Random Walks on Finite Groups

Laurent Saloff-Coste*

Summary. Markov chains on finite sets are used in a great variety of situations to approximate, understand and sample from their limit distribution. A familiar example is provided by card shuffling methods. From this viewpoint, one is interested in the "mixing time" of the chain, that is, the time at which the chain gives a good approximation of the limit distribution. A remarkable phenomenon known as the cut-off phenomenon asserts that this often happens abruptly so that it really makes sense to talk about "the mixing time". Random walks on finite groups generalize card shuffling models by replacing the symmetric group by other finite groups. One then would like to understand how the structure of a particular class of groups relates to the mixing time of natural random walks on those groups. It turns out that this is an extremely rich problem which is very far to be understood. Techniques from a great variety of different fields – Probability, Algebra, Representation Theory, Functional Analysis, Geometry, Combinatorics – have been used to attack special instances of this problem. This article gives a general overview of this area of research.

1 Introduction 264

2 Background and Notation 267

2.1 Finite Markov Chains 267
2.2 Invariant Markov Chains on Finite Groups 270

3 Shuffling Cards and the Cut-off Phenomenon 272

3.1 Three examples of card shuffling 272
3.2 Exact Computations 274
3.3 The Cut-off Phenomenon 277

4 Probabilistic Methods 281

4.1 Coupling 281
4.2 Strong Stationary Time 285

5 Spectrum and Singular Values 289

* Research supported in part by NSF grant DMS 0102126

5.1 General Finite Markov Chains . . . 289
5.2 The Random Walk Case . . . 292
5.3 Lower Bounds . . . 293

6 Eigenvalue Bounds Using Paths . . . 296

6.1 Cayley Graphs . . . 296
6.2 The Second Largest Eigenvalue . . . 297
6.3 The Lowest Eigenvalue . . . 300
6.4 Diameter Bounds, Isoperimetry and Expanders . . . 302

7 Results Involving Volume Growth Conditions . . . 308

7.1 Moderate Growth . . . 308
7.2 Nilpotent Groups . . . 311
7.3 Nilpotent Groups with many Generators . . . 312

8 Representation Theory for Finite Groups . . . 315

8.1 The General Set-up . . . 315
8.2 Abelian Examples . . . 317
8.3 Random Random Walks . . . 323

9 Central Measures and Bi-invariant Walks . . . 325

9.1 Characters and Bi-invariance . . . 325
9.2 Random Transposition on the Symmetric Group . . . 326
9.3 Walks Based on Conjugacy Classes of the Symmetric Group . . . 328
9.4 Finite Classical Groups . . . 331
9.5 Fourier Analysis for Non-central Measures . . . 334

10 Comparison Techniques . . . 335

10.1 The min-max Characterization of Eigenvalues . . . 335
10.2 Comparing Dirichlet Forms Using Paths . . . 336
10.3 Comparison for Non-symmetric Walks . . . 339

References . . . 340

1 Introduction

This article surveys what is known about the convergence of random walks on finite groups, a subject to which Persi Diaconis gives a marvelous introduction in [27]. In the early twentieth century, Markov, Poincaré and Borel discussed the special instance of this problem associated with card shuffling where the underlying group is the symmetric group S_{52}. Two early references are to Emile Borel [15] and K.D. Kosambi and U.V.R. Rao [95]. The early literature focuses mostly on whether or not a given walk is ergodic: for card shuffling, ergodicity means that the deck gets mixed up after many shuffles.

Once ergodicity is established, the next task is to obtain quantitative estimates on the number of steps needed to reach approximate stationarity. Of course, this requires precise models and the choice of some sort of distance between probability distributions.

Consider the shuffling method used by good card players called riffle shuffling. At each step, the deck is cut into two packs which are then riffled together. A model was introduced by Gilbert and Shannon in a 1955 Bell Laboratories technical memorandum. This model was later rediscovered and studied independently by Reeds in an unpublished work quoted in [27]. Around 1982, Aldous [1] proved that $\frac{3}{2}\log_2 n$ riffle shuffles are necessary and sufficient to mix up n cards, as n goes to infinity. A complete analysis of riffle shuffles was finally obtained in 1992 by Bayer and Diaconis [13], who argue that seven riffle shuffles are reasonable to mix up a deck of 52 cards.

A widespread misconception is to consider that the problem of the convergence of ergodic random walks (more generally, ergodic Markov chains) is solved by the Perron–Frobenius theorem which proves convergence to stationarity at an exponential rate controlled by the spectral gap (i.e., the gap between 1 and the second largest eigenvalue in modulus). To understand the shortcomings of this classical result, consider the Gilbert–Shannon–Reeds model for riffle shuffles. Its spectral gap is $1/2$, independently of the number of cards (see the end of Section 3.2). This does not tell us how many times n cards should be shuffled, let alone 52 cards. Spectral gap estimates are an important part of the study of ergodic random walks but, taking seriously the practical question "how many times should 52 cards be shuffled to mix up the deck?" and generalizing it to random walks on finite groups lead to richer and deeper mathematical problems. What is known about these problems is the subject of this article.

At first sight, it is not entirely clear that the question "how many times should 52 cards be shuffled to mix up the deck?" makes mathematical sense. One reason it does is that stationarity is often reached abruptly. This important fact, called the cut-off phenomenon, was discovered by Aldous, Diaconis and Shahshahani [1, 50] and formalized by Aldous and Diaconis [5, 30]. In their 1981 article [50], Diaconis and Shahshahani use the representation theory of the symmetric group (and hard work) to give the first complete analysis of a complex ergodic random walk: random transposition on the symmetric group. Their main finding is that it takes $t_n = \frac{1}{2}n\log n$ random transpositions to mix up a deck of n cards. More precisely, for any $\varepsilon > 0$, after $(1-\varepsilon)t_n$ random transpositions the deck is far from being well mixed whereas after $(1+\varepsilon)t_n$ random transpositions the deck is well mixed, when n is large enough. This is the first example of the cut-off phenomenon. The riffle shuffle model gives another example. Even for $n = 52$, the cut-off phenomenon for riffle shuffles is visible. See Table 1 in Section 3.3.

It is believed that the cut-off phenomenon is widespread although it has been proved only for a rather small number of examples. One of the most interesting problems concerning random walks on finite groups is to prove

or disprove the cut-off phenomenon for natural families of groups and walks. Focusing on walks associated with small sets of generators, one wants to understand how group theoretic properties relate to the existence or non-existence of a cut-off and, more generally, to the behavior of random walks. For instance, in any simple finite group, most pairs of elements generate the group (see, e.g., [130]). Is it true that any finite simple group G contains a pair of generators such that the associated random walk has a cut-off with a cut-off time of order $\log |G|$ as $|G|$ grows to infinity? Is it true that most walks based on two generators in a simple finite group behave this way? As the cut-off phenomenon can be very hard to establish, one often has to settle for less, for instance, the order of magnitude of a possible cut-off time.

In 2001, Diaconis and Holmes were contacted by a company that builds shuffling machines for the gambling industry. It turns out that these machines use a shuffling scheme that closely resembles one that they considered independently and without the least idea that it could ever be of practical value: see [37]. Besides shuffling and its possible multi-million-dollar applications for the gambling industry, random walks on finite groups are relevant for a variety of applied problems. Diaconis [27] describes connections with statistics. Random walks are a great source of examples for the general theory of finite Markov chains [3, 124, 131] and can sometimes be used to analyze by comparison Markov chains with fewer symmetries (see, e.g., [38]). It relates to Monte-Carlo Markov Chain techniques and to problems in theoretical computer science as described in [94, 131]. Random walks provided the first explicit examples of expander graphs [108], a notion relevant to the construction of communication networks, see, e.g., [98]. In [55], Durrett discusses the analysis of families of random walks modeling the scrambling of genes on a chromosome by reversal of sequences of various lengths.

One perspective to keep in mind is that the study of random walks on finite groups is part of the more general study of invariant processes on groups. See, e.g., [125]. This direction of research relates to many different fields of mathematics. In particular, probability, finite and infinite group theory, algebra, representation theory, number theory, combinatorics, geometry and analysis, all have contributed fundamental ideas and results to the study of random walks on groups. This is both one of the difficulties of the subject and one of its blessings. Indeed, the deep connections with questions and problems coming from other areas of mathematics are one of the exciting aspects of the field.

The author is not aware of any previous attempt thoroughly to survey techniques and results concerning the convergence of random walks on finite groups. The book of Diaconis [27] has played and still plays a crucial role in the development of the subject. The survey [45] by Diaconis and Saloff-Coste served as a starting point for this article but has a narrower focus. Several papers of Diaconis [28, 31, 32] survey some specific directions such as riffle shuffle or the developments arising from the study of random transpositions. Some examples are treated and put in the context of general finite Markov

chains in [3, 124, 131]. The excellent book [98] and the survey article [99] connect random walks to problems in combinatorics, group theory and number theory as does the student text [136].

This survey focuses exclusively on quantitative rates of convergence. Interesting questions such as hitting times, cover times, and other aspects of random walks are not discussed at all although they are related in various ways to rates of convergence. See [3, 27]. Important generalizations of random walks on groups to homogeneous spaces, Gelfand pairs, hypergroups and other structures, as well as Markov chains on groups obtained by deformation of random walks are not discussed. For pointers in these directions, see [14, 16, 17, 27, 29, 31, 32, 36, 41].

2 Background and Notation

2.1 Finite Markov Chains

Markov kernels and Markov chains. A *Markov kernel* on a finite set $\mathcal{X}$ is a function $K : \mathcal{X} \times \mathcal{X} \to [0,1]$ such that $\sum_y K(x,y) = 1$. Given an initial probability measure ν, the associated *Markov chain* is the discrete-time stochastic process $(X_0, X_1, \dots)$ taking values in $\mathcal{X}$ whose law $\mathbb{P}_\nu$ on $\mathcal{X}^{\mathbb{N}}$ is given by

$$\mathbb{P}_\nu(X_i = x_i, 0 \le i \le n) = \nu(x_0)K(x_0,x_1)\cdots K(x_{n-1},x_n). \tag{2.1}$$

We will use $\mathbb{P}_x$ to denote the law of the Markov chain $(X_n)_{n\ge 0}$ starting from $X_0 = x$, that is, $\mathbb{P}_x = \mathbb{P}_{\delta_x}$. One can view K as a stochastic matrix – the transition matrix – whose rows and columns are indexed by $\mathcal{X}$. We associate to K a *Markov operator* – also denoted by K – which acts on functions by $Kf(x) = \sum_y K(x,y)f(y)$ and on measures by $\nu K(A) = \sum_x \nu(x)K(x,A)$.

