Alea 4, 359–392 (2008)

Refined estimates for some basic random walks on the symmetric and alternating groups

L. Saloff-Coste and J. Zúñiga

Department of Mathematics, Cornell University, Malott Hall, Ithaca, NY 14850
E-mail address: lsc@math.cornell.edu

Department of Mathematics, Cornell University, Malott Hall, Ithaca, NY 14850
E-mail address: zuniga@math.cornell.edu

Abstract. We give refined estimates for the discrete time and continuous time versions of some basic random walks on the symmetric and alternating groups S_n and A_n. We consider the following models: random transposition, transpose top with random, random insertion, and walks generated by the uniform measure on a conjugacy class. In the case of random walks on S_n and A_n generated by the uniform measure on a conjugacy class, we show that in continuous time the ℓ^2-cutoff has a lower bound of $(n/2)\log n$. This result, along with the results of Müller, Schlage-Puchta and Roichman, demonstrates that the continuous time version of these walks may take much longer to reach stationarity than its discrete time counterpart.

1. Introduction

This work is concerned with some basic random walks on the symmetric group, S_n, and the alternating group, A_n. Specifically, we are interested in the following models: (a) Random transposition and transpose top with random; (b) walks generated by the uniform measure on a conjugacy class, e.g., 4-cycles or k_n-cycles with k_n an increasing function of n; (c) random insertion. Although these walks have been studied extensively, we obtain here results that either improved upon known estimates or complement those estimates.

The convergence of the random transposition walk on S_n was studied by Diaconis and Shahshahani (1981). We present a technical improvement of their fundamental result. This is motivated by the role played by random transposition in the comparison technique of Diaconis and Saloff-Coste (1994): any improvement upon the ℓ^2 convergence of the random transposition walk has consequences for a wealth of

Received by the editors April 18, 2008; accepted August 13, 2008.

2000 *Mathematics Subject Classification.* Primary 60J99, Secondary 60J10, 60J27.

Key words and phrases. Convergence of Markov chains, random walks, cutoff phenomenon

L. Saloff-Coste is partially supported by NSF grant DMS 0603806 and J. Zúñiga is partially supported by NSF grants DMS 0603886 and DMS 0306194.

other walks. We will illustrate this by obtaining the best known result for the random insertion walk. These results are also used in Saloff-Coste and Zúñiga (2007) to study certain time in-homogeneous versions of the random insertion walk and this was indeed our original motivation for developing the results presented here. For an overview of results connected to the random transposition walk, see Diaconis (2003).

The transpose top with random walk is an interesting example mentioned in Flatto et al. (1985) and in Diaconis (1991) but details of its ℓ^2 analysis have never appeared in print. (This walk should not be confused with the more classical top to random walk studied in Diaconis et al., 1992.) The estimates concerning this walk that are proved here are used in Saloff-Coste and Zúñiga (2007) to obtain the best known convergence bounds for a class of time in-homogeneous processes called semi-random transpositions.

Random walks associated with conjugacy classes other than the class of transpositions have been studied by Müller and Schlage-Puchta (2007), Schlage-Puchta (2008), Lulov and Pak (2002), Roichman (1996) and Roussel (1999). For most of those walks, we show that ℓ^2 convergence occurs at very different times for the discrete time process and the continuous time process. Although this phenomenon is simple to understand a posteriori, it is a bit surprising at first and is often overlooked.

Let us briefly describe our notation. On a finite group G with identity element e, the random walk started at e driven by a given probability measure q is the process $X_n = \xi_1 \cdot \dots \cdot \xi_n$ where the ξ_i are independent G-valued random variables with distribution q. The distribution of X_n is $q^{(n)}$, the convolution of q with itself, n times. Any such walk admits the uniform measure u as an invariant measure. It is reversible if and only if $q(x) = q(x^{-1})$ for all x. The walks studied here all have this property. We are mostly interested in the quantity (χ-square distance)

$$d_2(q^{(n)}, u) = \left(|G| \sum_G |q^{(n)} - u|^2 \right)^{1/2}, \quad u \equiv 1/|G|.$$

This is always an upper bound for $2\|q^{(n)} - u\|_{\mathrm{TV}}$ where

$$\|q - p\|_{\mathrm{TV}} = \sup_A \{q(A) - p(A)\}$$

is the total variation distance between the probability measures p and q.

Given such a discrete time process, we also consider the associated continuous time process whose distribution at time $t \in [0, \infty)$ is given by

$$h_t(x) = h_{q,t}(x) = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} q^{(n)}(x).$$

We now state some of the results proved in this work. Random transposition is the walk on the symmetric group $G = S_n$ driven by $q = q_{\mathrm{RT}}$ where

$$q_{\mathrm{RT}}(\tau) = \begin{cases} 2/n^2 & \text{if } \tau = (i,j),\ 1 \le i, j \le n,\ \ i \ne j, \\ 1/n & \text{if } \tau = e \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.1. *Let q be the random transposition measure on the group S_n, $n > 14$. For any $c \ge 0$ and $t \ge \frac{n}{2}(\log n + c)$, we have $d_2(q^{(t)}, u) \le 2e^{-c}$.*

Diaconis and Shahshahani (1981) prove this result with an unspecified constant B instead of 2 in front of e^{-c} and for large enough n. In this paper their approach is refined to obtain the bound stated above. We also prove a similar result in continuous time which turns out to be somewhat more difficult. Having good control of $d_2(h_{q_{\mathrm{RT}},t}, u)$ is very useful in connection with the comparison techniques of Diaconis and Saloff-Coste (1994). See Section 4.3 where this is used to study the random insertion walk.

Transpose top with random is the process driven by $q(\tau) = 1/n$ if $\tau \in \{(1,i), i = 1, \dots, n\}$ (where $(1,1) = e$) and 0 otherwise.

Theorem 1.2. *Let q be the transpose top with random measure on the group S_n. For any $c \geq 0$ and $t \geq n(\log n + c)$, we have $d_2(q^{(t)}, u) \leq \sqrt{2}e^{-c}$.*

To illustrate our results concerning walks driven by conjugacy classes, consider the measure $q_{\mathbf{c}_n}$ which, for each n, is uniform on $\mathbf{c}_n \subset S_n$, the conjugacy class of all cycles of odd length $k_n = 2m_n + 1$. The corresponding walk is on A_n.

Theorem 1.3. *Fix $\epsilon \in (0,1)$ and set $t_n = \frac{n}{2}\log n$. Referring to the continuous time process with distribution $h_{\mathbf{c}_n,t} = h_{q_{\mathbf{c}_n},t}$ associated to the cycle walk on A_n described above, if m_n tends to infinity with n, we have (with $u_n \equiv 1/|A_n| = 2/n!$)*

$$\lim_{n\to\infty} d_2(h_{\mathbf{c}_n,(1+\epsilon)t_n}, u_n) = 0 \text{ and } \lim_{n\to\infty} d_2(h_{\mathbf{c}_n,(1-\epsilon)t_n}, u_n) = \infty.$$

This result shows an ℓ^2-cutoff at time $(n/2)\log n$. When $k_n < cn$ for some $c \in (0,1)$, Roichman (1996) shows that the associated discrete time process has a mixing time in ℓ^2 of order $(n/k_n)\log n$. Roichman's results are improved in Müller and Schlage-Puchta (2007) and Schlage-Puchta (2008). As $k_m = 2m_n + 1 \to \infty$, the discrete mixing time $(n/k_n)\log n$ is much smaller than the continuous cutoff time $(n/2)\log n$. The explanation is simple. Consider the eigenvalues of the walk driven by $q_{\mathbf{c}_n}$, that is, the eigenvalues of the convolution operator $f \to f * q_{\mathbf{c}_n} : \ell^2(u_n) \to \ell^2(u_n)$, call these eigenvalues α_i. In continuous time, the ℓ^2 cutoff time is controlled by the very large number of very small eigenvalues. These small eigenvalues contribute significantly in continuous time because they appear in the form $e^{-t(1-\alpha_i)}$. In discrete time, these small eigenvalues do not contribute much since they appear in the form α_i^t. Although the explanation is simple, verifying that this is indeed the case is not an easy task. We will prove similar results for general conjugacy classes.

2. Review and notation

We refer the reader to Diaconis (1988) and Saloff-Coste (2004) for careful introduction to random walks on finite groups. We briefly review some of the needed material below.

2.1. *Cutoffs.* Many examples of random walks on groups that have been studied demonstrate a unique behavior called the cutoff phenomenon. This was first studied in Aldous (1983), Aldous and Diaconis (1986) and Diaconis and Shahshahani (1981). See also Diaconis (1996), Chen and Saloff-Coste (2008), Saloff-Coste (1997) and Saloff-Coste (2004).

Definition 2.1. Let $(G_n)_0^\infty$ be a sequence of finite groups and denote by u_n the uniform measure on G_n. For each $n \geq 0$ consider the random walk on G_n driven by the measure q_n. The sequence $((G_n, q_n))_0^\infty$ is said to have total variation cutoff (resp. ℓ^2) if there is a sequence $(t_n)_0^\infty$ with $t_n \to \infty$ such that for any $\epsilon \in (0,1)$

(1) if $k_n = (1+\epsilon)t_n$ then $d_{\mathrm{TV}}(p_n^{(k_n)}, u_n) \to 0$ (resp. $d_2(p_n^{(k_n)}, u_n) \to 0$);

(2) if $k_n = (1-\epsilon)t_n$ then $d_{\mathrm{TV}}(p_n^{(k_n)}, u_n) \to 1$ (resp. $d_2(p_n^{(k_n)}, u_n) \to \infty$).

The sequence $((G_n, q_n))_0^\infty$ is said to demonstrate a total variation (resp. ℓ^2) precutoff if there exist constants $0 < a < b$ such that

(1) $\liminf_{n\to\infty} d_{\mathrm{TV}}(p_n^{(at_n)}, u_n) > 0$ (resp. $\liminf_{n\to\infty} d_2(p_n^{(at_n)}, u_n) > 0$);

(2) $\lim_{n\to\infty} d_{\mathrm{TV}}(p_n^{(bt_n)}, u_n) = 0$ (resp. $\liminf_{n\to\infty} d_2(p_n^{(bt_n)}, u_n) = 0$).

Similar definitions apply in continuous time. Diaconis and Shahshahani (1981) prove that the random transposition walk on S_n has a cutoff (both in total variation and ℓ^2) at time $(n/2)\log n$. For a overview of other results in this direction, see Diaconis (1996) and Saloff-Coste (2004).

2.2. *Eigenvalues and representation theory.* It is well known that for reversible finite Markov chains, the χ-square distance can be expressed in terms of eigenvalues and eigenfunctions. (See, e.g., Saloff-Coste, 1997). For a reversible random walk on a finite group G driven by q, the expression simplifies and the eigenvectors drop out. If we let β_i, $i = 0, \dots, |G|-1$, be the eigenvalues of the operator of convolution by q acting on $\ell^2(G)$, in non-increasing order and repeated according to multiplicity, we have

$$d_2(q^{(t)}, u)^2 = \sum_{i=1}^{|G|-1} \beta_i^{2t} \quad\text{and}\quad d_2(h_t, u)^2 = \sum_{i=1}^{|G|-1} e^{-2t(1-\beta_i)}. \tag{2.1}$$

Representation theory provides a tool that can be helpful to compute eigenvalues. We give a very brief review of these methods. All the material in this section can be found in greater detail in Diaconis (1988) and Sagan (2001). A *representation* of a finite group G on a vector space V is a homomorphism $\rho : G \to GL(V)$ where $GL(V)$ is the group of general linear transformations of V. We say that ρ has dimension d_ρ where d_ρ is equal to the dimension of V. Let $W \subset V$, if $\rho W = W$ then $\rho|_W$ is called a *sub-representation* of ρ. A representation ρ is called *irreducible* if it admits no nontrivial sub-representation. The *character* of a representation ρ at $s \in G$ is $\chi_\rho = \mathrm{Tr}(\rho(s))$. Characters are constant under conjugation, i.e. for any $x, y \in G$ then

$$\chi_\rho(x^{-1}yx) = \chi_\rho(y).$$

For $f : G \to \mathbb{R}$, the *Fourier transform* of f at ρ is

$$\widehat{f}(\rho) = \sum_{s\in G} f(s)\rho(s).$$

The Fourier transform converts convolution of functions into multiplication of matrices (or composition of linear maps) $\widehat{f * g}(\rho) = \widehat{f}(\rho)\widehat{g}(\rho)$. If G is a finite group and if f, g are any two functions taking values on G then the *Plancharel formula* relates the convolution of f and g at e to the Fourier transform as follows

$$f * g(e) = \sum_{s\in G} f(s^{-1})g(s) = \frac{1}{|G|}\sum_{\rho} \mathrm{d}_\rho \mathrm{Tr}(\widehat{f}(\rho)\widehat{g}(\rho))$$

where $|G|$ is the order of G and the sum is over all (equivalent classes of) irreducible representations of G. In what follows $\rho \neq 1$ means that ρ is not the trivial representation. The Plancharel formula is used to obtain the following proposition.

Proposition 2.2. *Let G be a finite group equipped with a probability measure q satisfying $q(x) = q(x^{-1})$, $x \in G$. We have*

$$d_2(q^{(t)}, u)^2 = \sum_{\rho \neq 1} \mathrm{d}_\rho \operatorname{Tr}(\widehat{q}(\rho)^{2t}). \tag{2.2}$$

In general, it is very difficult to estimate $\operatorname{Tr}(\widehat{q}(\rho)^t)$. However, in the case were q is a class function, i.e., $q(x^{-1}yx) = q(y)$, for all $x, y \in G$, a celebrated lemma of Schur provides a nice analysis. If ρ is an irreducible representation and $(\mathcal{C}_j)_1^m$ are the conjugacy classes of the group G then $\widehat{q}(\rho)$ is a constant multiple of the identity matrix. This yields

$$\widehat{q}(\rho) = I_{\mathrm{d}_\rho} \cdot \left(\sum_{j=1}^{m} q(\mathcal{C}_j) \frac{\chi_\rho(c_j)}{\mathrm{d}_\rho} \right)$$

where $c_j \in \mathcal{C}_j$. For a proof of this fact see Diaconis (1988) and Diaconis (1991). The next proposition now follows.

