

ARTICLE

# On random walks and switched random walks on homogeneous spaces

Elvira Moreno<sup>1</sup>  and Mauricio Velasco<sup>2,\*</sup> 

<sup>1</sup>Computing and Mathematical Sciences, California Institute of Technology, Department of Computing and Mathematical Sciences, 1200 E. California Blvd., MC 305-16, Pasadena, CA 91125-2100, USA and <sup>2</sup>Departamento DE Matemáticas, Universidad de los Andes, Carrera 1 No. 18a 10, Edificio H, Primer Piso, 111711 Bogotá, Colombia

\*Corresponding author. Email: [mvelasco@uniandes.edu.co](mailto:mvelasco@uniandes.edu.co)

(Received 22 November 2021; revised 13 September 2022; accepted 14 October 2022;  
first published online: 28 November 2022)

## Abstract

We prove new mixing rate estimates for the random walks on homogeneous spaces determined by a probability distribution on a finite group  $G$ . We introduce the switched random walk determined by a finite set of probability distributions on  $G$ , prove that its long-term behaviour is determined by the Fourier joint spectral radius of the distributions, and give Hermitian sum-of-squares algorithms for the effective estimation of this quantity.

**Keywords:** Mixing rates; Random walks; Switched random walks; Joint spectral radius; Hermitian polynomial norms

**2020 MSC codes:** Primary: 60J10, 05E10; Secondary: 90C23

## 1. Introduction

A person shuffles a deck of  $n$  cards. Her shuffling method is specified by a probability distribution  $Q$  on the permutation group  $S_n$ . More concretely, at stage  $j = 1, \dots, N$ , the person takes the deck in some position  $\nu$  in  $S_n$  and reshuffles it to position  $g_j\nu \in S_n$ , where  $g_j$  is a random element of  $S_n$  selected according to the distribution  $Q$  and sampled independently from the chosen  $g_s$  for  $s < j$ . The resulting process is called a *random walk* on the group  $G = S_n$ . These processes have been the focus of much work, masterfully explained in the book [D]. Under common assumptions on the distribution  $Q$ , such processes approach the uniform distribution on  $G$  as  $N$  increases (i.e., the deck of cards gets *evenly* mixed). A key quantitative question is to determine how quickly this occurs. More precisely, one wishes to bound the total variation distance between the distribution  $Q^N$  of the process after  $N$  stages and the uniform distribution  $U$ , where the *total variation* distance is defined as

$$\|Q^N - U\|_{\text{TV}} := \max_{A \subseteq G} |Q^N(A) - U(A)|.$$

More generally, if the group  $G$  acts on a set  $X$  and  $x_0$  is an element of  $X$ , then the probability distribution  $Q$  on  $G$  determines a random walk  $(h_k)_{k \in \mathbb{N}}$  on  $X$  via the formula  $h_j := g_j g_{j-1} \dots g_1 \cdot x_0$ .

In the first part of this article (Sections 2 and 3), we study the behaviour of such random walks on sets  $X$  where  $G$  acts transitively using the modules  $\mathbb{C}X$  over the group ring  $\mathbb{C}G$  which such actions determine (see Section 3.1 for details). In order to describe our results precisely, we first establish some notation. Assume the finite group  $G$  has distinct irreducible representations  $(V_j, \rho_j)$  for  $j = 1, \dots, k$  and denote by  $\text{triv}$  the trivial representation. Furthermore, let  $\mathbb{C}X$  be the

permutation representation associated to the action of  $G$  on  $X$ , that is, the space with basis given by the symbols  $\{e_x : x \in X\}$  together with the natural action of  $G$  (i.e.,  $e_g \cdot e_x := e_{g(x)}$ ), and let  $\bar{u} = \frac{1}{|X|} \sum_{x \in X} u_x \in \mathbb{C}X$  represent the uniform distribution on  $X$ . Our first result is a bound on the average squared total variation:

**Theorem 1.1.** *Let  $Q$  be a probability distribution on  $G$  and let  $X$  be a  $G$ -homogeneous space. If  $q := \sum_{g \in G} Q(g)e_g \in \mathbb{C}G$ , then the following inequalities hold:*

$$\frac{1}{|X|} \sum_{x \in X} \|q \cdot e_x - \bar{u}\|_{TV}^2 \leq \frac{1}{4} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)\|_{Fb}^2$$

$$\frac{1}{|X|} \sum_{x \in X} \|q \cdot e_x - \bar{u}\|_{TV}^2 \geq \frac{1}{4|X|} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)\|_{Fb}^2,$$

where the matrix  $\hat{Q}(\rho_j)$  denotes the value of the Fourier transform of  $Q$  in the representation  $\rho_j$ ,  $\|A\|_{Fb}^2 := \text{Tr}(AA^*)$ , and  $m(V_j, \mathbb{C}X)$  denotes the multiplicity of the irreducible representation  $V_j$  in the  $\mathbb{C}G$ -module  $\mathbb{C}X$ . Furthermore, if  $\|q \cdot e_x\|_2$  is independent of  $x \in X$ , then for every  $x \in X$  we have

$$\|q \cdot e_x - \bar{u}\|_{TV}^2 \leq \frac{1}{4} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)\|_{Fb}^2$$

$$\|q \cdot e_x - \bar{u}\|_{TV}^2 \geq \frac{1}{4|X|} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)\|_{Fb}^2.$$

The first part of Theorem 1.1 implies the existence of *deterministic* initial states which are better than average and worse than average with respect to mixing, proving that the sum appearing in the theorems above is a fine estimator of the mixing rate. More precisely,

**Corollary 1.2.** *For every integer  $N$ , there exist initial states  $r$  and  $s$  in  $X$  satisfying*

$$\|q^N \cdot e_r - \bar{u}\|_{TV}^2 \leq \frac{1}{4} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)^N\|_{Fb}^2$$

$$\|q^N \cdot e_s - \bar{u}\|_{TV}^2 \geq \frac{1}{4|X|} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)^N\|_{Fb}^2.$$

The second part of Theorem 1.1 specializes, when  $X = G$ , to the celebrated Diaconis-Shahshahani Upper bound Lemma introduced in [DS], but leads to improved estimates of the total variation distance whenever  $X \neq G$ . The reason for this improvement is that the multiplicities appearing in Theorem 1.1 are  $m(V_j, \mathbb{C}X)$ , which are no larger and typically strictly smaller than the multiplicities  $\dim(V_j)$  appearing in the Upper Bound Lemma. For instance, this improvement occurs whenever  $G$  acts transitively on  $X$  and  $|X| < |G|$  in the following corollary

**Corollary 1.3.** *Suppose  $Q$  is a probability distribution on  $G$  which is constant on the conjugacy classes of  $G$  and let  $Q = \sum_{j=1}^k a_j \chi_j$  be its unique representation as a sum of characters. If  $q^{(N)} := \sum_{g \in G} Q^N(g)e_g \in \mathbb{C}G$ , then for any  $G$ -set  $X$  and any  $x_0 \in X$  we have*

$$\|q^{(N)} \cdot e_{x_0} - \bar{u}_X\|_{TV}^2 \leq \frac{1}{4} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \dim(V_j) \left( \frac{a_j |G|}{\dim(V_j)} \right)^{2N}.$$

In Section 3.3 we apply these methods to estimate convergence rates for random walks on tabloids, illustrating connections between such estimates and the Kostka numbers of combinatorial representation theory.

The random walks on homogeneous spaces described so far are often easy to simulate on a computer, even in cases where the group  $G$  is huge (this occurs for instance whenever the support of the distribution  $Q$  is small compared to the size of the group) and therefore give us effective means of approximating the uniform distribution on  $G$  by simulating the walk for  $N$  stages. Such simulations allow us to understand what *typical elements* of the homogeneous space (i.e., elements uniformly chosen at random) look like, providing us with a ‘statistical’ description of a finite group or of a large homogeneous space. Our next result makes this idea precise by giving us a bound on the error resulting from using the random walk at stage  $N$  to estimate the true average of a function on  $G$ . The theorem provides a key application for the estimates of total variation obtained in Theorem 1.1.

**Theorem 1.4.** *Assume  $Z_1, \dots, Z_M$  are  $M$  independent copies of the  $N$ -th stage of the random walk on  $G$  defined by  $Q$ ,  $f : G \rightarrow \mathbb{C}$  is any function with  $\max_{g \in G} |f(g)| \leq 1$ , and  $\epsilon > 0$ . If  $\|Q^N - U\|_{TV} \leq \epsilon$ , then the following inequality holds*

$$\mathbb{P} \left\{ \left| \mathbb{E}_U(f) - \frac{1}{M} \sum_{i=1}^M f(Z_i) \right| \geq \epsilon \right\} \leq 2 \exp \left( - \frac{M (\epsilon - \|Q^N - U\|_{TV})^2}{2} \right).$$

A concrete application of Theorem 1.4 for estimating the average features of travelling salesman tours is discussed in Example 3.16.

In the second part of this article (Section 4), we introduce a generalization of the random walk model. The random walk model for card shuffling has a strong assumption, namely that the probability distribution of allowed moves is assumed to be the same at every stage. While this may accurately describe the behaviour of a proficient card mixer, it may not be adequate for describing many real-life mixing behaviours. A more general approach would be to assume that the mixer has a collection of distributions  $Q_1, \dots, Q_m$  on  $G$  and uses them in some order  $w_1 w_2 \dots$  with  $w_i \in [m]$  to shuffle the deck (where the chosen order is perhaps unknown even to the mixer). We call these more complicated processes *switched random walks* by analogy with *switched dynamical systems* [AJ, JAPR], which motivated our definition. We ask the following basic questions about the switched random walk defined by a set of distributions  $Q_1, \dots, Q_m$ :

- (1) Does the deck get evenly mixed *regardless of the order in which the  $Q_i$ 's are used*? When the answer is yes, we say that the set  $\{Q_1, \dots, Q_m\}$  has the *adversarial mixing property*.
- (2) If  $\{Q_1, \dots, Q_m\}$  has the adversarial mixing property, then we would like quantitative estimates of how quickly mixing occurs in the *worst case*. In other words, we wish to estimate the maximum total variation distance between the distribution of the process after  $N$ -stages and the uniform distribution on the permutations of the deck.

The methods developed for random walks can sometimes be extended to the switched setting. For instance, Theorem 1.1 easily implies the following corollary, where  $X$  is any  $G$ -set and  $x_0 \in X$ .

**Corollary 1.5.** *Suppose  $Q_1, \dots, Q_m$  are probability distributions on  $G$  that are constant on conjugacy classes and let  $Q_i = \sum_{j=1}^k a_j^i \chi_j$  be their unique representations as sums of characters. For a word  $w = w_1 w_2 \dots w_N$  with  $w_i \in [m]$  let  $Q^{(w)}$  be the convolution  $Q^{(w)} := Q_{w_1} * \dots * Q_{w_N}$ . If  $q^w := \sum_{g \in G} Q^{(w)}(g) e_g$ , then the following inequality holds*

$$\|q^w \cdot e_{x_0} - \bar{u}_X\|_{TV}^2 \leq \frac{1}{4} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \dim(V_j) \left( \frac{|G|}{\dim(V_j)} \right)^{2N} \left( \prod_{i=1}^N a_j^i \right)^2.$$

The assumption that the  $Q_i$  are constant on conjugacy classes makes the dynamics much simpler because, via the Fourier transform, they are reduced to the problem of understanding the behaviour of products of commuting matrices. To study the general non-commutative case, we introduce the *Fourier joint spectral radius* of a set of distributions  $Q_1, \dots, Q_m$  on  $G$  relative to a  $G$ -set  $X$ , defined as

$$\bar{\omega}_X(Q_1, \dots, Q_m) := \max_{\rho_j \in \mathbb{C}X, \rho_j \neq \text{triv}} (\text{jsr}(\widehat{Q}_1(\rho_j), \dots, \widehat{Q}_m(\rho_j))),$$

where the maximum is taken over the irreducible representations  $\rho_j$  of  $G$  appearing with non-zero multiplicity in  $\mathbb{C}X$  and the symbol jsr denotes the joint spectral radius of a set of matrices (see Section 4.1 for preliminaries on joint spectral radii). Our next result proves that the Fourier spectral radius captures the asymptotic worst case behaviour of the total variation distance to the uniform distribution.

