symmetric functions. It is recorded in the expository account [19].

5.1. Overview

Let $\mathbb{C}[G]$ denote the group algebra for $G = GL_n(q)$. This is the space of functions $f : G \to \mathbb{C}$, with addition defined $(f+g)(s) = f(s) + g(s)$ and multiplication defined by

$$f_1 * f_2(s) = \sum_{t \in G} f_1(t) f_2(st^{-1}).$$

Equivalently, $\mathbb{C}[G] = \text{span}\{g \mid g \in G\}$ and we can write an element $f = \sum_g c_g g$ for $c_g \in \mathbb{C}$, so $f(g) = c_g$.

Define elements in $\mathbb{C}[G]$:

$$\mathbb{1}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}} x, \quad \text{and } T_\omega = \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}\omega\mathcal{B}} x, \quad \text{for } \omega \in S_n.$$

Note that if $b \in \mathcal{B}$, then

$$b\mathbb{1}_{\mathcal{B}} = \mathbb{1}_{\mathcal{B}} b = \mathbb{1}_{\mathcal{B}}, \quad \mathbb{1}_{\mathcal{B}}^2 = \mathbb{1}_{\mathcal{B}}.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

If $g \in \mathcal{B}\omega\mathcal{B}$, so $g = b_1\omega b_2$ with $b_1, b_2 \in \mathcal{B}$, then

$$\begin{aligned}\mathbb{1}_{\mathcal{B}} g \mathbb{1}_{\mathcal{B}} &= \mathbb{1}_{\mathcal{B}} b_1 \omega b_2 \mathbb{1}_{\mathcal{B}} = \mathbb{1}_{\mathcal{B}} \omega \mathbb{1}_{\mathcal{B}} \\ &= \frac{1}{|\mathcal{B}|^2} \sum_{b_1, b_2 \in \mathcal{B}} b_1 \omega b_2 = \frac{1}{|\mathcal{B}|^2} \frac{|\mathcal{B}|^2}{|\mathcal{B}\omega\mathcal{B}|} \sum_{x \in \mathcal{B}\omega\mathcal{B}} x \\ &= \frac{|\mathcal{B}|}{|\mathcal{B}\omega\mathcal{B}|} \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}\omega\mathcal{B}} x = \frac{|\mathcal{B}|}{|\mathcal{B}\omega\mathcal{B}|} \frac{1}{|\mathcal{B}|} T_\omega = q^{-I(\omega)} T_\omega.\end{aligned}$$

The Hecke algebra is $H_n(q) = \mathbb{1}_{\mathcal{B}} \mathbb{C}[G] \mathbb{1}_{\mathcal{B}}$ and has basis $\{T_\omega \mid \omega \in S_n\}$. Note that $H_n(q)$ are all functions in $\mathbb{C}[G]$ which are $\mathcal{B}$-$\mathcal{B}$ invariant, that is, $f(b_1 g b_2) = f(g)$.

Now, let P be the transition matrix for G defined by multiplying by a random transvection. We can also write this

$$P = \frac{1}{|\mathcal{T}_{n,q}|} \sum_{T \in \mathcal{T}_{n,q}} M_T,$$

where M_T is the transition matrix 'multiply by T'. In other words, $M_T(x, y) = \mathbf{1}(y = Tx)$.

Then P defines a linear transform on the space $\mathbb{C}[G]$, with respect to the basis $\{g \mid g \in G\}$. The matrix M_T is just the function: Multiply by T in the group algebra. This means P is equivalent to multiplication by $\frac{1}{|\mathcal{T}_{n,q}|} D$, with $D = \sum_{T \in \mathcal{T}_{n,q}} T$ as an element in $\mathbb{C}[G]$. The Markov chain lumped to $S_n = \mathcal{B}\backslash GL_n(q)/\mathcal{B}$ is then equivalent to multiplication by D on $H_n(q)$. Since D is the sum of all elements in a conjugacy class, it is in the center $Z(\mathbb{C}[G])$. This means if $g \in \mathcal{B}\omega\mathcal{B}$, then

$$\mathbb{1}_{\mathcal{B}}(Dg)\mathbb{1}_{\mathcal{B}} = D\mathbb{1}_{\mathcal{B}}\omega\mathbb{1}_{\mathcal{B}} = (D\mathbb{1}_{\mathcal{B}})q^{-I(\omega)}T_\omega.$$

In conclusion, to determine how D acts in $H_n(q)$, we can compute $D\mathbb{1}_{\mathcal{B}}$. The remainder of the section proves the following.

Theorem 5.1. *Let $D = \sum_{T \in \mathcal{T}_{n,q}} T \in \mathbb{C}[G]$. Then,*

$$D\mathbb{1}_{\mathcal{B}} = \Big((n-1)q^{n-1} - [n-1]_q\Big)\mathbb{1}_{\mathcal{B}} + (q-1) \sum_{1 \le i < j \le n} q^{n-1-(j-i)} T_{s_{ij}},$$

where s_{ij} is the transposition switching i and j.

5.2. Row reduction

Let $G = GL_n(q)$ and $\mathcal{B}$ be the upper-triangular matrices. For $1 \le i \le n-1$ and $c \in \mathbb{F}_q$, define $y_i(c) \in G$ by

$$y_i(c) = \begin{array}{c} \\ i \\ i+1 \end{array} \begin{array}{c} \begin{array}{ccccccc} & & & i & i+1 & & \end{array} \\ \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & 1 & & \\ & & & 1 & 0 & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ & & & & & & & 1 \end{pmatrix} \end{array}.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

That is, multiplication on the left by $y_i(c)$ acts by adding c times the ith row to the $(i+1)$th row, then permuting the i and $i+1$ rows.

Let $\omega \in S_n$. We can write the reduced word $\omega = s_{i_1} \dots s_{i_\ell}$, for $s_i = (i, i+1)$ the simple reflections. Then,

$$\mathcal{B}\omega\mathcal{B} = \{y_{i_1}(c_1) \dots y_{i_\ell}(c_\ell)\mathcal{B} \mid c_1, \dots, c_\ell \in \mathbb{F}_q\},$$

and $|\mathcal{B}\omega\mathcal{B}|/|\mathcal{B}| = q^{I(\omega)}$ and $G = \bigsqcup_{\omega \in S_n} \mathcal{B}\omega\mathcal{B}$. This provides a very useful way for determining the double coset that a matrix M belongs to, which just amounts to performing row reduction by multiplying by different matrices $y_i(c)$.

5.3. Transvections

For every transvection $T \in \mathcal{T}_{n,q}$, we can perform row operations, multiplying by $y_i(c)$, to determine which double coset T belongs to.

A transvection is defined by two vectors $\mathbf{a}, \mathbf{v} \in \mathbb{F}_q^n$ with $\mathbf{v}^\top \mathbf{a} = 0$, with the last nonzero entry of $\mathbf{v}$ normalized to be 1. Assume this entry is at position j. Assume the first nonzero entry of $\mathbf{a}$ is at position i. That is, the vectors look like:

$$\begin{aligned} \mathbf{a} &= (0, \dots, 0, a_i, a_{i+1}, \dots, a_n)^\top, \quad a_i \neq 0, \\ \mathbf{v} &= (v_1, \dots, v_{j-1}, 1, 0, \dots, 0)^\top. \end{aligned}$$

Let $T_{\mathbf{a},\mathbf{v}} = I + \mathbf{v}\mathbf{a}^\top$. We consider the possible cases for i and j to prove the following result.