Proposition 2.3. *Let G be a finite group and q a probability measure on G satisfying $q(x^{-1}) = q(x)$, $x \in G$. If q is constant on conjugacy classes then*

$$\begin{aligned} d_2(q^{(t)}, u)^2 &= \sum_{\rho \neq 1} \mathrm{d}_\rho^2 \left(\sum_{j=1}^{m} q(\mathcal{C}_j) \frac{\chi_\rho(c_i)}{\mathrm{d}_\rho} \right)^{2t} \quad \text{and} \\ d_2(h_{\mathcal{C},t}, u)^2 &= \sum_{\rho \neq 1} \mathrm{d}_\rho^2 \exp\left(-2t \left(1 - \sum_{j=1}^{m} q(\mathcal{C}_j) \frac{\chi_\rho(c_i)}{\mathrm{d}_\rho} \right) \right). \end{aligned} \tag{2.3}$$

To connect more directly representation theory with the usual spectral decomposition, let $\rho : G \to GL(V)$ be a representation of G on a finite vector space V equipped with an invariant Hermitian product $\langle \cdot, \cdot \rangle$. Fix a probability measure q and consider the linear transformation $\widehat{q}(\rho) : V \to V$. Suppose e_i, e_j are unit vectors in V and that e_i is an eigenvector of $\widehat{q}(\rho)$ with eigenvalue γ_i. Set $\phi_{i,j,\rho}(x) = \langle \rho(x) e_i, e_j \rangle$. We claim that $\phi_{i,j,\rho}$ is an eigenfunction for $f \mapsto f * q$ with

$$f * q(x) = \sum_y f(xy^{-1}) q(y)$$

on $\ell^2(G)$ with eigenvalue γ_j. Indeed,

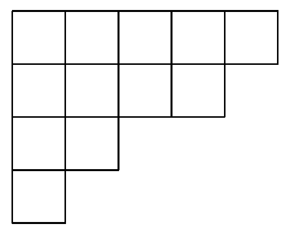
$$\begin{aligned} \phi_{i,j,\rho} * q(x) &= \sum_y q(y) \langle \rho(xy^{-1}) e_i, e_j \rangle = \left\langle \rho(x) e_i, \sum_y q(y) \rho(y) e_j \right\rangle \\ &= \langle \rho(x) e_i, \widehat{q}(\rho) e_j \rangle = \gamma_j \langle \rho(x) e_i, e_j \rangle = \gamma_j \phi_{i,j,\rho}(x). \end{aligned}$$

Now, if q is symmetric and thus $\widehat{q}(\rho)$ is diagonalizable in an orthonormal basis $(e_i)_1^{\mathrm{d}_\rho}$ then the construction above yields d_ρ eigenvalues and d_ρ^2 orthonormal eigenvectors in $\ell^2(G)$, each eigenvalue having multiplicity d_ρ. Furthermore, if ρ, ρ', are two inequivalent irreducible representations the corresponding eigenvectors are orthogonal (some of the eigenvalues may be the same). A proof of the orthogonality of $\phi_{i,j,\rho}$ is given in Corollary 4.10 of Knapp (1996). Hence, this produces $|G|$ orthonormal eigenfunctions since $\sum_\rho \mathrm{d}_\rho^2 = |G|$ where the sum is taken over all (equivalent classes of) irreducible representations.

For future reference we mention the well known fact that irreducible representations on S_n are indexed by the Young diagram with n boxes (see Sagan, 2001).

Definition 2.4. Let $\lambda = (\lambda_1, \ldots, \lambda_m)$ be a partition of n so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and $\sum_{i=1}^{m} \lambda_i = n$. λ is called a Young diagram of n boxes and λ_i denotes the number of boxes in the i-th row of the diagram.

FIGURE 2.1. The Young diagram for $\lambda = (5, 4, 2, 1)$



The association of an irreducible representation to a Young diagram will provide a key tool to calculate the normalized character $\chi_\rho(\cdot)/\mathrm{d}_\rho$ of an irreducible representation ρ and the eigenvalues of many of the walks we study. This technique is illustrated in the following sections.

3. Transpose top with random

Consider the following shuffling method of a deck of n cards: pick a card uniformly at random from the deck and transpose it with the top card. This shuffling scheme is described by the measure q on the symmetric group $G = S_n$ where

$$q(\tau) = \begin{cases} 1/n & \text{if } \tau = (1, j),\ 1 \leq j \leq n \\ 0 & \text{otherwise.} \end{cases}$$

This walk is called transpose top with random.

In order to establish an upper bound for the ℓ^2 mixing time, the tools from group representation presented in Section 2.2 are used to calculate the eigenvalues of q. Most of the needed computations are in Flatto et al. (1985) and the procedure is outlined in Diaconis (1991) where it is stated that transpose top with random has a cutoff time of $n \log n$. The following theorem gives a more precise upper bound. This result is used in Saloff-Coste and Zúñiga (2007) to study a class of time inhomogeneous chains called semi-random transpositions.

Theorem 3.1. *Let q be the transpose top with random measure on the group S_n. If $n \geq 1$, $c \geq 0$, and $t \geq n(\log n + c)$*

$$d_2(q^{(t)}, u) \leq \sqrt{2}\, e^{-c}, \quad d_2(h_{q,t}, u) \leq \sqrt{2}\, e^{-c}.$$

Proof. By Proposition 2.2,

$$d_2(q^{(t)}, u)^2 = \sum_{\rho \neq 1} \mathrm{d}_\rho \mathrm{Tr}(\widehat{q}(\rho)^{2t}).$$

Even though q is not constant on conjugacy classes Diaconis (1991) notes that q is invariant under conjugation by elements of S_{n-1} where

$$S_{n-1} = \{\tau \in S_n | \tau(1) = 1\}.$$

Using this fact it is shown that $\widehat{q}(\rho)$ is a diagonal matrix (with real entries). See Flatto et al. (1985) and Diaconis (1991). Therefore

$$\mathrm{Tr}\left(\widehat{q}(\rho)^{2t}\right) = \sum_{i=1}^{\mathrm{d}_\rho} \alpha_i^{2t} \leq \mathrm{d}_\rho \alpha_1^{2t} \tag{3.1}$$

where $\alpha_1 \geq \cdots \geq \alpha_{\mathrm{d}_\rho}$ are the eigenvalues of $\widehat{q}(\rho)$.

To compute α_i, consider $M = \sum_{i=2}^n \rho((1,i))$. Let $\lambda = (\lambda_1, \ldots, \lambda_m)$ be the Young diagram associated to the irreducible representation ρ. Let $\sigma_1 \geq \cdots \geq \sigma_{\mathrm{d}_\rho}$ be the eigenvalues of M. In Flatto et al. (1985) it is shown that for $1 \leq i \leq \mathrm{d}_\rho$ then

$$\sigma_i = \lambda_i - i.$$

The multiplicity of each σ_i is also described in Flatto et al. (1985). We do not need this for the present proof but, to give an example, if $\lambda = (n-1, 1)$ then the eigenvalues of M are $\sigma_1 = n-2$ with multiplicity $n-2$ and $\sigma_2 = -1$ with multiplicity 1.

As

$$\widehat{q}(\rho) = \sum_{\tau \in G} q(\tau)\rho(\tau) = \sum_{i=1}^n \left(\frac{1}{n}\right) \rho((1,i)) = \frac{M + \rho(e)}{n} = \frac{M+I}{n}$$

where I is the identity matrix of dimension d_ρ, we easily obtain the eigenvalues α_i, $1 \leq i \leq \mathrm{d}_\rho$: $\alpha_i = (\sigma_i + 1)/n$. In particular, $\alpha_1 = \lambda_1/n$.

Denote by ρ_λ the irreducible representation associated to a partition λ and by $\rho_\lambda = 1$ the trivial representation with corresponds to $\lambda = (n)$. Equation (3.1) yields

$$d_2(q^{(t)}, u)^2 \leq \sum_{\rho_\lambda \neq 1} \mathrm{d}_{\rho_\lambda}^2 \left(\frac{\lambda_1}{n}\right)^{2t} = \sum_{j=1}^{n-1} \sum_{\substack{\rho_\lambda \\ \lambda_1 = n-j}} \mathrm{d}_\lambda^2 \left(\frac{\lambda_1}{n}\right)^{2t}. \tag{3.2}$$

In Flatto et al. (1985) and Diaconis (1988) it is shown that for $l \geq 1$

$$\sum_{\substack{\rho_\lambda \\ \lambda_1 = l}} \mathrm{d}_{\rho_\lambda}^2 \leq \binom{n}{l}^2 (n-l)!. \tag{3.3}$$

It follows that for $c \geq 0$ and $t \geq n(\log n + c)$

$$\begin{aligned} d_2(q^{(t)}, u)^2 &\leq \sum_{j=1}^{n-1} \left(\frac{n!}{(n-j)!}\right)^2 \left(\frac{1}{j!}\right) \left(1 - \frac{j}{n}\right)^{2t} \\ &\leq \sum_{j=1}^{n-1} n^{2j} \left(\frac{1}{j!}\right) e^{-2j \log n} e^{-2jc} = (e-1)e^{-2c} \leq 2e^{-2c}. \end{aligned}$$

For the continuous time process, we have, similarly,

$$\begin{aligned} d_2(h_{q,t},u)^2 &\leq \sum_{\rho_\lambda \neq 1} \mathrm{d}_{\rho_\lambda}^2 e^{-2t(1-\alpha_1)} = \sum_{j=1}^{n-1} \sum_{\substack{\rho_\lambda \\ \lambda_1 = n-j}} \mathrm{d}_\lambda^2 \exp\{-2t(1-\lambda_1/n)\} \\ &= \sum_{j=1}^{n-1} \sum_{\substack{\rho_\lambda \\ \lambda_1 = n-j}} \mathrm{d}_\lambda^2 e^{-2tj/n} \leq \sum_{j=1}^{n-1} \left(\frac{n!}{(n-j)!}\right)^2 \left(\frac{1}{j!}\right) e^{-2tj/n} \end{aligned}$$

where the last inequality follows from (3.3). Again, if $n \geq 1$, $c \geq 0$ and $t \geq n(\log n + c)$ then

$$d_2(h_{q,t},u)^2 \leq \sum_{j=1}^{n-1} n^{2j} \left(\frac{1}{j!}\right) e^{-2j\log n} e^{-2jc} \leq 2e^{-2c}.$$

□

The next proposition shows that transpose top with random has ℓ^2 and total variation cutoffs at time $n \log n$.

Proposition 3.2. *Let q be the transpose top with random measure on S_n. For any sequence $(k_n)_0^\infty$ such that $(k_n - n\log n)/n$ tends to $-\infty$ as n tends to ∞ then*

$$d_2(q^{(k_n)}, u) \to \infty \;\; \textit{and} \;\; d_{\mathrm{TV}}(q^{(k_n)}, u) \to 1.$$

Proof. For the ℓ^2 bound, we observe that Flatto et al. (1985) also gives a description of the multiplicity of the eigenvalues. In particular, if $\lambda = (n-1, 1)$ then the eigenvalue $1 - 1/n$ of $\widehat{q}(\rho_\lambda)$ has multiplicity $n-2$. Since $\mathrm{d}_\lambda = n-1$ we get that

$$d_2(q^{(k)}, u)^2 \geq (n-1)(n-2)(1-1/n)^{2k}$$

from which the desired ℓ^2 statement easily follows.

Remark 3.3. Let $\varphi(\sigma)$ be the number of fixed points of σ. One can check by direct inspection that

$$f(\sigma) = \left(\frac{n-1}{n-2}\right)^{1/2} \times \begin{cases} \varphi(\sigma) - 2 & \text{if } \sigma(1) = 1 \\ \varphi(\sigma) - 1 + \frac{1}{n-1} & \text{if } \sigma(1) \neq 1 \end{cases}$$

is a normalized eigenfunction (for convolution by q) with eigenvalue $1 - 1/n$. Its value at e is $f(e)^2 = (n-1)(n-2)$. This gives a entirely elementary proof of the ℓ^2 lower bound since $d_2(q^{(k)}, u)^2 \geq (1 - 1/n)^{2k} f(e)^2$. The previous inequality results from the fact that one can write the χ-square distance in terms of eigenvalues and eigenfuctions. (See,e.g., Saloff-Coste, 1997).

The proof of the lower bound for total variation follows mostly an argument used in Aldous and Diaconis (1986) to give a lower bound for random transposition (and for the top to random insertion shuffle). Let

$$A_j = \{\sigma \in S_n : \varphi(\sigma) \geq j\} \tag{3.4}$$

with φ as defined above. Then

$$d_{\mathrm{TV}}(q^{(k_n)}, u) \geq q^{(k_n)}(A_j) - u(A_j).$$

Calculating $u(A_j)$ is equivalent to calculating the probability of at least j matches in the classical matching problem. Feller (1968) gives a closed form solution for $u(A_j)$. Using this we get the following estimate for $j \geq 2$

$$u(A_j) = \sum_{m=j}^{n} \frac{1}{m!} \left(\sum_{v=0}^{n-m} \frac{(-1)^v}{v!} \right) \leq e^{-1} \left(\frac{1}{(j-1)!} \right). \tag{3.5}$$

Next we bound $q^{(k_n)}(A_j)$ from below. Consider the experiment where successive balls are droped independently and uniformly at random into n boxes. Let $B_{j,k}$ be the event that after dropping k balls there are at least j empty boxes. Then

$$q^{(k_n)}(A_{j-1}) \geq P(B_{j,k_n}).$$

Let V_l be the number of balls dropped when exactly l boxes are filled. We have

$$P(B_{j,k_n}) = P(V_{n-j} \geq k_n) \geq 1 - P(V_{n-j} \leq k_n).$$

We would like to show that for any fixed j, $P(V_{n-j} \leq k_n) \to 0$ as $n \to \infty$. We have

$$V_{n-j} = (V_{n-j} - V_{n-j-1}) + (V_{n-j-1} - V_{n-j-2}) + \cdots + (V_2 - V_1) + V_1.$$

The $V_{i+1} - V_i$ are independent random variables with geometric distribution

$$P\left\{V_{i+1} - V_i = l\right\} = \left(\frac{n-i}{n}\right)\left(1 - \frac{n-i}{n}\right)^{l-1}, \quad l \geq 1.$$

Hence

$$E(V_{i+1} - V_i) = \frac{n}{n-i} \text{ and } \operatorname{Var}(V_{i+1} - V_i) = \left(\frac{n}{n-i}\right)^2 \left(1 - \frac{n-i}{n}\right).$$

It follows that

$$E(V_{n-j}) \quad = \quad \sum_{i=1}^{n-j-1} \frac{n}{n-i} \geq \int_0^{n-j-1} \frac{n}{n-x}\, dx \geq n \log\left(\frac{n}{j+1}\right)$$

and

$$\begin{aligned} \operatorname{Var}(V_{n-j}) \quad &= \quad \sum_{i=1}^{n-j-1} \frac{n^2}{(n-i)^2} - \frac{n^2}{n(n-i)} \leq \sum_{i=1}^{n-j-1} \frac{n^2}{(n-i)^2} \\ &\leq \quad \int_1^{n-j} \left(\frac{n}{n-x}\right)^2 dx \leq \frac{n^2}{j}. \end{aligned}$$

By assumption $k_n = n \log n - nc_n$ and $c_n \to \infty$ as $n \to \infty$. If we assume, as we may, that $c_n > \log(j+1)$ then Chebyshev's inequality gives

$$\begin{aligned} P(V_{n-j} \leq k_n) \quad &= \quad P(V_{n-j} \leq n \log n - nc_n) \\ &\leq \quad P(n(c_n - \log(j+1)) \leq |(V_{n-j}) - E(V_{n-j})|) \\ &\leq \quad \frac{\operatorname{Var}(V_{n-j})}{n^2(c_n - \log(j+1))^2} \leq \frac{1}{j(c_n - \log(j+1))^2}. \end{aligned}$$

This yields

$$\lim_{n\to\infty} d_{\mathrm{TV}}(q^{(k_n)}, u) \geq \lim_{n\to\infty} (P(B_{j+1,k_n}) - u(A_j)) \geq 1 - e^{-1}\left(\frac{1}{(j-1)!}\right).$$

Since j is arbitrary the desired result follows. □

Corollary 3.4. *Let h_t be the distribution for the continuous time process associated to the transpose top with random measure q. For any sequence $(k_n)_0^\infty$ such that $(k_n - n\log n)/n$ tends to $-\infty$ as n tends to ∞ we have*

$$d_2(h_{k_n}, u) \to \infty \text{ and } d_{\mathrm{TV}}(h_{k_n}, u) \to 1.$$

Proof. The ℓ^2 bound follows from the same argument used above. In the case of the total variation bound, one can show that for A_j defined in (3.4) then $h_{k_n}(A_j) \to 1$. A sight modification of the proof of Proposition 3.2 gives that for $\alpha \in (1/2, 1)$

$$\lim_{n\to\infty} q^{k_n + k_n^\alpha}(A_j) = 1.$$

Combining the limit above with the fact that

$$\lim_{n\to\infty} \sum_{t=0}^{k_n+k_n^\alpha} e^{-k_n}\frac{k_n^t}{t!} = \lim_{n\to\infty} P\left(\frac{X_n - k_n}{\sqrt{k_n}} \le k_n^{\alpha-1/2}\right) = 1$$

where X_n is a Poisson random variable with parameter k_n gives us the desired result. □

4. Random transpositions

4.1. *Discrete time.* Consider the following measure $q = q_{\mathrm{RT}}$ on the group $G = S_n$,

$$q(\tau) = \begin{cases} 2/n^2 & \text{if } \tau = (i,j),\ 1 \le i,j \le n,\ \ i \ne j, \\ 1/n & \text{if } \tau = id, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The measure q models the shuffle of a deck of n cards where one picks two cards independently and uniformly at random and transposes them. The random transposition shuffle has been shown to demonstrate cutoff at $(n/2)\log n$, see Diaconis (1988), Diaconis (1991) and Diaconis and Shahshahani (1981).