**Theorem 1.6.** *For a word  $w = w_1 w_2 \dots w_N$  with  $w_i \in [m]$  let  $Q^{(w)}$  be the convolution  $Q^{(w)} := Q_{w_1} * \dots * Q_{w_N}$ . If  $q^w := \sum_{g \in G} Q^{(w)}(g) e_g \in \mathbb{C}G$ , then the following equality holds for every  $G$ -set  $X$ ,*

$$\lim_{N \rightarrow \infty} \left( \max_{x_0, w: |w|=N} \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_{TV} \right)^{\frac{1}{N}} = \bar{\omega}_X(Q_1, \dots, Q_m),$$

where the maximum is taken over all words  $w$  of length  $N$  and all initial states  $x_0 \in X$ . Furthermore, determining whether a set of distributions  $Q_1, \dots, Q_m$  has the adversarial mixing property on  $X$  is equivalent to determining whether the inequality  $\bar{\omega}_X(Q_1, \dots, Q_m) < 1$  holds.

The computation of the jsr of a set of matrices is a rather difficult task and we expect this difficulty to also extend to Fourier jsrs. For instance, it is known that it is *undecidable* whether the jsr of a pair of matrices is at most one [BT] and it is not known whether checking if the strict inequality holds is decidable. It is therefore a question of much interest to devise methods for estimating joint spectral radii (even with the knowledge that they are bound to fail in some cases). Recent work by Ahmadi, de Klerk, and Hall [AdKH] gives a hierarchy of polynomial norms that can be used to produce a sequence of converging upper bounds to the jsr of a set of matrices. In the final section (Section 4.2) of this article we extend their results to norms on complex vector spaces that are expressible as Hermitian sums of squares, allowing us to estimate Fourier spectral radii via Hermitian semidefinite programming.

## 2. Preliminaries

### 2.1 Representation theory of finite groups

Throughout the article,  $G$  will denote a finite group. By a *representation* of  $G$  we mean a pair  $(V, \rho)$  where  $V$  is a finite-dimensional vector space over the complex numbers and  $\rho : G \rightarrow GL(V)$  is a group homomorphism. A *morphism* between the representations  $(V_1, \rho_1)$  and  $(V_2, \rho_2)$  is a linear map  $\psi : V_1 \rightarrow V_2$  with the property that  $\psi \circ \rho_1(g) = \rho_2(g) \circ \psi$  for every  $g \in G$ . A subspace  $W \subseteq V$  is an *invariant subspace* if  $\rho(g)(W) \subseteq W$  for all  $g \in G$ . A representation  $(V, \rho)$  is *irreducible* if its only invariant subspaces are 0 and  $W$ . An *invariant inner product* on a representation  $V$  is a Hermitian inner product satisfying  $\langle \rho(g)(u), \rho(g)(v) \rangle = \langle u, v \rangle$  for all  $g \in G$  and  $u, v \in V$ . Throughout the article, we will use the following fundamental facts about complex representations of finite groups (see [[FH], Chapter 1] for proofs):

- (1) Every finite group  $G$  has finitely many pairwise non-isomorphic irreducible representations, which we will denote by  $V_1, \dots, V_k$ .

- (2) The irreducible representations are the building blocks of all representations in the sense that for any representation  $(\Lambda, \rho)$  there is an isomorphism of representations

$$\Lambda \cong \bigoplus_{j=1}^k V_j^{m_j},$$

where the integers  $m_j$ , called multiplicities, are uniquely specified. We write  $m(V_j, V) := m_j$ .

- (3) Every irreducible representation has an invariant inner product, unique up to multiplication by a positive real number, and we fix a basis  $B_j$  for each  $V_j$ , orthonormal with respect to this product. In this basis, the matrices  $[\rho(g)]_{B_j}$  are unitary. If  $\langle \cdot, \cdot \rangle$  is an invariant inner product on a representation  $\Lambda$ , then there is an orthonormal basis for  $\Lambda$ , compatible with the isomorphism in (2), with respect to which the maps of the representation are simultaneously block-diagonal of the form

$$[\rho_\Lambda(g)]_B = \bigoplus_{j=1}^k \left( [\rho_{V_j}(g)]_{B_j} \otimes I_{m_j} \right),$$

where  $I_{m_j}$  is the  $m_j \times m_j$  identity matrix.

- (4) The *character* of a representation  $V$  is a function  $\chi_V : G \rightarrow \mathbb{C}$  given by  $\chi_V(g) = \text{Tr}(\rho(g))$ . Characters are constant functions in the conjugacy classes of  $G$  and the characters of the irreducible representations  $V_j$  form an orthonormal basis for such functions (under the inner product  $\langle f, h \rangle := \sum_{g \in G} f(g)\overline{h(g)} / |G|$ ).

**2.2 The group ring and the Fourier transform**

The *group algebra* of  $G$ , denoted by  $\mathbb{C}G$ , is the complex vector space with basis given by the symbols  $\{e_g : g \in G\}$ , endowed with the multiplication  $e_g \cdot e_h := e_{g \cdot h}$ , where the dot in the right-hand side expression is the product in  $G$ . This is an associative and generally non-commutative algebra of dimension  $|G|$ . If  $(\Lambda, \rho)$  is a representation of  $G$ , then there is a linear map  $\phi : \mathbb{C}G \rightarrow \text{Hom}(\Lambda, \Lambda)$  which sends  $\sum a_g e_g$  to the map sending  $w \in \Lambda$  to  $\sum a_g \rho(g)(w)$ . This map transforms the product in  $\mathbb{C}G$  into the composition of linear maps and makes  $\Lambda$  into a  $\mathbb{C}G$ -module. It is easy to see that there is a correspondence between  $\mathbb{C}G$ -modules and representations of  $G$ . In particular, the group algebra is itself a representation of  $G$  via left-multiplication by defining  $\rho_{\mathbb{C}G}(g)(e_h) = e_g \cdot e_h$ . The following are three very useful facts about this representation.

- (1) The representation  $\mathbb{C}G$  is isomorphic to the representation  $\mathbb{C}[G]$  defined as the collection of complex-valued functions  $f : G \rightarrow \mathbb{C}$  endowed with the contragradient action  $\rho^*(g)f(x) = f(g^{-1}x)$ . We will use this isomorphism throughout. It is given explicitly by mapping a function  $Q : G \rightarrow \mathbb{C}$  to the element  $q := \sum_{g \in G} Q(g)e_g$ .
- (2) If  $q_1$  and  $q_2$  are the elements of  $\mathbb{C}G$  corresponding to functions  $Q_1$  and  $Q_2$ , then their product  $q_1 q_2 \in \mathbb{C}G$  corresponds to the *convolution*  $Q_1 * Q_2$  of  $Q_1$  and  $Q_2$ , defined as

$$Q_1 * Q_2(h) = \sum_{g \in G} Q_1(hg^{-1})Q_2(g).$$

- (3) There is an isomorphism of representations  $\mathbb{C}G \cong \bigoplus_{j=1}^k V_j^{\dim(V_j)}$  and in particular  $m(V_j, \mathbb{C}G) = \dim(V_j)$ .

- (4) There is an isomorphism of algebras  $\phi : \mathbb{C}G \rightarrow \bigoplus_{j=1}^k \text{Hom}(V_j, V_j)$  called the *Fourier transform*, which is the map sending a function  $f$  to the map  $\bigoplus_{j=1}^k \hat{f}(\rho_j)$ , where

$$\hat{f}(\rho_j) := \sum_{g \in G} f(g)[\rho_j(g)]_{B_j}.$$

See [[FH], Exercise 3.32] for basic properties of the Fourier transform.

### 3. Random walks on homogeneous spaces and modules over group rings

A *homogeneous space* for  $G$  is a set  $X$  endowed with a transitive action of  $G$ . In this section, we study random walks induced on  $X$  by random walks on  $G$ . More precisely, any probability distribution  $Q$  on  $G$  and initial state  $x_0 \in X$  define a stochastic process  $(h_k)_{k \geq 1}$  on  $X$ , as follows:

- (1) Let  $g_1, g_2, \dots$  be a sequence of independent and identically distributed random elements of  $G$  having distribution  $Q$ .
- (2) Define the random variable  $h_j := g_j \dots g_2 g_1(x_0)$ .

There are two natural questions about the process  $h_N$ :

- (1) What is the distribution of  $h_N$ ?
- (2) How does the distribution of  $h_N$  vary as  $N$  grows? Since the action of  $G$  in  $X$  is transitive, it should be fairly common (for instance when  $Q$  assigns sufficiently large probability to all elements of  $G$ ) that the process mixes  $X$  evenly. More quantitatively we ask: What is the total variation distance between the distribution of  $h_N$  and the uniform distribution on  $X$ ?

We will address the questions above using the module  $\mathbb{C}X$  over the group ring  $\mathbb{C}G$  defined by an action, borrowing the maxim of modern commutative algebra of placing a greater emphasis on modules. The material is organized as follows: Section 3.1 introduces the basic theory, Section 3.2 contains the resulting convergence bounds and clarifies their relationship with previous work, Section 3.3 discusses random walks on tabloids, illustrating how tools from combinatorial representation theory can be used for estimating mixing rates. Finally, Section 3.4 discusses the application of mixing rates for obtaining ‘statistical’ descriptions of homogeneous spaces and a detailed analysis of the space of travelling salesman tours.

For use throughout the section, recall that the *total variation* distance between two probability distributions  $P$  and  $Q$  on  $X$  is given by

$$\|P - Q\|_{\text{TV}} := \max_{A \subseteq X} |P(A) - Q(A)| = \frac{1}{2} \sum_{x \in X} |P(x) - Q(x)|.$$

#### 3.1 Random walks and modules over the group ring

Let  $\mathbb{C}X$  be a vector space with basis given by the set of symbols  $S := \{e_x : x \in X\}$ . The action of  $G$  on  $X$  makes  $\mathbb{C}X$  into a  $\mathbb{C}G$ -module via the map  $\phi : \mathbb{C}G \rightarrow \text{Hom}(\mathbb{C}X, \mathbb{C}X)$  given by  $\phi(e_g)(e_x) := e_{g(x)}$ . We will denote this action by  $e_g \cdot e_x$ . The following proposition shows that the module structure can be used to compute the distributions of our random walks.

**Lemma 3.1.** *If  $Q$  is a probability distribution on  $G$  and  $t$  is a probability distribution on  $X$ , then the equality*

$$\sum_{x \in X} \mathbb{P}\{W = x\}e_x = q \cdot t$$

holds, where  $q := \sum_{g \in G} Q(g)e_g$ ,  $t := \sum_{x \in X} T(x)e_x$ , and  $W := g(z)$  is the random variable obtained by choosing  $g \in G$  and  $z \in X$  independently with distributions  $Q$  and  $T$ , respectively. In particular, the distribution of  $h_N$  is given by  $q^N \cdot e_{x_0} \in \mathbb{C}X$ .

**Proof.** The independence between  $g$  and  $z$  implies the following equality for any  $\alpha \in X$

$$\mathbb{P}\{W = \alpha\} = \sum_{x \in X} \sum_{g \in G: g(x) = \alpha} Q(g)T(x).$$

It follows that

$$\sum_{\alpha \in X} \mathbb{P}\{W = \alpha\}e_\alpha = \sum_{\alpha \in X} \sum_{x \in X} \sum_{g \in G: g(x) = \alpha} Q(g)H(x)e_\alpha = \sum_{x \in X} \sum_{g \in G} Q(g)H(x)e_{g(x)} = q \cdot t.$$

The last claim follows from the associativity relation  $(q_1q_2) \cdot h = q_1 \cdot (q_2 \cdot h)$ , which holds for all  $q_1, q_2 \in \mathbb{C}G$  and  $h \in \mathbb{C}X$  because  $\mathbb{C}X$  is a  $\mathbb{C}G$ -module. □

The previous lemma is useful because it allows us to use the representation theory of  $\mathbb{C}X$  to compute the probability distribution of  $h_N$ . Henceforth, we endow  $\mathbb{C}X$  with the Hermitian inner product which satisfies  $\langle e_x, e_y \rangle = \delta_{xy}$ . This product is invariant because  $G$  acts on  $X$  by permutations.