The *iterated kernel* $K_n(x,y)$ is defined inductively by

$$K_1(x,y) = K(x,y) \quad \text{and} \quad K_n(x,y) = \sum_{z\in\mathcal{X}} K_{n-1}(x,z)K(z,y). \tag{2.2}$$

Given $X_0 = x$, the law of X_n is the probability measure $A \mapsto K_n(x,A)$, $A \subset \mathcal{X}$. From this definition it follows that (X_i) has the *Markov property* : the future depends on the past only through the present. More precisely, let $\tau : \mathcal{X}^{\mathbf{N}} \to \{0,1,\dots\} \cup \{\infty\}$ be a random variable such that the event $\{\tau \le n\}$ depends only on $X_0, \dots, X_n$ (i.e., a stopping time). Then, conditional on $\tau < \infty$ and $X_\tau = x$, $(X_{\tau+i})_{i\ge 0}$ is a Markov chain with kernel K started at x and is independent of $X_0, \dots, X_\tau$.

There is also an $\mathcal{X}$-valued continuous-time Markov process $(X_t)_{t\ge 0}$ which evolves by performing jumps according to K with independent exponential(1) holding times between jumps. This means that $X_t = X_{N_t}$ where N_t has

a Poisson distribution with parameter t. Thus, starting from $X_0 = x$, the law of X_t is given by the familiar formula

$$H_t(x,\cdot) = \mathrm{e}^{-t}\sum_0^\infty \frac{t^n}{n!}K_n(x,\cdot). \tag{2.3}$$

In terms of Markov operators, this continuous-time process is associated with the Markov semigroup $H_t = \mathrm{e}^{-t(I-K)}$, $t \geq 0$, where I denotes the identity operator.

The invariant measure and time reversal. A probability distribution π is *invariant* for K if $\pi K = \pi$. Given an invariant distribution π for K and $p \in [1,\infty)$, set

$$\|f\|_p = \left(\sum_x |f(x)|^p \pi(x)\right)^{1/p}, \quad L^p(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \|f\|_p < \infty\},$$

where $\|f\|_\infty = \max_{\mathcal{X}} |f|$. Then K is a contraction on each $L^p(\pi)$. Define

$$K^*(x,y) = \frac{\pi(y)K(y,x)}{\pi(x)}. \tag{2.4}$$

The kernel K^* is Markov and has the following interpretation: Let $(X_n)_{0\leq n\leq N}$ be a Markov chain with kernel K and initial distribution π. Set $Y_n = X_{N-n}$, $0 \leq n \leq N$. Then $(Y_n)_{0\leq n\leq N}$ is a Markov chain with kernel K^* and initial distribution π. Thus K^* corresponds to the chain obtained from (X_n) by time reversal. The Markov kernel K^* is also the kernel of the adjoint of the operator K acting on $L^2(\pi)$. Clearly, $K^* = K$ if and only if

$$\forall\, x,y \in \mathcal{X}, \quad \pi(x)K(x,y) = \pi(y)K(y,x). \tag{2.5}$$

When (K,π) satisfies (2.5), one says that K is *reversible* with respect to π and that π is a *reversible measure* for K. Equation (2.5) is also called the *detailed balance condition* in the statistical mechanics literature.

Ergodic chains. A Markov kernel K is *irreducible* if, for any two states x, y there exists an integer $n = n(x,y)$ such that $K_n(x,y) > 0$. A state x is called *aperiodic* if $K_n(x,x) > 0$ for all sufficiently large n. If K is irreducible and has an aperiodic state then all states are aperiodic. We will mostly be interested in irreducible, aperiodic chains.

Theorem 2.1. *Let K be an irreducible Markov kernel on a finite state space $\mathcal{X}$. Then K admits a unique invariant distribution π and*

$$\forall\, x,y \in \mathcal{X}, \quad \lim_{t\to\infty} H_t(x,y) = \pi(y).$$

Assume further that K is aperiodic. Then the chain is ergodic, that is,

$$\forall\, x,y \in \mathcal{X}, \quad \lim_{n\to\infty} K_n(x,y) = \pi(y).$$

For irreducible K, the unique invariant distribution is also called the *stationary* (or equilibrium) probability.

In practice, one is interested in turning the qualitative conclusion of Theorem 2.1 into more quantitative assertions. To this end some sort of distance between probability measures must be chosen. The *total variation distance* between two probability measures μ, ν on $\mathcal{X}$ is defined as

$$d_{\mathrm{TV}}(\mu,\nu) = \|\mu-\nu\|_{\mathrm{TV}} = \sup_{A\subset\mathcal{X}} \{\mu(A)-\nu(A)\}. \tag{2.6}$$

It gives the maximum error made when using μ to approximate ν. Next, consider the $L^p(\pi)$-distances relative to a fixed underlying probability measure π on $\mathcal{X}$. In the cases of interest here, π will be the invariant distribution of a given Markov chain under consideration. Given two probability distributions μ, ν with respective densities f, g with respect to π, set

$$d_{\pi,p}(\mu,\nu) = \|f-g\|_p = \left(\sum_{x\in\mathcal{X}} |f(x)-g(x)|^p \pi(x)\right)^{1/p} \tag{2.7}$$

and $d_{\pi,\infty}(\mu,\nu) = \max\{|f-g|\}$. Setting $\mu(f) = \sum f\mu$ and $p=1$, we have

$$d_{\pi,1}(\mu,\nu) = 2d_{\mathrm{TV}}(\mu,\nu) = 2\|\mu-\nu\|_{\mathrm{TV}} = \max_{\|f\|_\infty=1} \{|\mu(f)-\nu(f)|\} \tag{2.8}$$

which is independent of the choice of π. For $p=2$,

$$d_{\pi,2}(\mu,\nu) = \left(\sum_{x\in\mathcal{X}} \left|\frac{\mu(x)}{\pi(x)} - \frac{\nu(x)}{\pi(x)}\right|^2 \pi(x)\right)^{1/2}.$$

Note that Jensen's inequality shows that $p \mapsto d_{\pi,p}$ is a non-decreasing function. In particular,

$$2d_{\mathrm{TV}}(\mu,\nu) \le d_{\pi,2}(\mu,\nu) \le d_{\pi,\infty}(\mu,\nu). \tag{2.9}$$

The following is one of the most useful basic results concerning ergodic chains. It shows and explains why exponentially fast convergence is the rule if the chain converges at all.

Proposition 2.2. *Let K be a Markov kernel with invariant probability distribution π. Then, for any fixed $1 \le p \le \infty$, $n \mapsto \sup_{x\in\mathcal{X}} d_{\pi,p}(K_n(x,\cdot),\pi)$ is a non-increasing sub-additive function. In particular, if*

$$\sup_{x\in\mathcal{X}} d_{\pi,p}(K_m(x,\cdot),\pi) \le \beta$$

for some fixed *integer m and some $\beta \in (0,1)$ then*

$$\forall\, n \in \mathbb{N}, \quad \sup_{x\in\mathcal{X}} d_{\pi,p}(K_n(x,\cdot),\pi) \le \beta^{\lfloor n/m \rfloor}.$$

See, e.g., [1, 3, 5, 124].

2.2 Invariant Markov Chains on Finite Groups

Random walks. Let G be a finite group with identity element e. Let $|G|$ be the order (i.e., the number of elements) of G. Let p be a probability measure on G. The *left-invariant random walk* on G driven by p is the Markov chain with state space $\mathcal{X} = G$ and transition kernel

$$K(x, y) = p(x^{-1}y).$$

As $\sum_x p(x^{-1}y) = \sum_x p(x) = 1$, any such chain admits the normalized counting measure (i.e., uniform distribution) $u \equiv 1/|G|$ as invariant distribution. Moreover, $u \equiv 1/|G|$ is a reversible measure for p if and only if p is symmetric, i.e., $p(x) = p(x^{-1})$ for all $x \in G$.

Fix an initial distribution ν. Let $(\xi_i)_0^\infty$ be a sequence of independent G-valued random variables, with ξ_0 having law ν and ξ_i having law p for all $i \geq 1$. Then the left-invariant random walk driven by p can be obtained as

$$X_n = \xi_0 \xi_1 \dots \xi_n.$$

The iterated kernel $K_n(x, y)$ defined at (2.2) is given by the *convolution power*

$$K_n(x, y) = p^{(n)}(x^{-1}y)$$

where $p^{(n)}$ is the n-fold convolution product $p * \cdots * p$ with

$$f * g(x) = \sum_{z \in G} f(z) g(z^{-1}x) = \sum_{z \in G} f(xz^{-1}) g(z).$$

For any initial distribution ν, we have $\mathbb{P}_\nu(X_n = x) = \nu * p^{(n)}(x)$. The associated Markov operator K acting on functions is then given by

$$Kf(x) = f * \check{p}(x)$$

where $\check{p}(x) = p(x^{-1})$. The law of the associated continuous-time process defined at (2.3) satisfies $H_t(x, y) = H_t(x^{-1}y)$ where

$$H_t(x) = H_t(e, x) = \mathrm{e}^{-t} \sum_0^\infty \frac{t^n}{n!} p^{(n)}(x). \tag{2.10}$$

The adjoint K^* of the operator K on $L^2(G)$ (i.e., L^2 with respect to the normalized counting measure) is

$$K^* f = f * p.$$

This means that the time reversal of a random walk driven by a measure p is driven by the measure $\check{p}$. Referring to the walk driven by p, we call the walk driven by $\check{p}$ the *reverse walk*. Observe that we always have

$$d_{u,s}(p^{(n)}, u) = d_{u,s}(\check{p}^{(n)}, u). \tag{2.11}$$

In words, the distance to stationarity measured in terms of any of the distances $d_{u,s}$ is the same for a given random walk and for its associated reverse walk. By (2.8), this applies to the distance in total variation as well.