Theorem 4.1. (Diaconis and Shahshahani, 1981) *Let q be the random transposition measure on the group S_n then there exists a positive universal constant B such that for any $c \ge 0$ and $t \ge \frac{n}{2}(\log n + c)$ then*

$$2d_{\mathrm{TV}}(q^{(t)}, u) \le d_2(q^{(t)}, u) \le Be^{-c}.$$

One of the aims of this section is to get a more precise estimate on the constant B in the theorem above.

Proposition 4.2. *Let q be the random transposition measure on S_n. For $n \ge 14$, $c \ge 0$, and $t \ge \frac{n}{2}(\log n + c)$ then Theorem 4.1 holds with*

$$B^2 \le 2 + \varphi(n) \le 4$$

where $\varphi(n) \to 0$ as $n \to \infty$.

Let t_n be the smallest integer larger or equal to $(n/2)\log n$. Then the result above and an easy lower bound discussed below imply that

$$1 \le \lim_{n\to\infty} d_2(q^{(t_n)}, u) \le 2.$$

It is quite rare to be able to capture the mixing time of a chain with such precision.

Proof. Let $\mathcal{C} \subset S_n$ be the conjugacy class of transpositions, $\tau \in \mathcal{C}$ be a transposition, and

$$r(\rho_\lambda) = \frac{\chi_{\rho_\lambda}(\tau)}{\mathrm{d}_{\rho_\lambda}}.$$

Proposition 2.3 gives

$$d_2(q^{(t)}, u)^2 = \sum_{\rho_\lambda \neq 1} \mathrm{d}^2_{\rho_\lambda} \left(\frac{1}{n} + \frac{n-1}{n} r(\rho_\lambda) \right)^{2t}. \tag{4.2}$$

In Diaconis and Shahshahani (1981) it is shown that

$$r(\rho_\lambda) \leq \begin{cases} 1 - \frac{2(n-\lambda_1)(\lambda_1+1)}{n(n-1)} & \text{if } \lambda_1 \geq n/2 \\ \frac{\lambda_1 - 1}{n-1} & \text{always.} \end{cases} \tag{4.3}$$

It follows from equations (3.3), (4.2), and (4.3) that

$$\begin{aligned} d_2(q^{(t)}, u)^2 &= \sum_{j=1}^{n-1} \sum_{\substack{\rho_\lambda \\ \lambda_1 = n-j}} \mathrm{d}^2_{\rho_\lambda} \left(\frac{1}{n} + \frac{n-1}{n} r(\rho_\lambda) \right)^{2t} \\ &\leq \sum_{j=1}^{\frac{n}{2}} \left(\frac{n!}{(n-j)!} \right)^2 \frac{1}{j!} \left(1 - \frac{2j}{n} \left(1 - \frac{j-1}{n} \right) \right)^{2t} \\ &\quad + \sum_{j=\frac{n}{2}}^{n-1} \left(\frac{n!}{(n-j)!} \right)^2 \frac{1}{j!} \left(1 - \frac{j}{n} \right)^{2t}. \end{aligned}$$

Note that for $1 \leq j \leq \frac{n}{2}$ we have that $1 - \frac{2j}{n}\left(1 - \frac{j-1}{n}\right) \leq 1 - \frac{2}{n} \leq e^{-2/n}$. So for $t \geq (n/2)(\log n + c)$

$$d_2(q^{(t)}, u)^2 \leq e^{-2c} \left(\sum_{j=1}^{n/2} A_j + \sum_{j=n/2}^{n-1} B_j \right),$$

where

$$A_j = \left(\frac{n!}{(n-j)!} \right)^2 \frac{1}{j!} \left(1 - \frac{2j}{n} \left(1 - \frac{j-1}{n} \right) \right)^{n \log n} \tag{4.4}$$

$$B_j = \left(\frac{n!}{(n-j)!} \right)^2 \frac{1}{j!} \left(1 - \frac{j}{n} \right)^{n \log n}. \tag{4.5}$$

Consider the following two technical propositions.

Proposition 4.3. *Set* $\varphi_0(n) = \sum_{j=1}^{\lfloor n/4 \rfloor} A_j$ *and* $\varphi_1(n) = \sum_{j=\lceil n/4 \rceil}^{\lfloor n/2 \rfloor} A_j$. *For* $n \geq 14$

$$\varphi_0(n) \leq 2 \quad \textit{and} \quad \varphi_1(n) \leq \exp\left\{ 2 - \frac{1}{6} n \log n \right\}.$$

Proposition 4.4. *Set* $\varphi_2(n) = \sum_{j=\lceil n/2 \rceil}^{n} B_j$. *For* $n \geq 9$

$$\varphi_2(n) \leq \exp\left\{ 1 - \frac{3}{1000} n \log n \right\}.$$

Propositions 4.3 and 4.4 give that for $n \geq 14$

$$d_2(q^t, u)^2 = e^{-2c}(\varphi_0 + \varphi_1 + \varphi_2) \leq e^{-2c}(2 + \varphi_1(14) + \varphi_2(14)) \leq 4e^{-2c}.$$

It also follows that $\varphi_1 \to 0$ and $\varphi_2 \to 0$ as $n \to \infty$. □

Next we will show the proofs of the propositions above.

Proof of Proposition 4.3. Let A_j be as in equation (4.4), the ratio between two consecutive terms is

$$\frac{A_{j+1}}{A_j} = \exp\{f_n(j) + g_n(j)\}$$

where

$$\begin{aligned} f_n(j) &= 2\log(n-j) - \log(j+1) \\ g_n(j) &= n \log n \log\left(\frac{n^2 - 2(j+1)n + 2j(j+1)}{n^2 - 2jn + 2j(j-1)}\right). \end{aligned}$$

Taking derivatives gives

$$\begin{aligned} f_n'(j) &= -\frac{2}{n-j} - \frac{1}{j+1} \\ g_n'(j) &= \frac{4(n\log n)(2jn - 2j^2 - n)}{(n^2 - 2jn + 2j^2 - 2j)(n^2 - 2jn - 2n + 2j^2 + 2j)}. \end{aligned}$$

Note that for $1 \leq j \leq n/4$ and $n \geq 4$ we have that $f_n''(j) = \frac{1}{(j+1)^2} - \frac{2}{(n-j)^2} \geq 0$. Furthermore, $g_n''(j) \geq 0$ for $1 \leq j \leq n/2$. The last inequality holds since for $1 \leq j \leq n/2$ the numerator of g_n' is a positive increasing function of j and the denominator is a positive decreasing function of j.

Set $h_n = f_n + g_n$. For $1 \leq j \leq n/4$ the function h_n is continuous and has positive second derivative. It follows that h_n is convex for said values of j, which implies that

$$h_n(j) \leq \max\{h_n(1), h_n(n/4)\}.$$

Consider the following estimates.

$$\begin{aligned} h_n(1) &= 2\log(n-1) - \log 2 + n(\log n)\log\left(1 - \frac{2n-4}{n(n-2)}\right) \\ &\leq 2\log(n-1) - \log 2 - n(\log n)\left(\frac{2n-4}{n(n-2)}\right) \\ &\leq 2\log(n-1) - \log 2 - 2\log n \end{aligned}$$

$$\begin{aligned} h_n(n/4) &= 2\log\left(\frac{3n}{4}\right) - \log\left(\frac{n+4}{4}\right) + n(\log n)\log\left(1 - \frac{8}{5n-4}\right) \\ &\leq 2\log\left(\frac{3n}{4}\right) - \log\left(\frac{n+4}{4}\right) - n(\log n)\left(\frac{8}{5n-4}\right) \\ &\leq 2\log\left(\frac{3n}{4}\right) - \log\left(\frac{n+4}{4}\right) - \frac{8\log n}{5} \\ &\leq 2\log 3 - \log 4 + \frac{2}{5}\log n - \log(n+4) \end{aligned}$$

For $n \geq 2$ $h_n(1)$ and $h_n(n/4)$ are decreasing functions of n less than $-\log 2$. Since $A_1 = n^2\,(1-2/n)^{n\log n} \leq 1$, it follows that for $1 \leq j \leq n/4$

$$A_j \leq (1/2)^{j-1} A_1 \leq (1/2)^{j-1}.$$

We can now state the first part of Proposition 4.3

$$\varphi_0(n) = \sum_{j=0}^{\frac{n}{4}} A_j \leq \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j = 2.$$

Next we bound A_j for $n/4 \leq j \leq n/2$. It is not hard to show that $f_n'''(j) \leq 0$, so for the values of j above

$$f_n'(j) \geq \min\{f_n'(n/4), f_n'(n/2)\}.$$

Note that

$$f_n'(n/4) = -\frac{4(5n+8)}{3n(n+4)} \quad \text{and} \quad f_n'(n/2) = -\frac{2(3n+4)}{n(n+2)}.$$

For $n \geq 14$, $f_n'(n/2) \geq f_n'(n/4)$. Recall that for $1 \leq j \leq n/2$ we had that $g_n'' \geq 0$. It follows that for $n/4 \leq j \leq n/2$,

$$h_n'(j) = f_n'(j) + g_n'(j) \geq f_n'(n/4) + g_n'(n/4) \geq 0.$$

Above we showed that $h_n(n/4) \leq 0$. For $n \geq 3$ we have that $h_n(n/2) = 2\log\left(\frac{n}{2}\right) - \log\left(\frac{n}{2}+1\right) \geq 0$, so there must be a unique point $x \in [n/4, n/2]$ such that $h_n(x) = 0$. If $n/4 \leq j \leq x$ then $(A_{j+1}/A_j) \leq 1$. If $x \leq j \leq n/2$ then $(A_{j+1}/A_j) \geq 1$. So for $n/4 \leq j \leq n/2$

$$A_j \leq \max\left\{A_{\frac{n}{4}}, A_{\frac{n}{2}}\right\}.$$

In Feller (1968) a proof of Stirling's formula shows that

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq e^{\frac{1}{12n}}\sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{4.6}$$

To determine the largest value among $A_{\frac{n}{4}}$ and $A_{\frac{n}{2}}$ we consider the ratio

$$\begin{aligned}
\frac{A_{\frac{n}{4}}}{A_{\frac{n}{2}}} &= \left(\frac{(\frac{n}{2})!}{(\frac{3n}{4})!}\right)^2 \left(\frac{(\frac{n}{2})!}{(\frac{n}{4})!}\right) \left(\frac{5n-4}{4n-8}\right)^{n\log n} \\
&\leq \left(\frac{2\sqrt{2}e^{\frac{1}{4n}}}{3}\right) \left(\frac{e^{\frac{n}{2}}4^n}{n^{\frac{n}{2}}3^{\frac{3n}{2}}}\right) \left(\frac{n^{\frac{n}{4}}}{e^{\frac{n}{4}}}\right) \left(\frac{5n-4}{4n-8}\right)^{n\log n} \\
&= \left(\frac{2\sqrt{2}e^{\frac{1}{4n}}}{3}\right) \left(\frac{4}{3^{\frac{3}{2}}}\right)^n \left(\frac{e^{\frac{1}{4}}}{n^{\frac{1}{4}}}\right)^n n^{n\log\left(\frac{5n-4}{4n-8}\right)} = \left(\frac{2\sqrt{2}e^{\frac{1}{4n}}}{3}\right) \exp\{l(n)\}
\end{aligned}$$

where $l(n) = n\left(\log\left(\frac{4}{3^{3/2}}\right) + \frac{1}{4}\right) + n\log n\left(\log\left(\frac{5n-4}{4n-8}\right) - \frac{1}{4}\right)$. For $n \geq 47$ we have that $\left(\log\left(\frac{5n-4}{4n-8}\right) - \frac{1}{4}\right) \leq 0$ which implies that $l(n) \leq 0$. If $5 \leq n \leq 47$ one can check that $l(n) \leq 1$. So for $n \geq 5$ we have that $(A_{\frac{n}{4}}/A_{\frac{n}{2}}) \leq e$ which in turn implies that $\sum_{j=n/4}^{n/2} A_j \le (e/4)nA_{\frac{n}{2}}$. By using Stirling's formula to estimate $A_{\frac{n}{2}}$ we get

$$\begin{aligned}
\left(\frac{en}{4}\right) A_{\frac{n}{2}} &= \left(\frac{en}{4}\right)\left(\frac{n!}{(\frac{n}{2})!}\right)^2 \left(\frac{1}{(\frac{n}{2})!}\right)\left(\frac{n-2}{2n}\right)^{n\log n} \\
&= \left(\frac{en}{4}\right)\left(\frac{e^{\frac{1}{12n}}\sqrt{2}n^{\frac{n}{2}}2^{\frac{n}{2}}}{e^{\frac{n}{2}}}\right)^2 \left(\frac{2^{\frac{n}{2}}e^{\frac{n}{2}}}{n^{\frac{n}{2}}\sqrt{\pi n}}\right)\left(\frac{n-2}{2n}\right)^{n\log n} \\
&= \left(\frac{e^{1+\frac{1}{6n}}\sqrt{n}}{2\sqrt{\pi}}\right)\left(\frac{n^{\frac{n}{2}}2^{\frac{3n}{2}}}{e^{\frac{n}{2}}}\right)\left(\frac{n-2}{2n}\right)^{n\log n} \\
&= \left(\frac{e^{1+\frac{6}{n}}}{2\sqrt{\pi}}\right)\exp\{nf(n)\log n\}
\end{aligned}$$

where $f(n) = \left(\frac{3\log 2 - 1}{2}\right)(\log n)^{-1} + \frac{1}{2n} + \frac{1}{2} + \log\left(\frac{n-2}{2n}\right)$. Computing the derivative gives us that