Proposition 5.2. *Let $T_{\mathbf{a},\mathbf{v}} = I + \mathbf{v}\mathbf{a}^\top$ be the transvection defined by nonzero vectors $\mathbf{a}, \mathbf{v}$ with $\mathbf{v}^\top\mathbf{a} = 0$ and the last nonzero entry of $\mathbf{v}$ equal to 1. Let j be the index of the last nonzero entry of $\mathbf{v}$ and i the index of the first nonzero entry of $\mathbf{a}$. If $i > j$, then $T_{\mathbf{a},\mathbf{v}} \in \mathcal{B}$. If $i < j$, then*

$$T_{\mathbf{a},\mathbf{v}} \in \mathcal{B}s_{j-1} \cdots s_{i+1}s_i s_{i+1} \cdots s_{j-1}\mathcal{B}.$$

The case $i = j$ does not occur.

Proof. **Case 1** $i > j$**.** If $i > j$, then $T_{\mathbf{a},\mathbf{v}} \in \mathcal{B}$. To see this, suppose $k > l$. Then

$$T_{\mathbf{a},\mathbf{v}}(k,l) = (\mathbf{v}\mathbf{a}^\top)_{k,l} = v_k a_l = 0$$

because $v_k a_l$ can only be nonzero if $k \leq j < i \leq l$. Now, we can count how many transvections satisfy this. There are $(q-1)q^{n-i}$ choices for $\mathbf{a}$ (because a_i must be nonzero) and q^{j-1} choices for $\mathbf{v}$ (because v_j is fixed at 1). In total for this case, there are

$$\begin{aligned} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (q-1)q^{n-i}q^{j-1} &= (q-1)\sum_{j=1}^{n-1} q^{j-1}(q^{n-j-1} + q^{n-j-2} + \dots + q + 1) \\ &= (q-1)\sum_{j=1}^{n-1} q^{j-1}\frac{q^{n-j}-1}{q-1} = \sum_{j=1}^{n-1}(q^{n-1} - q^{j-1}) \\ &= (n-1)q^{n-1} - (1 + q + \dots + q^{n-1}) = (n-1)q^{n-1} - [n-1]_q. \end{aligned}$$

Case 2 $i = j$**.** In this case, $\mathbf{v}^\top\mathbf{a} = a_i \neq 0$. This does not satisfy the condition $\mathbf{v}^\top\mathbf{a} = 0$, so this case cannot occur.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Case 3 $i < j$**.** The transvection is of the form

$$T_{\mathbf{a},\mathbf{v}} = \begin{array}{c} \\ 1 \\ \vdots \\ i \\ \vdots \\ j \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & \dots & i & \dots & j & \dots & n \end{array} \\ \left(\begin{array}{c|c|c} Id & * & * \\ \hline & N_{\mathbf{a},\mathbf{v}} & * \\ \hline & & Id \end{array}\right) \end{array},$$

where

$$N_{\mathbf{a},\mathbf{v}} = \begin{pmatrix} 1+a_iv_i & a_{i+1}v_i & a_{i+2}v_i & \cdots & a_jv_i \\ a_iv_{i+1} & 1+a_{i+1}v_{i+1} & a_{i+2}v_{i+1} & \cdots & a_jv_{i+1} \\ \vdots & & & & \vdots \\ a_i & a_{i+1} & & \cdots & 1+a_j \end{pmatrix}.$$

Then,

$$T_{\mathbf{a},\mathbf{v}} = y_{j-1}(v_{j-1}) \dots y_{i+1}(v_{i+1}) \cdot y_i(1+a_i^{-1}) \cdot y_{i+1}(-a_{i+1}a_i^{-1}) \dots y_{j-1}(-a_{j-1}a_i^{-1}) \cdot A,$$

where A is the upper-triangular matrix

$$A = \left(\begin{array}{c|c|c} Id & * & * \\ \hline 0 & b & * \\ \hline 0 & 0 & Id \end{array}\right), \quad \text{with } b = \begin{pmatrix} a_i & a_{i+1} & \dots & a_{j-1} & -1-a_j \\ 0 & 1 & \dots & 0 & -v_{i+1} \\ 0 & 0 & \ddots & & \vdots \\ 0 & 0 & 0 & 1 & -v_{j-1} \\ 0 & 0 & 0 & 0 & -a_i^{-1} \end{pmatrix}.$$

See Appendix A for details of this row reduction calculation.

To count how many transvections fit Case 3: There are

- q^{n-i} choices for $a_{i+1}, \dots, a_n$.
- $q-1$ choices of a_i.
- q^{i-1} choices of $v_1, \dots, v_{i-1}$.
- q^{j-1-i} choices of $v_i, \dots, v_{j-1}$ to satisfy $a_iv_i + \dots + v_{j-1}a_{j-1} + a_j = 0$.

The total is

$$\begin{aligned} q^{n-i}(q-1)q^{i-1}q^{j-1-i} &= (q-1)q^{n-i+j-2} = (q-1)q^{n-1-(j-i)}q^{2(j-1-i)+1} \\ &= (q-1)q^{n-1-(j-i)}q^{I(s_{ij})}. \end{aligned}$$

□

Proposition 5.2 now enables the proof of Theorem 5.1.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Proof of Theorem 5.1. Let $\mathbb{C}[G]$ be the group algebra of $G = GL_n(q)$, and $Z(\mathbb{C}[G])$ the center. Then since transvections are a conjugacy class, the sum of transvections is in the center,

$$D = \sum_{\mathbf{a},\mathbf{v}} T_{\mathbf{a},\mathbf{v}} \in Z(\mathbb{C}[G]).$$

Since D commutes with every element of $\mathbb{C}[G]$, we can compute

$$\begin{aligned} D\mathbb{1}_{\mathcal{B}} = D\mathbb{1}_{\mathcal{B}}^2 = \mathbb{1}_{\mathcal{B}} D \mathbb{1}_{\mathcal{B}} &= \sum_{\mathbf{a},\mathbf{v}} \mathbb{1}_{\mathcal{B}} T_{\mathbf{a},\mathbf{v}} \mathbb{1}_{\mathcal{B}} \\ &= \sum_{\substack{\mathbf{a},\mathbf{v}:\\ i>j}} \mathbb{1}_{\mathcal{B}} T_{\mathbf{a},\mathbf{v}} \mathbb{1}_{\mathcal{B}} + \sum_{\substack{\mathbf{a},\mathbf{v}:\\ i<j}} T_{\mathbf{a},\mathbf{v}} \mathbb{1}_{\mathcal{B}} = \sum_{\substack{\mathbf{a},\mathbf{v}:\\ i>j}} \mathbb{1}_{\mathcal{B}} + \sum_{\substack{\mathbf{a},\mathbf{v}:\\ i<j}} \mathbb{1}_{\mathcal{B}} s_{ij} \mathbb{1}_{\mathcal{B}} \\ &= ((n-1)q^{n-1} - [n-1]_q)\mathbb{1}_{\mathcal{B}} + \sum_{1\le i<j\le n} (q-1) q^{n-1-(j-i)} q^{I(s_{ij})} \mathbb{1}_{\mathcal{B}} s_{ij} \mathbb{1}_{\mathcal{B}} \\ &= ((n-1)q^{n-1} - [n-1]_q)\mathbb{1}_{\mathcal{B}} + (q-1) \sum_{1\le i<j\le n} q^{n-1-(j-i)} T_{s_{ij}}. \end{aligned}$$

□

6. A single coset lumping

This section develops the correspondence between $\mathcal{B}\backslash GL_n(q)/\mathcal{B}$ double cosets and flags and describes the random transpositions Markov chain in this setting. This description is useful for analyzing specific features of the Markov chain. If $\{\omega_t\}_{t\ge 0}$ is the induced chain on S_n, then $\{\omega_t(1)\}_{t\ge 0}$ is a process on $\{1,\dots,n\}$. Thinking of $\omega \in S_n$ as a deck of cards, this is the evolution of the label of the 'top card'. It is the lumping of the chain on S_n onto cosets S_n/S_{n-1}. The main result, Lemma 6.11 below, shows that the top card takes only a bounded number of steps to reach stationarity.