**Lemma 3.2.** *Let  $Q$  be any complex-valued function on  $G$  and  $q := \sum_{g \in G} Q(g)e_g$ . There exists an  $|X| \times |X|$  unitary matrix  $W$  such that*

$$W[\phi(q)]_S W^* = \bigoplus_{j=1}^k \widehat{Q}(\rho_j) \otimes I_{m(V_j, \mathbb{C}X)},$$

where  $\widehat{Q}$  denotes the Fourier transform of  $Q$ .

**Proof.** Since the inner product we defined in  $\mathbb{C}X$  is  $G$ -invariant, we can use it to construct an orthonormal basis  $B$  for  $\mathbb{C}X$  compatible with the decomposition of  $\mathbb{C}X = \bigoplus V_i^{m(V_i, \mathbb{C}X)}$  as a representation of  $G$ . Letting  $W$  be the change of basis matrix from the basis  $S = \{e_x : x \in X\}$  to the basis  $B$ , we see that  $W$  is unitary and that for every  $g \in G$  the equality

$$W[\rho_{\mathbb{C}X}(g)]_S W^* = \bigoplus_{j=1}^k [\rho_j(g)]_{B_j} \otimes I_{m(V_j, \mathbb{C}X)}$$

holds. Since  $\phi$  is linear, we conclude that

$$\begin{aligned} W\phi(q)W^* &= \sum_{g \in G} Q(g) \bigoplus_{j=1}^k [\rho_j(g)]_{B_j} \otimes I_{m(V_j, \mathbb{C}X)} \\ &= \bigoplus_{j=1}^k \left( \sum_{g \in G} Q(g)[\rho_j(g)]_{B_j} \right) \otimes I_{m(V_j, \mathbb{C}X)}, \end{aligned}$$

which agrees with the claimed formula by definition of Fourier transform. □

**Remark 3.3.** In order to use Lemma 3.2, it is extremely useful to be able to decompose  $\mathbb{C}X$  into irreducibles. This process is often simplified by the following two facts:

- (1) If  $x_0 \in X$  is any point and  $H \subseteq G$  is the stabilizer of  $x_0$ , then the permutation module  $\mathbb{C}X$  is isomorphic to the induced representation of  $G$  defined by the trivial representation of  $H$  [[FH], Example 3.13].



(2) In particular, the multiplicities with which the irreducible representations  $V_j$  appear in  $\mathbb{C}X$  are determined by the Frobenius Reciprocity Theorem [[FH], Corollary 3.20], which implies that

$$m(V_j, \mathbb{C}X) = \frac{1}{|H|} \sum_{h \in H} \chi_{V_j}(h).$$

**3.2 Mixing rates**

For any  $G$ -homogeneous space  $X$ , we let  $\bar{u} \in \mathbb{C}X$  be the element corresponding to the uniform distribution, that is,  $\bar{u} := \frac{1}{|X|} \sum_{x \in X} e_x$ . Note that  $\bar{u}_G \cdot e_{x_i} = \bar{u}$  for every  $x_i \in X$ . We endow  $\mathbb{C}X$  with the total variation norm  $\|h\|_{TV} := \frac{1}{2} \sum_{x \in X} |h(x)|$  and endow  $\text{Hom}(\mathbb{C}X, \mathbb{C}X)$  with the Frobenius norm  $\|A\|_{Fb} := \text{Tr}(AA^*)$ . We are now ready to prove Theorem 1.1, our main tool for establishing convergence estimates on homogeneous spaces. The key observation behind the proof is that even though the total variation norm is not unitarily invariant, it can be bounded on average by a unitarily invariant norm. This allows us to choose the convenient orthonormal basis for our operators by means of Lemma 3.2.

**Proof of Theorem 1.1.** The equality

$$\sum_{x \in X} \|q \cdot e_x - \bar{u}\|_2^2 = \|\phi(q) - \phi(\bar{u}_G)\|_{Fb}^2$$

holds, since both sides equal the sum of the squares of the entries of the matrix  $\phi(q) - \phi(\bar{u}_G)$ . Since the Frobenius norm is unitarily invariant, we can compute the right-hand side of this equality in any orthonormal basis. In particular, using the basis from Lemma 3.2 yields

$$\|\phi(q) - \phi(\bar{u}_G)\|_{Fb}^2 = \sum_{j=1}^k m(V_j, \mathbb{C}X) \|\widehat{Q - U}(\rho_j)\|_{Fb}^2,$$

where  $U$  is the uniform distribution on  $G$ . For any probability distribution  $Q$  on  $G$ , we know that  $\widehat{Q}(\text{triv}) = 1$  and for the uniform distribution  $U$  we know that  $\widehat{U}(\rho_j) = 0$  for all  $\rho_j \neq \text{triv}$ . We thus conclude that the following equality holds

$$\sum_{x \in X} \|q \cdot e_x - \bar{u}\|_2^2 = \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\widehat{Q}(\rho_j)\|_{Fb}^2. \tag{1}$$

The Cauchy–Schwarz inequality and the fact that  $\|\bullet\|_2 \leq \|\bullet\|_1$  imply that for every  $x \in X$  we have

$$\frac{1}{4} \|q \cdot e_x - \bar{u}\|_2^2 \leq \|q \cdot e_x - \bar{u}\|_{TV}^2 \leq \frac{1}{4} |X| \|q \cdot e_x - \bar{u}\|_2^2. \tag{2}$$

Averaging the inequalities in (2) over  $X$ , we obtain

$$\frac{1}{4|X|} \sum_{x \in X} \|q \cdot e_x - \bar{u}\|_2^2 \leq \frac{1}{|X|} \sum_{x \in X} \|q \cdot e_x - \bar{u}\|_{TV}^2 \leq \frac{1}{4} \sum_{x \in X} \|q \cdot e_x - \bar{u}\|_2^2.$$

Combining these inequalities with identity (1) we complete the proof of the two inequalities in our main claim.

Furthermore, if  $\|q \cdot e_x\|_2^2$  is independent of  $x$ , then the same is true of  $\|q \cdot e_x - \bar{u}\|_2^2$ , since this quantity equals  $\|q \cdot e_x\|_2^2 - 2\langle q \cdot e_x, \bar{u} \rangle + \|\bar{u}\|_2^2$ , which is independent of  $x$  because  $\langle q \cdot e_x, \bar{u} \rangle =$



$1/|X|$  for all  $x$ . As a result, for every  $x \in X$  we have

$$\|q \cdot e_x - \bar{u}\|_2^2 = \frac{1}{|X|} \sum_{y \in X} \|q \cdot e_y - \bar{u}\|_2^2$$

and we can replace the leftmost and rightmost terms in (2) for averages, yielding

$$\frac{1}{4|X|} \sum_{y \in X} \|q \cdot e_y - \bar{u}\|_2^2 \leq \|q \cdot e_x - \bar{u}\|_{TV}^2 \leq \frac{1}{4} \sum_{y \in X} \|q \cdot e_y - \bar{u}\|_2^2.$$

This result, combined with (1), completes the proof. □

**Remark 3.4.** There are two cases where the condition that  $\|q \cdot e_x\|_2$  be independent of  $X$  occurs automatically because  $q \cdot e_{x_2}$  is obtained from  $q \cdot e_{x_1}$  by rearranging the order of the coefficients for every  $x_1, x_2 \in X$ . This happens

- (1) If  $X = G$  because the set of coefficients of  $q \cdot e_x$  for any  $x$  is exactly the set of values of  $Q$ .
- (2) If  $Q$  is constant on conjugacy classes of  $G$ . This holds because the equality  $q \cdot e_{x_2} = \sum_{g \in G} Q(g)e_{g(x_2)} = \sum_{y \in X} c_y e_y$  implies that  $c_y$  equals the sum of the  $Q$ -probabilities of the set  $A$  of  $g \in G$  with  $g(x_2) = y$ . It follows that if  $rx_1 = x_2$ , then the conjugates  $r^{-1}gr$  for  $g$  in  $A$  are the group elements which map  $x_1$  to  $r^{-1}y$ . Since  $Q$  is conjugation-invariant, we have that  $c_y$  is the coefficient of  $r^{-1}y$  in  $q \cdot e_{x_1}$ . We conclude that  $q \cdot e_{x_1}$  is a permutation of  $q \cdot e_{x_2}$ .

**Remark 3.5.** More generally, it can be shown that if  $T$  is a subgroup of  $G$  which acts transitively on  $X$  and  $Q$  is a probability distribution which is constant on the conjugacy classes of  $T$ , then:

- (1) The convolution  $Q^{(N)}$  is also constant on the conjugacy classes of  $T$  and
- (2)  $\|q \cdot e_x\|_2$  is independent of  $x \in X$ .

So in this case the bound from Theorem 1.1 holds for every initial state  $x_0 \in X$ .

**Remark 3.6.** The proof of Theorem 1.1 uses the same approach as that of the celebrated Upper Bound Lemma [[D], pag. 24] for random walks on a group  $G$ . The lemma was introduced in the early 80s and is still a key tool in the state-of-the-art analysis of Markov chains (see for instance [B, BN]).

The use of averaging allows us to extend the result to arbitrary homogeneous spaces, increasing the range of applications, and to improve the coefficients in the inequality by replacing  $\dim(V_j)$  with the typically smaller  $m(V_j, \mathbb{C}X)$ .

**Remark 3.7.** Theorem 1.1 should be compared with the Upper Bound Lemma for homogeneous spaces (UBLH) from [[D], pag. 53]. The UBLH applies to random walks defined by distributions  $Q$  on  $G$  which are right-invariant under the subgroup  $H \leq G$  which stabilizes  $x_0 \in X$  (i.e., the distributions are forced to satisfy  $Q(gh) = Q(g)$  for all  $h \in H$ ) and the bound depends on the Fourier transforms of the induced distributions on  $X$  and not on the Fourier transforms of the original distributions. Due to the restriction to right- $H$ -invariant distributions and the presence of Fourier transforms of induced distributions, the UBLH has a smaller range of applicability than Theorem 1.1. On the other hand, the UBLH gives inequalities valid for every initial state  $x_0$ , albeit at the cost of using larger constants than the  $m(V_j, \mathbb{C}X)$  of Theorem 1.1. It follows that the Theorems are strictly incomparable and that it may be more convenient to use one or the other, depending on the intended application.

**Proof of Corollary 1.2.** Theorem 1.1 provides us with lower and upper bounds for the average of the squared total variation over the starting points of  $X$ . The average of a set of real numbers is at

least the smallest one and at most the largest one, proving the existence of the good and bad initial states  $r$  and  $s$  in  $X$ . □

**Remark 3.8.** Theorem 1.1 and Markov’s inequality imply that the relation

$$\left| \left\{ x \in X : \|q^N e_x - \bar{u}\|_{TV} \geq \alpha \right\} \right| \leq \frac{|X|}{4\alpha^2} \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\hat{Q}(\rho_j)^N\|_{Fb}^2$$

holds for every  $\alpha > 0$ , allowing us to prove that most (and even all, when the right-hand side is  $< 1$ ) initial states mix well. In the special case where  $Q$  is right-invariant under the stabilizer of a point  $x_0 \in X$ , this inequality is weaker than the UBLH, which provides a bound for every initial state. However, the inequality above applies to arbitrary (not necessarily right-invariant) distributions  $Q$ .