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

Note that $f' \ge 0$ for $n > 2$, so $f(n) \le \lim_{n\to\infty} f(n) = \frac{1}{2} - \log 2$. We can now concluded that for $n \ge 5$

$$\begin{aligned}
\varphi_1(n) &\le \left(\frac{en}{4}\right) A_{\frac{n}{2}} \le \left(\frac{e^{1+\frac{6}{n}}}{2\sqrt{\pi}}\right)\exp\left\{(n\log n)\left(\frac{1}{2} - \log 2\right)\right\} \\
&\le \frac{e^3}{2\sqrt{\pi}}\exp\left\{-\frac{1}{6}n\log n\right\} \le \exp\left\{2 - \frac{1}{6}n\log n\right\}.
\end{aligned}$$

□

Proof of Proposition 4.4. Let B_j be as in equation (4.5). If $n > 2$ and $n/2 \le j \le n$ we can estimate the ratio of B_j and B_{j+1} by

$$\frac{B_{j+1}}{B_j} = \frac{(n-j)^2}{(j+1)}\left(1 - \frac{1}{n-j}\right)^{n\log n} \le 2n\left(1 - \frac{2}{n}\right)^{n\log n} \le \frac{2}{n}.$$

We get that $B_j \le (2/n)^{j-n/2}B_{\frac{n}{2}}$. It follows that

$$\varphi_2(n) = \sum_{j=\frac{n}{2}}^{n} B_j \le B_{\frac{n}{2}}\sum_{j=\frac{n}{2}}^{n}(2/n)^{j-\frac{n}{2}} \le B_{\frac{n}{2}}\sum_{j=0}^{\infty}(2/n)^j = \frac{B_{\frac{n}{2}}}{1-(2/n)}.$$

Using Stirling's formula we can bound $B_{\frac{n}{2}}$ to get

$$\begin{aligned}
\left(\frac{1}{1-2/n}\right)B_{\frac{n}{2}} &= \left(\frac{1}{1-2/n}\right)\left(\frac{n!}{(\frac{n}{2})!}\right)^2\left(\frac{1}{(\frac{n}{2})!}\right)\left(\frac{1}{2}\right)^{n\log n} \\
&\le \left(\frac{1}{1-2/n}\right)\left(\frac{2e^{\frac{1}{6n}}}{\sqrt{\pi n}}\right)\left(\frac{n^{\frac{n}{2}}2^{\frac{3n}{2}}}{e^{\frac{n}{2}}}\right)\left(\frac{1}{2}\right)^{n\log n} \\
&= \left(\frac{1}{1-2/n}\right)\left(\frac{2e^{\frac{1}{6n}}}{\sqrt{\pi}}\right)\exp\{n(\log n)b(n)\}
\end{aligned}$$

where $b(n) = -\frac{1}{2n} + \left(\frac{3\log 2 - 1}{2}\right)(\log n)^{-1} + \frac{1}{2} + \log\left(\frac{1}{2}\right)$. Taking derivatives gives that

$$b'(n) = \frac{(\log n)^2 - n(3\log 2 - 1)}{2n^2(\log n)^2} \ \text{ so } \ b'(n) \le 0 \ \text{ for } n \ge 1.$$

For $n \ge 9$ we have that $b(n) \le b(9) < -\frac{3}{1000}$. Furthermore, for $n \ge 9$ the function

$$g(n) = \left(\frac{1}{1-2/n}\right)\left(\frac{2e^{\frac{1}{6n}}}{\sqrt{\pi}}\right)\exp\left\{-\frac{3}{1000}n\log n\right\}$$

is decreasing. So for $n \ge 9$

$$\varphi_2(n) \le g(n) \le \exp\left\{1 - \frac{3}{1000}n\log n\right\}.$$

□

A lower bound for the χ-square distance is obtain by writing $d_2(q^{(k)}, u)^2 \ge (n-1)^2(1-2/n)^{2k}$ which uses the term associated to the Young diagram $(n-1,1)$. Alternatively, let $\varphi(\sigma)$ be the function with denotes the number of fixed points of σ. One can check by inspection that $\varphi - 1$ is a normalized eigenfunction associated with the eigenvalue $(1-2/n)$. This gives the same ℓ^2 lower bound.

Concerning total variation lower bounds, Diaconis (1988) shows that for any $c > 0$ and $t \ge (n/2)(\log n - c)$

$$\lim_{n\to\infty} d_{\mathrm{TV}}(q^{(t)}, u) \ge 1/e - e^{-e^{-2c}}$$

A slight modification of the argument used in Diaconis (1988) (as presented above in the proof of Proposition 3.2) yields the following proposition.

Proposition 4.5. *Let q be the random transposition measure on the group S_n. For any sequence k_n such that $(2k_n - n\log n)/n$ tends to $-\infty$ as n tends to ∞, we have*

$$\lim_{n\to\infty} d_2(q^{(k_n)}, u) = \infty \ \ \textit{and} \ \ \lim_{n\to\infty} d_{\mathrm{TV}}(q^{(k_n)}, u) = 1.$$

4.2. *Random transposition in continuous time.* This section is devoted to the continuous time version of random transposition. There is no proof in the literature that the continuous time random transposition shuffle has a ℓ^2 cutoff at time $(n/2)\log n$. One reason is that the fact that it does not automatically follow from the discrete time result is often overlooked. In fact, getting an upper bound in the continuous time case turns out to be somewhat more difficult than in the discrete case. The difficulty comes from handling the contribution of the small eigenvalues of q. Compare with what is proved below for conjugacy classes with less fixed points, e.g., 4-cycles. One very good reason to want to have a good ℓ^2 upper-bound in continuous time for random transposition is that it yields better results when used with the comparison technique of Diaconis and Saloff-Coste (1994) to study other chains. See Section 4.3 below.

Proposition 4.6. *Let h_t be the law of the continuous time process associated to the random transposition measure q. If $n \ge 10$, $c \ge 2$ then for $t \ge (n/2)(\log n + c)$*

$$2d_{\mathrm{TV}}(h_t, u) \le d_2(h_t, u) \le e^{-(c-2)}.$$

Moreover, if t_n is any sequence of time such that $(2t_n - n\log n)/n$ tends to $-\infty$ as n tends to ∞, we have

$$\lim_{n\to\infty} d_{\mathrm{TV}}(h_{t_n}, u) = 1, \quad \lim_{n\to\infty} d_2(h_{t_n}, u) = \infty.$$

Let us observe that we are not able to show that $d_2(h_{(n/2)\log n}, u)$ is bounded above independently of n (compare with the discrete time case).

Proof. The lower bound in ℓ^2 follows from the same argument used in the discrete time case. The lower bound in total variation is known. See, e.g., in Saloff-Coste (1994). We focus on the upper bound in ℓ^2.

Let $\mathcal{C} \subset S_n$ be the conjugacy class of transpositions. Proposition 2.3 implies that

$$d_2(h_t, u)^2 = \sum_{\rho_\lambda \neq 1} \mathrm{d}^2_{\rho_\lambda} \exp\left\{-2t\left(1 - \frac{1}{n} - \frac{n-1}{n} r(\rho_\lambda)\right)\right\} \tag{4.7}$$

where $r(\rho_\lambda) = \chi_{\rho_\lambda}(\tau)/\mathrm{d}_{\rho_\lambda}$ and τ is a transposition. Using equations (3.3), (4.7), and (4.3) we get that for $t \geq (n/2)(\log n + c)$

$$\begin{aligned} d_2(h_t, u)^2 &= \sum_{j=1}^{n-1} \sum_{\substack{\rho_\lambda \neq 1 \\ \lambda_1 = n-j}} \mathrm{d}^2_{\rho_\lambda} \exp\left\{-2t\left(1 - \frac{1}{n} - \frac{n-1}{n} r(\rho_\lambda)\right)\right\} \\ &\leq \sum_{j=1}^{n/2} \left(\frac{n!}{(n-j)!}\right)^2 \frac{1}{j!} \exp\left\{-2t\left(\frac{2j}{n}\right)\left(1 - \frac{j-1}{n}\right)\right\} \\ &\quad + \sum_{j=n/2}^{n-1} \left(\frac{n!}{(n-j)!}\right)^2 \frac{1}{j!} \exp\left\{-2t\left(\frac{j}{n}\right)\right\} \\ &\leq \sum_{j=1}^{n/2} \left(\frac{n!}{(n-j)!}\right)^2 \frac{1}{j!} \exp\left\{-2j(\log n + c)\left(1 - \frac{j-1}{n}\right)\right\} \\ &\quad + \sum_{j=n/2}^{n-1} \left(\frac{n!}{(n-j)!}\right)^2 \frac{1}{j!} \exp\left\{-j(\log n + c)\right\} \end{aligned}$$

Note that for $c \geq 2$ and $j \leq n/2$ we have $-2cj\left(1 - \frac{j-1}{n}\right) \leq -2c - 2j + 4$. It follows that for $t \geq (n/2)(\log n + c)$

$$d_2(h_t, u)^2 \leq e^{-2(c-2)} \left(\sum_{j=1}^{n/2} A_j + \sum_{j=n/2}^{n} B_j\right)$$

where

$$A_j = \left(\frac{n!}{(n-j)!}\right)^2 \frac{1}{j!} \exp\left\{-2j\log n\left(1 - \frac{j}{n}\right) - 2j\right\} \tag{4.8}$$

$$B_j = \left(\frac{n!}{(n-j)!}\right) \frac{1}{j!} \exp\{-j\log n - 2j\}. \tag{4.9}$$

Consider the following technical lemmas.

Lemma 4.7. *For $n \geq 10$ then $\sum_{j=1}^{n/4} A_j \leq 2/3$ and $\sum_{j=n/4}^{n/2} A_j \leq 1/4$.*

Lemma 4.8. *Set $\gamma(n) = \sum_{j=n/2}^{n} B_j$. For $n \geq 2$*

$$\gamma(n) \leq 2\left(\frac{2}{e}\right)^{\frac{3n}{2}}.$$

It follows from the lemmas above that for $n \geq 10$

$$\begin{aligned} d_2(h_t,u)^2 &\leq e^{-2(c-2)}\left(\sum_{j=1}^{n/4} A_j + \sum_{j=n/4}^{n/2} A_j + \gamma(10)\right) \\ &\leq e^{-2(c-2)}\left(2/3 + 1/4 + 2(2/e)^{15}\right) \\ &\leq e^{-2(c-2)}. \end{aligned}$$

□

Proof of Lemma 4.7. Let A_j be as in equation (4.8). For $1 \leq j < n/2$ the ratio of two consecutive terms is given by

$$\frac{A_{j+1}}{A_j} = \frac{(n-j)^2}{(j+1)}\exp\left\{-\left(\frac{2\log n}{n}\right)(n-2j-1)-2\right\} = \exp\{f_n(j)\}$$

where

$$f_n(j) = 2\log(n-j) - \log(j+1) - \left(\frac{2\log n}{n}\right)(n-2j-1) - 2. \tag{4.10}$$

Taking derivatives gives

$$\begin{aligned} f_n'(j) &= -\frac{2}{n-j} - \frac{1}{j+1} + \frac{4\log n}{n} \\ f_n''(j) &= -\frac{2}{(n-j)^2} + \frac{1}{(j+1)^2}. \end{aligned}$$

Let $n \geq 4$ and $1 \leq x \leq n/4$. For these values of n and x we get that f_n is convex since f_n'' is a decreasing function and $f_n''(x) \geq f_n''(n/4) \geq 0$.

$$\frac{A_{x+1}}{A_x} = \exp\left(\max\left\{f_n(1), f_n\left(\frac{n}{4}\right)\right\}\right).$$

If $n \geq 2$ we have the estimates

$$\begin{aligned} f_n(1) &= 2\log(n-1) - \log 2 - \left(\frac{2\log n}{n}\right)(n-3) - 2 \leq -\log 2 \\ f_n(n/4) &= 2\log\left(\frac{3n}{4}\right) - \log\left(\frac{n}{4}+1\right) - \left(\frac{2\log n}{n}\right)\left(\frac{n}{2}-1\right) - 2 \leq 2\log\left(\frac{3}{4}\right). \end{aligned}$$

Since $-\log 2 \leq 2\log(3/4)$, we get that $A_x \leq (9/16)^{x-1}$. It now follows that

$$\sum_{j=1}^{\frac{n}{4}} A_j \leq A_1 \sum_{j=0}^{\infty}\left(\frac{9}{16}\right)^j = \left(\frac{16}{7}\right)A_1.$$

For $n \geq 4$ we get $A_1 = n^2 \exp\left\{-\frac{2(n-1)}{n}\log n - 2\right\} \leq 2e^{-2}$. This gives that

$$\sum_{j=1}^{\frac{n}{4}} A_j \leq (32/7)e^{-2} \leq 2/3.$$

For the next part of the proof let $n \geq 10$. Recall that $f_n^{''}$ is a decreasing function, which implies that for $n/4 \leq j \leq n/2$

$$f_n'(j) \geq \min\{f_n'(n/4), f_n'(n/2)\} \geq 0$$

where the last inequality holds since $n \geq 10$. Since f_n is an increasing function with $f_n(n/4) \leq 0$ and $f_n(n/2) \geq 0$ then there exists a unique point $z \in [n/4, n/2]$ such that $f_n(z) = 0$. It follows that if $n/4 \leq j \leq z$ then $A_j \leq A_{n/4}$ and if $z \leq j \leq n/2$ then $A_j \leq A_{n/2}$. Combining these two inequalities gives us that for $n/4 \leq j \leq n/2$

$$A_j \leq \max\{A_{n/4}, A_{n/2}\}.$$

To compare $A_{n/4}$ and $A_{n/2}$ we use Stirling's formula (4.6). For $n \geq 2$

$$\begin{aligned} A_{n/4} &= \left(\frac{n!}{\left(\frac{3n}{4}\right)!}\right)^2 \left(\frac{1}{\left(\frac{n}{4}\right)!}\right) \exp\left\{-\left(\frac{3n}{8}\right)\log n - \frac{n}{2}\right\} \\ &\leq \left(\frac{e^{\frac{1}{6n}} 4\sqrt{2}}{3\sqrt{\pi n}}\right) n^{-\frac{n}{8}} \left(\frac{4}{3}\right)^{\frac{3n}{2}} 4^{\frac{n}{4}} e^{-\frac{3n}{4}} \leq n^{-\frac{n}{8}} \left(\frac{4}{3}\right)^{\frac{3n}{2}} 4^{\frac{n}{4}} e^{-\frac{3n}{4}} \\ A_{n/2} &= \left(\frac{n!}{\left(\frac{n}{2}\right)!}\right)^2 \left(\frac{1}{\left(\frac{n}{2}\right)!}\right) \exp\left\{-\left(\frac{n}{2}\right)\log n - n\right\} \\ &\leq \left(\frac{2e^{\frac{1}{6n}}}{\sqrt{\pi n}}\right) 2^{\frac{3n}{2}} e^{-\frac{3n}{2}} \leq 2^{\frac{3n}{2}} e^{-\frac{3n}{2}}. \end{aligned}$$