Remark 6.1. If Q is *any* probability measure which defines a random walk on G, then the process induced by multiplication on the left by Q on left cosets G/K, for any subgroup $K \subset G$, is always Markov. This is the special case of Proposition 2.4 with $H = \{id\}$ (similarly, if Q defines a random walk by multiplication on the *right*, then it always induces a Markov chain on *right* cosets $H\backslash G$). There are many examples of random walks lumped to single cosets, for example, [26], for which the properties of Q are used to analyze the mixing times of the lumped Markov chain.

6.1. Flag representation

The subgroup $\mathcal{B}$ gives rise to the quotient $GL_n(q)/\mathcal{B}$. This may be pictured as the space of 'flags'.

Definition 6.2. Here, a flag F consists of an increasing sequence of subspaces $F = F_1 \subset F_2 \subset \ldots \subset F_n$ with $\dim(F_i) = i$. The *standard flag* is

$$E = \langle \mathbf{e}_1 \rangle \subset \langle \mathbf{e}_1, \mathbf{e}_2 \rangle \ldots \subset \langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle.$$

Indeed, $GL_n(q)$ operates transitively on flags, and the subgroup fixing the *standard flag* is exactly $\mathcal{B}$. There is a useful notion of 'distance' between two flags F, F' which defines a permutation.

Definition 6.3. Let F, F' be two flags. The *Jordan–Holder permutation* $\omega(F, F')$ is a permutation $\omega = \omega(F, F') \in S_n$ defined by $\omega(i) = j$ if j is the smallest index such that

$$F_{i-1} + F'_j \supseteq F_i.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

The Jordan–Holder permutation distance satisfies

$$\omega(F, F') = \omega(F', F)^{-1} \quad \text{and} \quad \omega(F, F') = \omega(MF, MF') \quad \text{for } M \in GL_n(q). \tag{6.1}$$

A thorough development with full proofs is in [1].

Lemma 6.4. *For $M \in GL_n(q)$ and E the standard flag,*

$$M \in \mathcal{B}\omega\mathcal{B} \iff \omega = \omega(E, ME).$$

6.2. Top label chain

This representation is useful for analyzing a further projection of the chain on S_n: Let $\{\omega_t\}_{t\geq 0}$ denote the Markov chain on S_n. Let $P_1(\cdot,\cdot)$ denote the marginal transition probabilities of the first position. That is,

$$P_1(j,k) = \mathbb{P}(\omega_{t+1}(1) = k \mid \omega_t(1) = j), \quad j,k \in \{1,\dots,n\}.$$

If ω is distributed as π_q, then the marginal distribution on $\{1,\dots,n\}$ of the first card $\omega(1)$ is

$$\pi_{q,1}(j) := P(\omega(1) = j) = \sum_{\omega:\omega(1)=j} \frac{q^{I(\omega)}}{[n]_q!} = \frac{q^{j-1}}{1+q+\ldots+q^{n-1}} = \frac{q^{j-1}}{[n]_q}.$$

Lemma 6.5. *For $1 \leq i, j \leq n$,*

$$P_1(i,j) = \begin{cases} \dfrac{(q-1)^2 q^{n+j-3}}{(q^n-1)(q^{n-1}-1)}, & i \neq j, \\ \dfrac{(q^{n-1}-1)^2 + (q-1)^2(q^{i-1}-1)q^{n-2}}{(q^n-1)(q^{n-1}-1)}, & i = j. \end{cases}$$

Remark 6.6. Observe that for $i \neq j$ the transition probabilities can be written

$$P_1(i,j) = \frac{q^{j-1}}{(q^n-1)/(q-1)}\left(\frac{(q-1)q^{n-2}}{q^{n-1}-1}\right) = \pi_{q,1}(j)\cdot\left(\frac{(q-1)q^{n-2}}{q^{n-1}-1}\right)$$
$$P_1(j,j) = \pi_{q,1}(j)\cdot\left(\frac{(q-1)q^{n-2}}{q^{n-1}-1}\right) + \left(1 - \frac{(q-1)q^{n-2}}{q^{n-1}-1}\right).$$

Write $p := ((q-1)q^{n-2})/(q^{n-1}-1)$. This provides another description of the Markov chain: At each step, flip a coin which gives heads with probability p, tails with probability $1-p$. If heads, move to a random sample from $\pi_{q,1}$. Otherwise, don't move.

Remark 6.7. Though the Markov chain P_1 on $\{1,\dots,n\}$ was defined via lumping from a chain on the group $GL_n(q)$ with $q > 1$ a prime power, note that the transitions are well defined even for $q < 1$. If $q < 1$, then the Mallows measure π_q is concentrated at the identity permutation and $P_1(i,j)$ is largest for $j = 1$. Note also that the description in Remark 6.6 is also valid since $p = ((q-1)q^{n-2})/(q^{n-1}-1) < 1$ for all $q > 0$.

Proof. If $\omega(1) = i$, then a flag representing the double coset has $F_1 = \mathbf{e}_i$. Recall that a transvection T defined by vectors $\mathbf{v}, \mathbf{a}$ has $T(\mathbf{e}_i) = \mathbf{e}_i + a_i\mathbf{v}$. The first coordinate in the new permutation is the smallest j such that

$$\mathbf{e}_i + a_i\mathbf{v} \subset \langle e_1, \dots, e_j \rangle.$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

There are two cases when this smallest j is equal to i:

1. $a_i = 0$, and $\mathbf{v}$ can be anything: There are $(q^{n-1} - 1)$ possibilities for $\mathbf{a}$ such that $a_i = 0$. Then there are $(q^{n-1} - 1)/(q - 1)$ possibilities for $\mathbf{v}$.
2. $a_i \neq 0$ and $\mathbf{v}$ is such that $a_i v_i \neq -1$ and $v_k = 0$ for all $k > i$. Note that if $i = 1$, then this is not possible. Since the nonzero entry at the largest index in $\mathbf{v}$ is normalized to be 1, there are two further possibilities:
   - If $v_i = 1$, then there are $(q - 2)$ possibilities for a_i, $(q^{i-1} - 1)$ possibilities for the rest of $\mathbf{v}$ (note that it's not allowed for the rest of $\mathbf{v}$ to be 0 because then we could not get $\mathbf{a}^\top \mathbf{v} = 0$), and then q^{n-2} possibilities for the rest of $\mathbf{a}$.
   - If $v_i = 0$, then there are $(q^{i-1} - 1)/(q - 1)$ possibilities for $\mathbf{v}$ and then $(q - 1)q^{n-2}$ possibilities for $\mathbf{a}$.

In summary, if $i \neq 1$, then

$$P_1(i,i) = \frac{1}{|\mathcal{T}_{n,q}|}\left(\frac{(q^{n-1}-1)(q^{n-1}-1)}{q-1} + (q^{i-1}-1)(q-2)q^{n-2} + \frac{(q^{i-1}-1)(q-1)q^{n-2}}{q-1}\right)$$
$$= \frac{(q^{n-1}-1)^2 + (q-1)^2(q^{i-1}-1)q^{n-2}}{(q^n-1)(q^{n-1}-1)}.$$

Now, if $i \neq j$ we consider the two possibilities.