One can also consider right-invariant random walks. The right-invariant random walk driven by p has kernel $\widetilde{K}(x,y) = p(yx^{-1})$ and, in the notation introduced above, it can be realized as $\widetilde{X}_n = \xi_n \dots \xi_1 \xi_0$. The iterated kernel $\widetilde{K}_n(x,y)$ is given by $\widetilde{K}_n(x,y) = p^{(n)}(yx^{-1})$. Under the group anti-isomorphism $x \mapsto x^{-1}$, the left-invariant random walk driven by a given probability measure p transforms into the right-random walk driven by $\check{p}$. Hence, it suffices to study left-invariant random walks.

Ergodic random walks. The next proposition characterizes irreducibility and aperiodicity in the case of random walks. It has been proved many times by different authors. Relatively early references are [143, 144].

Proposition 2.3. *On a finite group G, let p be a probability measure with support $\Sigma = \{x \in G : p(x) > 0\}$.*

- *The chain driven by p is irreducible if and only if Σ generates G, i.e., any group element is the product of finitely many elements of Σ.*
- *Assuming Σ generates G, the random walk driven by p is aperiodic if and only if Σ is not contained in a coset of a proper normal subgroup of G.*

To illustrate this proposition, let $G = S_n$ be the symmetric group on n letters and p the uniform distribution on the set $\Sigma = \{(i,j) : 1 \le i < j \le n\}$ of all transpositions. As any permutation can be written as a product of transpositions, this walk is irreducible. It is not aperiodic since $\Sigma \subset (1,2)A_n$ and the alternating group A_n is a proper normal subgroup of S_n.

If the random walk driven by p is aperiodic and irreducible then, by Theorem 2.1, its iterated kernel $K_n(x,y) = p^{(n)}(x^{-1}y)$ converges for each fixed $x \in G$ to its unique invariant measure which is the uniform measure $u \equiv 1/|G|$. By left invariance, there is no loss of generality in assuming that the starting point x is the identity element e in G and one is led to study the difference $p^{(n)} - u$. This brings some useful simplifications. For instance, $d_{u,s}(K_n(x,\cdot), u)$ is actually independent of x and is equal to

$$d_{u,s}(p^{(n)}, u) = |G|^{1-1/s} \left(\sum_{y \in G} \left| p^{(n)}(y) - 1/|G| \right|^s \right)^{1/s}$$

for any $s \in [1, \infty]$ with the usual interpretation if $s = \infty$. From now on, for random walks on finite groups, we will drop the reference to the invariant measure u and write d_s for $d_{u,s}$. Proposition 2.2 translates as follows.

Proposition 2.4. *For any $s \in [1,\infty]$ and any probability measure p, the function $n \to d_s(p^{(n)},u)$ is non-increasing and sub-additive. In particular, if $d_s(p^{(m)},u) \le \beta$ for some fixed integer m and $\beta \in (0,1)$ then*

$$\forall\, n \in \mathbb{N}, \quad d_s(p^{(n)},u) \le \beta^{\lfloor n/m \rfloor}.$$

To measure ergodicity, we will mostly use the total variation distance $\|p^{(k)} - u\|_{\text{TV}}$ and the L^2-distance $d_2(p^{(k)},u)$. Note that d_2 also controls the a priori stronger distance d_∞. Indeed, noting that $p^{(2k)} - u = (p^{(k)} - u) * (p^{(k)} - u)$ and using the Cauchy–Schwarz inequality and (2.11), one finds that

$$d_\infty(p^{(2k)},u) \le d_2(p^{(k)},u)^2$$

with equality in the symmetric (i.e, reversible) case where $p = \check{p}$.

3 Shuffling Cards and the Cut-off Phenomenon

3.1 Three examples of card shuffling

Modeling card shuffling. That shuffling schemes can be modeled by Markov chains has been clearly recognized from the beginning of Markov chain theory. Indeed, card shuffling appears as one of the few examples given by Markov in [104]. It then appears in the works of Poincaré and Borel. See in particular [15], and the excellent historical discussion in [92]. Obviously, from a mathematical viewpoint, an arrangement of a deck of cards can be thought of as a permutation of the cards. Also, a shuffling is obviously a permutation of the cards. There is however an intrinsic difference between an arrangement of the cards and a shuffling: an arrangement of the cards relates face values to positions whereas, strictly speaking, a shuffling is a permutation of the positions. By a good choice of notation, this difference somehow disapears but this might introduce some confusion. Thus we now spell out in detail one of the possible equivalent ways to model shufflings using random walks on S_n, $n = 52$. We view the symmetric group S_n as the set of all bijective maps from $\{1,\dots,n\}$ to itself equipped with composition. Hence, for $\sigma,\theta \in S_n$, $\sigma\theta = \sigma\circ\theta$. One of several ways to describe a permutation σ is as an n-tuple $(\sigma_1,\dots,\sigma_n)$ where $\sigma(i) = \sigma_i$.

To simplify, think of the 52 cards as marked from 1 to 52. An arrangement of the deck can described as a 52-tuple giving the face values of the cards in order from top to bottom. Thus we can identify the arrangement of the deck $(\sigma_1,\dots,\sigma_{52})$ with the permutation $\sigma : i \mapsto \sigma(i) = \sigma_i$ in S_{52}. In this notation, the deck corresponding to a permutation σ has card i in position $\sigma^{-1}(i)$ whereas $\sigma(i)$ gives the value of the card in position i. In particular, the deck in order is represented by the identity element. Now, from a card shuffling perspective, we want permutations to act on positions, not on face values. One easily checks that, in the present notation, this corresponds to

multiplication on the right in S_{52}. Indeed, if the arrangement of the deck is σ and we transpose the top two cards then the new arrangement of the deck is $\sigma \circ \tau$ with $\tau = (1, 2)$ since $\sigma \circ \tau$ is $(\sigma_2, \sigma_1, \sigma_3, \ldots, \sigma_{52})$.

Typically, shuffling cards proceeds by repeating several times a fixed procedure where some randomness occurs. This can now be modeled by a measure p on S_{52} which describes the shuffling procedure as picking a permutation θ according to p and changing the arrangement σ of the deck to $\sigma\theta = \sigma \circ \theta$. Thus the shuffling scheme whose elementary steps are modeled by p corresponds to the left-invariant random walk on S_{52} driven by p. By invariance, we can always assume that we start from the identity permutation, that is, with the deck in order. Then, the distribution of the deck after n shuffles is given by $p^{(n)}$. Let us describe three examples.

The Borel–Chéron shuffle. In [15, pages 8–10 and 254–256], Borel and Chéron consider the following shuffling method: remove a random packet from the deck and place it on top. The corresponding permutations are $\pi_{a,b}$, $1 < a \leq b \leq n = 52$, given by

$$\begin{pmatrix} 1 & 2 & \cdots & b-a+1 & b-a+2 & \cdots & b & b+1 & \cdots & 52 \\ a & a+1 & \cdots & b & 1 & \cdots & a-1 & b+1 & \cdots & 52 \end{pmatrix}$$

where the first row indicates position and the second row gives the value of the cards in that position after $\pi_{a,b}$ if one starts with a deck in order. The removed packet is random in the sense that $p(\pi) = 0$ unless $\pi = \pi_{a,b}$ for some $1 < a \leq b \leq n$ in which case $p(\pi) = \binom{n}{2}^{-1}$ (a slightly different version is considered in [42]).

The crude overhand shuffle. In this example, the player holds the deck in the right hand and transfers a first block of cards from the top of the deck to the left hand, then a second block of cards, and finally all the remaining cards. This is then repeated many times. The randomness comes from the size of the first and second block, say a and b. With our convention, the corresponding permutation $\sigma_{a,b}$ is

$$\begin{pmatrix} 1 & 2 & \cdots & 51-a-b & 52-a-b & \cdots & 52-a-1 & 52-a & \cdots & 51 & 52 \\ a+b+1 & a+b+2 & \cdots & 52 & a+1 & \cdots & a+b & 1 & \cdots & a-1 & a \end{pmatrix}.$$

In this case, it is natural to take $p(\sigma) = 0$ unless $\sigma = \sigma_{a,b}$ for some $1 \leq a \leq n = 52$ and $0 \leq b \leq n - a$, in which case $p(\sigma_{a,b}) = 1/[n(n + 1 - a)]$. Other overhand shuffles are described in [116, 44].

The riffle shuffle or dovetail shuffle. Consider the way serious players shuffle cards. The deck is cutinto two packs (of roughly equal sizes) and the two packs are riffled together. A model was introduced by Gilbert and Shannon (see Gilbert [66]) and later, independently, by Reeds [118]. In this model, the cut is made according to a binomial distribution: the k top cards are cut

with probability $\binom{n}{k}/2^n$, $n = 52$. The two packets are then riffled together in such a way that the cards drop from the left or right heaps with probability proportional to the number of cards in each heap. Thus, if there are a and b cards remaining in the left and right heaps, then the chance the next card will drop from the left heap is $a/(a+b)$. This describes a probability p_{RS} on the symmetric group. Experiments reported in Diaconis' book [27] indicate that this model describes well the way serious card players shuffle cards. It is interesting to note that the inverse shuffle – i.e., the shuffle corresponding to the measure $\check{p}_{\text{RS}}$ – is simple to describe: starting from the bottom, each card is removed from the deck and placed randomly on one of two piles, left or right, according to an independent sequence of Bernoulli random variables (probability 1/2 for right and left). Finally, the right pile is put on top.

3.2 Exact Computations

The analysis of riffle shuffles. This section focuses on the riffle shuffle model p_{RS} of Gilbert, Shannon and Reeds, the GSR model for short. How many GSR shuffles are needed to mix up a deck of n cards? To make this question precise, let us use the total variation distance between the uniform distribution u on the symmetric group S_n and the distribution $p_{\text{RS}}^{(k)}$ after k shuffles. The question becomes: how large must k be for $\|p_{\text{RS}}^{(k)} - u\|_{\text{TV}}$ to be less than some fixed $\varepsilon > 0$? As far as shuffling cards is concerned, a value of ε a little below 0.5 seems quite reasonable to aim for. Bayer and Diaconis [13] give the following remarkably precise analysis of riffle shuffles.