It follows that

$$nA_{n/4} \leq \exp\{\phi_1(n)\} \text{ and } nA_{n/2} = \exp\{\phi_2(n)\},$$

where

$$\begin{aligned} \phi_1(n) &= \log n - \left(\frac{n}{8}\right)\log n - \left(\frac{3n}{4}\right) + \left(\frac{3n}{2}\right)\log\left(\frac{4}{3}\right) + \left(\frac{n}{4}\right)\log 4 \\ \phi_2(n) &= \log n - \left(\frac{3n}{2}\right) + \left(\frac{3n}{2}\right)\log 2. \end{aligned}$$

For $n \geq 10$ we have $\phi_1(n) \leq 0$ and $\phi_2(n) \leq 0$ which implies that

$$\sum_{j=n/4}^{n/2} A_j \leq \left(\frac{1}{4}\right) \max\left\{nA_{\frac{n}{4}}, nA_{\frac{n}{2}}\right\} \leq \frac{1}{4}.$$

□

Proof of Lemma 4.8. Let $n/2 \leq j \leq n$ and B_j be as in equation (4.9). As usual, consider we consider the ratio between two consecutive terms

$$\frac{B_{j+1}}{B_j} = \frac{(n-j)^2}{(j+1)} \exp\{-\log n - 2\} \leq \left(\frac{n}{2}\right) \exp\{-\log n - 2\} \leq \frac{1}{2}.$$

Note that $B_j \leq \left(\frac{1}{2}\right)^{n/2-j} B_{n/2}$, which implies that

$$\gamma(n) = \sum_{j=n/2}^{n} B_j \leq B_{n/2} \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j = 2B_{n/2}.$$

Since $B_{\frac{n}{2}} = A_{\frac{n}{2}} \leq (2/e)^{\frac{3n}{2}}$ then for $n \geq 2$ we have that $\gamma(n) \leq 2(2/e)^{\frac{3n}{2}}$. □

4.3. *Random Insertions.* In the random insertion shuffle for a deck of n cards, one picks out a random card and inserts it back into the deck at a random position. This shuffle is modeled by the measure q on the S_n given by

$$q(\tau) = \begin{cases} 1/n & \text{if } \tau = e \\ 2/n^2 & \text{if } \tau = c_{i,j} \text{ s.t. } 1 \le i,j \le n \text{ and } |i-j| = 1 \\ 1/n^2 & \text{if } \tau = c_{i,j} \text{ s.t. } 1 \le i,j \le n \text{ and } |i-j| > 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

where $c_{i,j}$ denotes the cycle created by taking the card in position i and inserting it into position j. A formal definition is given by

$$c_{i,j} = \begin{cases} e & \text{if } i = j \\ (j, j-1, \ldots, i+1, i) & \text{if } 1 \le i < j \le n \\ (j, j+1, \ldots, i-1, i) & \text{if } 1 \le j < i \le n. \end{cases}$$

Random insertion is the first of the shuffles discussed in this paper for which it is not known whether there is a total variation cutoff or not although it is strongly believed that there is one. The results of Chen (2006) and Chen and Saloff-Coste (2008) show that there is a cutoff in ℓ^2 but the exact cutoff time is not known. What is known and follows from Diaconis and Saloff-Coste (1994) is that there is a pre-cutoff (in both total variation and ℓ^2) at time $n \log n$. Finding the precise ℓ^2 cutoff time and proving a cutoff in total variation are challenging open problems that have been investigated (but not solved) by Uyemura-Reyes (2002).

Theorem 4.9. (Diaconis and Saloff-Coste, 1994; Uyemura-Reyes, 2002) *Let q be the random insertion measure on S_n defined above. For $c > 0$ and $t \ge 4n(\log n + c)$ there exists a constant B such that*

$$d_2(q^{(t)}, u) \le Be^{-c}.$$

For any sequence (k_n) such that $(2k_n - n \log n)/n$ tends to $-\infty$ as n tends to ∞ then

$$d_{\mathrm{TV}}(q^{(k_n)}, u) \to 1 \ \text{ and } \ d_2(q^{(k_n)}, u) \to \infty.$$

In Diaconis and Saloff-Coste (1994) the mixing time in Theorem 4.9 is shown to be $\mathcal{O}(n \log n)$ while in Uyemura-Reyes (2002) the more precise upper bound given in Theorem 4.9 is shown. The proof of the upper bound in Theorem 4.9 relies on the comparison techniques developed in Diaconis and Saloff-Coste (1993).

Definition 4.10. Let V be a state space equipped with a Markov kernel K with reversible measure ν. The Dirichlet form associated to (K, ν) is

$$\begin{aligned} \mathcal{E}_{K,\nu}(f,g) &= \langle (I-K)f, g\rangle_\nu = \sum_{x\in V} [(I-K)f(x)]g(x)\nu(x) \\ &= \frac{1}{2} \sum_{x,y \in V} (f(x)-f(y))(g(x)-g(y))\nu(x)K(x,y) \end{aligned}$$

where $f, g \in \ell^2(\nu, V)$. In the case where V is a finite group and $p(x^{-1}y) = K(x,y)$ we set $\mathcal{E}_{p,\nu} = \mathcal{E}_{K,\nu}$.

Diaconis and Saloff-Coste show the following theorem.

Theorem 4.11. (Diaconis and Saloff-Coste, 1993) *Let q and $\tilde{q}$ be the probability measures on a finite group G. Set $\mathcal{E} = \mathcal{E}_{q,u}, \tilde{\mathcal{E}} = \mathcal{E}_{\tilde{q},u}$ and $\beta_i, \tilde{\beta}_i$, $0 \leq i \leq |G| - 1$ to be the associated Dirichlet forms and eigenvalues of q and $\tilde{q}$ respectively. Let $\tilde{h}_t$ to be the law at time t of the continuous time process associated with $\tilde{q}$. If there exists a constant A such that $\tilde{\mathcal{E}} \leq A\mathcal{E}$ then*

$$d_2(q^{(t)}, u)^2 \leq \beta_-^{2t_1}(1 + d_2(\tilde{h}_{t_2/A}, u)^2) + d_2(\tilde{h}_{t/A}, u)^2$$

where $t = t_1 + t_2 + 1$ and $\beta_- = \max\{0, -\beta_{|G|-1}\}$.

Let q and $\tilde{q}$ be the measures for the random insertion shuffle and the random transposition shuffle respectively. In his thesis, Uyemura-Reyes shows that $A = 4$ is the smallest constant such that $\tilde{\mathcal{E}} \leq A\mathcal{E}$. By noting that $\beta_- = 0$ we get

$$d_2(q^{(t)}, u)^2 \leq d_2(\tilde{h}_{t/4}, u)^2. \tag{4.12}$$

Equation (4.12) gives the following corollary to Proposition 4.6.

Corollary 4.12. *Let q be the random insertion measure on S_n defined above. If $n \geq 10$, $c \geq 2$ and $t \geq 2n(\log n + c)$ then*

$$d_2(q^{(t)}, u)^2 \leq e^{-(c-2)}.$$

For any sequence (k_n) such that $(2k_n - n \log n)/n$ tends to $-\infty$ as n tends to ∞ then

$$\lim_{n\to\infty} d_{\mathrm{TV}}(q^{(k_n)}, u) = 1.$$

Proof. The upper bound results as a corollary to Proposition 4.6 after applying equation (4.12). The improvement by a factor of 2 compared to Theorem 4.9 is due to the use of the continuous time random transposition process in the comparison inequality (4.12).

Uyemura-Reyes also proves the total variation lower bound in his thesis but his proof uses a rather sophisticated argument involving results concerning the longest increasing subsequence of a permutation. We give an alternative proof of this result based on a technique due to Wilson (2004).

First note the following result of Uyemura-Reyes. Set ρ to be the permutation representation. Let q to be the random insertion measure and Q its associated Markov kernel such that $Q(x, y) = q(x^{-1}y)$. In Uyemura-Reyes (2002) it is shown that the Fourier transform $\widehat{q}(\rho)$ has an eigenvector $v = (v_0, \ldots, v_{n-1})$ where

$$\widehat{q}(\rho)v = \left(1 - \frac{1}{n}\right)v \quad \text{and} \quad v_i = 1 - \frac{2i}{n-1}.$$

As noticed at the end of Section 2.2, it follows that $f_\rho(\sigma) = \langle \rho(\sigma)v, v\rangle$, $\sigma \in S_n$, is an eigenvector of Q with associated eigenvalue $(1 - 1/n)$.

Computing $f_\rho(\sigma)$, one gets

$$\begin{aligned} f_\rho(\sigma) &= \left\langle \sum_{i=0}^{n-1}\left(1-\frac{2i}{n-1}\right)e_{\sigma(i)}, \sum_{j=0}^{n-1}\left(1-\frac{2j}{n-1}\right)e_j\right\rangle \\ &= \sum_{j=0}^{n-1}\left(1-\frac{2\sigma(j)}{n-1}\right)\left(1-\frac{2j}{n-1}\right) \\ &= \sum_{j=0}^{n-1} 1-\frac{2j}{n-1}-\frac{2\sigma(j)}{n-1}+\frac{4\sigma(j)j}{(n-1)^2} \\ &= -n+\frac{4}{(n-1)^2}\sum_{j=0}^{n-1}\sigma(j)j. \end{aligned} \tag{4.13}$$

Therefore

$$f^2(\sigma) = n^2 - \frac{8n}{(n-1)^2}\sum_{j=0}^{n-1}\sigma(j)j + \frac{16}{(n-1)^4}\sum_{i,j=0}^{n-1}\sigma(i)\sigma(j)ij$$

and

$$\begin{aligned} \sum_{\sigma\in S_n} f^2(\sigma) &= n!n^2 - \frac{8n}{(n-1)^2}\sum_{j=0}^{n-1} j \sum_{\sigma\in S_n}\sigma(j) + \frac{16}{(n-1)^4}\sum_{i,j=0}^{n-1} ij \sum_{\sigma\in S_n}\sigma(i)\sigma(j) \\ &= n!n^2 - \frac{8n[(n-1)!]}{(n-1)^2}\left(\frac{n(n-1)}{2}\right)^2 + \frac{16[(n-2)!]}{(n-1)^4}\left(\frac{n(n-1)}{2}\right)^4 \\ &= n!n^2 - \frac{8n^3(n-1)!}{4} + n^4(n-2)! = n!\left(\frac{n^2}{n-1}\right) \end{aligned}$$

Next we estimate the supremum norm of the discrete square gradient of f_ρ defined in (4.13). The discrete square gradient of the function g with respect to the kernel K is given by the equation

$$|\nabla g(x)|^2 = \frac{1}{2}\sum_y |g(x)-g(y)|^2 K(x,y).$$

Calculating the discrete square gradient for f_ρ gives us

$$\begin{aligned} |\nabla f_\rho(\sigma)|^2 &\le \frac{16}{n^2(n-1)^4}\sum_{i,j=0}^{n-1}\left|\sum_{k=0}^{n-1}\sigma^{-1}(k)\left(k-c_{ij}(k)\right)\right|^2 \\ &\le \frac{16}{n^2(n-1)^2}\sum_{i,j=0}^{n-1}\sum_{k=0}^{n-1}|k-c_{i,j}(k)|^2 \end{aligned}$$

where c_{ij} is defined in (4.12). To calculate $k-c_{ij}(k)$ we consider the following two cases.

Case 1 If $i<j$

$$k-c_{ij}(k) = \begin{cases} i-j & \text{if } k=i \\ 1 & \text{if } i<k\le j \\ 0 & \text{otherwise.} \end{cases}$$

Case 2 If $j < i$

$$k - c_{ij}(k) = \begin{cases} i-j & \text{if } k = i \\ -1 & \text{if } j \le k < i \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned} |\nabla f(\sigma)|^2 &\le \frac{16}{n^2(n-1)^2} \sum_{i<j} \sum_{k=0}^{n-1} |k - c_{ij}(k)|^2 + \sum_{j<i} \sum_{k=0}^{n-1} |k - c_{ij}(k)|^2 \\ &= \frac{16}{n^2(n-1)^2} \sum_{i<j} \left((i-j)^2 + (j-i)\right) + \sum_{j<i} \left((i-j)^2 + (i-j)\right) \\ &= \frac{16}{n^2(n-1)^2} \sum_{i,j=0}^{n-1} (i-j)^2 + |i-j| \\ &\le \frac{32}{n^2(n-1)^2} \sum_{i,j=0}^{n-1} (i-j)^2 \le 32. \end{aligned}$$

Lemma 4 of Wilson (2004) along with the estimate above imply the stated lower bound in total variation.

□

5. Random walks driven by conjugacy classes.

5.1. *Review of some discrete time results.* In Section 4 we considered the random walk on S_n driven by the conjugacy class of transpositions. More generally, one can study random walks driven by a fixed conjugacy class. Recall that $\mathcal{C}$ is a conjugacy class of a group G if for some $x \in G$ we have that $\mathcal{C} = \{gxg^{-1} : \forall g \in G\}$.

Throughout this section, $\mathcal{C}$ will refer to a conjugacy class in S_n and $\operatorname{supp}(\mathcal{C})$ will denote the support size of $\mathcal{C}$, that is, the number of points that are not fixed under the action of an element in $\mathcal{C}$. Conjugacy classes of the symmetric group S_n are described by the cycle structure of their elements which is often given by a tuple of non-increasing integers greater than or equal to 2 and with sum at most n. For instance, in S_n with $n \ge 8$, the tuple $(4, 2, 2)$ describes the conjugacy class $\mathcal{C}$ of those permutations that are the product of two transpositions and one 4-cycle, all with disjoint supports. In this example, $\operatorname{supp}(\mathcal{C}) = 8$.

If $\mathcal{C}$ consists of odd permutations, that is, permutations which can be written as a product of an odd number of transpositions, then $\mathcal{C}$ generates S_n. If $\mathcal{C}$ is even, that is, any element in $\mathcal{C}$ can be written as the product of an even number of transpositions and $\mathcal{C} \neq \{e\}$ then it generates the alternating group A_n. Set $q_{\mathcal{C}}$ to be the measure

$$q_{\mathcal{C}}(\sigma) = \begin{cases} \frac{1}{\#\mathcal{C}} & \text{if } \sigma \in \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

where $\#\mathcal{C}$ denotes the number of elements in $\mathcal{C}$. When $\mathcal{C}$ is an odd conjugacy class the random walk driven by $q_{\mathcal{C}}$ is be periodic and $q_{\mathcal{C}}^t$ is supported on A_n when t is even, and on $S_n \backslash A_n$ otherwise. In this case, it is convenient to study the random walk on A_n driven by $q_{\mathcal{C}}^2$ to avoid periodicity.