1. $i > j$: This transition will occur if $a_i \cdot v_i = -1$ and $\mathbf{v}$ is such that $v_j > 0$ and $v_k = 0$ for all $j < k < i$. Since the $\mathbf{v}$ is normalized so that the entry at the largest index is 1, this requires $v_i = 1$ and $a_i = -1$. There are then $(q - 1)q^{j-1}$ such $\mathbf{v}$, and for each $\mathbf{v}$ there are q^{n-2} possibilities for $\mathbf{a}$. This gives in total
$$(q - 1)q^{n+j-3}.$$
2. $i < j$: This transition will occur if $a_i \neq 0$, $v_j = 1$ and $v_k = 0$ for $k > j$. There are q^{j-1} possibilities for $\mathbf{v}$ and then $(q - 1)q^{n-2}$ possibilities for $\mathbf{a}$, so again in total
$$(q - 1)q^{n+j-3}.$$

In summary, for any $i \neq j$,

$$P_1(i,j) = \frac{(q-1)q^{n+j-3}}{|\mathcal{T}_{n,q}|} = \frac{(q-1)^2 q^{n+j-3}}{(q^n-1)(q^{n-1}-1)}.$$

□

Lemma 6.8. *Let P_1 be the Markov chain from Lemma 6.5 with stationary distribution $\pi_{q,1}$. Then P_1 has eigenvalues $\{1, \beta\}$ with*

$$\beta = \frac{q^{n-2}-1}{q^{n-1}-1},$$

which is equal to $\beta_{(n-1,1)}$ from Section 3.3. The eigenvalue β has multiplicity $n - 1$. If $q > 1$ is fixed and n large, then $\beta \sim 1/q$.

Proof. Call $P_1(i,j) = c_j$, and remember this is constant across all $i \neq j$. Then $P_1(i,i) = p_i = 1 - \sum_{j \neq i} c_j$. Let M be the matrix with column j constant c_j, so the rows are constant $[c_1, \ldots, c_n]$ and let $r = \sum c_j$ be the constant row sum. Note we can write our transition matrix as

$$T = (1 - r)I + M.$$

Note that the diagonal entries of this matrix match what we want since $T(i,i) = (1-r)+c_i = 1-\sum_{j\neq i} c_j = p_i$.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Since M has constant columns/rows it has null space of dimension $n-1$. That is, we can find $n-1$ vectors $v_2, v_3, \ldots v_n$ such that $vM = 0$. Then

$$vT = v((1-r)I + M) = (1-r)v$$

(and remember of course the other eigenvector is the stationary distribution, with eigenvalue 1). This says that there is only one eigenvalue $\beta = 1-r$, with multiplicity $n-1$—eigenfunctions are basis of null space of M. The eigenvalue β is

$$\beta = 1 - \sum_j \frac{(q-1)^2 q^{n+j-3}}{(q^n-1)(q^{n-1}-1)} = 1 - \frac{q^{n-2}(q-1)}{q^{n-1}-1} = \frac{q^{n-2}-1}{q^{n-1}-1}.$$

Note that this is the largest eigenvalue of the full chain on S_n. □

Remark 6.9. This lumping comes from the embedding $S_{n-1} \subset S_n$ as all permutations which fix the first coordinate. Then the coset space S_n/S_{n-1} consists of equivalence classes of permutations which have the same label in the first element. Similarly, we could consider the embedding $S_{n-2} \subset S_n$ as all permutations which fix the first two coordinates. This would induce a Markov chain on the space of $\{(a,b) : 1 \le a, b, \le n, a \ne b\}$; we have not attempted to find the more complicated transitions of this chain.

6.3. Mixing time

The mixing time of the Markov chain P_1 is very fast; the chain reaches stationarity within a constant number of steps (the constant depends on q but not on n). This can be proven using a strong stationary time.

Strong stationary time

An *strong stationary time* for a Markov chain is a random stopping time τ at which the chain is distributed according to the stationary distribution. That is, if $(X_t)_{t\ge 0}$ is a Markov chain on Ω with stationary distribution π, then τ satisfies

$$\mathbb{P}(\tau = t, X_\tau = y \mid X_0 = x) = \mathbb{P}(\tau = t \mid X_0 = x)\pi(y), \quad x, y \in \Omega. \tag{6.2}$$

In words, X_τ has distribution π and is independent from τ. See Section 6.4 of [43].

A strong stationary time is very powerful for bounding convergence time. Intuitively, once the strong stationary time is reached the chain has mixed, so the mixing time is bounded by the random time τ. This idea is formalized in the following.

Proposition 6.10 (Proposition 6.10 from [43]). *If τ is a strong stationary time for a Markov chain P on state space Ω, then for any time t,*

$$\max_{x\in\Omega} \|P^t(x,\cdot) - \pi\|_{TV} \le \max_{x\in\Omega} \mathbb{P}(\tau > t \mid X_0 = x).$$

For most Markov chains it is very difficult to find an obvious strong stationary time. A simple one exists for the chain P_1, using the alternate description from Remark 6.6.

To restate the alternate description, let $p := ((q-1)q^{n-2})/(q^{n-1}-1)$. Let $\{R_t\}_{t\ge 1}$ be a sequence Bernoulli(p) random variables. The Markov chain $\{X_t\}_{t\ge 0}$ defined by P_1 can be coupled with the random variables R_t by: Given X_t,

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

1. If $R_{t+1} = 1$, sample $Z \sim \pi_{q,1}$ and set $X_{t+1} = Z$.
2. If $R_{t+1} = 0$, set $X_{t+1} = X_t$.

Lemma 6.11. *With $\{R_t\}_{t\geq 1}$ defined above, the random time $\tau = \inf\{t > 0 : R_t = 1\}$ is a strong stationary time for $\{X_t\}_{t\geq 0}$. If*

$$t = c \cdot \frac{q^{n-1} - 1}{(q-1)q^{n-2}}, \quad \textit{with } c > 0,$$

then

$$\mathbb{P}(\tau > t) < e^{-c}.$$

Proof. By the alternate description of the Markov chain P_1, whenever $R_t = 1$, the next state X_t is a sample from $\pi_{q,1}$. That is,

$$X_t \mid \{R_t = 1\} \quad \text{is distributed as } \pi_{q,1}.$$

Then, by equation (6.2),

$$\mathbb{P}(\tau = t, X_\tau = y \mid X_0 = x) = \mathbb{P}(\tau = t \mid X_0 = x)\pi_{q,1}(y),$$

and τ is a strong stationary time. The time τ is a geometric random variable with parameter p. Note also that τ is independent of the starting state X_0. Then,

$$\mathbb{P}(\tau > t) = (1-p)^t \leq e^{-pt} = \exp\left(-t\frac{(q-1)q^{n-2}}{(q^{n-1}-1)}\right) \leq e^{-c},$$

for $t = c \cdot \frac{q^{n-1}-1}{(q-1)q^{n-2}}$. □

7. Some extensions

The main example treated above has $G = GL_n(q)$ and $H = K$ the Borel subgroup. As explained in Theorem 1.2 for *any* finite group, *any* subgroups H, K, and *any* H-conjugacy invariant probability measure Q on G ($Q(hsh^{-1}) = Q(s)$ for all $s \in G, h \in H$), the walk on G generated by Q, lumped to double cosets $H\backslash G/K =: \mathcal{X}$ gives a Markov chain on $\mathcal{X}$ with transition kernel

$$K(x, A) = Q(HAKx^{-1}), \quad A \subset \mathcal{X},$$

and stationary distribution the image of Haar measure on G.

There are many possible choices of G, H, K and Q. This gives rise to the problem of making choices and finding interpretations that will be of interest. This section briefly describes a few examples: Gelfand pairs, contingency tables, the extension from $GL_n(q)$ to finite Chevally groups and a continuous example $\mathcal{O}_{n-1}\backslash\mathcal{O}_n/\mathcal{O}_{n-1}$. We have high hopes that further interesting examples will emerge.

7.1. Parabolic subgroups of GL_n

In [39], the authors enumerate the double cosets of $GL_n(q)$ generated by parabolic subgroups.

Definition 7.1. Let $\alpha = (\alpha_1, \ldots, \alpha_k)$ be a partition of n. The parabolic subgroup $P_\alpha \subset GL_n(q)$ consists of all invertible block upper-triangular matrices with diagonal block sizes $\alpha_1, \ldots, \alpha_k$.