Theorem 3.1. *If a deck of n cards is shuffled k times with*

$$k = \frac{3}{2}\log_2 n \ + c,$$

then for large n

$$\|p_{\text{RS}}^{(k)} - u\|_{\text{TV}} = 1 - 2\Phi\left(\frac{-2^{-c}}{4\sqrt{3}}\right) + O\left(\frac{1}{n^{1/4}}\right),$$

where

$$\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \mathrm{e}^{-s^2/2} ds.$$

A weaker form of this result was proved earlier in [1].

To people studying finite Markov chains, the fact that Theorem 3.1 can be proved at all appears like a miracle. Consider for instance the following "neat riffle shuffle" model proposed by Thorpe (see [27, 137]). For a deck of $n = 2k$ cards, cut the deck into two piles of exactly k cards each and put in positions $2j$ and $2j-1$ the j-th card of each of the two piles in random order. No reasonable quantitative analysis of this shuffle is known.

The idea used by Bayer and Diaconis to analyze repeated riffle shuffles is elementary. Given an arrangement of a deck of cards, a *rising sequence* is

a maximal subset of cards of this arrangement consisting of successive face values displayed in order. For example, the arrangement $2, 4, 3, 9, 1, 6, 7, 8, 5$, consists of 1; $2, 3$; $4, 5$; $6, 7, 8$ and 9. Note that the rising sequences form a partition of the deck. Denote by r the number of rising sequences of an arrangement of the deck. By extension, we also say that r is the number of rising sequences of the associated permutation. Now, it is a simple observation that, starting from a deck in order, one riffle shuffle produces permutations having at most 2 rising sequences. In fact (see [13]), the riffle shuffle measure p_{RS} is precisely given by

$$p_{\text{RS}}(\sigma) = 2^{-n} \binom{n+2-r}{n}$$

where r is the number of rising sequences of σ and $\binom{m}{n} = 0$ when $m < n$.

The next step is to define the notion of an m-riffle shuffle which generalizes the above 2-riffle shuffle. In an m-riffle shuffle, the deck is cut into m parts which are then riffled together. It is easier to define a reverse m-riffle shuffle: hold the deck, face down and create m piles by dealing the deck in order and turning the cards face up on a table. For each card, pick a pile uniformly at random, independently from all previous picks. When all the cards have been distributed, assemble the piles from left to right and turn the deck face down. Let $p_m = p_{m\text{-RS}}$ be the probability measure corresponding to an m-riffle shuffle. Diaconis and Bayer show that

$$p_m(\sigma) = m^{-n} \binom{n+m-r}{n}$$

where r is again the number of rising sequences. Moreover, they show that following an m-riffle shuffle by an ℓ-riffle shuffle produces exactly an $m\ell$-riffle shuffle, that is, $p_\ell * p_m = p_{m\ell}$. Thus the distribution $p_{\text{RS}}^{(k)}$ of a deck of n cards after k GSR riffle shuffles is given by

$$p_{\text{RS}}^{(k)}(\sigma) = 2^{-kn} \binom{n+2^k-r}{n}. \tag{3.1}$$

From there, the proof of Theorem 3.1 consists in working hard to obtain adequate asymptotics and estimates. Formula (3.1) allows us to compute the total variation distance exactly for $n = 52$. This is reported (to three decimal places) in Table 1.

Table 1. The total variation distance for k riffle shuffles of 52 cards

k	1	2	3	4	5	6	7	8	9	10
$\|p_{\text{RS}}^{(k)} - u\|_{\text{TV}}$	1.000	1.000	1.000	1.000	0.924	0.614	0.334	0.167	0.085	0.043

Top to random shuffles. There are not many examples of shuffles where the law after k shuffles can be explicitly computed as above. In [34], the authors study a class of shuffles that they call top to random shuffles. In a top m to random shuffle, the top m cards are cut and inserted one at a time at random in the remaining $n-m$ cards. Call q_m the corresponding probability measure. In particular, q_1 is called the top to random measure. Note the similarity with the riffle shuffle: a top to random shuffle can be understood as a riffle shuffle where exactly one card is cut off.

Given a probability measure μ on $\{0,1,\dots,n\}$, set

$$q_\mu = \sum_0^n \mu(i) q_i. \tag{3.2}$$

Further variations are considered in [34]. In some cases, an exact formula can be given for the convolutions of such measures and this leads to the following theorem.

Theorem 3.2. *Let a, n, $a \le n$, be two integers. Let μ be a probability on $\{0,\dots,a\}$ with positive mean m. On S_n, consider the probability measure q_μ at* (3.2)*. Then, for large n and*

$$k = \frac{n}{m}\log n \ + c,$$

we have $\|q_\mu^{(k)} - u\|_{\mathrm{TV}} = f(c) + o(1)$ where f is a positive function such that $f(c) \le (1/2)\mathrm{e}^{-2c}$ for $c > 0$ and $f(c) = 1 - \exp\left(-\mathrm{e}^{-c} + o(1)\mathrm{e}^{-c}\right)$ for $c < 0$.

Diagonalization. The riffle shuffles and top to random shuffles described above, as well as variants and generalizations discussed in [60, 61], have remarkable connections with results in algebra. These connections explain in part why an exact formula exists for repeated convolution of these measures. See [13, 32, 34, 40, 60, 61].

In particular, the convolution operators corresponding to the m-riffle shuffle measures p_m and the top to random measures q_m are diagonalizable with eigenvalues that can be explicitly computed. For instance, for the GSR measure $p_{\mathrm{RS}} = p_2$, the eigenvalues are the numbers 2^{-i} with multiplicity the number of permutations having exactly $n-i$ cycles, $i = 0,\dots,n-1$. For the top to random measure $q = q_1$, the eigenvalues are i/n, $i = 0,1,\dots,n-2,n$, and the multiplicity of i/n is exactly the number of permutations having i fixed points. However, these results do not seem to be useful to control convergence to stationarity. Curiously, the eigenvalues of top to random have been computed independently for different reasons by different authors including Wallach (Lie algebra cohomology) and Phatafod (linear search). See the references in [32, 34].

3.3 The Cut-off Phenomenon

Cut-off times. Table 1, Theorem 3.1 and Theorem 3.2 all illustrate a phenomenon first studied by Aldous and Diaconis [5] and called the *cut-off phenomenon* [30] (in [5], the term threshold phenomenon is used instead).

To give a precise definition, consider a family of finite groups G_n, each equipped with its uniform probability measure u_n and with another probability measure p_n which induces a random walk on G_n.

Definition 3.3. *We say that the* cut-off phenomenon *holds (in total variation) for the family* $((G_n, p_n))$ *if there exists a sequence* (t_n) *of positive reals such that*

(a) $\lim_{n\to\infty} t_n = \infty$*;*
(b) For any $\varepsilon \in (0,1)$ *and* $k_n = [(1+\varepsilon)t_n]$, $\lim_{n\to\infty} \|p_n^{(k_n)} - u_n\|_{\mathrm{TV}} = 0$*;*
(c) For any $\varepsilon \in (0,1)$ *and* $k_n = [(1-\varepsilon)t_n]$, $\lim_{n\to\infty} \|p_n^{(k_n)} - u_n\|_{\mathrm{TV}} = 1$.

We will often say, informally, that (G_n, p_n) has a (total variation) cut-off at time t_n. For possible variants of this definition, see [30, 124].

Theorem 3.1 shows that the GSR riffle shuffle measure p_{RS} on S_n has a cut-off at time $\frac{3}{2}\log_2 n$. Similarly, Theorem 3.2 shows that the top to random measure q_1 on S_n has a cut-off at time $n \log n$. Note that if (t_n) and (t'_n) are cut-off times for the same family $((G_n, p_n))$, then $t_n \sim t'_n$ as n tends to infinity. Table 2 below lists most examples known to have a cut-off.

Definition 3.4. *For any probability measure* p *on a finite group* G*, set*

$$T(G,p) = T(G,p,1/(2\mathrm{e})) = \inf\left\{k : \|p^{(k)} - u\|_{\mathrm{TV}} \le 1/(2\mathrm{e})\right\} \tag{3.3}$$

where $T(G,p,\varepsilon) = \inf\left\{k : \|p^{(k)} - u\|_{TV} \le \varepsilon\right\}$*. We call* $T(G,p)$ *the total variation* mixing time *(mixing time for short) of the random walk driven by* p*.*

Thus $T(G,p)$ is the number of steps needed for the given random walk to be $1/(2\mathrm{e})$-close to the uniform distribution in total variation. The arbitrary choice of $\varepsilon = 1/(2\mathrm{e})$ (any $\varepsilon \in (0, 1/2)$ would do) is partially justified by Proposition 2.4 which shows that

$$\forall\, k \in \mathbb{N}, \quad 2\|p^{(k)} - u\|_{\mathrm{TV}} \le \mathrm{e}^{-\lfloor k/T(G,p)\rfloor}.$$

To relate the last definition to the notion of cut-off, let $((G_n, p_n))$ be a family of random walks having a (t_n)-cut-off. Then, for any $\varepsilon \in (0,1)$,

$$T(G_n, p_n, \varepsilon) \sim T(G_n, p_n) \sim t_n \quad \text{as } n \text{ tends to } \infty.$$

Thus, if (G_n, p_n) presents a cut-off, one can always take the cut-off time to be $t_n = T(G_n, p_n)$ and one often says that the cut-off time t_n is "the time needed to reach equilibrium".