The mixing time of these random walks was studied in Lulov and Pak (2002), Müller and Schlage-Puchta (2007), Roichman (1996) and Schlage-Puchta (2008), among other works. See the discussion in Saloff-Coste (2004). For simplicity, we describe some of the known results in the case of even conjugacy classes. The same results hold in the odd case, modulo periodicity. In Schlage-Puchta (2008) it is shown that any sequence $(A_n, q_{\mathcal{C}_n})$ has a total variation cutoff at time

$$t_1(n) = \inf\{k : q^k_{\mathcal{C}_n}(\varphi) \le \log n\}$$

where $\varphi(\sigma)$ is the number of fixed points of $\sigma \in S_n$ and $q^k_{\mathcal{C}}(\varphi)$ is the expected value of φ taken according to the measure $q^k_{\mathcal{C}}$. It is well known , see Diaconis (1988) and Sagan (2001), that

$$\varphi(\cdot) - 1 = \chi_{(n-1,1)}(\cdot) = n - 1 - \operatorname{supp}(\cdot).$$

This implies that $\varphi - 1$ is an eigenfunction of $q_{\mathcal{C}}$ with eigenvalue $\left(\frac{\chi_{(n-1,1)}(\mathcal{C})}{n-1}\right)$. Thus we can rewrite $t_1(n)$ as

$$t_1(n) = \inf\left\{k : (n-1)\left(1 - \frac{\operatorname{supp}(\mathcal{C}_n)}{n-1}\right)^k + 1 \le \log n\right\}.$$

When $\operatorname{supp}(\mathcal{C}_n)$ is not too large (e.g., $\operatorname{supp}(\mathcal{C}_n)/n = o(1)$) then

$$t_1(n) \sim (n/\operatorname{supp}(\mathcal{C}_n)) \log n$$

and when $\operatorname{supp}(\mathcal{C}_n)$ is very large then $t_1(n)$ is $\mathcal{O}(1)$.

Assuming that $\operatorname{supp}(\mathcal{C}_n) \le n - 1$, Müller and Schlage-Puchta (2007) show that the random walk driven by $q_{\mathcal{C}_n}$ has an ℓ^2 pre-cutoff at time $t_2(n)$ where

$$\left| t_2(n) - \frac{2\log n}{\log(n/(n - \operatorname{supp}(\mathcal{C}_n) + 1))} \right| \le 3.$$

As in the total variation case, when $\operatorname{supp}(\mathcal{C}_n)$ is not too large then

$$t_2(n) \sim (n/\operatorname{supp}(\mathcal{C}_n)) \log n$$

and when $\operatorname{supp}(\mathcal{C}_n)$ is large we get the at $t_2(n)$ is $\mathcal{O}(1)$. Here, we will focus on the continuous time process associated to $q_{\mathcal{C}_n}$.

Corollary 4.1 in Chen (2006) implies that the continuous time process driven by $q_{\mathcal{C}_n}$ has a total variation mixing time bounded above by that of the discrete time process. Arguments similar to those in Chapter 4 of Roussel (1999) give a lower bound for the continuous time process that is comparable to the upper bound just mentioned. In particular, for $\operatorname{supp}(\mathcal{C}_n) \le (n-1)/(\log(n-1)+1)$, these arguments show that the continuous time chain associated to $q_{\mathcal{C}_n}$ has a total variation cutoff at time $t_1(n)$.

Perhaps surprisingly, we show below that, in ℓ^2, the mixing time of the continuous time process has a lower bound of $(n/2)\log n$ for any conjugacy class with $\operatorname{supp}(\mathcal{C}_n) \ge 2$. A matching upper bound is shown when $\operatorname{supp}(\mathcal{C}_n) \to \infty$ as $n \to \infty$ as well as for the conjugacy class of 4-cycles.

5.2. *ℓ^2 lower bounds in continuous time.* Through out this section $\mathcal{C}_n$ is a conjugacy class in S_n (or A_n) and $c_n \in \mathcal{C}_n$ is an arbitrary fixed element in $\mathcal{C}_n$. Recall that $\operatorname{supp}(\mathcal{C}_n)$ is $n - \varphi(c_n)$ where $\varphi(\cdot)$ is the number of fixed points.

Theorem 5.1. *For each n, set*

$$t_n = \frac{n}{2}\log n.$$

For any odd conjugacy class $\mathcal{C}_n \subset S_n$ with supp$(\mathcal{C}_n) \geq 2$, and any $\epsilon \in (0,1)$

$$\lim_{n\to\infty} d_2(h_{\mathcal{C}_n,(1-\epsilon)t_n}, u_n) = \infty.$$

In order to understand the mixing time of these continuous time processes we will again rely on (2.3) and we will need to estimate the dimensions and characters of some of the irreducible representations of S_n. The following well known definitions and results will help us understand these quantities.

Definition 5.2. Let λ be a Young diagram with n boxes, as usual, we denote this by $\lambda \vdash n$. The hook at the cell (i,j) is defined as the set of boxes $H_{i,j}$ where

$$H_{i,j} = \{(i,l) : (i,l) \in \lambda, l \geq j\} \cup \{(k,j) : (k,j) \in \lambda, k \geq i\}.$$

$H_{i,j}$ has hook length $h_{i,j} = |H_{i,j}|$.

Theorem 5.3. (The Hook formula) *Let λ be a Young diagram with n boxes. Set d_λ to be the dimension of the irreducible representation associated to λ. Then*

$$\mathrm{d}_\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}} \tag{5.2}$$

With the hook formula we can now get an estimate on the dimension of some representations of S_n.

Lemma 5.4. *Let $n \in \mathbb{N}$ and $\lambda \vdash n$ be a Young diagram. If λ fits into a rectangle of $s \times t$ boxes, then*

$$\mathrm{d}_\lambda \geq \left(\frac{n}{e(s+t-1)}\right)^n.$$

Proof. Note that any hook in λ will be of hook length at most $s+t-1$. The inequality then follows from Stirling's formula in (4.6) and hook formula (5.2). □

We will use the following rather non-trivial bound on character ratios.

Theorem 5.5. (Rattan and Śniady, 2008) *Let $a > 0$ be a fixed constant, and let $\lambda \vdash n$ be a Young diagram with at most $a\sqrt{n}$ rows and columns. Then there exists a constant $D = D(a)$ such that*

$$\left|\frac{\chi_\lambda(\sigma)}{\mathrm{d}_\rho}\right| \leq \left(\frac{D\max\{1, |\sigma|^2/n\}}{\sqrt{n}}\right)^{|\sigma|}$$

for any $\sigma \in S_n$ and where $|\sigma|$ is the minimal number of transpositions needed to write σ as a product of transpositions.

Recall that, for any $\sigma \in S_n$ which is not the identity then $|\sigma| \leq \mathrm{supp}(\sigma)$.

Proof of Theorem 5.1. The idea behind this proof is to write the desired ℓ^2 distance as in equation (2.3) and find an irreducible representation which has large dimension and small character. The representations that are useful in this respect turn out to be those that have an approximately square shape.

Let $\lambda_n \vdash n$ be a Young diagram that fits into a box of side $\lceil\sqrt{n}\rceil$, so that λ_n looks almost like a square. By Lemma 5.4 and the fact that $\lceil\sqrt{n}\rceil \leq 2\sqrt{n}$ we get that

$$\mathrm{d}_{\lambda_n} \geq \left(\frac{n}{2e\lceil\sqrt{n}\rceil}\right)^n \geq \left(\frac{\sqrt{n}}{4e}\right)^n. \tag{5.3}$$

If $\mathcal{C}_n$ is a conjugacy class with $\mathrm{supp}(\mathcal{C}_n) \geq \sqrt{n}$ and $c_n \in \mathcal{C}_n$, Roichman (1996, Theorem 1) yields a positive constant $q < 1$ such that

$$\left| \frac{\chi_{\lambda_n}(c_n)}{\mathrm{d}_{\lambda_n}} \right| \leq q^{\mathrm{supp}(\mathcal{C}_n)} \leq q^{\sqrt{n}}.$$

If $2 \leq \mathrm{supp}(\mathcal{C}_n) \leq \sqrt{n}$, Theorem 5.5 implies that

$$\left| \frac{\chi_{\lambda_n}(c_n)}{\mathrm{d}_{\lambda_n}} \right| \leq \frac{D}{\sqrt{n}}.$$

In either case, we have that

$$\left| \frac{\chi_{\lambda_n}(c_n)}{\mathrm{d}_{\lambda_n}} \right| = o(1).$$

Using (5.3), we obtain that, for any $\epsilon > 0$,

$$\mathrm{d}_{\lambda_n}^2 \left\{ -(1-\epsilon)\, n \log n \left(1 - \frac{\chi_{\lambda_n}(c_n)}{\mathrm{d}_{\lambda_n}} \right) \right\} \geq \left(\frac{n}{16e^2} \right)^n \exp\left\{ -(1-\epsilon)\, n \log n\, (1 + o(1)) \right\}.$$

It now follows from (2.3)

$$\lim_{n\to\infty} d_2(h_{\mathcal{C}_n,(1-\epsilon)t_n}, u_n) \geq \lim_{n\to\infty} \mathrm{d}_{\lambda_n} \exp\left\{ -(1-\epsilon)t_n \left(1 - \frac{\chi_{\lambda_n}(c_n)}{\mathrm{d}_{\lambda_n}} \right) \right\} = \infty$$

as desired. □

Using the same ideas as in the proof of Theorem 5.1 we get the following result.

Theorem 5.6. *Let $\mathcal{C}_n$ be a conjugacy class in A_n with $\mathrm{supp}(\mathcal{C}_n) \geq 2$, and set $\overline{u}_n$ to be the uniform measure on A_n. For any $\epsilon \in (0,1)$ and $t_n = \frac{n}{2}\log n$*

$$\lim_{n\to\infty} d_2(h_{\mathcal{C}_n,(1-\epsilon)t_n}, \overline{u}_n) = \infty.$$

Remark 5.7. For $\epsilon > 0$, it is interesting to consider the discrete time chain driven by

$$\tilde{q}_{\mathcal{C}_n,\epsilon}(\sigma) = \begin{cases} \epsilon & \text{if } \sigma = e \\ \frac{1-\epsilon}{\#\mathcal{C}_n} & \text{if } \sigma \in \mathcal{C}_n \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

When $\epsilon = 1/2$, this is often called the lazy chain associated to $q_{\mathcal{C}_n}$. The arguments used in the proof of Theorem 5.1 show that the random walk driven by $\tilde{q}_{\mathcal{C}_n,\epsilon}$ will have a ℓ^2 mixing time lower bound of $(n/2)\log_{1/\epsilon}(n)$.

In Saloff-Coste (2004) it is conjectured (Conjecture 9.3) that both the total variation mixing time and the ℓ^2 mixing time of the random walk driven by $\tilde{q}_{\mathcal{C}_n}$ will have an upper bound of $(2n/\mathrm{supp}(\mathcal{C}_n))\log n$. While this is true for in the case of total variation, the results above show that the bound does not hold for ℓ^2.

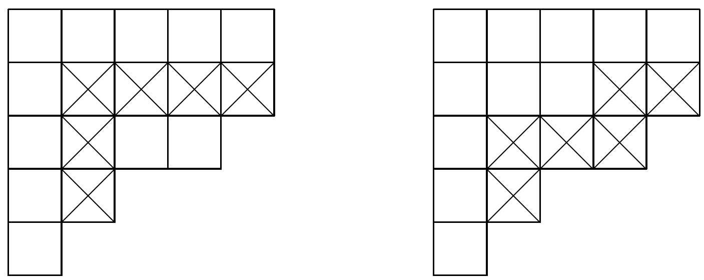
The proof of Theorem 5.1 relies on the character estimates of Rattan and Śniady (2008) and Roichman (1996). While these estimates are very useful, one can construct simple Young diagrams and use the Murnaghan-Nakayama Rule below to get estimates on the values of characters at k-cycles, for infinitely many k. This gives a much more accessible proof of a weaker version of Theorems 5.1 and 5.6. For further details on the following definitions see Section 4.10 in Sagan (2001).

Definition 5.8. A skew hook ξ in a Young diagram is a collection of boxes that result from the projection of a regular hook along the right boundary of a Young diagram.

The leg length of a skew hook ξ is denote by $ll(\xi)$ with

$$ll(\xi) = \text{the number of rows of } \xi - 1.$$

FIGURE 5.2. A hook and its corresponding skew hook of leg length 2.



Theorem 5.9. (Murnaghan-Nakayama Rule) *If λ is a partition of n and $\alpha \in S_n$ such that α has cycle type $(\alpha_1, \alpha_2, \ldots, \alpha_i)$, then we have*

$$\chi_\lambda(\alpha) = \sum_{\xi} (-1)^{ll(\xi)} \chi_{\lambda\backslash\xi}(\alpha\backslash\alpha_1) \tag{5.5}$$

where the sum runs over all skew hooks ξ of λ having α_1 cells and $\chi_{\lambda\backslash\xi}(\alpha\backslash\alpha_1)$ denotes that character of the representation $\lambda\backslash\xi$ evaluated at an element of cycle type $\alpha\backslash\alpha_1$.

It is important to remark that when using the Murnaghan-Nakayama rule, if it is impossible to remove a skew hook of the right size then the part of the sum corresponding to that skew hook is zero. A good source for more information on the Murnaghan-Nakayama rule and skew hooks is Sagan (2001).

Lemma 5.10. *For $m \in \mathbb{N}$ set $n = m(m+1)/2$. Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m) \vdash n$ be a triangular Young diagram such that $\lambda_i = m - i + 1$. Let c_k be a cycle of length k. If $k = 4i + 1$ for $i = 1, 2, \ldots$ then $\chi_\lambda(c_k) > 0$. If k is even then $\chi_\lambda(c_k) = 0$.*

Proof. The Murnaghan-Nakayama rule implies that

$$\chi_\lambda(c_k) = \sum_{|\xi|=k} (-1)^{ll(\xi)} \mathrm{d}_{\lambda\backslash\xi}.$$

Any hook in λ composed of must have even leg length by construction so it follows that $\chi_\lambda(c_k) > 0$. The second part of the proof follows directly from the Murnaghan-Nakayama rule and the fact that every hook in λ will have odd hook length, making it impossible to remove a skew-hook of even length. □

Using the dimension and character estimates from (5.3) and Lemma 5.10, one can replicate the ideas in the proof of Theorem 5.1. If $\mathbf{c}_k$ denotes here the conjugacy class of cycles of length k, for any $\epsilon > 0$ and $t_n = \frac{n}{2} \log n$, we have

(1) if k_n is even then

$$\lim_{n\to\infty} d_2(h_{\mathbf{c}_{k_n},(1-\epsilon)t_n}, u_n) = \infty$$

(2) if k_n is odd and $k_n = 4i_n + 1$ for $i_n = 1, 2, 3, \ldots$ then

$$\lim_{n\to\infty} d_2(h_{\mathbf{c}_{k_n},(1-\epsilon)t_n}, \overline{u}_n) = \infty.$$

5.3. *Total variation upper bounds in continuous time.* As we mentioned at the beginning of this section, the mixing time of the continuous time process $h_{\mathcal{C}_n,t_n}$ will depend on whether one considers the total variation or the ℓ^2 distance. In this section we derive a total variation upper bound of type $(n/\mathrm{supp}(\mathcal{C}_n))\log n$ for the continuous time process associated to $q_{\mathcal{C}_n}$. In the next section, we shall show that the ℓ^2 mixing time has an upper bound of $(n/2)\log n$ for the continuous time process when $\mathrm{supp}(\mathcal{C}_n) \to \infty$.