**3.3 Example: Random walks on tabloids**

Fix positive integers  $n, a, b$  with  $a \geq b$  and  $a + b = n$ . Suppose we have a set of  $n$  of cards and that these are placed face up forming a single row. We permute the cards by independently sampling permutations according to a fixed distribution  $Q$  on  $S_n$ , and acting with these permutations on the row of cards. After  $N$  stages, we split the row of cards into two disjoint sets  $A$  and  $B$ , consisting of the first  $a$  cards and the remaining  $b$  cards, respectively (reading the row of cards from left to right). We ask: How near-uniform is the set  $A$ , or equivalently, how random is the set partition  $(A, B)$ ? In this section, we define random walks on tabloids, which generalize this problem, and discuss tools suitable for their analysis.

**3.3.1 Preliminaries: partitions, tableaux, and tabloids**

A partition of a positive integer  $n$  is a nonincreasing sequence  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$  with  $n = \sum \lambda_i$ . Partitions are partially ordered by the *dominance ordering*, defined as  $\lambda \leq \mu$  if  $\sum_{i \leq j} \lambda_i \leq \sum_{i \leq j} \mu_i$  for all  $j = 1, \dots, n$ . A *tableaux* with shape  $\lambda$  is a finite collection of boxes, arranged in left-justified rows of sizes  $\lambda_1, \dots, \lambda_k$ , and containing the integers  $1, \dots, n$  without repetitions. Two tableaux of the same shape  $\lambda$  are *row-equivalent* if the sets of elements in each of their corresponding rows coincide. A *tabloid* is a row-equivalence class of tableaux.

**Example 3.9.** The sequence  $\lambda := (3, 3, 2, 1)$  is a partition of 9. The following two tableaux are row-equivalent and therefore members of the same tabloid.

1	2	3	2	1	3
4	5	6	5	6	4
7	8		8	7	
9			9		

This tabloid is encoded by the ordered set partition  $(\{1, 2, 3\}, \{4, 5, 6\}, \{7\}, \{8\})$ , which keeps track of the set of elements in each row.

A *generalized tableaux* of shape  $\lambda$  is a tableaux  $T$  of shape  $\lambda$  filled with elements of  $\{1, \dots, n\}$  where repetitions are allowed. The content of such a tableaux is a vector  $(\mu_1, \dots, \mu_n)$ , where  $\mu_i$  is the number of copies of the integer  $i$  appearing in  $T$  for  $i = 1, \dots, n$ . A *semi-standard tableaux* is a generalized tableaux where the labels are weakly increasing along rows and strictly increasing along columns.

**Example 3.10.** If we fix the content  $\mu = (2, 2, 1)$  then the set of all semi-standard tableaux of shapes  $(3, 2)$  and  $(4, 1)$  with content  $\mu$  are given by

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 2 \\ \hline 2 & 3 & \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline 1 & 1 & 3 \\ \hline 2 & 2 & \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 2 \\ \hline 3 & & & \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 3 \\ \hline 2 & & & \\ \hline \end{array}$$

3.3.2 Conjugation-invariant distributions and random walks on tabloids

Continuing with the example introduced at the beginning of Section 3.3, any partition  $\lambda$  of  $n$  having  $k$  parts allows us to partition our row of  $n$  cards into  $k$  sets  $(A_1, \dots, A_k)$  where  $A_1$  consists of the first  $\lambda_1$  cards, the set  $A_2$  consists of the next  $\lambda_2$  cards (those in positions  $\lambda_1 + 1, \dots, \lambda_1 + \lambda_2$  along the row),  $A_3$  consists of the next  $\lambda_3$  cards, etc. We can then ask: how near-uniform is the resulting set partition  $(A_1, \dots, A_k)$  after  $N$  stages of our random walk? In this section, we address this problem when the distribution  $Q$  is constant on conjugacy classes by applying the tools introduced in the article. Our main result is Corollary 3.14, which gives bounds on the mixing rate of the process described in the first paragraph, that is, when  $\lambda$  is any partition with at most two parts.

To begin the analysis, note that if  $X$  denotes the set of tabloids of shape  $\lambda$  with the natural action of  $S_n$  by permutations (see Section 3.3.1 for basic definitions), then the process above coincides with the random walk on the homogeneous space  $X$  defined by the probability distribution  $Q$  on  $S_n$ . The corresponding permutation module  $\mathbb{C}X$  is well-known and plays a distinguished role in Young’s construction of the irreducible representations of the group  $S_n$  (see [[S], Chapter 2]). It is common in the literature to refer to these modules as *the permutation modules*  $M^\lambda$  and we will do so throughout this section. The following theorem [[S], Theorem 2.11.2] describes their structure, where  $S^\lambda$  denotes the irreducible representation of  $S_n$  corresponding, via Young’s construction, to the partition  $\lambda$  (see [[S], Definition 2.3.4] for an explicit description of  $S^\lambda$ ).

**Lemma 3.11** (Young’s rule). *Let  $\mu$  be a partition of  $n$ . The following isomorphism of representations holds*

$$M^\mu \cong \bigoplus_{\lambda: \lambda \geq \mu} (S^\lambda)^{\oplus K_{\lambda, \mu}},$$

where the sum runs over the partitions  $\lambda$  of  $n$  with  $\lambda \geq \mu$  in the dominance ordering of partitions and  $K_{\lambda, \mu}$  is the Kostka number of  $(\lambda, \mu)$ , that is, the number of semi-standard tableaux of shape  $\lambda$  and content  $\mu$ .

In order to understand the behaviour of Markov chains defined by conjugation-invariant probability distributions, we need to express such distributions as sums of characters. To this end, we will use orthogonality of characters together with Frobenius’ remarkable character formula [[FH], Theorem 4.10], which gives a combinatorial description of the characters of  $S^\lambda$ . Recall that the conjugacy class of a permutation  $\sigma \in S_n$  is specified by its cycle type, namely the sequence  $(n_1, \dots, n_n)$ , where  $n_j$  equals the number of  $j$ -cycles appearing in the unique decomposition of  $\sigma$  as a product of disjoint cycles.

**Lemma 3.12** (Frobenius character formula). *If  $\lambda = \lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0$  is a partition of  $n$  then the value of the character of  $S^\lambda$  in the conjugacy class  $(n_1, \dots, n_n)$  is given by the coefficient of the monomial  $x^{\ell(\lambda)}$  in the polynomial  $\Delta \cdot P_n \in \mathbb{C}[x_1, \dots, x_n]$ , where*

$$P_n(x) := \prod_{j=1}^n (x_1^j + \dots + x_n^j)^{n_j}, \Delta := \prod_{1 \leq i < j \leq n} (x_i - x_j),$$

and  $\ell(\lambda) = (\lambda_1 + n - 1, \lambda_2 + n - 2, \dots, \lambda_k + n - k, \dots, \lambda_{n-1} + 1, \lambda_n)$ .

**Lemma 3.13.** *Let  $\lambda = a \geq b$  be a partition of  $n$ . If  $C_k$  denotes the conjugacy class of a  $k$ -cycle in  $S_n$ , then*

$$\chi_{S^\lambda}(C_k) = \left[ \binom{n-k}{a} - \binom{n-k}{a+1} \right] + \left[ \binom{n-k}{b} - \binom{n-k}{b-1} \right]$$

and furthermore,

$$\dim(S^\lambda) = \binom{n}{a} - \binom{n}{a+1} = \binom{n}{b} - \binom{n}{b-1}.$$

**Proof.** Since  $\Delta$  is the determinant of the Vandermonde matrix with  $i, j$  entry given by  $x_{n-i+1}^j$  we know that

$$\Delta = \sum_{\sigma \in S_n} \text{sgn}(\sigma) x_n^{\sigma(1)-1} \dots x_1^{\sigma(n)-1}$$

and in particular, only two terms have exponents which are componentwise smaller than  $\ell(\lambda) = (a+n-1, b+n-2, n-3, n-4, \dots, 1, 0)$ , namely

$$x_1^{n-1} x_2^{n-2} x_3^{n-3} \dots x_{n-2}^2 x_{n-1}^1 x_n^0 \text{ and } -x_1^{n-2} x_2^{n-1} x_3^{n-3} \dots x_{n-2}^2 x_{n-1}^1 x_n^0.$$

Since  $P_{C_k}(x) = A(x)(x_1^k + \dots + x_n^k)$ , where  $A(x) = (x_1 + \dots + x_n)^{n-k}$ , we conclude from Frobenius' character formula and the observation from the previous paragraph that  $\chi_{S^\lambda}(C_k)$  is given by

$$[P(x)]_{(a,b)} - [P(x)]_{(a+1,b-1)},$$

where we have removed the  $n-2$  trailing zeroes in our notation for exponents. Since  $P(x)$  factors, this quantity equals

$$[A(x)]_{(a-k,b)} + [A(x)]_{(a,b-k)} - ([A(x)]_{(a-k+1,b-1)} + [A(x)]_{(a+1,b-k-1)}).$$

Each of these coefficients can be computed by the multinomial theorem, yielding  $\binom{n-k}{b} + \binom{n-k}{a} - (\binom{n-k}{b-1} + \binom{n-k}{a+1})$ . Similarly, the dimension of  $S^\lambda$  is given by the value of its character at the identity element, given by

$$[(x_1 + \dots + x_n)^n]_{(a,b)} - [(x_1 + \dots + x_n)^n]_{(a+1,b-1)},$$

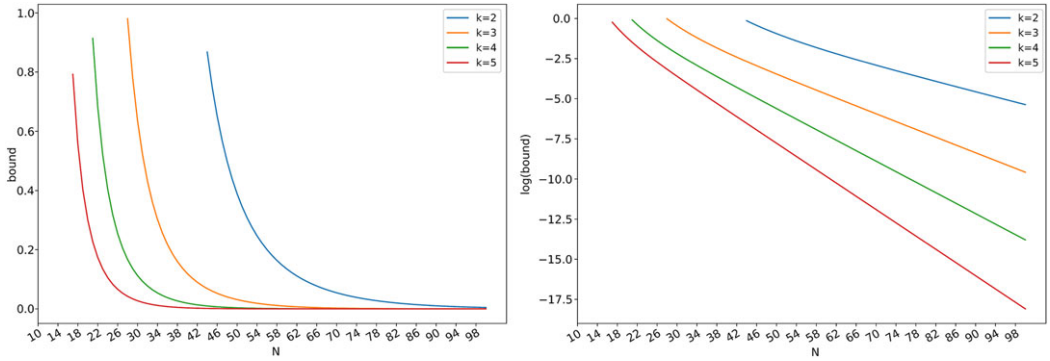
proving the claim. □

The following corollary estimates the rate of convergence to the uniform distribution of the pure-cycle random walk on the set of disjoint pairs  $(A, B)$  of sizes  $a$  and  $b$ , respectively, with  $A \cup B = [n]$ . An explicit illustration of these upper bounds (for  $n = 52$  and  $\lambda = 26 \geq 26$  is shown in Figure 1).

**Corollary 3.14.** *Let  $Q$  be the probability distribution which samples  $k$ -cycles uniformly. If  $\mathbb{C}X$  is the permutation module corresponding to a partition  $\lambda = a \geq b$  with  $a + b = n$  and  $q^{(N)} := \sum_{g \in G} Q^N(g) e_g \in \mathbb{C}G$ , then for any  $x_0 \in X$ , the quantity  $\|q^{(N)} \cdot u_{x_0} - \bar{u}_X\|_{TV}^2$  is bounded above by*

$$\frac{1}{4} \sum_{t=1}^b \frac{\left( \left[ \binom{n-k}{n-t} - \binom{n-k}{n+1-t} \right] + \left[ \binom{n-k}{t} - \binom{n-k}{t-1} \right] \right)^{2N}}{\left( \binom{n}{t} - \binom{n}{t-1} \right)^{2N-1}}.$$

**Proof.** By Lemma 3.11, the decomposition of  $\mathbb{C}X$  into  $S_n$ -irreducibles is given by  $\mathbb{C}X = \bigoplus_{t=0}^b S^{(a+b-t,t)}$  and in particular, no representation appears with multiplicity greater than one. For notational convenience we write  $V_t := S^{(a+b-t,t)}$  throughout this proof. Since the distribution



**Figure 1.** Upper bound for total variation to uniformity for  $\lambda : 26 \geq 26$  with  $n = 52$  (the number of cards on a regular deck) and  $k = 2, 3, 4, 5$  (y axis in standard (left) and logarithmic (right) scales). The set  $X$  has around  $4.9 \times 10^{14}$  elements.