Section 4 of [39] shows that if α, β are two partitions of n, the double cosets $P_\alpha\backslash GL_n(q)/P_\beta$ are indexed by contingency tables with row sums α and column sums β. Proposition 4.37 of [39] contains

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

a formula for the size of a double coset which corresponds to the table X, with $\theta = 1/q$,

$$\theta^{-n^2+\sum_{1\le i<i'\le I,1\le j<j'\le J} X_{ij}X_{i'j'}}(1-\theta)^n \cdot \frac{\prod_{i=1}^{I}[\alpha_i]_\theta!\prod_{j=1}^{J}[\beta_j]_\theta!}{\prod_{i,j}[X_{ij}]_\theta!}, \quad I = |\alpha|, J = |\beta|. \tag{7.1}$$

Dividing equation (7.1) by $(1-\theta)^n$ and sending $\theta \to 1$ recovers the usual Fisher–Yates distribution for partitions α, β.

For $\alpha = (n-1,1), \beta = (1^n)$, the contingency tables $P_\alpha \backslash GL_n(q)/P_\beta$ are uniquely determined by the position of the single 1 in the second row. That is, the space is in bijection with the set $\{1,2,\dots,n\}$. If X^j denotes a table with the 1 in the second row in column j, that is, $X^j_{2j} = 1, X^j_{2k} = 0$ for all $k \neq j$, then equation (7.1) becomes

$$q^{-n+n^2-(j-1)}(q-1)^n \cdot [n-1]_{1/q}! = q^{\binom{n}{2}-(j-1)}\frac{q^n - q^{n-1}}{q^n-1} \cdot \prod_{k=1}^{n}(1-q^k),$$

which uses

$$\begin{aligned}[n-1]_{1/q}! &= \frac{((1/q)^{n-1}-1)((1/q)^{n-2}-1)\dots(1/q-1)}{(1/q-1)^{n-1}} = \prod_{i=1}^{n-1}\frac{q^i-1}{q^i-q^{i-1}} \\ &= \frac{q^n-q^{n-1}}{q^n-1}\cdot\prod_{i=1}^{n}\frac{q^i-1}{q^i-q^{i-1}} = \frac{q^n-q^{n-1}}{q^n-1}\cdot\frac{1}{q^{\binom{n}{2}}(q-1)^n}\cdot\prod_{k=1}^{n}(1-q^k).\end{aligned}$$

Dividing by $|GL_n(q)| = (q-1)^n q^{\binom{n}{2}} \cdot [n]_q$ gives

$$q^{-(j-1)} \cdot \frac{(q^n-q^{n-1})(q-1)}{(q^n-1)^2(q-1)^n}\prod_{k=1}^{n}(q^k-1).$$

Note that this distribution is equal to $\pi_{q,1}$ from Section 6 if we map $j \to n-j$, that is, index the double cosets by $n, n-1, \dots, 1$ instead of $1,2,\dots,n$. The 'follow the top card' chain in Section 6 is then equivalent to the induced chain on double cosets $P_\alpha \backslash GL_n(q)/P_\beta$ from random transvections on $GL_n(q)$.

It remains an open problem to further investigate probability distributions and Markov chains on the double cosets of $GL_n(q)$ from parabolic subgroups. In a reasonable sense, for finite groups of Lie type, parabolic subgroups or close cousins are the only systematic families which can occur; see [57].

7.2. Gelfand pairs

A group G with subgroup H is a *Gelfand pair* if the convolution of H-bi-invariant functions is commutative. Probability theory for Gelfand pairs was initiated by Letac [42] and Sawyer [55] who studied the induced chain on

$$GL_n(\mathbb{Z}_p)\backslash GL_n(\mathbb{Q}_p)/GL_n(\mathbb{Z}_p)$$

as simple random walk on a p-ary tree. Many further examples of finite Gelfand pairs appear in [23], [15], [12]. These allow analysis of classical problems such as the Bernoulli–Laplace model of diffusion and natural walks on phylogenetic trees. The literature cited above contains a large number of concrete examples waiting to be interpreted. For surprising examples relating Gelfand pairs, conjugacy class walks on a 'dual group' and 'folding', see [41].

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

We also note that Gelfand pairs occur more generally for compact and noncompact groups. For example, $\mathcal{O}_n/\mathcal{O}_{n-1}$ is Gelfand and the spherical functions become the spherical harmonics of classical physics (this example is further discussed in Section 7.5 below). Gelfand pairs are even useful for large groups such as S_∞ and U_∞, which are not locally compact; see [8] and [47] for research in this direction.

7.3. Contingency tables

Simper [59] considers the symmetric group S_n with parabolic subgroups S_λ and S_μ, for $\lambda = (\lambda_1, \ldots, \lambda_I), \mu = (\mu_1, \ldots, \mu_J)$ partitions of n. Then $S_\lambda \backslash S_n / S_\mu$ can be interpreted as $I \times J$ contingency tables $\{T_{ij}\}_{1 \le i \le I, 1 \le j \le J}$ with row sums λ and column sums μ. Such tables appear in every kind of applied statistical work. See [24] Section 5 for references. The stationary distribution,

$$\pi(T) = \frac{1}{n!} \prod_{i,j} \frac{\lambda_i! \mu_j!}{T_{ij}!},$$

is the familiar Fisher–Yates distribution underlying 'Fisher's exact test for independence'. Markov chains with the Fisher–Yates distribution as stationary were studied in [25]. If Q is the random transposition measure on S_n, [59] uses knowledge of the Q chain to give an eigen-analysis of the chain induced by Q on contingency tables.

7.4. Chevalley groups

Let G be a finite Chevalley group (a finite simple group of Lie type). These come equipped with natural notions of Borel subgroups B and Weyl groups W. The Bruhat decomposition

$$G = \bigsqcup_{\omega \in W} \mathcal{B} \omega \mathcal{B}$$

is in force, and conjugacy invariant probabilities Q on G induce Markov chains on W. Let U be a minimal unipotent conjugacy class in G ([11], Chapter 5) and Q the uniform probability on U. Conjugacy invariance implies that convolving by Q induces an element of $\mathrm{End}_G(G/\mathcal{B})$. This may be identified with the Hecke algebra of B-bi-invariant functions. James Parkinson has shown us that the argument of Section 2.4 (for the computation of D in the Hecke algebra) goes through for a general Chevalley groups G over a finite field $\mathbb{F}_q$. Although a similar formula holds in full generality, for simplicity we will state it in the equal parameter case (nontwisted) and when G is not of type C_n with character lattice equal to the root lattice. Let θ be the highest root, and let $\mathcal{X}_\theta = \{x_\theta(c) \mid c \in \mathbb{F}_q\}$ be the corresponding root subgroup. The conjugacy class of $x_\theta(1)$ is an analogue, for general Chevalley groups, of the conjugacy class of transvections in $GL_n(q)$. The sum of the elements in the conjugacy class of $x_\theta(1)$ provides a Markov chain on $\mathcal{B}\backslash G/\mathcal{B}$ which acts the same way as

$$D = (q-1) \sum_{\alpha \in \Phi_l^+} q^{\frac{1}{2}(I(s_\theta) - I(s_\alpha))} (1 + T_{s_\alpha}),$$

where θ is the highest root, Φ_l^+ is the set of positive long roots, and s_β denotes the reflection in the root β and T_w is the element of the Iwahori–Hecke algebra for $\mathcal{B}\backslash G/\mathcal{B}$ corresponding to the element w in the Weyl group.

Following the ideas in [18], the T_{s_α} can be interpreted via the Metropolis algorithm applied to the problem of sampling from the stationary distribution $\pi(x) = Z^{-1} q^{I(x)}$ on W by choosing random generators. We have not worked out any further examples but would be pleased if someone would.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

7.5. A continuous example

Most of the generalities above extend to compact groups G and closed subgroups $H, K \neq G$. Then, $X = H\backslash G/K$ is a compact space and an H-conjugacy invariant probability Q on G induces a Markov chain on X.