Table 2. Total variation cut-offs

G	p	(CO)	§	Ref
$\mathbb{Z}_2^d$	$p(e_i) = 1/(d+1)$	$\frac{d}{4}\log d$	8.2	[35, 27, 28]
$\mathbb{Z}_2^d$	random spatula	$\frac{d}{8}\log d$	8.2	[138]
$\mathbb{Z}_n^d$	$p(e_i) = 1/(d+1),\ \ d \to \infty$	$\frac{d\log d}{2(1-\cos 2\pi/n)}$	8.2	[44, 47]
$\mathbb{Z}_2^d$	most k-sets, $k > d$	$T(d,k)$	8.2	[140]
abelian	most k-sets, $k = \lfloor(\log\lvert G\rvert)^s\rfloor$, $s > 1$	$\frac{s}{s-1}\frac{\log\lvert G\rvert}{\log k}$	8.3	[54, 87]
S_n	GSR riffle shuffle, p_{RS}	$\frac{3}{2}\log_2 n$	3.2	[13]
S_n	top m to random, q_m	$\frac{n}{m}\log n$	3.2	[34]
S_n	random transposition, p_{RT}	$\frac{n}{2}\log n$	9.2	[50, 27]
S_n	transpose $(1,i)$, $p_\star$	$n\log n$	9.2	[28, 59]
S_n	lazy small odd conjugacy classes $C = (2),\ (4),\ (3,2),\ (6),\ (2,2,2)$	$\frac{2n}{\lvert C\rvert}\log n$	9.2	[59, 122]
A_n	small even S_n conjugacy classes $(3),\ (2,2),\ (5),\ (4,2),\ (3,3),\ (7)$	$\frac{n}{\lvert C\rvert}\log n$	9.2	[59, 122]
A_n	random m-cycle, m odd $m > n/2,\ n-m \to \infty$	$\frac{\log n}{\log(n/(n-m))}$	9.2	[103]
$G \wr S_n$	random transposition with independent flips	$\frac{n}{2}\log n$	9.2	[128, 129]
$G \wr S_n$	random transposition with paired flips	$\frac{n}{2}\log n$	9.2	[128, 129]
$SL_n(\mathbb{F}_q)$	random transvections	n	9.2	[86]

$$T(d,k) \sim \begin{cases} (d/4)\log(d/(k-d)) & \text{if } k-d = o(d) \\ a_\eta d & \text{if } k = (1+\eta)d \\ d/\log_2(k/d) & \text{if } d/k = o(1). \end{cases}$$

One can easily introduce the notion of L^s-mixing time and L^s-cut-off, $1 < s \le \infty$, by replacing $2\|p_n^{(k_n)} - u_n\|_{\mathrm{TV}}$ by $d_s(p_n^{(k_n)}, u_n)$ in Definitions 3.4, 3.3. In Definition 3.3(c), one should require that $\lim_{n\to\infty} d_s(p_n^{(k_n)}, u_n) = \infty$. In this survey, we will focus mostly on mixing time and cut-off in total variation but we will also make significant use the L^2-distance d_2.

Cut-off and group structure. Not all natural families of walks have a cut-off. For instance, the walk on $G_n = \mathbb{Z}/n\mathbb{Z}$ driven by the uniform measure on $\{-1,0,1\}$ does not present a cut-off. For this walk, it takes k of order n^2 to have $\|p_n^{(k)} - u_n\|_{\mathrm{TV}}$ close to $1/2$. It then takes order n^2 additional steps to go

down to 1/4, etc. In particular, for any integer $k > 0$,

$$0 < \liminf_{n\to\infty} \|p_n^{(kn^2)} - u_n\|_{\mathrm{TV}} \le \limsup_{n\to\infty} \|p_n^{(kn^2)} - u_n\|_{\mathrm{TV}} < 1.$$

See Sections 7.2 and 8.2 below.

Trying to understand which walks on which families of groups have a cut-off is one of the difficult open problems concerning random walks on finite groups. To be meaningful, this question should be made more precise. One possibility is to focus on walks driven by the uniform measure on minimal generating sets, i.e., generating sets that do not contain any proper generating sets (one might allow here the inclusion of inverses to have reversible walks and of the identity to cure periodicity problems). For instance, the set $\Sigma = \{(1, i) : 1 < i \le n\}$ (where $(1, i)$ means transpose 1 and i) is a minimal generating set of S_n and in this case one may want to consider the "transpose top and random" measure $p_\star$, i.e., the uniform probability measure on $\{e\} \cup \Sigma$. Fourier analysis can be used to show that $(S_n, p_\star)$ has a cut-off at time $n \log n$, see Section 9.5 below. For another example, take $\Sigma = \{\tau, c\}$ where $\tau = (1, 2)$ and c is the long cycle $(1, 2 \ldots, n)$ in S_n. These two elements generate S_n and this is obviously a minimal generating set. Let $p_{\tau,c}$ denotes the uniform measure on $\{\tau, c\}$. It is known that, for odd n, $cn^3 \log n \le T(S_n, p_{\tau,c}) \le Cn^3 \log n$ (see [45, 142] and Section 10). It is conjectured that this walk has a cut-off.

Problem 3.5. Is it true that most natural families (S_n, p_n) where p_n is uniform on a minimal generating set of S_n have a cut-off?

Problem 3.6. Is it true that most natural families (G_n, p_n) where each G_n is a simple group and p_n is uniform on a minimal generating set of G_n have a cut-off?

Problem 3.7. What is the range of the possible cut-off times for walks on the symmetric group S_n based on minimal generating sets? (known examples have the form $t_n = cn^a \log n$ with a a small integer)

Unfortunately, these problems seem extremely difficult to attack. It is known that about 3/4 of all pairs of permutations in S_n generate S_n [52] but no one seems to know how to study the associated random walks, let alone to prove or disprove the existence of a cut-off. The situation is similar for all finite simple groups (almost all pairs in a finite simple group generate the group [130]). One of the only satisfactory results in this direction is a negative result which will be discussed in Section 7.2 and says that reversible walks (with holding) based on minimal generating sets in groups of order p^a (such groups are necessarily nilpotent) with a bounded and p any prime do not present a cut-off. Instead, such walks behave essentially as the simple random walk (with holding) on the circle group $\mathbb{Z}_n$.

Precut-off. The cut-off phenomenon is believed to be widespread but it has been proved only in a rather limited number of examples, most of which are recorded in Table 2. Indeed, to prove that a cut-off occurs, one needs to understand the behavior of the walk before and around the time at which it reaches equilibrium and this is a difficult question. In [124], further versions of the cut-off phenomenon are discussed that shed some light on this problem. Let us point out that there are many families of walks $((G_n, p_n))$ for which the following property is known to be satisfied.

Definition 3.8. *We say that the family (G_n, p_n) presents a* precut-off *if there exist a sequence t_n tending to infinity with n and two constants $0 < a < b < \infty$ such that*

$$\lim_{n\to\infty} \|p_n^{(bk_n)} - u_n\|_{TV} = 0 \quad \textit{and} \quad \lim_{n\to\infty} \|p_n^{(ak_n)} - u_n\|_{TV} > 0.$$

Table 3. Precut-offs

G	p	(PCO)	§	Ref
S_n	adjacent transposition p_{AT}	$n^3 \log n$	4.1, 5.3, 10.2	[42, 141]
S_n	ℓ-adjacent transposition, $p_{\ell\text{-AT}}$	$(n^3/\ell^2)\log n$	10.2	[55]
S_n	nearest neighbors transposition on a square grid	$n^2 \log n$	10.2	[42, 141]
S_n	random insertion	$n \log n$	10.2	[42]
S_n	Borel-Chéron random packet to top	$n \log n$	3.1, 10.2	[42]
S_n	random inversion	$n \log n$	10.2	[55]
S_n	neat overhand shuffle, i.e., reverse top to random	$n \log n$	10.2	[42]
S_n	crude overhand shuffle	$n \log n$	3.1, 10.2	[42]
S_n	Rudvalis shuffle, i.e., top to $n-1$ or n	$n^3 \log n$	4.1	[31, 85]
S_n	uniform on $e, (1,2)$, top to bottom, botton to top	$n^3 \log n$	4.1, 10.2	[31, 85]
A_n	S_n conjugacy classes $C=(c_1,\dots,c_\ell)$, $\lvert C\rvert = c_1+\cdots+c_\ell = m \ll n$	$\frac{n}{m}\log n$	9.2	[119]
$U_m(q)$	$E_{i,j}(a), a \in \mathbb{Z}_q, 1 \le i < j \le m$	$m^2 \log m$	4.2	[114]
Lie type	small conjugacy classes	n=rank(G)	9.2	[68]
$\mathbb{Z}_2^d \rtimes \mathbb{Z}_d$	perfect shuffles	d^2	4.2	[138]
$SL_n(\mathbb{Z}_q)$	$A^{\pm}, B^{\pm}$, r prime, n fixed	$\log q$	6.4	[46, 98]

Thus, if a family $((G_n, p_n))$ presents a precut-off at time t_n, there exist two constants $0 < c \leq C < \infty$ such that, for each $\varepsilon > 0$ small enough and all n large enough,

$$ct_n \leq T(G_n, p_n, \varepsilon) \leq Ct_n.$$

The notion of precut-off captures the order of magnitude of a possible cut-off, but it is unknown whether or not families having a precut-off must have a cut-off. In many cases, it is conjectured that they do. The Borel-Chéron shuffle and the crude overhand shuffle described in Section 3.1 are two examples of shuffles for which a precut-off has been proved (with $t_n = n \log n$, see [42] and Section 10). Another example is the adjacent transposition walk driven by the uniform probability measure p_{AT} on $\{e\} \cup \{(i, i+1) : 1 \leq i < n\}$. This walk satisfies a precut-off at tine $n^3 \log n$ ([42, 141]). In all these cases, the existence of a cut-off is conjectured. See [30, 141] and Table 3. Solutions to the variants of Problems 3.5, 3.6 and 3.7 involving the notion of precut-off instead of cut-off would already be very valuable results.

4 Probabilistic Methods

Two probabilistic methods have emerged that produce quantitative estimates concerning the convergence to stationarity of finite Markov chains: *coupling* and *strong stationary times*. Coupling is the most widely known and used. Strong stationary times give an alternative powerful approach. Both involve the construction and study of certain "stopping times" and have theoretical and practical appeal. In particular, a stationary time can be interpreted as a perfect sampling method. These techniques are presented below and illustrated on a number of examples of random walks. The books [3, 27] are excellent references, as are [1, 4, 5]. When these techniques work, they often lead to good results through very elegant arguments. The potential user should be warned that careful proofs are a must when using these techniques. Experience shows that it is easy to come up with "obvious" couplings or stationary times that end up not being coupling or stationary times at all. Moreover, these two techniques, especially strong stationary times, are not very robust. A good example of a walk that has not yet been studied using coupling or stationary time is random insertion on the symmetric group: pick two positions i, j uniformly independently at random, pull out the card in position i and insert it in position j. This walk has a precut-off at time $n \log n$, see Section 10 and Table 3.