$Q$  is constant on conjugacy classes, it can be written uniquely as a sum of irreducible characters, and we determine its coefficients  $a_t$  with respect to the characters  $\chi_{V_t}$  for  $t = 0, \dots, b$ . By orthogonality of characters we have

$$a_t = \langle Q, \chi_{V_t} \rangle = \frac{1}{|S_n|} \sum_{g \in G} Q(g) \chi_{V_t}(g) = \frac{1}{|S_n|} \chi_{V_t}(\tau),$$

where  $\tau$  is any  $k$ -cycle. By Lemma 3.13 we know that

$$a_t = \frac{1}{|S_n|} \left( \left[ \binom{n-k}{a+b-t} - \binom{n-k}{a+b+1-t} \right] + \left[ \binom{n-k}{t} - \binom{n-k}{t-1} \right] \right)$$

and that  $\dim(V_t) = \binom{n}{t} - \binom{n}{t-1}$ . The claim now follows from Corollary 1.3. □

**Remark 3.15.** The previous corollary shows that Theorem 1.1 can be applied more generally than the Upper Bound Lemma on  $S_n$ , since selecting a transposition uniformly at random *does not* mix  $S_n$  (because only even transpositions can be reached in even stages), while it does converge to the uniform distribution on the homogeneous space  $X$ .

### 3.4 A concentration inequality

In this section we prove Theorem 1.4 and illustrate its applicability by analysing random walks on travelling salesman tours.

**Proof of Theorem 1.4.** Let  $\mathbb{E}_{Q^N}(\bullet)$  denote the expected value with respect to the distribution of the process after  $N$  stages. By the triangle inequality and the definition of total variation, the following inequalities hold

$$\begin{aligned} \left| \mathbb{E}_U(f) - \frac{1}{M} \sum_{i=1}^M f(Z_j) \right| &\leq \left| \mathbb{E}_U(f) - \mathbb{E}_{Q^N}(f) \right| + \left| \mathbb{E}_{Q^N}(f) - \frac{1}{M} \sum_{i=1}^M f(Z_j) \right| \\ &\leq \|Q^N - U\|_{TV} + \left| \mathbb{E}_{Q^N}(f) - \frac{1}{M} \sum_{i=1}^M f(Z_j) \right| \end{aligned}$$

and we conclude that

$$\mathbb{P} \left\{ \left| \mathbb{E}_U(f) - \frac{1}{M} \sum_{i=1}^M f(Z_i) \right| \geq \epsilon \right\} \leq \mathbb{P} \left\{ \left| \mathbb{E}_{Q^N}(f) - \frac{1}{M} \sum_{i=1}^M f(Z_i) \right| \geq \beta \right\},$$

where  $\beta = \epsilon - \|Q^N - U\|_{TV}$ . Since  $\beta > 0$ , the right-hand side is bounded by  $2 \exp(-M\beta^2/2)$  from Hoeffding’s inequality for bounded random variables [[H], Theorem 2], proving the break claim.  $\square$

**Example 3.16.** Suppose  $X$  is the set of tours through a fixed set of cities  $1, \dots, n$  and let  $\ell(x)$  be the total distance travelled in tour  $x$ . The set  $X$  has  $(n - 1)!$  elements and finding a tour of least total length is the classical travelling salesman problem (TSP), of much interest in combinatorial optimization. It is often desirable to know the *average cost*  $\mathbb{E}_U(f)$  of functions  $f$  among all possible tours. This is especially challenging if the function depends nonlinearly on the tour. When  $f$  is bounded, Theorem 1.4 gives us a natural approach for obtaining such estimates via simulations, together with error bars on such estimates. In this example we give some nonlinear functions whose averages are of interest for the TSP and show some combinatorial techniques that can be used for obtaining the necessary total variation bounds.

Motivated by the simulated annealing approach [[M], Section 4.4.3] for solving the TSP, define for a fixed real number  $\beta$  the probability distribution  $\pi_\beta(x)$  on  $X$ , given by the formula  $\pi_\beta(x) = e^{-\beta\ell(x)}/C_\beta$ , where  $C_\beta(x) := \sum_{x \in X} \pi_\beta(x)$  is the associated partition function. Note that  $\pi_\beta(x)$  assigns higher probability to shorter tours and it is easy to see that in the limit  $\beta \rightarrow \infty$  the  $\pi_\beta(x)$ -average length

$$\bar{\ell}_\beta := \sum_{x \in X} \ell(x)\pi_\beta(x)$$

converges to the length of the shortest tour.

To estimate  $\bar{\ell}_\beta$  define the functions  $a_\beta(x) := \ell(x)e^{-\beta x}$ ,  $c_\beta(x) = e^{-\beta\ell(x)}$  and the numbers  $A_\beta := \mathbb{E}_U[a(x)]$ ,  $C_\beta := \mathbb{E}_U[b(x)]$ , and note that  $\bar{\ell}_\beta = A_\beta/C_\beta$ . It is natural to estimate  $A_\beta$  and  $C_\beta$  via their sample averages

$$\hat{A}_\beta := \frac{1}{M} \sum_{j=1}^M a_\beta(Z_j) \quad \hat{C}_\beta := \frac{1}{M} \sum_{j=1}^M c_\beta(Z_j)$$

which are easily computable with simulations. The convexity of the function  $y/x$  in the positive orthant, together with Theorem 1.4, now imply the following corollary, which gives error bounds on these estimates. In the expressions below,  $D$  is any upper bound on the length of tours (for instance the sum of the  $n$  largest pairwise distances among cities).

**Corollary 3.17.** *Let  $\epsilon, \eta > 0$  be real numbers. Choose  $N$  large enough so that*

$$\|Q^{(N)} - U\|_{TV} < \frac{\epsilon e^{-2\beta D}}{D^2}$$

and choose  $M$  large enough so that

$$2 \exp \left( - \frac{M \left( \frac{\epsilon e^{-2\beta D}}{D^2} - \|Q^{(N)} - U\|_{TV} \right)^2}{2} \right) < \eta.$$

If we use  $M$  independent samples of the  $N$ -th stage of the random walk defined by the distribution  $Q$  to compute the estimates  $\hat{A}_\beta, \hat{C}_\beta$ , then the following inequalities hold with probability at least  $1 - 2\eta$

$$-2\epsilon + \frac{\hat{A}_\beta}{\hat{C}_\beta} \leq \bar{\ell}_\beta = \frac{A_\beta}{C_\beta} \leq \frac{\hat{A}_\beta}{\hat{C}_\beta} + 2\epsilon.$$

In order to use the previous corollary one needs good total variation bounds. For conjugation-invariant probability distributions  $Q$ , such bounds follow from Corollary 1.3 provided we have good estimates of the involved multiplicities. We now illustrate such multiplicity calculations assuming, for simplicity, that  $n$  is a prime number.

The stabilizer of the cycle  $(1, 2, \dots, n)$  is the cyclic subgroup  $\mathbb{Z}/n\mathbb{Z}$  generated by the cycle  $(1, 2, \dots, n)$ . By primality of  $n$ , this subgroup is contained in only two conjugacy classes of  $S_n$ , namely that of the identity and that of a single  $n$ -cycle  $c$ .

By Remark 3.3, we conclude that for any partition  $\lambda$  of  $n$  we have

$$m(S^\lambda, \mathbb{C}X) = \frac{1}{n} (\dim(S^\lambda) + (n - 1)\chi_{S^\lambda}(c)).$$

We then use the hook length formula and the Murnaghan-Nakayama rule to compute the terms in the right-hand side of the expression above. Recall that a *hook* in a partition  $\lambda$  is a choice of a box in the corresponding Young diagram together with all boxes to the right in the same row and all boxes below in the same column. The total number of boxes in a hook is called its length. A *skew-hook* in a partition  $\lambda$  is a connected collection of boundary boxes in its Young diagram whose removal leaves a smaller Young diagram. The total number of boxes in a skew-hook is called its length. There is a bijective correspondence between hooks and skew-hooks which preserves length (illustrated in [[FH], Problem 4.45]).

The hook length allows us to compute the dimensions of the  $S^\lambda$  using the remarkable hook length formula [[FH], Formula 4.12]

$$\dim(S^\lambda) = \frac{n!}{\prod_h \text{Hook Length}(h)},$$

where the product is taken over all hooks contained in  $\lambda$ . Furthermore, the value of a character  $\chi_{S^\lambda}(g)$  can be computed from the Murnaghan-Nakayama rule, which states [[FH], Problem 4.45] that if  $g = jh$ , where  $j$  is an  $m$ -cycle and  $h$  is a permutation disjoint from  $j$ , then

$$\chi_{S^\lambda}(g) = \sum (-1)^{r(\mu)} \chi_{S^\mu}(h).$$

Above, the sum is taken over all partitions  $\mu$  of  $n - m$  obtained from  $\lambda$  by removing a skew-hook of length  $m$  and  $r(\mu)$  is the number of vertical steps in the skew-hook.

In our special case,  $g = c$  is a cycle of length  $n$  and  $\lambda$  is a partition of  $n$ . Therefore, the Murnaghan-Nakayama rule implies that  $\chi_{S^\lambda}(c) = 0$  unless  $\lambda$  contains at least one hook of length  $n$ . This occurs only if  $\lambda$  is itself a hook, meaning that  $\lambda$  is a partition of the form  $1 + a, 1^b$ , where  $a$  and  $b$  are positive integers with  $1 + a + b = n$ . For such partitions the sum above becomes

$$\chi_{S^\lambda}(c) = (-1)^b + (-1)^{b-1} + \dots + (-1)^0,$$

so the absolute value of  $\chi_{S^\lambda}(c)$  is bounded by one. Furthermore, the hook length formula gives  $\dim(S^\lambda) = \binom{n-1}{b}$ , and we conclude that the inequality

$$\frac{m(S^\lambda, \mathbb{C}X)}{\dim(S^\lambda)} \leq \frac{1}{n} \left( 1 + \frac{n-1}{\binom{n-1}{a}} \right)$$

holds for every partition. In particular, for every representation other than the trivial and the sign representations (which occur when  $a = n$  and  $b = 0$  respectively), we have

$$\frac{m(S^\lambda, \mathbb{C}X)}{\dim(S^\lambda)} \leq \frac{1}{n} \left( 1 + \frac{n-1}{\binom{n-1}{a}} \right) \leq \frac{2}{n}.$$

For the trivial and alternating representations, direct calculation shows that the equalities  $m(S^1, \mathbb{C}X) = m(S^{(n)}, \mathbb{C}X) = 1$  hold. These observations allow us to leverage existing mixing rate



estimates for conjugation-invariant random walks on  $S_n$  to obtain improved mixing rate estimates for  $X$ . Applying the well-known transposition walk estimates from [[D], Theorem 5], we obtain the following corollary, which provides the total variation bounds needed for Corollary 3.17.

**Corollary 3.18.** *For a prime number  $n$ , let  $X$  be the set of tours of cities  $1, \dots, n$ , and let  $Q$  be the probability distribution on  $S_n$  given by  $Q(id) = 1/n$  and  $Q(\tau) = 2/n^2$  for all transpositions  $\tau$ . If  $q = \sum_{g \in S_n} Q(g)e_g$ , then the following inequality holds for every initial state  $x_0 \in X$  and every stage  $N = \frac{n \log(n)}{2} + cn$  for  $c > 0$*

$$\|q^{(N)} \cdot x_0 - \bar{u}\|_{TV}^2 \leq \frac{2a^2 e^{-4c}}{n} + \frac{(2/n - 1)^{2N}}{4},$$

where  $a$  is a universal constant (i.e., independent of the values of  $n$  and  $N$ ).