To consider a simple example, let $G = \mathcal{O}_n$, the usual orthogonal group over $\mathbb{R}$ and $H = K = \mathcal{O}_{n-1}$ embedded as all $m \in \mathcal{O}_n$ fixing the 'north pole' $e_1 = (1, 0, 0 \ldots, 0)^\top$. Then, $\mathcal{O}_n/\mathcal{O}_{n-1}$ can be thought of as the $(n-1)$-sphere $\mathcal{S}_{n-1}$. The double coset space $\mathcal{O}_{n-1}\backslash\mathcal{O}_n/\mathcal{O}_{n-1}$ codes up the 'latitude'. Consider the sphere $\mathcal{O}_n/\mathcal{O}_{n-1}$ defined by 'circles' orthogonal to e_1 (see Figure 1).

Then $\mathcal{O}_{n-1}\backslash\mathcal{O}_n/\mathcal{O}_{n-1}$ simply codes which circle contains a given point on the sphere. Thus, $\mathcal{O}_{n-1}\backslash\mathcal{O}_n/\mathcal{O}_{n-1}$ may be identified with $[-1, 1]$.

Represent a uniformly chosen point on the sphere as $x = z/\|z\|$ with $z = (z_1, z_2, \ldots, z_n)$ independent standard normals. The latitude is $z_1/\|z\|$ and so $\pi(x)$ is the distribution of the square root of a $\beta(1/2, (n-1)/2)$ distribution on $[-1, 1]$. When $n = 3$, $\pi(x)$ is uniform on $[-1, 1]$ (theorem of Archimedes).

One simple choice for a driving measure Q on $\mathcal{O}_n$ is 'random reflections'. In probabilistic language, this is the distribution of $I - 2UU^\top$, with U uniform on $\mathcal{S}_{n-1}$. There is a nice probabilistic description of the induced walk on the sphere.

Lemma 7.2. *The random reflections measure on $\mathcal{S}_{n-1}$ has the following equivalent description:*

- *From $x \in \mathcal{S}_{n-1}$, pick a line ℓ through x uniformly (see Figure 2). With probability 1, ℓ intersects the sphere in a unique point y. Move to y.*

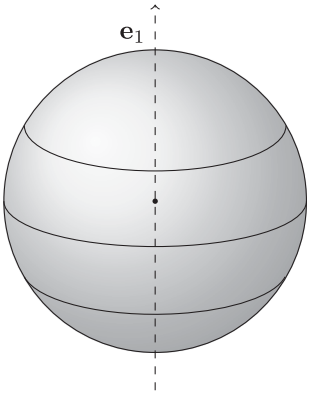


Figure 1. *The space $\mathcal{O}_n/\mathcal{O}_{n-1}$ is defined by the circles on the sphere orthogonal to e_1.*

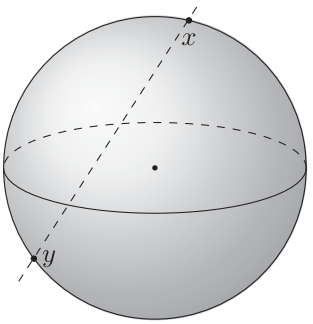


Figure 2. *Illustration of procedure from Lemma 7.2.*

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

Remark 7.3. As the lemma shows, there is a close connection between the walk generated from Q and the popular 'princess and monster' algorithm. See [13]. These algorithms proceed in general convex domains. We know all the eigenvalues of the walk on the sphere and can give sharp rates of convergence.

The induced chain on $[-1,1] \cong \mathcal{O}_{n-1}\backslash\mathcal{O}_n/\mathcal{O}_{n-1}$ is obtained by simply reporting the latitude of y. Thanks to Sourav Chatterjee for the following probabilistic description. For simplicity, it is given here for $n = 3$ (so $\pi(x)$ is uniform on $[-1,1]$).

Lemma 7.4. *For $n = 3$, the Markov chain on $[-1,1]$ described above is; from $x_0 \in [-1,1]$, the chain jumps to*

$$X_1 := (1 - 2U_1^2)x_0 + 2(\cos(\pi U_2))|U_1|\sqrt{1 - U_1^2}\sqrt{1 - x_0^2},$$

where U_1, U_2 are i.i.d. uniform on $[-1,1]$ random variables. (We have checked that the uniform distribution is stationary using Monte Carlo.)

Remark 7.5. For a detailed analysis of the random reflections walk on $\mathcal{O}_n$ (including all the eigenvalues), see [51].

A. Row reduction for $GL_n(q)$

Explicit row reduction can be performed on a general transvection to determine which $\mathcal{B}$-$\mathcal{B}$ double coset it lies in. This allows computing transition probabilities using the combinatorics of possible transvections, which is elementary but tedious. This section proves the following result, used in Section 5.2.

Proposition A.1. *Let $T_{\mathbf{a},\mathbf{v}} = I + \mathbf{v}\mathbf{a}^\top$ be the transvection defined by nonzero vectors $\mathbf{a}, \mathbf{v}$ with $\mathbf{v}^\top\mathbf{a} = 0$ and the last nonzero entry of $\mathbf{v}$ equal to 1. If $j > i$, then*

$$T_{a,v} \in Bs_{j-1}\cdots s_{i+1}s_i s_{i+1}\cdots s_{j-1}B$$

exactly when the last nonzero entry of $\mathbf{v}$ is v_j and the first nonzero entry of a is $\mathbf{a}_i$.

The transvection corresponding to $\mathbf{a}, \mathbf{v}$ is

$$T_{\mathbf{a},\mathbf{v}}(\mathbf{x}) = x + \mathbf{v}(\mathbf{a}^\top\mathbf{x}) \qquad \text{so that} \qquad T_{\mathbf{a},\mathbf{v}}(\mathbf{e}_i) = \mathbf{e}_i + a_i\mathbf{v},$$

and the ith column of $T_{\mathbf{a},\mathbf{v}}$ is $a_i\mathbf{v}$ except with an extra 1 added to the ith entry. So

$$T_{\mathbf{a},\mathbf{v}} = \begin{pmatrix} 1 + a_1v_1 & a_2v_1 & a_3v_1 & \cdots & a_nv_1 \\ a_1v_2 & 1 + a_2v_2 & a_3v_2 & \cdots & a_nv_2 \\ \vdots & & & & \vdots \\ a_1v_n & a_2v_n & & \cdots & 1 + a_nv_n \end{pmatrix} = 1 + \left(a_iv_j\right)_{1\le i,j\le n}.$$