4.1 Coupling

Let K be a Markov kernel on a finite set $\mathcal{X}$ with invariant distribution π. A coupling is simply a sequence of pairs of $\mathcal{X}$-valued random variables

(X_n^1, X_n^2) such that each marginal sequence (X_n^i), $i = 1, 2$, is a Markov chain with kernel K. These two chains will have different initial distributions, one being often the stationary distribution π. The pair (X_n^1, X_n^2) may or may not be Markovian (in most practical constructions, it is). Given the coupling (X_n^1, X_n^2), consider

$$T = \inf\left\{n : \forall\, k \geq n,\ \ X_k^1 = X_k^2\right\}.$$

Call T the *coupling time* (note that T is not a stopping time in general).

Theorem 4.1. *Denote by μ_n^i the distribution of X_n^i, $i = 1, 2$. Then*

$$d_{\mathrm{TV}}(\mu_n^1, \mu_n^2) \leq \mathbb{P}(T > n).$$

This is actually a simple elementary result (see, e.g., [1, 3, 27]) but it turns out to be quite powerful. For further developments of the coupling technique for finite Markov chains, see [3] and the references therein. For relations between coupling and eigenvalue bounds, see, e.g., [18].

Specializing to random walks on finite groups, we obtain the following.

Theorem 4.2. *Let p a probability measure on a finite group G. Let (X_n^1, X_n^2) be a coupling for the random walk driven by p with (X_n^1) starting at the identity and (X_n^2) stationary. Then*

$$d_{\mathrm{TV}}(p^{(n)}, u) \leq \mathbb{P}(T > n).$$

One theoretical appeal of coupling is that there always exists a coupling such that the inequalities in the theorems above are in fact equalities (see the discussions in [3, 27] and the references given there). Hence the coupling technique is exactly adapted to the study of convergence in total variation. In practice, Theorem 4.2 reduces the problem of estimating the total variation distance between a random walk and the uniform probability measure on G to the construction of a coupling for which $\mathbb{P}(T > n)$ can be estimated. This is best illustrated and understood by looking at some examples.

Coupling for random to top [1, 4, 27]. TS[12] Consider the random to top shuffling scheme where a card is chosen at random and placed on top. Obviously, this is the inverse shuffle of top to random. On S_n, this is the walk driven by the uniform measure on the cycles $c_i = (1, 2, \ldots, i)$, $i = 1, \ldots, n$. To construct a coupling, imagine having two decks of cards. The first one is in some given order, the second one is perfectly shuffled. Pick a card at random in the first deck, say, the tenth card. Look at is face value, say, the ace of spades. Put it on top and put a check on its back. In the second deck, find the ace of spades and put it on top. At each step, repeat this procedure. This produces a pair of sequences of S_n-valued random variables (X_k^1, X_k^2) corresponding respectively to the arrangements of each of the decks of cards. Obviously, (X_k^1) is a random walk driven by the random to top measure p.

TS[12] Where should these quotations be placed?

Editor's or typesetter's annotations (will be removed before the final TEX run)

The same is true for X_n^2 because choosing a position in the deck uniformly at random is equivalent to choosing the face value of a card uniformly at random. Say we have a match if a card value has the same position in both decks. This coupling has the following property: any checked card stays matched with its sister card for ever and each time an unchecked card is touched in the first deck, it is checked and matched with its sister card. Note however that matches involving an unchecked card from the first deck might be broken along the way. In any case, the coupling time T is always less or equal to T', the first time all cards in the first deck have been checked. A simple application of the well-known coupon collector's problem gives $\mathbb{P}(T' > k) \leq n\mathrm{e}^{-k/n}$. This, combined with a matching lower bound result, shows that random to top (and also top to random) mixes in about $n \log n$ shuffles, a result which compares well with the very precise result of Theorem 3.2.

Coupling for random transposition [1, 27]$_{\text{TS}}^{12}$**.** For n cards, the random transposition shuffle involves choosing a pair of positions (i, j) uniformly and independently at random in $\{1, \ldots, n\}$ and switching the cards at these positions. Thus, the random transposition measure p_{RT} is given by

$$p_{\text{RT}}(\tau) = \begin{cases} 2/n^2 & \text{if } \tau = (i,j),\ 1 \leq i < j \leq n, \\ 1/n & \text{if } \tau = e, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Obviously, choosing uniformly and independently at random a position i and a face value V and switching the card in position i with the card with face value V gives an equivalent description of this measure. Given two decks, we construct a coupling by picking i and V uniformly and independently. In each deck, we transpose the card in position i with the card with face value V. In this way, the number of matches never goes down and at least one new match is created each time the cards with the randomly chosen face value V are in different positions in the two decks and the cards in the randomly chosen position i have distinct face values. Let (Z_k) denote the Markov process on $\{0, \ldots, n\}$ started at n with transition probabilities $K(i, i-1) = (i/n)^2$, $K(i, i) = 1 - (i/n)^2$. Let $T' = \inf\{k : Z_k = 0\}$. Then, it is not hard to see that $\mathbb{E}(T) \leq \mathbb{E}(T') \leq 2n^2$ where T is the coupling time. By Theorem 4.2, we obtain $d_{\text{TV}}(p_{\text{RT}}^{(k)}, u) \leq \mathbb{E}(T)/k \leq 2n^2/k$ and the subadditivity of $k \mapsto 2d_{\text{TV}}(p_{\text{RT}}^{(k)}, u)$ yields $d_{\text{TV}}(p_{\text{RT}}^{(k)}, u) \leq \mathrm{e}^{1-k/(12n^2)}$. This shows that $T(S_n, p_{\text{RT}}) \leq 36n^2$. Theorem 9.2 below states that (S_n, p_{RT}) presents a cut-off at time $t_n = \frac{1}{2} n \log n$. Convergence after order n^2 steps is the best that has been proved for random transposition using coupling.

Coupling for adjacent transposition [1]$_{\text{TS}}^{12}$**.** Consider now the shuffling scheme where a pair of adjacent cards are chosen at random and switched. The adjacent transposition measure on S_n, call it p_{AT}, is the uniform measure on $\{e, (1,2), \ldots, (n-1, n)\}$. Set $\sigma_0 = e$ and $\sigma_i = (i, i+1)$, $1 \leq i < n$. To construct a coupling, consider two decks of cards. Call A the set containing 0 and all

positions $j \in \{1, \dots, n-1\}$ such that neither the cards in position j nor the cards in position $j+1$ are matched in those decks. List A as $\{j_0, j_1, \dots, j_\ell\}$ in order. Let J be a uniform random variable in $\{0, \dots, n-1\}$ and set

$$J^* = \begin{cases} J & \text{if } J \notin A \\ j_{k+1} & \text{if } J = j_k \in A \text{ with the convention that } \ell + 1 = 0. \end{cases}$$

The coupling is produced by applying σ_J to the first deck and σ_{J^*} to the second deck. As J^* is uniform in $\{0, \dots, n-1\}$, this indeed is a coupling. To analyze the coupling time, observe that matches cannot be destroyed and that, for any face value, the two cards with this face value always keep the same relative order (e.g., if the ace of spades is higher in the first deck than in the second deck when we start, this stays the same until they are matched. Call T_i' the first time card i reaches the bottom of the deck (in the deck in which this card is initially higher) and set $T' = \max_i\{T_i'\}$. Then the coupling time T is bounded above by T'. Finally, any single card performs a symmetric simple random walk on $\{1, \dots, n\}$ with holding probability $1 - 2/n$ except at the endpoints where the holding probability is $1 - 1/n$. Properly rescaled, this process converges weakly to reflected Brownian motion on $[0, 1]$ and the hitting time of 1 starting from any given point can be analyzed. In particular, there are constants $A, a > 0$ such that, for any i and any $s > 0$, $\mathbb{P}(T_i' > sn^3) \le A\mathrm{e}^{-as}$. Hence, for C large enough, $\mathbb{P}(T > Cn^3 \log n) \le An\mathrm{e}^{-aC \log n} \le (2\mathrm{e})^{-1}$. This shows that $T(S_n, p_{\text{AT}}) \le Cn^3 \log n$. A matching lower bound is given at the end of Section 5.3. Hence (S_n, p_{AT}) presents a precut-off at time $t_n = n^3 \log n$. See also Theorem 10.4 and [141].

Other couplings. Here we briefly describe further examples of random walks for which reasonably good couplings are known:

- Simple random walk on the hypercube $\{0, 1\}^n$ as described in Section 8.2. See [1, 27, 105].
- The GSR riffle shuffle described in Section 3.2. See [1] for a coupling showing that $2 \log_2 n$ riffle shuffles suffice to mix up n cards.
- Overhand shuffles [1, 116]. An overhand shuffle is a shuffle where the deck is divided into k blocks and the order of the blocks are reversed. Pemantle [116] gives a coupling analysis of a range of overhand shuffle models showing that, in many reasonable cases, order $n^2 \log n$ shuffles suffice to mix up n cards whereas at least order n^2 are necessary. Note however that the crude overhand shuffle discussed in Section 3.1 has a precut-off at time $t_n = n \log n$.
- The following shuffling method is one of those discussed in Borel and Chéron [15]: take the top card and insert it at random, take the bottom card and insert it a random. The coupling described above for random to top can readily be adapted to this case. See [1, 27].
- Slow shuffles. At each step, either stay put or transpose the top two cards or move the top card to the bottom, each with probability 1/3. It is not

hard to construct a coupling showing that order $n^3 \log n$ shuffles suffice to mix up the cards using this procedure. Rudvalis (see [27, p. 90]) proposed another shuffle as a candidate for the slowest shuffle. At each step, move the top card either to bottom or second to bottom each with probability 1/2. Hildebrand gives a coupling for this shuffle in his Ph. D Thesis [85] and shows that order $n^3 \log n$ such shuffles suffice. For these slow shuffles and related variants, Wilson [142] proves that order $n^3 \log n$ shuffles are necessary to mix up n cards.