**Remark 3.19.** When using Corollary 3.17, one is given  $\epsilon, \beta > 0$ ,  $D$  equals the sum of the  $n$  largest pairwise distances among cities, and one wishes to choose  $N$  so that  $\|Q^{(N)} - U\|_{TV} \leq \epsilon e^{-2\beta D}/D^2$ . Note that guaranteeing this bound via Corollary 3.18 requires choosing  $N = \frac{n \log(n)}{2} + cn$  for a sufficiently large multiple  $c = Kn$  of  $n$ . More precisely, the bound is satisfied when there is a quadratic dependence of  $N$  on  $n$  (a similar statement holds for  $M$ ).

#### 4. Switched random walks

If  $Q_1, \dots, Q_m$  are probability distributions on  $G$  and  $X$  is a  $G$ -set, then a choice of initial state  $x_0 \in X$  determines a dynamical system which we call a *switched random walk* on  $X$ . The switched random walk is a family of random variables  $h_k^{(w)}$ , as  $w$  ranges over all words  $w_1 w_2 \dots$  with  $w_i \in [m]$  and  $k \in \mathbb{N}$ , which describe the state of our system at time  $k$  when the mixing strategies  $Q_i$  have been switched according to the word  $w$  (i.e., where strategy  $Q_{w_i}$  has been used at stage  $i$  for  $i \leq k$ ). More formally, the switched random walk starting at  $x_0 \in X$  is constructed as follows:

- (1) Sample  $[m] \times \mathbb{N}$  independent elements of  $G$  with  $g_{i,j}$  having distribution  $Q_i$ .
- (2) For each word  $w_1 w_2 \dots w_N$  of length  $N$ , having letters  $w_i \in [m]$ , and for  $k = 1, \dots, N$ , let  $h_k^w := g_{w_N, N} \dots g_{w_2, 2} g_{w_1, 1}(x_0)$ .

We say that the switched random walk *converges to the uniform distribution* if  $X$  gets evenly mixed as time passes, regardless of the initial state  $x_0 \in X$  and the order in which the  $Q_i$  have been chosen. Quantitatively, this means that

$$\lim_{N \rightarrow \infty} \left( \max_{x_0 \in X, w: |w|=N} \|q^w \cdot e_{x_0} - \bar{u}\|_{TV} \right) = 0,$$

where  $q^w$  is defined as the product  $q^{w_1} q^{w_2} \dots q^{w_n}$  and  $q^j := \sum_{g \in G} Q_j(g)e_g$ , so that  $q^w \cdot e_{x_0}$  encodes the distribution of  $h_N^w$ .

**Remark 4.1.** Note that we are interested in studying the worst case mixing behaviour of the dynamical system, in which the probability distributions  $Q_i$  are chosen adversarially. This case is very different in nature from the ‘typical’ random walk, by which we mean selecting one of the matrices uniformly at random at each stage. The latter case reduces to the standard random walk with probability distribution  $Q := \sum_{i=1}^m Q_i$  and can be studied with the techniques described in the previous sections.

In this section, we address the problem of convergence of switched random walks. We define the *Fourier joint spectral radius* relative to  $X$  of a set of distributions  $Q_1, \dots, Q_m$ , denoted by  $\bar{\omega}_X(Q_1, \dots, Q_m)$ , and prove Theorem 1.6, which shows that this quantity captures the long-term

behaviour of switched random walks (see Section 4.1). The effective estimation of Fourier jsrs is discussed in the final section 4.2.

**4.1 Fourier joint spectral radii.**

If  $A_1, \dots, A_m$  are a set of  $n \times n$  matrices with complex entries, then their joint spectral radius, introduced by Rota and Strang in [RS], is defined as

$$\text{jsr}(A_1, \dots, A_m) := \lim_{N \rightarrow \infty} \max_{w: |w|=N} \|A_{w_1} A_{w_2} \dots A_{w_N}\|^{1/N}.$$

It is known that this limit always exists and that its value is independent of the chosen matrix norm.

**Definition 4.2.** The Fourier joint spectral radius of the distributions  $Q_1, \dots, Q_m$  on  $G$  relative to  $X$  is the number

$$\bar{\omega}_X(Q_1, \dots, Q_m) := \max_{\text{triv} \neq \rho_j \in \mathbb{C}X} \{ \text{jsr}(\widehat{Q}_1(\rho_j), \dots, \widehat{Q}_m(\rho_j)) \},$$

where the maximum is taken over all nontrivial irreducible representations  $\rho_j$  of  $G$  that appear in the module  $\mathbb{C}X$  and jsr denotes the joint spectral radius of a set of matrices.

Next, we prove Theorem 1.6, which establishes that the worst case asymptotic behaviour of the switched random walk determined by the probability distributions  $Q_1, \dots, Q_m$  is encoded in their Fourier joint spectral radius. Note that the mixing rate of the worst case behaviour is an upper bound for that of the typical behaviour of the switched random walk, where at each step the distribution is chosen uniformly at random from the set  $\{Q_1, \dots, Q_m\}$ .

**Proof of Theorem 1.6.** Since any two norms in a finite-dimensional vector space are equivalent, we know that there exist positive constants  $C_1$  and  $C_2$  such that the following inequality holds for every word  $w$  of length  $N$  and every  $x_0 \in X$

$$C_1 \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_2 \leq \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_1 \leq C_2 \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_2.$$

Therefore, the equality in the Theorem is equivalent to

$$\lim_{N \rightarrow \infty} \left( \max_{x, w: |w|=N} \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_2 \right)^{1/N} = \bar{\omega}_X(Q_1, \dots, Q_m).$$

For any word  $w$  of length  $N$ , the inequality

$$\max_{x_0 \in X} \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_2 \geq \sqrt{\frac{\sum_{x \in X} \|q^{(w)} \cdot e_x - \bar{u}\|_2^2}{|X|}} = \frac{\|q^{(w)} - \bar{u}_G\|_{\text{Fb}}}{\sqrt{|X|}}$$

holds, and by the submultiplicativity of the Frobenius norm, so does the following

$$\max_{x_0 \in X} \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_2 \leq \|q^{(w)} - \bar{u}_G\|_{\text{Fb}}.$$

The equality of the Theorem is therefore equivalent to

$$\lim_{N \rightarrow \infty} \left( \max_{w: |w|=N} \|q^{(w)} - \bar{u}_G\|_{\text{Fb}} \right)^{1/N} = \bar{\omega}_X(Q_1, \dots, Q_m).$$

By Lemma 3.2 and by the orthogonal invariance of the Frobenius norm, we know that for every word  $w$  of length  $N$  the following equality holds

$$\|q^{(w)} - \bar{u}_G\|_{\text{Fb}}^2 = \sum_{V_j \neq \text{triv}} m(V_j, \mathbb{C}X) \|\widehat{Q}(\rho_j)_{w_1} \widehat{Q}(\rho_j)_{w_2} \dots \widehat{Q}(\rho_j)_{w_N}\|_{\text{Fb}}^2.$$

Taking  $N$ th roots on both sides and letting  $R$  denote the number of irreducible representations of  $G$  appearing in  $\mathbb{C}X$ , we obtain the inequality

$$\|q^{(w)} - \bar{u}_G\|_2^{\frac{1}{N}} \leq \left( R \max_{\rho_j \neq \text{triv}} m(V_j, \mathbb{C}X) \right)^{1/2N} \left( \max_{\text{triv} \neq \rho_j \in \mathbb{C}X} \max_{w: |w|=N} \|\hat{Q}(\rho_j)_{w_1} \hat{Q}(\rho_j)_{w_2} \dots \hat{Q}(\rho_j)_{w_N}\|_{\text{Fb}}^2 \right)^{1/2N},$$

and therefore

$$\max_{w: |w|=N} \|q^{(w)} - \bar{u}\|_2^{\frac{1}{N}} \leq \left( R \max_{\rho_j \neq \text{triv}} m(V_j, \mathbb{C}X) \right)^{1/2N} \left( \max_{\text{triv} \neq \rho_j \in \mathbb{C}X} \max_{w: |w|=N} \|\hat{Q}(\rho_j)_{w_1} \hat{Q}(\rho_j)_{w_2} \dots \hat{Q}(\rho_j)_{w_N}\|_{\text{Fb}}^2 \right)^{1/2N}.$$

Letting  $N \rightarrow \infty$ , we obtain

$$\lim_{N \rightarrow \infty} \left( \max_{w: |w|=N} \|q^{(w)} - \bar{u}\|_2 \right)^{1/N} \leq \bar{\omega}_X(Q_1, \dots, Q_m).$$

For the opposite inequality, note that for every irreducible representation  $V_t$  appearing in  $\mathbb{C}X$  and every word  $w$  of length  $N$  we have

$$\|\hat{Q}(\rho_t)_{w_1} \hat{Q}(\rho_t)_{w_2} \dots \hat{Q}(\rho_t)_{w_N}\|_{\text{Fb}} \leq \|q^{(w)} - \bar{u}_G\|_{\text{Fb}}.$$

Therefore, taking  $N$ th roots, maximizing over  $w$ , and letting  $N \rightarrow \infty$ , we obtain

$$\text{jsr} \left( \hat{Q}_1(\rho_t), \dots, \hat{Q}_m(\rho_t) \right) \leq \lim_{N \rightarrow \infty} \max_{w: |w|=N} \|q^{(w)} - \bar{u}_G\|_2^{1/N}.$$

We conclude that the right-hand side is bounded below by  $\bar{\omega}_X(Q_1, \dots, Q_m)$ , proving the equality in the theorem. Since the total variation distance between two probability distributions is bounded by one, the equality

$$\lim_{N \rightarrow \infty} \left( \max_{x, w: |w|=N} \|q^{(w)} \cdot e_{x_0} - \bar{u}\|_{\text{TV}} \right)^{\frac{1}{N}} = \bar{\omega}_X(Q_1, \dots, Q_m)$$

implies that  $\bar{\omega}_X(Q_1, \dots, Q_m) \leq 1$ . We will show that  $Q_1, \dots, Q_m$  has the adversarial mixing property if and only if the strict inequality holds. If  $\bar{\omega}_X(Q_1, \dots, Q_m) < 1$  and  $\alpha$  is any real number with  $\bar{\omega}_X(Q_1, \dots, Q_m) < \alpha < 1$ , then there exists an integer  $N_0$  such that for every initial  $x_0 \in X$  and every word  $w$  of length  $N \geq N_0$  we have

$$\|q^{(w)} \cdot e_{x_0} - \bar{u}\|_{\text{TV}} \leq \alpha^N.$$

This proves the convergence to the uniform distribution, since  $\alpha^N$  converges exponentially to zero. Conversely, if  $\bar{\omega}_X(Q_1, \dots, Q_m) = 1$ , then there exists a representation  $\rho_t$  appearing in  $\mathbb{C}X$  such that  $\text{jsr}(\hat{Q}_1(\rho_t), \dots, \hat{Q}_m(\rho_t)) = 1$ . By [B], Theorem 2],  $\limsup \max_w \Lambda(\prod_{w_i} \hat{Q}_{w_i}(\rho_t)) = 1$ , where  $\Lambda(A)$  denotes the magnitude of the largest eigenvalue of the matrix  $A$ . As a result, given  $\epsilon > 0$ , there exists a sequence of integers  $n_j$  and words  $w^j$  of length  $n_j$  such that  $\hat{Q}_{w_1}(\rho_t) \dots \hat{Q}_{w_{n_j}}(\rho_t)$  has an eigenvalue of size at least  $(1 - \epsilon)$  and therefore its Frobenius norm is at least  $1 - \epsilon$ . By Corollary 1.2, for every such word there exists an initial state  $x_j \in X$  such that

$$\|q^{(w^j)} \cdot e_{x_j} - \bar{u}\|_{\text{TV}} \geq \sqrt{\frac{(1 - \epsilon)^2}{4|X|}}.$$

Since  $X$  is finite, there is an initial state  $x^*$  which appears infinitely many times among the  $x_j$ 's, and we conclude that the switched random walk determined by  $Q_1, \dots, Q_m$  starting from  $x^*$  does not converge.  $\square$

At this point the sceptical reader may wonder whether the theory is trivial in the sense that the equality  $\bar{\omega}_X(Q_1, \dots, Q_m) = \max_j (\bar{\omega}_X(Q_j))$  holds in general (case in which switching the random walk can never be worse than permanently using some of its defining distributions). The following example shows that this is not the case even for  $3 \times 3$  doubly stochastic matrices.