Proposition 5.11. *Let $\mathcal{C}_n$ be an even conjugacy class and $\overline{u}_n$ to be the uniform measure on A_n. Let T_n be the total variation cutoff time of $q_{\mathcal{C}_n}$ (in discrete time) and assume that $T_n \to \infty$. Then, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} d_{\mathrm{TV}}(h_{\mathcal{C}_n,(1+\epsilon)T_n}, \overline{u}_n) = 0$$

Proof. Let

$$\begin{aligned} T^d_{n,\epsilon} &= \inf\{t \geq 0 : d_{\mathrm{TV}}(q^{(t)}_{\mathcal{C}_n}, \overline{u}_n) \leq \epsilon\} \\ T^c_{n,\epsilon} &= \inf\{t \geq 0 : d_{\mathrm{TV}}(h_{\mathcal{C}_n,t}, \overline{u}_n) \leq \epsilon\}. \end{aligned}$$

Corollary 4.1 in Chen (2006) shows that for any $\delta \in (0,1)$, $\epsilon > 0$ and $\eta \in (0, \epsilon)$ there exists an integer $N = N(\delta, \eta)$ such that

$$(1-\delta)T^c_{n,\epsilon} \leq T^d_{n,\eta} \ \text{ for all } n \geq N.$$

In particular, for any $\epsilon > 0$ we can find a $\delta \in (0,1)$ and an $N_1 = N_1(\delta, \eta)$ such that for all $n \geq N$

$$T^c_{n,\eta} \leq \sqrt{1+\epsilon}\ T^d_{n,\eta/2}.$$

From Schlage-Puchta (2008) we know that the random walk driven by $q_{\mathcal{C}_n}$ has cutoff, hence for any $\epsilon > 0$ and $\eta \geq 0$ there exists an $N_2 = N_2(\epsilon, \eta)$ such that for all $n \geq N_2$

$$T^d_{n,\eta/2} \leq \sqrt{1+\epsilon}\ T_n.$$

Combining the inequalities above gives that for any $\epsilon > 0$ and $\eta > 0$ there exists an $N = \max\{N_1, N_2\}$ such that for all $n \geq N$

$$T^c_{n,\eta} \leq (1+\epsilon)T_n.$$

The desired result follows.

□

Remark 5.12. In the case of the lazy random walk $\tilde{q}_{\mathcal{C}_n,1/2}$ defined in (5.4), one can show that the total variation mixing time is bounded by approximately twice that of the discrete time process $q_{\mathcal{C}_n}$. (This is a more general phenomenon.) We only treat the case when $\mathcal{C}_n$ is an even conjugacy class. Note that

$$d_{\mathrm{TV}}(\tilde{q}^{(t_n)}_{\mathcal{C}_n}, \overline{u}_n) = \sum_{k=0}^{t} 2^{-t_n}\binom{t_n}{k} d_{\mathrm{TV}}(q^{(k)}_{\mathcal{C}_n}, \overline{u}_n).$$

For any constant $D > 0$ set $\mathcal{I}_n = [0, t_n/2 - D\sqrt{t_n}] \cup [t_n/2 + D\sqrt{t_n}, t_n]$. Then we have that

$$\sum_{k \in A_n} 2^{-t_n}\binom{t_n}{k} d_{\mathrm{TV}}(q_{\mathcal{C}_n}^{(k)}, \overline{u}_n) \leq \sum_{k \in A_n} 2^{-t_n}\binom{t_n}{k}.$$

By the central limit theorem the right hand side tends to 0 as D tends to ∞.

Outside of the set $\mathcal{I}_n$ we get that

$$\begin{aligned}\sum_{k \notin A_n} 2^{-t_n}\binom{t_n}{k} d_{\mathrm{TV}}(q_{\mathcal{C}_n}^{(k)}, \overline{u}_n) &\leq d_{\mathrm{TV}}(q_{\mathcal{C}_n}^{(t_n/2 - D\sqrt{t_n})}, \overline{u}_n) \sum_{k \notin A_n} 2^{-t_n}\binom{t_n}{k} \\ &\leq d_{\mathrm{TV}}(q_{\mathcal{C}_n}^{(t_n/2 - D\sqrt{t_n})}, \overline{u}_n).\end{aligned}$$

The arguments above shows that the cutoff time of the lazy walk is asymptotically $2T_n$. A similar argument would show that for any $\epsilon > 0$ the walk driven by $\tilde{q}_{\mathcal{C}_n,\epsilon}$ has a cutoff time asymptotically equal to $(1/(1-\epsilon))T_n$.

5.4. *Continuous time ℓ^2 upper bounds: supp$(\mathcal{C}_n) \to \infty$.* Section 5.2 shows that the ℓ^2 mixing time of $h_{\mathcal{C}_n,t}$ must be at least $(n/2)\log n$ for all non trivial conjugacy classes. We show that when supp$(\mathcal{C})$ goes to ∞ as $n \to \infty$ and for the conjugacy class of 4-cycles the continuous time random walk has an ℓ^2 cutoff at $(n/2)\log n$.

Theorem 5.13. *Let $\mathcal{C}_n$ be a conjugacy class such that* supp$(\mathcal{C}_n) \to \infty$ *as $n \to \infty$. For any $\epsilon > 0$, and $t_n = (n/2)\log n$*

(1) $\lim_{n\to\infty} d_2\left(h_{\mathcal{C}_n,(1+\epsilon)t_n}, u_n\right) = 0$ *if $\mathcal{C}_n$ is odd.*
(2) $\lim_{n\to\infty} d_2\left(h_{\mathcal{C}_n,(1+\epsilon)t_n}, \overline{u}_n\right) = 0$ *if $\mathcal{C}_n$ is even.*

Proof. Let $\mathcal{C}_n$ be an odd conjugacy class. Set $(\beta_i)_0^{n!-1}$ to be the eigenvalues associated to the measure $q_{\mathcal{C}_n}$ and $\lambda_i = 1 - \beta_i$. From (2.1) we know that

$$\begin{aligned} d_2(h_{\mathcal{C}_n,(1+\epsilon)t_n}, u_n)^2 &= \sum_{i=1}^{n!-1} e^{-2(1+\epsilon)t_n\lambda_i} \\ &= \sum_{\lambda_i \leq 1-1/w} e^{-2t_n(1+\epsilon)\lambda_i} + \sum_{\lambda_i \geq 1-1/w} e^{-2t_n(1+\epsilon)\lambda_i}.\end{aligned}$$

We will use the following Calculus inequality.

Claim 5.14. *For $w \geq 4$ and $0 \leq x \leq 1 - 1/w$ we have that $2\log(1-x) \geq -wx$.*

For $1/3 \geq \epsilon > 0$, $w = (1+\epsilon)/\epsilon \geq 4$, so by the claim above and (2.1) we obtain

$$\begin{aligned} d_2(h_{\mathcal{C}_n,(1+\epsilon)t_n}, u_n)^2 &\leq \sum_{1/w \leq \beta_i} \beta_i^{\epsilon 4 t_n} + n! e^{-2t_n(1-1/w)(1+\epsilon)} \\ &= \sum_{1/w \leq \beta_i} \beta_i^{\epsilon 4 t_n} + n! e^{-n\log n}\end{aligned}$$

We know that the eigenvalues of $q_{\mathcal{C}_n}$ are just the normalized characters $\chi_\rho(c_n)/\mathrm{d}_\rho$, $c_n \in \mathcal{C}_n$, that occur with multiplicity d_ρ^2. Let ρ_1 and ρ_2 be the trivial and sign representations respectively. When $\mathcal{C}_n$ is odd $\chi_{\rho_2}(c_n)/\mathrm{d}_{\rho_2} = -1$, so the character associated to the sign representation does not contribute to the sum of eigenvalues above. Furthermore, (see, e.g., Müller and Schlage-Puchta, 2007, Lemma 2)

$$d_2(q_{\mathcal{C}_n}^{(2t)}, \overline{u}_n)^2 = \frac{1}{2}\sum_{\rho \neq \rho_1,\rho_2} \mathrm{d}_\rho^2 \left(\frac{\chi_\rho(c_n)}{\mathrm{d}_\rho}\right)^{4t}. \tag{5.6}$$

It now follows that

$$\begin{aligned} d_2(h_{\mathcal{C}_n,(1+\epsilon)t_n}, u_n)^2 &\leq \sum_{\rho \neq \rho_1, \rho_2} \mathrm{d}_\rho^2 \left(\frac{\chi_\rho(c_n)}{\mathrm{d}_\rho} \right)^{\epsilon 4 t_n} + n! e^{-n \log n} \\ &\leq 2 d_2(q_{\mathcal{C}_n}^{(\epsilon 2 t_n)}, \overline{u}_n)^2 + n! e^{-n \log n}. \end{aligned}$$

In Müller and Schlage-Puchta (2007) it is shown that there exists a fixed constant $D > 0$ such that for t_n even and $t_n \geq (Dn/\mathrm{supp}(\mathcal{C}_n)) \log n$ then $d_2(q_{\mathcal{C}_n}^{(t_n)}, \overline{u}_n) \to 0$ as $n \to \infty$. Since $\mathrm{supp}(\mathcal{C}_n) \to \infty$ as $n \to \infty$ then for large enough n we have that $\epsilon n \log n \geq (Dn/\mathrm{supp}(\mathcal{C}_n)) \log n$ and the desired result follows. The case when $\mathcal{C}_n$ is an even conjugacy class can be treated in a similar way. □

Remark 5.15. Let $\tilde{q}_{\mathcal{C}_n}$ be the lazy chain defined in (5.4). In the remark after Theorem 5.6 it is noted that the random walk driven by $\tilde{q}_{\mathcal{C}_n}$ will have a ℓ^2 lower bound on the mixing time of $(n/2)\log_2(n)$. A matching upper bound for conjugacy classes $\mathcal{C}_n$ such that $\mathrm{supp}(\mathcal{C}_n) \to \infty$ as $n \to \infty$ follows from an argument similar to the proof of Theorem 5.13.

5.5. *ℓ^2 continuous time upper bound: 4-cycles.* The next theorem gives a sharp ℓ^2 upper bound for the 4-cycle walk. In the case when $\mathrm{supp}(\mathcal{C}_n) \to \infty$ we relied on the (rather deep) results of Roichman (1996), Müller and Schlage-Puchta (2007) and Schlage-Puchta (2008) concerning the discrete time case to obtain a continuous time result matching our lower bound. This technique does not work for conjugacy classes with fixed support size. We conjecture that, with out any restriction on $\mathrm{supp}(\mathcal{C}_n)$, $(n/2)\log n$ is a ℓ^2 cutoff time for the family $(h_{\mathcal{C}_n,t})$. Note however that there is no reasons to hope for a proof simpler than that for random transposition. In discrete time, the only cases with fixed support size for which the ℓ^2 cutoff time has been determined are the cases of support size at most 6 (and the 7-cycles) treated in Roussel (1999) and Roussel (2000). Using the techniques of Roussel (1999) and Roussel (2000) one can probably treat the corresponding continuous time processes, but this will be hard work. Here we focus on the 4-cycle walk. The reason is that we are able to reduce most technical computations to those already done above for transposition. We note that is is unlikely such reduction would work easily for 3-cycles and other even conjugacy classes (see Roussel, 1999 and Roussel, 2000).

Recall that the conjugacy class of 4-cycles is denoted by $\mathbf{c}_4$. We let c_4 be a given 4-cycle.

Theorem 5.16. *For $n \geq 11$, $c \geq 2$ and $t \geq (n/2)(\log n + c)$*

$$d_2\left(h_{\mathbf{c}_4,t}, u_n\right) \leq e^{-(c-2)}$$

We will use (2.3) again and bound $\chi_\rho(c_4)/\mathrm{d}_\rho$, $c_4 \in \mathbf{c}_4$, with the same upper bounds that we used for $\chi_\rho(\tau)/\mathrm{d}_\rho$, $\tau \in \mathbf{c}_2$, in the case of transpositions in Proposition 4.6. In order to do this we will need the following definitions and lemmas.

Definition 5.17. If $\lambda' = (\lambda'_1, \ldots, \lambda'_j)$ and $\lambda = (\lambda_1, \ldots, \lambda_k)$ are two Young diagrams such that $\sum_{i=1}^{j} \lambda'_i = \sum_{i=1}^{k} \lambda_i = n$ and it is possible to get from λ to λ' by moving boxes up to the right then we say that $\lambda' \geq \lambda$.

Definition 5.18. Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ denote a Young diagram such that $\sum_{i=1}^m \lambda_i = n$. For any integer $l \geq 0$

$$M_{\lambda,2l} = \sum_{j=1}^{m} \left\{ (\lambda_j - j)^l (\lambda_j - j + 1)^l - j^l (j-1)^l \right\}.$$

Lemma 5.19. *Let λ' and λ be two Young diagrams associated to irreducible representations of S_n. If $\lambda' \geq \lambda$ then $M_{\lambda',2l} \geq M_{\lambda,2l}$ for all $l \geq 0$.*

Proof. It suffices to show that $M_{\lambda',2l} \geq M_{\lambda,2l}$ for that case when $a < b$ and $\lambda'_a = \lambda_a + 1$, $\lambda'_b = \lambda_b - 1$ and $\lambda'_c = \lambda_c$ for $c \neq a, b$. In this case,

$$\begin{aligned} M_{\lambda',2l} - M_{\lambda,2l} &= (\lambda_a - a + 1)^l \left\{ ((\lambda_a - a + 1) + 1)^l - ((\lambda_a - a + 1) - 1)^l \right\} \\ &\quad + (\lambda_b - b)^l \left\{ (\lambda_b - b - 1)^l - (\lambda_b - b + 1)^l \right\}. \end{aligned}$$

Set $x = \lambda_a - a + 1$ and $y = \lambda_b - b$ then $n \geq x \geq y \geq 1 - n$ and $M_{\lambda',2l} - M_{\lambda,2l} = f_{x,y}(l)$ where

$$f_{x,y}(l) = x^l \left\{ (x+1)^l - (x-1)^l \right\} + y^l \left\{ (y-1)^l - (y+1)^l \right\}.$$

Diaconis (1988) shows that $f_{x,y}(1) \geq 0$ for $n \geq x \geq y \geq 1 - n$ which implies that $M_{\lambda',2} \geq M_{\lambda,2}$. We will show the general case by induction. Assume that $f_{x,y}(l) \geq 0$ then

$$\begin{aligned} f_{x,y}(l+1) &= x^{l+1} \left\{ (x+1)^{l+1} - (x-1)^{l+1} \right\} + y^{l+1} \left\{ (y-1)^{l+1} - (y+1)^{l+1} \right\} \\ &= x^2 \left(x^l \left\{ (x+1)^l - (x-1)^l \right\} \right) + y^2 \left(y^l \left\{ (y-1)^l - (y+1)^l \right\} \right) \\ &\quad + x^{l+1} \left\{ (x+1)^l + (x-1)^l \right\} - y^{l+1} \left\{ (y+1)^l + (y-1)^l \right\} \\ &\geq x^{l+1} \left\{ (x+1)^l + (x-1)^l \right\} - y^{l+1} \left\{ (y+1)^l + (y-1)^l \right\}. \end{aligned}$$

The last inequality follows since $f_{x,y}(l) \geq 0$. To conclude that $f_{x,y}(l+1) \geq 0$ we must check the following three cases.