4.2 Strong Stationary Time

Separation. Given a Markov kernel K with invariant distribution π on a finite set $\mathcal{X}$, set

$$\mathbf{sep}_K(x,n) = \max_{y\in\mathcal{X}} \left(1 - \frac{K_n(x,y)}{\pi(y)}\right), \quad \mathbf{sep}_K(n) = \max_{x\in\mathcal{X}} \mathbf{sep}_K(x,n).$$

The quantity $\mathbf{sep}(n) = \mathbf{sep}_K(n)$ is called the *maximal separation* between K^n and π. As

$$d_{\mathrm{TV}}(K^n(x,\cdot),\pi) = \sum_{y:K^n(x,y)\le\pi(y)} (\pi(y) - K^n(x,y)),$$

it is easy to see that $d_{\mathrm{TV}}(K^n(x,\cdot),\pi) \le \mathbf{sep}_K(x,n)$. Thus separation always controls the total variation distance. Separation is an interesting alternative way to measure ergodicity. The function $n \mapsto \mathbf{sep}(n)$ is non-increasing and sub-multiplicative [3, 5]. As an immediate application of these elementary facts, one obtains the following Doeblin's type result: Assume that there exist an integer m and a real $c > 0$ such that, for all $x, y \in \mathbb{X}$, $K^m(x,y) \ge c\pi(y)$. Then $d_{\mathrm{TV}}(K^{nm}(x,\cdot),\pi) \le \mathbf{sep}(nm) \le (1-c)^n$ (this line of reasoning produces very poor bounds in general but an example where it is useful is given in [39]).

Let (X_k) be a Markov chain with kernel K. A *strong stationary time* is a randomized stopping time T for (X_k) such that

$$\forall\, k,\ \ \forall\, y \in \mathcal{X},\ \ \mathbb{P}(X_k = y/T = k) = \pi(y). \tag{4.2}$$

This is equivalent to say that X_T has distribution π and that the random variables T and X_T are independent. For a discussion of the relation between strong stationary time and coupling, see [5]. Relations between strong stationary time and eigenvalues are explored in [107]. Strong stationary times are related to the separation distance by the following theorem of Aldous and Diaconis [5, 3, 27].

Theorem 4.3. *Let T be a strong stationary time for the chain starting at $x \in \mathcal{X}$. Then*

$$\forall\, n, \quad \mathbf{sep}_K(x,n) \le \mathbb{P}_x(T > n).$$

Moreover there exists a strong stationary time such that the above inequality is an equality.

Separation for random walks. In the case of random walks on finite groups, separation becomes

$$\mathbf{sep}(k) = \mathbf{sep}_p(k) = \max_{x\in G}\left(1 - |G|p^{(k)}(x)\right).$$

The next theorem restates the first part of Theorem 4.3 and gives an additional result comparing separation and total variation distances in the context of random walks on finite groups. See [5] and the improvement in [23].

Theorem 4.4. *Let p be a probability measure on a finite group G. Then*

$$d_{\mathrm{TV}}(p^{(k)}, u) \le \mathbf{sep}(k)$$

and, provided $d_{\mathrm{TV}}(p^{(k)}, u) \le (|G|-1)/(2|G|)$,

$$\mathbf{sep}(2k) \le 2d_{\mathrm{TV}}(p^{(k)}, u).$$

Let T be a strong stationary time for the associated random walk starting at the identity e. Then

$$d_{\mathrm{TV}}(p^{(k)}, u) \le \mathbf{sep}(k) \le \mathbb{P}_e(T > k).$$

One can easily introduce the notion of separation cut-off (and precut-off): The family $((G_n, p_n))$ has a separation cut-off if and only if there exists a sequence s_n tending to infinity such that

$$\lim_{n\to\infty} \mathbf{sep}_{p_n}(\lfloor(1-\varepsilon)s_n\rfloor) = 1, \quad \lim_{n\to\infty} \mathbf{sep}_{p_n}(\lfloor(1+\varepsilon)s_n\rfloor) = 0.$$

Theorem 4.4 implies that if $((G_n, p_n))$ has both a total variation cut-off at time t_n and a separation cut-off at time s_n then $t_n \le s_n \le 2t_n$.

There is sometimes an easy way to decide whether a given strong stationary time is optimal (see [33, Remark 2.39]).

Definition 4.5. *Given an ergodic random walk (X_n) on G started at e and a strong stationary time T for (X_n), the group element x is called a* halting state *if $\mathbb{P}_e(X_k = x, T > k) = 0$, for all $k = 0, 1, \ldots$.*

Hence, a halting state is an element that cannot be reached before the strong stationary time T (observe that, of course, $\mathbb{P}_e(X_T = x) > 0$). Obviously, if there is a halting state, then T is a stochastically smallest possible strong stationary time. As for coupling, the power of strong stationary times is best understood by looking at examples.

Stationary time for top to random [27]$_{\mathbf{TS}}^{\mathbf{12}}$. Let q_1 denote the top to random measure on S_n. Consider the first time T_1 a card is inserted under the bottom card. This is a geometric waiting time with mean n. Consider the first time T_2 a second card is inserted under the original bottom card. Obviously $T_2 - T_1$ is a geometric waiting time with mean $n/2$, independent of

T_1. Moreover, the relative position of the two cards under the original bottom card is equally likely to be high-low or low-high. Pursuing this analysis, we discover that the first time T the bottom card comes on top and is inserted at random is a strong stationary time. Moreover $T = T_n = T_1 + (T_2 - T_1) + \cdots + (T_n - T_{n-1})$ where $T_i - T_{i-1}$ are independent geometric waiting time with respective means n/i. Hence $\mathbb{P}_e(T > k)$ can be estimated. In particular, it is bounded by $n\mathrm{e}^{-k/n}$. Hence Theorem 4.4 gives

$$d_{TV}(q_1^{(k)}, u) \leq \mathbf{sep}(k) \leq \mathbb{P}_e(T > k) \leq n\mathrm{e}^{-k/n}.$$

This is exactly the same bound as provided by the coupling argument described earlier. In fact, in this example, the coupling outlined earlier and the stationary time T above are essentially equivalent. This T is not an optimal stationary time but close. Let T' be the first time the card originally second to bottom comes to the top and is inserted. This T' is an optimal stationary time. It has a halting state: the permutation corresponding to the deck in exact reverse order. This example has both a total variation and a separation cut-off at time $t_n = n \log n$.

Stationary time for random transposition [27]$_{\text{TS}}^{12}$. We describe a strong stationary time constructed by A. Broder. Variants are discussed in [27, 106]. The construction involves checking the back of the cards as they are shuffled using repeated random transpositions. Recall that the random transposition measure p_{RT} defined at (4.1) can be described by letting the left and right hands choose cards uniformly and independently at random. If either both hands touch the same unchecked card or if the card touched by the left hand is unchecked and the card touched by the right hand is checked then check the back of the card touched by the left hand. Let T be the time that only one card remains unchecked. The claim is that T is a strong stationary time. See [27] for details. This stationary time has mean $2n \log n + O(\log n)$ and can be used to show that a little over $2n \log n$ random transpositions suffices to mix up a deck of n cards. This is better than what is obtained by the best known coupling, i.e., n^2. Theorem 9.2 and Matthews [106] show that (S_n, p_{RT}) has a total variation cut-off as well as a separation cut-off at time $\frac{1}{2} n \log n$.

Stationary time for riffle shuffle [27]$_{\text{TS}}^{12}$. Recall that the inverse of a riffle suffle can be described as follows. Consider a binary vector of length n whose entries are independent uniform $\{0, 1\}$-random variables. Sort the deck from bottom to top into a left pile and a right pile by using the above binary vector with 0 sending the card left and 1 sending the card right. When this is done, put the left pile on top of the right to obtain a new deck. A sequence of k inverse riffle shuffles can be described by a binary matrix with n rows and k columns where the (i, j)-entry describes what happens to the original i-th card during the j-th shuffle. Thus the i-th row describes in which pile the original i-th card falls at each of the k shuffles.

Let T be the first time the matrix above has distinct rows. Then T is a strong stationary time. Indeed, using the right to left lexicographic order on binary vectors, after any number of shuffles, cards with "small" binary vectors are on top of cards with "large" binary vectors. At time T all the rows are distinct and the lexicographic order sorts out the cards and describes uniquely the state of the deck. Because the entries are independent uniform $\{0,1\}$-variables, at time T, all deck arrangements are equally likely. Moreover, the chance that $T > k$ is the same as the probability that dropping n balls into 2^k boxes there is no box containing two or more balls. This is the same as the birthday problem and we have

$$\mathbb{P}_e(T > k) = 1 - \prod_1^{n-1} (1 - i2^{-k}).$$

Using Calculus, this proves a separation cut-off at time $2\log_2 n$. Indeed, this stationary time has a halting state: the deck in reverse order. Theorem 3.1 proves a variation distance cut-off at time $\frac{3}{2}\log_2 n$. See [1, 13, 27].

Stationary time on nilpotent groups. In his thesis [112], Pak used strong stationary times skillfully to study problems that are somewhat different from those discussed above. The papers [7, 21, 114] develop results for nilpotent groups (for a definition, see Section 7 below). Here is a typical example. Let $U_m(q)$ denote the group of all upper-triangular matrices with 1 on the diagonal and coefficients mod q where q is an odd prime. Let $E_{i,j}(a)$, $1 \le i < j \le m$, denote the matrix in $U_m(q)$ whose non-diagonal entries are all 0 except the (i,j)-entry which equals a. The matrices $E_{i,i+1}(1)$, $1 \le i < m$, generate $U_m(q)$. Consider the following two sets

$$\Sigma_1 = \{E_{i,i+1}(a) : a \in \mathbb{Z}_q,\ 1 \le i < m\}$$
$$\Sigma_2 = \{E_{i,j}(a) : a \in \mathbb{Z}_q,\ 1 \le i < j \le m\}.$$

and let p_1, p_2 denote the uniform probability on Σ_1, Σ_2 respectively. The article [114] uses the strong stationary time technique to prove that the walk driven by p_2 presents a precut-off at time $t_m = m^2 \log m$, uniformly in the two parameters m, q. In particular, there are constants C, c such that

$$cm^2 \log m \le T(U_m(q), p_2) \le Cm^2 \log m.$$

The results for the walk driven by p_1 are less satisfactory. In [21], the authors use a strong stationary time to show that if $q \gg m^2$ then

$$cm^2 \le T(U_m(q), p_1) \le Cm^2.$$

The best known result for fixed q is described in Section 7 below and says that $T(U_m(q), p_1) \le Cm^3$.