As an example of the row reduction, take $n = 5$ with $v_5 = 1, a_1 \neq 0$. Then,

$$T_{\mathbf{a},\mathbf{v}} = \begin{pmatrix} 1 + a_1v_1 & a_2v_1 & a_3v_1 & a_4v_1 & a_5v_1 \\ a_1v_2 & 1 + a_2v_2 & a_3v_2 & a_4v_2 & a_5v_2 \\ a_1v_3 & a_2v_3 & 1 + a_3v_3 & a_4v_3 & a_5v_3 \\ a_1v_4 & a_2v_4 & a_3v_4 & 1 + a_4v_4 & a_5v_4 \\ a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1 + a_5v_5 \end{pmatrix}$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

$$= y_4(v_4v_5^{-1})\begin{pmatrix} 1+a_1v_1 & a_2v_1 & a_3v_1 & a_4v_1 & a_5v_1 \\ a_1v_2 & 1+a_2v_2 & a_3v_2 & a_4v_2 & a_5v_2 \\ a_1v_3 & a_2v_3 & 1+a_3v_3 & a_4v_3 & a_5v_3 \\ a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \end{pmatrix}$$

$$= y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})\begin{pmatrix} 1+a_1v_1 & a_2v_1 & a_3v_1 & a_4v_1 & a_5v_1 \\ a_1v_2 & 1+a_2v_2 & a_3v_2 & a_4v_2 & a_5v_2 \\ a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \end{pmatrix}$$

$$= y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})y_2(v_2v_5^{-1})\begin{pmatrix} 1+a_1v_1 & a_2v_1 & a_3v_1 & a_4v_1 & a_5v_1 \\ a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 1 & 0 & 0 & -v_2v_5^{-1} \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \end{pmatrix}$$

$$= y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})y_2(v_2v_5^{-1})y_1(a_1^{-1}v_5^{-1}+v_1v_5^{-1})\begin{pmatrix} a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & -a_1^{-1}a_2 & -a_1^{-1}a_3 & -a_1^{-1}a_4 & z \\ 0 & 1 & 0 & 0 & -v_2v_5^{-1} \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \end{pmatrix},$$

with $(a_1^{-1}v_5^{-1}+v_1v_5^{-1})(1+a_5v_5)+z=a_5v_1$ so that

$$z = a_5v_1 - (a_1^{-1}v_5^{-1}+a_1^{-1}a_5+v_1v_5^{-1}+a_5v_1) = -a_1^{-1}a_5 - a_1^{-1}v_5^{-1} - v_1v_5^{-1}.$$

Thus,

$$T_{\mathbf{a},\mathbf{v}} = y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})y_2(v_2v_5^{-1})y_1(a_1^{-1}v_5^{-1}+v_1v_5^{-1})$$
$$\cdot\, y_2(-a_2a_1^{-1})\begin{pmatrix} a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 1 & 0 & 0 & -v_2v_5^{-1} \\ 0 & 0 & -a_3a_1^{-1} & -a_4a_1^{-1} & z-a_2v_2a_1^{-1}v_5^{-1} \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \end{pmatrix}$$
$$= y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})y_2(v_2v_5^{-1})y_1(a_1^{-1}v_5^{-1}+v_1v_5^{-1})$$
$$\cdot\, y_2(-a_2a_1^{-1})y_3(-a_3a_1^{-1})\begin{pmatrix} a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 1 & 0 & 0 & -v_2v_5^{-1} \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & -a_4a_1^{-1} & z-(a_2v_2+a_3v_3)a_1^{-1}v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \end{pmatrix}$$
$$= y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})y_2(v_2v_5^{-1})y_1(a_1^{-1}v_5^{-1}+v_1v_5^{-1})$$
$$\cdot\, y_2(-a_2a_1^{-1})y_3(-a_3a_1^{-1})y_4(-a_4a_1^{-1})\begin{pmatrix} a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 1 & 0 & 0 & -v_2v_5^{-1} \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \\ 0 & 0 & 0 & 0 & z-(a_2v_2+a_3v_3+a_4v_4)a_1^{-1}v_5^{-1} \end{pmatrix}$$

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

$$= y_4(v_4v_5^{-1})y_3(v_3v_5^{-1})y_2(v_2v_5^{-1})y_1(a_1^{-1}v_5^{-1} + v_1v_5^{-1})$$
$$\cdot\, y_2(-a_2a_1^{-1})y_3(-a_3a_1^{-1})y_4(-a_4a_1^{-1})\begin{pmatrix} a_1v_5 & a_2v_5 & a_3v_5 & a_4v_5 & 1+a_5v_5 \\ 0 & 1 & 0 & 0 & -v_2v_5^{-1} \\ 0 & 0 & 1 & 0 & -v_3v_5^{-1} \\ 0 & 0 & 0 & 1 & -v_4v_5^{-1} \\ 0 & 0 & 0 & 0 & -a_1^{-1}v_5^{-1} \end{pmatrix}.$$

Acknowledgments. The authors thank Sourav Chatterjee, David Craven, Jason Fulman, Michael Geline, Bob Guralnick, Jimmy He, Marty Isaacs, Zhihan Li, Martin Liebeck, Gabriel Navarro, James Parkinson, Mehrdad Shahshahani and Nat Thiem for helpful discussions. We also thank an anonymous referee for a rapid, useful report. P.D. was partially supported by NSF grant DMS-1954042. M.S. was supported by a National Defense Science & Engineering Graduate Fellowship and a Lieberman Fellowship from Stanford University.

Conflict of Interest. The authors have no conflict of interest to declare.

References

[1] H. Abels, 'The gallery distance of flags', *Order* **8**(1) (1991), 77–92, 1991.
[2] M. Aguiar, C. André, C. Benedetti, N. Bergeron, Z. Chen, P. Diaconis, A. Hendrickson, S. Hsiao, I. M. Isaacs, A. Jedwab, K. Johnson, G. Karaali, A. Lauve, T. Le, S. Lewis, H. Li, K. Magaard, E. Marberg, J.-C. Novelli, A. Pang, F. Saliola, L. Tevlin, J.-Y. Thibon, N. Thiem, V. Venkateswaran, C. R. Vinroot, N. Yan and M. Zabrocki, 'Supercharacters, symmetric functions in noncommuting variables, and related Hopf algebras', *Adv. Math.* **229**(4) (2012), 2310–2337.
[3] D. J. Aldous, 'Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists', *Bernoulli* (1999), 3–48.
[4] E. Artin, *Geometric Algebra* (Courier Dover Publications, 2016).
[5] J. M. Ball and J. Carr, 'The discrete coagulation-fragmentation equations: Existence, uniqueness, and density conservation', *Journal of Statistical Physics* **61**(1) (1990), 203–234.
[6] N. Berestycki, O. Schramm and O. Zeitouni, 'Mixing times for random k-cycles and coalescence-fragmentation chains', *The Annals of Probability* **39**(5) (2011), 1815–1843.
[7] J. Bertoin, 'Random Fragmentation and Coagulation Processes', vol. 102 (Cambridge University Press, Cambridge, UK, 2006).
[8] A. Borodin and G. Olshanski, 'Representations of the Infinite Symmetric Group, Cambridge Studies in Advanced Mathematics', vol. 160 (Cambridge University Press, Cambridge, 2017).
[9] S. Boyd, P. Diaconis, P. Parrilo and L. Xiao, 'Symmetry analysis of reversible Markov chains', *Internet Mathematics* **2**(1) (2005), 31–71.
[10] A. Bufetov and P. Nejjar, 'Cutoff profile of ASEP on a segment', *Probability Theory and Related Fields* **183**(1) (2022), 229–253.
[11] R. W. Carter, 'Simple Groups of Lie Type', vol. 22 (John Wiley & Sons, Hoboken, NJ, 1989).
[12] T. Ceccherini Silberstein, F. Scarabotti and F. Tolli, 'Harmonic Analysis on Finite Groups', vol. 108 (Cambridge University Press, Cambridge, 2008).
[13] F. Comets, S. Popov, G. M. Schütz and M. Vachkovskaia, 'Billiards in a general domain with random reflections', *Archive for Rational Mechanics and Analysis* **191**(3) (2009), 497–537.
[14] C. W. Curtis and I. Reiner, 'Representation Theory of Finite Groups and Associative Algebras' (AMS Chelsea Publishing, Providence, RI, 2006). Reprint of the 1962 original.
[15] P. Diaconis, 'Group Representations in Probability and Statistics, Institute of Mathematical Statistics Lecture Notes—Monograph Series', vol. 11 (Institute of Mathematical Statistics, Hayward, CA, 1988).
[16] P. Diaconis. 'Random Walks on Groups: Characters and Geometry', London Mathematical Society Lecture Note Series, vol **1**. (Cambridge University Press, Cambridge, 2003) 120–142.
[17] P. Diaconis and M. Malliaris, 'Complexity and randomness in the Heisenberg groups (and beyond)', *New Zealand Journal of Mathematics* **52**(Sep.) (2021), 403–426.
[18] P. Diaconis and A. Ram, 'Analysis of systematic scan Metropolis algorithms using Iwahori–Hecke algebra techniques', *Michigan Math. J.* **48** (2000), 157–190. Dedicated to William Fulton on the occasion of his 60$^{\text{th}}$ birthday.
[19] P. Diaconis, A. Ram and M. Simper, 'Double coset Markov chains', Preprint, 2022, arXiv:2208.10699.
[20] P. Diaconis and L. Saloff-Coste, 'Comparison theorems for reversible Markov chains', *The Annals of Applied Probability* **3**(3) (1993), 696–730.
[21] P. Diaconis and L. Saloff-Coste, 'What do we know about the Metropolis algorithm?' *Journal of Computer and System Sciences* **57**(1) (1998), 20–36, 1998.
[22] P. Diaconis and M. Shahshahani, 'Generating a random permutation with random transpositions', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **57**(2) (1981), 159–179.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