Case 1: $x \geq y \geq 0$. This case follows directly from the assumption $x \geq y$.

Case 2: $x \geq 0$ and $y \leq 0$. Note that in this case

$$\begin{aligned} x^{l+1} \left\{ (x+1)^l + (x-1)^l \right\} &\geq 0 \\ y^{l+1} \left\{ (y+1)^l + (y-1)^l \right\} &\leq 0. \end{aligned}$$

The last inequality follows from the fact that l and $l+1$ are odd and even numbers.

Case 3: $y \leq x \leq 0$. In this case let $\tilde{x} = -x$ and $\tilde{y} = -y$ then $\tilde{y} \geq \tilde{x} \geq 0$ and

$$\begin{aligned} & x^{l+1} \left\{ (x+1)^l + (x-1)^l \right\} - y^{l+1} \left\{ (y+1)^l + (y-1)^l \right\} = \\ & \tilde{y}^{l+1} \left\{ (\tilde{y}+1)^l + (\tilde{y}-1)^l \right\} - \tilde{x}^{l+1} \left\{ (\tilde{x}+1)^l + (\tilde{x}-1)^l \right\}. \end{aligned}$$

Case 3 now follows directly from Case 1. □

Lemma 5.20. *Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_j)$ denote a Young diagram such that $\sum \lambda_i = n$. Then*

$$M_{2l,\lambda} \leq n(\lambda_1 - 1)^l \lambda_1^{l-1}.$$

Proof.

$$\begin{aligned}M_{2l,\lambda} &= \sum_{i=1}^{j}(\lambda_j - j)^l(\lambda_j - j + 1)^l - j^l(j-1)^l \\ &\le \sum_{\lambda_j \ge j-1} (\lambda_j - j)^l(\lambda_j - j + 1)^l \\ &\quad + \sum_{\lambda_j < j-1} (\lambda_j - j)^l(\lambda_j - j + 1)^l - j^l(j-1)^l\end{aligned}$$

For $0 \le \lambda_j \le j-1$ it is true that $|\lambda_j - j| \le j$ and $|\lambda_j - j + 1| \le j - 1$ which implies that the second sum in the inequality above is negative. Therefore

$$M_{2l,\lambda} \le \sum_{\lambda_j \ge j-1} (\lambda_j - j)^l(\lambda_j - j + 1)^l \le n(\lambda_1 - 1)^l \lambda_1^{l-1}.$$

□

Lemma 5.21. *Let ρ be an irreducible representation of S_n and λ the associated Young diagram. For $n \ge 11$ the normalized character $r_4(\lambda) = \chi_\rho(c_4)/\mathrm{d}_\rho$ can be bounded as follows.*

$$r_4(\lambda) \le \begin{cases} 1 - \frac{2\lambda_1(n-\lambda_1)}{n(n-1)} & \text{if } \lambda_1 \ge n/2 \\ \frac{\lambda_1 - 1}{n-1} & \text{if } \lambda_1 \le n/2. \end{cases}$$

Proof. Set $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_j)$. Ingram (1950) shows that

$$\frac{n!}{(n-4)!} r_4(\lambda) = M_{4,\lambda} - 2(2n-3)M_{2,\lambda}. \tag{5.7}$$

Lemma 5.19 implies that $M_{2,\lambda} \ge M_{2,\lambda'}$ where $\lambda' = (\lambda_1, 1, 1, \ldots, 1)$. We get

$$\begin{aligned}M_{2,\lambda'} &= (\lambda_1 - 1)\lambda_1 + \sum_{j=2}^{n-\lambda_1} (1-j)(2-j) - j(j-1) \\ &= (\lambda_1 - 1)\lambda_1 - 2\sum_{j=1}^{n-\lambda_1-1} j \\ &= (\lambda_1 - 1)\lambda_1 - (n - \lambda_1 - 1)(n - \lambda_1).\end{aligned}$$

If $\lambda_1 \ge n/2$ then $M_{4,\lambda} \le M_{4,(\lambda_1, n-\lambda_1)}$. Note that

$$\begin{aligned}M_{4,(\lambda_1,n-\lambda_1)} &= (\lambda_1 - 1)^2\lambda_1^2 + (n-\lambda_1-1)^2(n-\lambda_1)^2 - 4 \\ &\le (\lambda_1 - 1)^2\lambda_1^2 + (n-\lambda_1-1)^2(n-\lambda_1)^2 \\ &= [(\lambda_1-1)\lambda_1 - (n-\lambda_1-1)(n-\lambda_1)]^2 + 2(\lambda_1-1)\lambda_1(n-\lambda_1-1)(n-\lambda_1).\end{aligned}$$

Hence if $\lambda_1 \ge n/2$, we have

$$\begin{aligned}M_{4,\lambda} - 2(2n-3)M_{2,\lambda} &= [(\lambda_1-1)\lambda_1 - (n-\lambda_1)(n-\lambda_1-1)] \\ &\quad \times[(\lambda_1-1)\lambda_1 - (n-\lambda_1)(n-\lambda_1-1) - 2(2n-3)] \\ &\quad +2(\lambda_1-1)\lambda_1(n-\lambda_1)(n-\lambda_1-1).\end{aligned}$$

Note that

$$(\lambda_1-1)\lambda_1 - (n-\lambda_1)(n-\lambda_1-1) - 2(2n-3) \le (n-2)(n-3).$$

It follows that

$$M_{4,\lambda} - 2(2n-3)M_{2,\lambda} \le (n-2)(n-3)[(\lambda_1-1)\lambda_1 - (n-\lambda_1)(n-\lambda_1-1)] + 2\lambda_1(\lambda_1-1)(n-\lambda_1)(n-\lambda_1-1). \tag{5.8}$$

If $\lambda_1 \ge n-1$ then $2\lambda_1(\lambda_1-1)(n-\lambda_1)(n-\lambda_1-1) = 0$. If $\lambda_1 \le n-2$ then $\lambda_1(\lambda_1-1) \le (n-2)(n-3)$. In either case, (5.8) gives that

$$\begin{aligned} r_4(\lambda) &\le \frac{(\lambda_1-1)\lambda_1 - (n-\lambda_1)(n-\lambda_1-1)}{n(n-1)} + \frac{2(n-\lambda_1)(n-\lambda_1-1)}{n(n-1)} \\ &= 1 - \frac{2\lambda_1(n-\lambda_1)}{n(n-1)}. \end{aligned}$$

Next, we show the second part of the inequality. By Lemma 5.20 and (5.7) we have that for $\lambda_1 \le n/2$

$$\begin{aligned} |r_4(\lambda)| &\le \frac{(n-4)!}{n!}\left[n(\lambda_1-1)^2\lambda_1 + 2(2n-3)n(\lambda_1-1)\right] \\ &= \left(\frac{\lambda_1-1}{n-1}\right)\left[\frac{(\lambda_1-1)\lambda_1 + 2(2n-3)}{(n-2)(n-3)}\right] \\ &\le \left(\frac{\lambda_1-1}{n-1}\right)\left[\frac{n^2/4+4n-6}{(n-2)(n-3)}\right] \le \frac{\lambda_1-1}{n-1}. \end{aligned}$$

The last inequality holds for $n \ge 11$. □

Proof of Theorem 5.16. Recall that

$$d_2(h_{\mathbf{c}_4,t}, u_n)^2 = \sum_{\lambda\ne 1} \mathrm{d}_\lambda^2 \exp\left\{-2t\left(1 - r_4(\lambda)\right)\right\}.$$

In order to obtain the desired $e^{-2(c-2)}$ constant we will bound the term corresponding to $\lambda = (n-1,1)$ separately. For $\lambda = (n-1,1)$ we get that

$$M_{2,(n-1,1)} = (n-2)(n-1) - 2 \text{ and } M_{4,(n-1,1)} = (n-2)^2(n-1)^2 - 4$$

which implies $r_4((n-1,1)) = 1 - 4/(n-1)$. So for $t \ge (n/2)(\log n + c)$,

$$d^2_{(n-1,1)} \exp\{-2t(1 - r_{(n-1,1)}(4))\} \le (n-1)^2 \exp\{-4(\log n + c)\} \le e^{-4c}/n^2.$$

Lemma 5.21 and equation (3.3) imply that for $t \ge (n/2)(\log n + c)$ we have $d_2(h_{t,4}, u_n)^2 \le e^{-4c}/n^2 + S_1 + S_2$ where

$$\begin{aligned} S_1 &= \sum_{j=2}^{n/2}\left(\frac{n!}{(n-j)!}\right)^2\left(\frac{1}{j!}\right)\exp\left\{-(\log n + c)\left(\frac{2j(n-j)}{n-1}\right)\right\} \\ S_2 &= \sum_{j=n/2}^{n-1}\left(\frac{n!}{(n-j)!}\right)^2\left(\frac{1}{j!}\right)\exp\left\{-j(\log n + c)\right\}. \end{aligned}$$

For a more detailed description on how to obtain the sums S_1 and S_2 see the proof of Proposition 4.2. For $c \ge 2$ we have that

(1) $-c(2j)(n-j)/(n-1) \le -2(c-2) - 2j$ when $2 \le j \le n/2$ and
(2) $-jc \le -2(c-2) - 2j$ when $j \ge 2$.

It follows that

$$d_2(h_{t,4}, u_n)^2 \le e^{-2(c-2)} \left(\frac{1}{n^2} + \sum_{j=1}^{n/2} A_j + \sum_{j=n/2}^{n} B_j \right)$$

where A_j and B_j are defined in equations (4.8) and (4.9). Lemmas 4.7 and 4.8 now imply that for $n \ge 11$

$$d_2(h_{t,4}, u_n)^2 \le e^{-2(c-2)} \left(\frac{1}{n^2} + \frac{2}{3} + \frac{1}{4} + 2\left(\frac{2}{e}\right)^{3n/2} \right) \le e^{-2(c-2)}.$$

□

References

D. Aldous. *Random walks on finite groups and rapidly mixing Markov chains*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin (1983). Seminar on probability, XVII.

D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly* **93** (5), 333–348 (1986).

G-Y Chen. *The cutoff phenomenon for finite Markov chains*. Ph.D. thesis, Cornell University (2006).

G-Y Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability* **13**, 26–78 (2008).

P. Diaconis. *Group representations in probability and statistics*. Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics (1988). Hayward, CA.

P. Diaconis. Finite Fourier Methods: Access to Tools. *Proceedings of Symposia in Applied Mathematics* **44** (1991). American Mathematical Society, Providence, RI.

P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* **93**, 1659–1664 (1996).

P. Diaconis. *Random walks on groups: characters and geometry*, volume 1 of *London Math. Soc. Lecture Note Ser.*, pages 120–142. Cambridge Univ. Press (2003). Groups St. Andrews 2001 in Oxford.

P. Diaconis, J. Fill and J. Pitman. Analysis of top to random shuffles. *Combin. Probab. Comput.* **1** (2), 135–155 (1992).

P. Diaconis and L. Saloff-Coste. Comparison techniques for random walk on finite groups. *Annals of Probability* **21** (4), 2131–2156 (1993).

P. Diaconis and L. Saloff-Coste. Random walks on finite groups: a survey of analytic techniques. *Probability measures on groups and related structures, Oberwolfach* **XI**, 44–75 (1994).

P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Geb.* **52** (2), 159–179 (1981).

W. Feller. *An introduction to probability theory and its applications. Volume 1.* John Wiley & Sons, Inc., New York-London-Sydney, Third edition (1968).

L. Flatto, Odlyzko A. M. and Wales D. B. Random Shuffles and Group Representations. *The Annals of Probability* **13** (1), 154–178 (1985).

R. E. Ingram. Some characters of the symmetric group. *Proc. Amer. Math. Soc.* **1**, 358–368 (1950).

A. Knapp. *Lie groups beyond an introduction*, volume 140 of *Progress in Mathematics*. Birkhauser Boston, Inc., Boston, MA (1996).

N. Lulov and I. Pak. Rapidly mixing random walks and bounds on characters of the symmetric group. *J. Algebraic Combin.* **16**, 151–163 (2002).

T. Müller and J. Schlage-Puchta. Character theory of symmetric groups, subgroup growth of Fuchsian groups, and random walks. *Adv. Math.* **213** (2), 919–982 (2007).

A. Rattan and P. Śniady. Upper bound on the characters of the symmetric goups for balanced Young diagrams and a generalized Frobenius Formula. *Advances in Mathematics* **218**, 673–695 (2008).

Y. Roichman. Upper bounds on the characters of the symmetric group. *Invent. Math.* **125** (3), 451–485 (1996).

S. Roussel. *Marches aléatoires sur le group symétrique*. Ph.D. thesis, Université Paul Sabatier, Toulouse III (1999).

S. Roussel. Phénoméne de cutoff pour certaines marches aléatoires sur le groupe syémetrique. *Colloquium Mathematicum* **86** (1), 111–135 (2000).

B.E. Sagan. *The Symmetric Group, Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer-Verlag, New York (2001).

L. Saloff-Coste. Precise estimates on the rate at which certain diffusions tend to equilibrium. *Mathematische Zeitschrift* **217** (4), 641–677 (1994).

L. Saloff-Coste. *Lectures on finite Markov chains*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin (1997). Lectures on probability theory and statistics (Saint-Flour, 1996).

L. Saloff-Coste. *Random walks on finite groups*, volume 110 of *Encyclopedia Math. Sci*, pages 261–346. Springer, Berlin (2004). Probability on discrete structures (H. Kesten, Ed.).

L. Saloff-Coste and J. Zúñiga. Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stocastic Processes and their Applications* **117**, 961–979 (2007).

J. Schlage-Puchta. Mixing Properties of Finite Permutation Groups (2008). In progress.

J. Uyemura-Reyes. *Random Walk, Semi-direct Products, and Card Shuffling*. Ph.D. thesis, Stanford University (2002).

D. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Prob.* **14** (1), 274–325 (2004).