**Example 4.3 (Non-triviality).** Consider the following two probability distributions in  $S_3$

	e	(23)	(12)	(123)	(132)	(13)
$Q_1$	2/8	1/8	1/8	2/8	1/8	1/8
$Q_2$	1/8	1/8	1/8	2/8	1/8	2/8

Their action in the permutation module  $M^{(2,1)}$  is given by the matrices

$$M_1 = \begin{pmatrix} 3/8 & 1/4 & 3/8 \\ 3/8 & 3/8 & 1/4 \\ 1/4 & 3/8 & 3/8 \end{pmatrix}, M_2 = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 3/8 & 3/8 & 1/4 \\ 3/8 & 3/8 & 1/4 \end{pmatrix}.$$

Their Fourier jsr relative to  $M^{(2,1)}$  is equal to the jsr of their Fourier transforms in the representation  $S^{(2,1)}$ , namely the matrices:

$$N_1 = \begin{pmatrix} 0.0625 & 0.108253 \\ -0.108253 & 0.0625 \end{pmatrix}, N_2 = \begin{pmatrix} -0.125 & 0 \\ -0.216506 & 0 \end{pmatrix}.$$

Their spectral radii  $\Lambda$  satisfy  $\Lambda(N_1) = 0.125$ ,  $\Lambda(N_2) = 0.125$  and  $\Lambda(N_1N_2) = 0.03125$ . As a result  $\Lambda(N_1N_2) > \max(\Lambda(N_1), \Lambda(N_2))^2$  and therefore

$$\bar{\omega}_{M^{(2,1)}}(M_1, M_2) \geq \Lambda(N_1N_2)^{1/2} > \max_i (\Lambda(N_i)) = \max(\bar{\omega}_{M^{(2,1)}}(M_1), \bar{\omega}_{M^{(2,1)}}(M_2)).$$

**Remark 4.4.** The example above notwithstanding, there are cases beyond the trivial situation of commuting matrices where switching *does not* make mixing more elusive. For instance, if the  $Q_i$  are *symmetric distributions*, in the sense that  $Q_i(g) = Q_i(g^{-1})$  for all  $g \in G$ , then the matrices  $\hat{Q}_i(\rho_j)$  are Hermitian and in particular, their spectral radius coincides with their operator norm  $\|\hat{Q}_i(\rho_j)\|$  (matrices with this property are called radial and have been classified [GZ]). Since the operator norm is submultiplicative, for any word  $w$  of length  $N$  we have the inequality

$$\|\hat{Q}_{w_i}(\rho_j) \cdots \hat{Q}_{w_N}(\rho_j)\|^{1/N} \leq \max_t \|\hat{Q}_t(\rho_j)\|,$$

which implies that the equality

$$\bar{\omega}_X(Q_1, \dots, Q_m) = \max_j (\bar{\omega}_X(Q_j))$$

holds for every  $G$ -set  $X$  and for symmetric distributions  $Q_1, \dots, Q_m$  or, more generally, for distributions whose Fourier transforms are radial.

**Remark 4.5.** By Birkhoff's Theorem, the convex hull of permutation matrices coincides with the set of doubly stochastic matrices. It follows that doubly stochastic matrices are precisely the possible random walks on  $M^{(n-1,1)}$  induced by a distribution  $Q$  on  $S_n$ . Since  $M^{(n-1,1)} = \text{triv} \oplus S^{(n-1,1)}$ , it follows that for a single distribution the number  $\bar{\omega}_{M^{(n-1,1)}}(Q)$  coincides with the SLEM (second

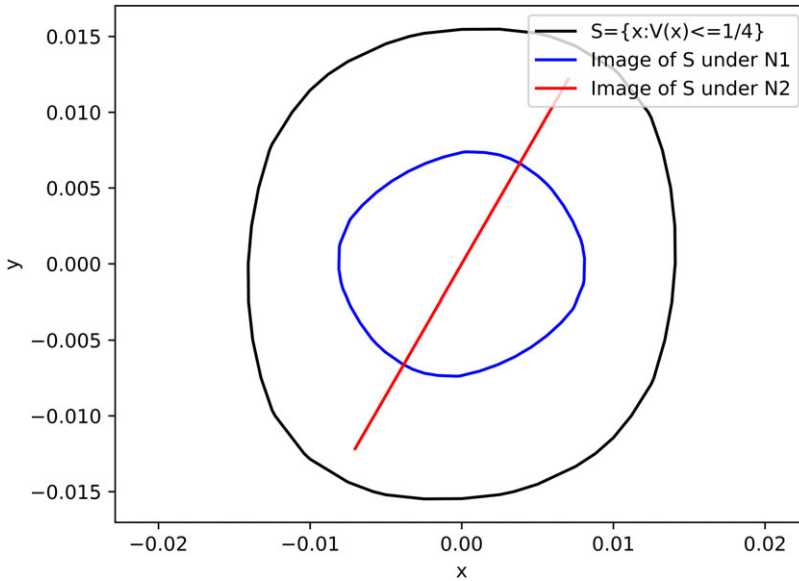


Figure 2. The ball of  $V$  with radius  $1/4$  contains the images under  $N_1$  and  $N_2$  of the unit ball of  $V$ .

largest eigenvalue in magnitude) of the chain studied in [BDX, BDPX] for the design of fast-mixing chains. We can think of Fourier spectral radii as a generalization of this quantity for several distributions and arbitrary symmetries (see Figure 2).

**4.2 Estimation of Fourier jsrs**

The computation of the jsr of a given set of matrices is a surprisingly difficult problem. As mentioned in the introduction, it is known to be *undecidable* whether the jsr of a pair of matrices is bounded by one and it is unknown whether checking if it is strictly bounded by one is decidable. Nevertheless, the seminal work of Parrilo and Jabjabadie [PJ], Ahmadi and Jungers [AJ] and Ahmadi, de Klerk and Hall [AdKH], among others, has provided us with sum-of-squares algorithms which are capable of approximating jsrs to arbitrary accuracy (albeit at an often significant computational effort which cannot be predicted in advance as the undecidability results show). In this section, we extend the results of [AdKH] to polynomial norms expressible via Hermitian sums of squares, allowing us to estimate join spectral radii for matrices with complex entries, the case of interest for the computation of Fourier jsrs. Note that the extension is not completely trivial, since a norm on the underlying real vector space of  $\mathbb{C}^n$  is not generally a complex norm because the equality  $\|\lambda x\| = |\lambda| \|x\|$  needs to hold for arbitrary *complex* numbers  $\lambda$ .

We begin by explaining the general approach for the estimation of jsrs via sums of squares. Recall that the jsr of a set  $A_1, \dots, A_m$  of  $n \times n$  matrices with complex entries is a limit which can be computed using any matrix norm. If we use a matrix norm  $\|\bullet\|_{op}$  which is induced by a norm  $\|\bullet\|$  on vectors in  $\mathbb{C}^n$ , then the submultiplicativity of induced norms implies that the inequality

$$jsr(A_1, \dots, A_m) \leq \max_j \|A_j\|_{op}$$

holds. A basic jsr result of Rota and Strang [RS] states that such inequalities give arbitrarily good estimates for jsr's in the sense that

$$jsr(A_1, \dots, A_m) = \inf_{\|\bullet\|} \left( \max_j \|A_j\|_{op} \right),$$

as the infimum runs over all norms in  $\mathbb{C}^n$ .

Over the real numbers, we know from results of [AdKH] that arbitrary norms can be uniformly approximated by polynomial norms (i.e., by norms of the form  $V(x) = f^{1/2d}(x)$ , where  $f$  is some sum-of-squares form of degree  $2d$ ). This proves that the optima over polynomial norms  $V$  of increasing degrees satisfying  $V(A_i x) \leq \gamma V(x)$  would eventually prove that a real number  $\gamma$  is an upper bound for the jsr of  $A_1, \dots, A_m$  when this is indeed the case. The following lemma extends this result to the complex numbers, proving that all norms on a complex vector space can be approximated via norms defined by Hermitian sums of squares (i.e., sums of squared norms of complex polynomials).

**Theorem 4.6** (Complex polynomial norms). *Let  $\|\bullet\|$  be a norm in  $\mathbb{C}^n$ . There exists a sequence  $\{F_{2d}\}_{d \in \mathbb{N}}$  of Hermitian sums of squares*

$$F_{2d}(z) := \sum_{i=1}^{N(2d)} w_i |\langle z, y_i \rangle|^{2d}$$

for  $N(2d) \in \mathbb{N}$ , points  $y_i \in \mathbb{C}^n$ , and real positive weights  $w_i$  for  $i = 1, \dots, N(2d)$ , which satisfy the following properties:

- (1) The function  $n_{2d}(z) := F_{2d}(z)^{\frac{1}{2d}}$  is a norm in  $\mathbb{C}^d$ .
- (2) The inequality  $n_{2d}(z) \leq \|z\|$  holds for  $z \in \mathbb{C}^d$ .
- (3) The sequence  $n_{2d}(z)$  converges to  $\|z\|$  uniformly on compact subsets of  $\mathbb{C}^d$ .

**Proof.** Let  $B \subseteq \mathbb{C}^n$  be the unit ball for  $\|\bullet\|$  and let  $B^\circ$  be its polar set

$$B^\circ := \{y \in \mathbb{C}^n : \|y\|_* \leq 1\},$$

where  $\|y\|_* := \sup_{x \in B} |\langle x, y \rangle|$  is the norm dual to  $\|\bullet\|$ . Let  $dy$  denote the Lebesgue measure in  $\mathbb{C}^n$  and define the measure  $\mu(A) := \frac{1}{\text{Vol}(B^\circ)} \int_{B^\circ \cap A} dy$ . Since  $\mu$  has compact support, the generalized Tchakaloff Theorem–complex case of Curto and Fialkow [[CF], Theorem 3.1] implies that for every degree  $2d$  there exist  $N_i(2d) \in \mathbb{N}$ , points  $y_i \in B^\circ$  and positive weights  $w_i$  for  $i = 1, \dots, N_i(2d)$  such that for every polynomial  $f(z, \bar{z})$  of degree  $2d$  the equality

$$\int_{\mathbb{C}^n} f(z, \bar{z}) d\mu(z) = \sum_{i=1}^{N(2d)} w_i f(y_i, \bar{y}_i)$$

holds. In particular, for every  $x \in \mathbb{C}^n$  we have

$$F_{2d}(x) := \int_{\mathbb{C}^n} |\langle x, y \rangle|^{2d} d\mu(z) = \frac{1}{\text{Vol}(B^\circ)} \int_{B^\circ} |\langle x, y \rangle|^{2d} dy = \sum_{i=1}^{N(2d)} w_i |\langle x, y_i \rangle|^{2d},$$

proving that  $F_{2d}(x)$  is a sum of Hermitian squares. Furthermore, if  $0 = F_{2d}(\alpha)$ , the integral expression implies that  $\langle \alpha, y \rangle = 0$  for all  $y \in B^\circ$ , so  $\|\alpha\| = 0$ . As a result, the linear map  $\phi : \mathbb{C}^n \rightarrow \mathbb{C}^{N(2d)}$  sending  $x$  to  $(\langle x, y_i \rangle)_i$  is injective. It follows that the function  $n_{2d}(x) := F_{2d}(x)^{\frac{1}{2d}}$ , which is obtained from the  $\ell_{2d}$ -norm  $(\sum |a_i|^{2d})^{1/2d}$  in  $\mathbb{C}^{N(2d)}$  by composition with the injective linear map  $\phi$ , is automatically a norm on  $\mathbb{C}^d$ , proving (1). For (2), note that the inequality  $\langle x, y \rangle \leq \|x\| \|y\|_*$  bounds the integral form of  $F_{2d}(x)$  from above by  $\|x\|^{2d}$ , yielding the inequality  $n_{2d}(x) \leq \|x\|$  for all  $x \in \mathbb{C}^n$ . It remains to show (3). Let  $S$  (resp  $S^*$ ) be the unit sphere for the norm  $\|\bullet\|$  (resp. for the norm  $\|\bullet\|_*$ ). Given  $\epsilon > 0$ , the compactness of  $S$  implies that there exist finitely many centres  $a_1, \dots, a_M$  in  $S$  such that balls of norm  $\|\bullet\|$  with radius  $\epsilon$  centred at the  $a_i$  cover  $S$ . For each  $i$ , let  $b_i \in S^*$  be such that  $\langle a_i, b_i \rangle = 1$  and define  $B_i^\circ := \{y \in B^\circ : \|y - b_i\|_* \leq \epsilon\}$ . We claim that for every  $\epsilon > 0$  there exists  $d$  such that  $n_{2d}(x) \geq 1 - 3\epsilon$  for every  $x' \in S$  simultaneously, proving (3). To verify this claim, take  $x' \in S$  and assume  $i$  is such that  $\|x' - a_i\| < \epsilon$ . For every  $y' \in B_i^\circ$  we have

$$|\langle x', y' \rangle| = |\langle a_i, b_i \rangle + \langle x' - a_i, y' \rangle - \langle a_i, b_i \rangle| \geq 1 - |\langle x', y' \rangle - \langle a_i, b_i \rangle| \geq 1 - 2\epsilon,$$

where the last inequality holds because

$$|\langle x', y' \rangle - \langle a_i, b_i \rangle| = |\langle x', y' - b_i \rangle + \langle x' - a_i, b_i \rangle| \leq \|y' - b_i\|_* + \|x' - a_i\| \leq 2\epsilon.$$