[23] P. Diaconis and M. Shahshahani, 'The subgroup algorithm for generating uniform random variables', *Probability in the Engineering and Informational Sciences* **1**(1) (1987), 15–32.
[24] P. Diaconis and M. Simper, 'Statistical enumeration of groups by double cosets', *Journal of Algebra* 607 (2022), 214–246. Special Issue dedicated to J. Saxl.
[25] P. Diaconis and B. Sturmfels, 'Algebraic algorithms for sampling from conditional distributions', *Annals of Statistics*, **26**(1) (1998), 363–397.
[26] P. W. Diaconis and S. P. Holmes, 'Matchings and phylogenetic trees', *Proceedings of the National Academy of Sciences* **95**(25) (1998), 14600–14602.
[27] R. Durrett, B. L. Granovsky and S. Gueron, 'The equilibrium behavior of reversible coagulation-fragmentation processes', *Journal of Theoretical Probability* **12**(2) (1999), 447–474.
[28] A. Eskenazis and E. Nestoridi, 'Cutoff for the Bernoulli–Laplace urn model with $o(n)$ swaps, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **56**(4) (2020), 2621–2639.
[29] A. Gladkich and R. Peled, 'On the cycle structure of Mallows permutations', *The Annals of Probability* **46**(2) (2018), 1114–1169.
[30] R. M. Guralnick, M. Larsen and P. H. Tiep, 'Character levels and character bounds', *in Forum of Mathematics, Pi*, vol. 8 (Cambridge University Press, Cambridge, 2020).
[31] T. Halverson and A. Ram, q-rook monoid algebras, Hecke algebras, and Schur-Weyl duality', *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 283 (Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 6) (2001), 224–250, 262–263.
[32] G. H. Hardy and S. Ramanujan, 'Asymptotic formulae in combinatory analysis', *Proceedings of the London Mathematical Society* s2-17(1) (1918), 75–115.
[33] J. He, T. Möller and T. Verstraaten, 'Cycles in Mallows random permutations', Preprint, 2022, arXiv:2201.11610.
[34] H. A. Helfgott, Á. Seress and A. Zuk, 'Random generators of the symmetric group: diameter, mixing time and spectral gap', *Journal of Algebra* **421** (2015), 349–368.
[35] M. Hildebrand, 'Generating random elements in $SL_n(F_q)$ by random transvections', *Journal of Algebraic Combinatorics* **1**(2) (1992), 133–150.
[36] E. Hingant and R. Yvinec, 'Quasi-stationary distribution and metastability for the stochastic Becker–Döring model', *Electronic Communications in Probability* **26** (2021), 1–14, 2021.
[37] R. Howe, A century of Lie theory', in *American Mathematical Society centennial publications*, vol. II *(Providence, RI,* 1988*)* (Amer. Math. Soc., Providence, RI, 1992), 101–320.
[38] G. James and A. Kerber, 'The Representation Theory of the Symmetric Group' (Cambridge University Press, Cambridge, 2009).
[39] S. N. Karp and H. Thomas, 'q-whittaker functions, finite fields, and Jordan forms', Preprint, 2022, arXiv:2207.12590.
[40] J. G. Kemeny and J. L. Snell, *Finite Markov Chains: With a New Appendix 'Generalization of a Fundamental Matrix'* (Springer, New York, NY, 1983).
[41] R. Lawther, 'Folding actions', *Bulletin of the London Mathematical Society* **25**(2) (1993), 132–144.
[42] G. Letac, 'Problèmes classiques de probabilité sur un couple de Gelfand', in *Analytical Methods in Probability Theory (Oberwolfach, 1980)*, Lecture Notes in Math., vol. 861 (Springer, Berlin-New York, 1981), 93–120.
[43] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, vol. 107 (American Mathematical Soc., 2017).
[44] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, vol. 10 (Springer, New York, NY, 2001).
[45] G. Lusztig, 'Bruhat decomposition and applications', Preprint, 2010, arXiv:1006.5004.
[46] I. G. Macdonald, *Symmetric Functions and Hall Polynomials* (Oxford University Press, Oxford, UK, 1998).
[47] Y. Neretin, 'Infinite-dimensional -adic groups, semigroups of double cosets, and inner functions on Bruhat–Tits buildings', *Izvestiya: Mathematics* **79** (2011), 08.
[48] J. Paguyo, 'Fixed points, descents, and inversions in parabolic double cosets of the symmetric group', Preprint, 2021, arXiv:2112.07728.
[49] C. Pang, 'Lumpings of algebraic Markov chains arise from subquotients', *Journal of Theoretical Probability* **32**(4) (2019), 1804–1844.
[50] J. Pitman, *Combinatorial Stochastic Processes: Ecole d'eté de probabilités de saint-flour xxxii-2002* (Springer, New York, NY, 2006).
[51] U. Porod, 'The cut-off phenomenon for random reflections', *The Annals of Probability* **24**(1) (1996), 74–96.
[52] A. Ram, 'Seminormal representations of Weyl groups and Iwahori–Hecke algebras', *Proceedings of the London Mathematical Society* **75**(1) (1997), 99–133.
[53] L. Saloff-Coste, 'Random walks on finite groups', in *Probability on Discrete Structures* (Springer, New York, NY, 2004), 263–346.
[54] L. Saloff-Coste and J. Zúñiga, 'Refined estimates for some basic random walks on the symmetric and alternating groups', Preprint, 2008, arXiv:0809.0688.
[55] S. Sawyer, 'Isotropic random walks in a tree', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **42**(4) (1978), 279–292.
[56] O. Schramm, 'Compositions of random transpositions', *Israel Journal of Mathematics* **147**(1) (2005), 221–243.
[57] G. M. Seitz, 'Small rank permutation representations of finite Chevalley groups', *Journal of Algebra* **28**(3) (1974), 508–517.
[58] J.-P. Serre, *Linear Representations of Finite Groups*, vol. 42. (Springer, New York, NY, 1977).

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press

[59] M. Simper, ' *Double coset Markov chains*', PhD thesis, Stanford University, 2022.
[60] M. Simper, 'Random transpositions on contingency tables', Preprint, 2022, arXiv:2208.10700.
[61] G. Strang, 'The algebra of elimination', in *Excursions in Harmonic Analysis, Volume* 3 (Springer, New York, NY, 2015), 3–22.
[62] M. Suzuki, 'Group Theory. I, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]', vol. 247 (Springer-Verlag, Berlin-New York, 1982). Translated from the Japanese by the author.
[63] L. Teyssier, 'Limit profile for random transpositions', *The Annals of Probability* **48**(5) (2020), 2323–2343, 2020.
[64] C. Zhong, 'Mallows permutation model: Sampling algorithms and probabilistic properties', PhD thesis, Stanford University, 2022.

https://doi.org/10.1017/fms.2022.106 Published online by Cambridge University Press