We conclude that for every  $x'$  with  $\|x'\| = 1$  we have

$$n_{2d}(x') \geq (1 - 2\epsilon) \left( \min_{i=1, \dots, m} \frac{\text{Vol}(B_i^\circ)}{\text{Vol}(B^\circ)} \right)^{\frac{1}{2d}}$$

and the right-hand side can be made larger than  $1 - 3\epsilon$  by choosing a sufficiently large  $d$ . □

The previous theorem shows that norms defined by sums of powers of Hermitian squares of linear forms are sufficiently general so as to approximate all norms. Now we will consider a relaxation which has the advantage of being expressible via Hermitian semidefinite programming. To this end, let  $\gamma \geq 0$  be any real number, let  $L(z)$  be a Hermitian polynomial in the variables  $z_1, z_2, \dots, z_n$ , and assume that the following conditions are met:

- (1)  $L(z)$  is a Hermitian sum-of-squares of forms of degree  $d$  in  $z_1, \dots, z_n$ . In particular,  $L$  is real-valued in  $\mathbb{C}^n$ .
- (2) There exists  $\epsilon > 0$  such that  $L(z) \geq \epsilon \|z\|^{2d}$ .
- (3) For every  $(z, w) \in \mathbb{C}^n \times \mathbb{C}^n$ , the inequality  $w^* H_L(z) w \geq 0$  holds, where  $H_L(z)$  denotes the (Hermitian) Hessian matrix of  $L(z)$ .
- (4) The inequalities  $L(A_j z) \leq \gamma^{2d} L(z)$  hold.

Theorem [[AdKH], Theorem 2.1] and condition (1), which guarantees the correct behaviour for scalar multiplication, imply that  $V(z) := L(z)^{\frac{1}{2d}}$  is a complex norm. By (4), its induced operator norm proves that  $\text{jsr}(A_1, \dots, A_m) \leq \gamma$ . Conversely, if  $\text{jsr}(A_1, \dots, A_m) < \gamma$ , then there exists an integer  $d$  such that  $F_{2d}(z)$  from Theorem 4.6 satisfies items (1),(3) and (4) with strict inequalities, so the set of  $L$ 's satisfying the above inequalities strictly are able to guarantee that  $\text{jsr}(A_1, \dots, A_m) < \gamma$  when this is the case. Quillen's positivity theorem [[BPT], Theorem 9.50], which states that every strictly positive byhomogeneous form  $f(z, \bar{z})$  becomes a sum of Hermitian squares (HSOS) when multiplied by  $\|z\|^{2r}$  for some sufficiently large integer  $r$ , can now be used to construct the desired hierarchy of Hermitian semidefinite programmes. More precisely, we have proven.

**Corollary 4.7.** *If  $\text{jsr}(A_1, \dots, A_m) < \gamma$ , then there exist integers  $r, d > 0$  and a real number  $\epsilon > 0$  such that the following Hermitian semidefinite programme is feasible:*

- (1)  $L(z)$  is a Hermitian sum-of-squares of forms of degree  $2d$  in  $z_1, \dots, z_n$ .
- (2) There exists  $\epsilon > 0$  such that  $\|z\|^{2r}(L(z) - \epsilon \|z\|^{2d})$  is HSOS.
- (3) The function  $\|(z, w)\|^{2r} w^* H_L(z) w$  is HSOS, where  $H_L(z)$  denotes the (Hermitian) Hessian matrix of  $L(z)$ .
- (4) The functions  $\|z\|^{2r}(\gamma^{2d} L(z) - L(A_j z))$  are HSOS.

Conversely, any  $L(z)$  satisfying the conditions above defines a norm  $V(z) := L(z)^{\frac{1}{2d}}$  which provides a proof that  $\text{jsr}(A_1, \dots, A_m) \leq \gamma$ .

Our final example illustrates the sum-of-squares approach for estimating Fourier jsrs. The values of the characters of the symmetric group are rational numbers and therefore we can limit ourselves to real sums of squares when computing Fourier jsrs for distributions on  $S_n$ .

**Example 4.8.** We wish to estimate the Fourier spectral radius of the distributions  $Q_1, Q_2$  defined in Example 4.3 relative to  $\mathbb{C}S_3$ . In the sign representation  $\rho_{\text{sgn}}$  the Fourier transforms satisfy



$\hat{Q}_1(\rho_{\text{sgn}}) = 1/4$  and  $\hat{Q}_2(\rho_{\text{sgn}}) = 0$ , so we know that

$$\bar{\omega}(Q_1, Q_2) = \max\left(\frac{1}{4}, \text{jsr}(N_1, N_2)\right),$$

where  $N_1$  and  $N_2$  are the Fourier transforms on the irreducible representation  $S^{(2,1)}$ , computed explicitly in Example 4.3. To estimate the jsr of  $N_1$  and  $N_2$  we solve the optimization problem above. This problem constructs a polynomial  $F(x)$  of degree  $2d = 4$  such that  $V(x) := F(x)^{\frac{1}{2d}}$  is a norm which certifies that the inequality  $\text{jsr}(N_1, N_2) \leq 0.25$  holds, proving that  $\bar{\omega}(Q_1, Q_2) = 1/4$ .

## Acknowledgments

We wish to thank Michael Hoegele, Mauricio Junca, and Pablo Parrilo for useful conversations during the completion of this work. Mauricio Velasco was partially supported by proyecto INV-2018-50-1392 from Facultad de Ciencias, Universidad de los Andes and by Colciencias-ECOSNord grant 80740-047-2019 (Problemas de momentos en control y optimización). We also wish to thank the two anonymous reviewers for comments leading to an improvement in the exposition of our results.

## References

- [AJ] Ahmadi, A. A. and Jungers, R. M. (2016) Lower bounds on complexity of Lyapunov functions for switched linear systems. *Nonlinear Anal. Hybrid Syst.* **21** 118–129. DOI [10.1016/j.nahs.2016.01.003](https://doi.org/10.1016/j.nahs.2016.01.003).
- [JAPR] Jungers, R. M., Ahmadi, A. A., Parrilo, P. A. and Roozbehani, M. (2017) A characterization of Lyapunov inequalities for stability of switched systems. *IEEE Trans. Automat. Control* **62**(6) 3062–3067. DOI [10.1109/TAC.2017.2671345](https://doi.org/10.1109/TAC.2017.2671345).
- [D] Diaconis, P. (1988) *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA, Institute of Mathematical Statistics Lecture Notes—Monograph Series, **11**.
- [BDPX] Boyd, S., Diaconis, P., Parrilo, P. and Xiao, L. (2009) Fastest mixing Markov chain on graphs with symmetries. *SIAM J. Optim.* **20**(2) 792–819. DOI [10.1137/070689413](https://doi.org/10.1137/070689413).
- [BDX] Boyd, S., Diaconis, P. and Xiao, L. (2004) Fastest mixing Markov chain on a graph. *SIAM Rev.* **46**(4) 667–689. DOI [10.1137/S0036144503423264](https://doi.org/10.1137/S0036144503423264).
- [BPT] Blekherman, G., Parrilo, P. A. and Thomas, R. R. (2013) *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics (SIAM), Mathematical Optimization Society, Philadelphia, PA, MOS-SIAM Series on Optimization, **13**.
- [BT] Blondel, V. D. and Tsitsiklis, J. N. (2000) The boundedness of all products of a pair of matrices is undecidable. *Syst. Control Lett.* **41**(2) 135–140. DOI [10.1016/S0167-6911\(00\)00049-9](https://doi.org/10.1016/S0167-6911(00)00049-9).
- [AdKH] Ahmadi, A. A., de Klerk, E. and Hall, G. (2019) Polynomial norms. *SIAM J. Optim.* **29**(1) 399–422. DOI [10.1137/18M1172843](https://doi.org/10.1137/18M1172843).
- [CF] Curto, R. E. and Fialkow, L. A. (2002) A duality proof of Tchakaloff's theorem. *J. Math. Anal. Appl.* **269**(2) 519–532. DOI [10.1016/S0022-247X\(02\)00034-3](https://doi.org/10.1016/S0022-247X(02)00034-3).
- [DS] Diaconis, P. and Shahshahani, M. (1981) Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57**(2) 159–179. DOI [10.1007/BF00535487](https://doi.org/10.1007/BF00535487).
- [FH] Fulton, W. and Harris, J. (1991) *Representation Theory, Graduate Texts in Mathematics*. Springer-Verlag, New York, A first course; Readings in Mathematics, **129**.
- [H] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Am. Statist. Assoc.* **58** 13–30.
- [PJ] Parrilo, P. A. and Jadbabaie, A. (2008) Approximation of the joint spectral radius using sum of squares. *Linear Algebra Appl.* **428**(10) 2385–2402. DOI [10.1016/j.laa.2007.12.027](https://doi.org/10.1016/j.laa.2007.12.027).
- [B] Breuillard, E. On the joint spectral radius, ArXiv preprint, posted on March 2021, <https://arxiv.org/abs/2103.09089>
- [M] Madras, N. (2002). *Lectures on Monte Carlo Methods*. American Mathematical Society, Providence, RI, Fields Institute Monographs, **16**,
- [B] Bernstein, M. (2018) A random walk on the symmetric group generated by random involutions. *Electron. J. Probab.* **23** PaperNo. 26, 28. DOI [10.1214/18-EJP140](https://doi.org/10.1214/18-EJP140).
- [BN] Bernstein, M. and Nestoridi, E. (2019) Cutoff for random to random card shuffle. *Ann. Probab.* **47**(5) 3303–3320. DOI [10.1214/19-AOP1340](https://doi.org/10.1214/19-AOP1340).
- [GZ] Goldberg, M. and Zwas, G. (1974) On matrices having equal spectral radius and spectral norm. *Linear Algebra Appl.* **8**(5) 427–434. DOI [10.1016/0024-3795\(74\)90076-7](https://doi.org/10.1016/0024-3795(74)90076-7).

- [RS] Rota, G.-C. and Strang, G. (1960) A note on the joint spectral radius. *Nederl. Akad. Wetensch. Proc. Ser. A* 63 = *Indag. Math.* 22 379–381, 1960.
- [BDX] Boyd, S., Diaconis, P. and Xiao, L. (2004) Fastest mixing Markov chain on a graph. *SIAM Rev.* 46(4) 667–689. DOI [10.1137/S0036144503423264](https://doi.org/10.1137/S0036144503423264).
- [S] Sagan, B. E. (2001). *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer-Verlag, New York, Graduate Texts in Mathematics, 203, 2nd,