Stopping time and semidirect products. In his thesis [138], Uyemura-Reyes develops a technique for walks on semidirect products which is closely related to the strong stationary time idea. Let H, K be two finite groups and $\phi : k \mapsto \phi_k$ a homomorphism from K to the automorphism group of H. The *semidirect product* $H \rtimes_\phi K$ is the group whose underlying set is $H \times K$ and whose product law is $(h_1, k_1)(h_2, k_2) = (h_1\phi_{k_1}(h_2), k_1k_2)$. By construction, H is normal in $H \rtimes_\phi K$. It follows that there is a natural projection from $H \rtimes_\phi K$ onto $K \cong (H \rtimes_\phi K)/H$. If p is a probability measure on $H \rtimes_\phi K$, let p_K denote its projection on K. Let (X_n) be the random walk on $H \rtimes_\phi K$ driven by p and write $X_n = (\zeta_n, \xi_n)$ with $\zeta_n \in H, \xi_n \in K$. Then (ξ_n) is a random walk on K driven by p_K. Consider a stopping time T for (X_n) which satisfies

$$\mathbb{P}_e(\zeta_n = h, \xi_n = k/T \le n) = \frac{1}{|H|}\mathbb{P}_e(\xi_n = k/T \le n). \tag{4.3}$$

Theorem 4.6. *Referring to the notation introduced above, let (X_n) be the random walk on $G = H \rtimes_\phi K$ driven by p and starting at the identity. Assume that T is a stopping time satisfying* (4.3)*. Then*

$$\|p^{(n)} - u_G\|_{\text{TV}} \le \|p_K^{(n)} - u_K\|_{\text{TV}} + 2\mathbb{P}_e(T > n).$$

Moreover,

$$\mathbf{sep}_p(n) \le \mathbf{sep}_{p_K}(n) + |K|\mathbb{P}_e(T > n).$$

We now describe two applications taken from [138]. See [77] for related results. Let $G = \mathbb{Z}_b^d \rtimes \mathbb{Z}_d$ where the action of $\mathbb{Z}_d$ is by circular shift of the coordinates in $\mathbb{Z}_b^d$. When $b = 2$, this example has a card shuffling interpretation. Given a deck of $2n$ cards, there are exactly two different perfect shuffles: cut the deck into two equal parts and interlace the two heaps starting either from the left or the right heap. When $2n = 2^d$ for some d, the subgroup of S_{2n} generated by the two perfect shuffles is isomorphic to $G = \mathbb{Z}_2^d \rtimes \mathbb{Z}_d$. One of the shuffles can be interpreted as $g_1 = (0, 1)$ and the other as $g_2(1_1, 1)$ where $0 = (0, \ldots, 0)$ and $1_1 = (1, 0, \ldots, 0)$ in $\mathbb{Z}_2^d$. Consider the simple random walk on $G = \mathbb{Z}_2^d \rtimes \mathbb{Z}_d$ driven by the probability p with $p(e) = 2p(g_1) = 2p(g_2) = 1/2$. Theorem 4.6 can be used to prove that $T(\mathbb{Z}_2^d \rtimes \mathbb{Z}_d, p) \le Cd^2$ ([138] also gives a matching lower bound).

For a second example, take $b = d$ and consider the probability measure p defined by $p(0, 0) = p(\pm 1_1, 0) = p(0, \pm 1) = p(\pm 1_1, 1) = p(\pm 1_1, -1) = 1/9$. Uyemura-Reyes uses Theorem 4.6 to prove the mixing time upper bound $T(\mathbb{Z}_d^d \rtimes \mathbb{Z}_d, p) \le Cd^3 \log d$. He also derives a lower bound of order d^3.

5 Spectrum and Singular Values

5.1 General Finite Markov Chains

Diagonalization. Let K be a Markov kernel with invariant distribution π on a finite set $\mathcal{X}$. Irreducibility and aperiodicity can be characterized in

terms of the spectrum of K on $L^2(\pi)$ where $L^2(\pi)$ denote the space of all complex valued functions equipped with the Hermitian scalar product $\langle f, g\rangle_\pi = \sum_x f(x)\overline{g(x)}\pi(x)$. Indeed, K is irreducible if and only if 1 is a simple eigenvalue whereas K is aperiodic if and only if any eigenvalue $\beta \neq 1$ satisfies $|\beta| < 1$.

If K and K^* commute, that is, if K viewed as an operator on $L^2(\pi)$ is normal, then K is diagonalizable in an orthonormal basis of $L^2(\pi)$. Let $(\beta_i)_{i\geq 0}$ be an enumeration of the eigenvalues, each repeated according to its multiplicity and let $(v_i)_{i\geq 0}$ be a corresponding orthonormal basis of eigenvectors. Note that in general, the β_i are complex numbers and the v_i complex valued functions. Without loss of generality, we assume that $\beta_0 = 1$ and $u_0 \equiv 1$. Then

$$\frac{K_n(x,y)}{\pi(y)} = \sum_{i\geq 0} \beta_i^n v_i(x)\overline{v_i(y)} \tag{5.1}$$

and

$$d_{\pi,2}(K_n(x,\cdot),\pi)^2 = \sum_{i\geq 1} |\beta_i|^{2n}|v_i(x)|^2. \tag{5.2}$$

Let us describe a simple but useful consequence of (5.2) concerning the comparison of the $L^2(\pi)$-distances to stationarity of the discrete and continuous Markov processes associated to a given reversible Markov kernel K. An application is given below at the end of Section 8.2.

Theorem 5.1. *Let (K,π) be a reversible Markov kernel on a finite set $\mathcal{X}$ and let H_t be as in* (2.3)*. Then*

$$d_{\pi,2}(K_n(x,\cdot),\pi)^2 \leq \beta_-^{2n_1}(1 + d_{\pi,2}(H_{n_2}(x,\cdot),\pi)^2) + d_{\pi,2}(H_n(x,\cdot),\pi)^2$$

where $n = n_1 + n_2 + 1$ and $\beta_- = \max\{0, -\beta_{\min}\}$, $\beta_{\min}$ being the smallest eigenvalue of K. Moreover,

$$d_{\pi,2}(H_{2n}(x,\cdot),\pi)^2 \leq (\pi(x)^{-1} - 1)\mathrm{e}^{-2n} + d_{\pi,2}(K_n(x,\cdot),\pi)^2.$$

Proof. The idea behind this theorem is simple: as (K,π) is reversible, it has real eigenvalues $1 = \beta_0 \geq \beta_1 \geq \cdots \geq \beta_{|\mathcal{X}|-1} \geq -1$. Viewed as an operator, H_t is given by $H_t = \mathrm{e}^{-t(I-K)}$ and has real eigenvalues $\mathrm{e}^{-t(1-\beta_i)}$, in increasing order, associated with the same eigenvectors as for K. Hence, using (5.2) and the similar formula for H_t, the statements of Theorem 5.1 follow from simple Calculus inequalities. See Lemma 3 and Lemma 6 in [42] for details. The factor $\pi(x)^{-1}$ appears because, using the same notation as in (5.2), we have $\sum_{i\geq 0}|v_i(x)|^2 = \pi(x)^{-1}$. □

Poincaré inequality. When (K,π) is reversible, an important classical tool to bound eigenvalues is the variational characterization of the first eigenvalue. Set

$$\mathcal{E}(f,g) = \langle (I-K)f, g\rangle_\pi = \sum_x [(I-K)f(x)]g(x)\pi(x). \tag{5.3}$$

This form is called the *Dirichlet form* associated to (K, π). A simple computation shows that

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{x,y}(f(x)-f(y))(g(x)-g(y))\pi(x)K(x,y). \tag{5.4}$$

Restricting attention to the orthogonal of the constant functions, we see that

$$\lambda_1 = 1-\beta_1 = \inf\left\{\frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)} : f \in L^2(\pi),\ \mathrm{Var}_\pi(f) \neq 0\right\} \tag{5.5}$$

where $\mathrm{Var}_\pi(f)$ denote the variance of f with respect to π, that is,

$$\mathrm{Var}_\pi(f) = \pi(f^2) - \pi(f)^2 = \frac{1}{2}\sum_{x,y}|f(x)-f(y)|^2\pi(x)\pi(y). \tag{5.6}$$

It follows that, for any $A \geq 1$, the inequality $\beta_1 \leq 1 - 1/A$ is equivalent to the so-called *Poincaré inequality*

$$\mathrm{Var}_\pi(f) \leq A\ \mathcal{E}(f,f).$$

The quantity $\lambda_1 = 1-\beta_1$ is called the *spectral gap* of (K,π). It is the second smallest eigenvalue of $I-K$. Some authors call $1/\lambda_1$ the *relaxation time*. It is a widespread misconception that the relaxation time contains all the information one needs to have good control on the convergence of a reversible Markov chain. What λ_1 gives is only the asymptotic exponential rate of convergence of $H_t - \pi$ to 0 as t tends to infinity.

Singular values. When K and its adjoint K^* do not commute, it seems hard to use the spectrum of K to get quantitative information on the convergence of $K_n(x,\cdot)$ to π. However, the *singular values* of K can be useful. For background on singular values, see [91, Chap. 18]. Consider the operators KK^* and K^*K. Both are self-adjoint on $L^2(\pi)$ and have the same eigenvalues, all non-negative. Denote the eigenvalues of K^*K in non-increasing order and repeated according to multiplicity by

$$\sigma_0^2 = 1 \geq \sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_{|\mathcal{X}|-1}^2$$

with $\sigma_i \geq 0$, $0 \leq i \leq |\mathcal{X}|-1$. Then, the non-negative reals σ_i are called the singular values of K. More generally, for each integer j, denote by $\sigma_i(j)$, $0 \leq i \leq |\mathcal{X}|-1$ the singular values of K^j and let also $v_{i,j}$ be the associated normalized eigenfunctions. Then we have

$$d_{\pi,2}(K_n(x,\cdot),\pi)^2 = \sum_{i\geq 1}\sigma_i(n)^2|v_{i,n}(x)|^2. \tag{5.7}$$

As $\sum_{i\geq 0}|v_{i,j}(x)|^2 = \pi(x)^{-1}$ and $\sigma_i(n) \leq \sigma_1(n) \leq \sigma_1^n$ (see [91, Th. 3.3.14]), we obtain

$$\forall\, n \in \mathbb{N},\quad d_{\pi,2}(K_n(x,\cdot),\pi)^2 \leq \left(\pi(x)^{-1}-1\right)\sigma_1^{2n}.$$

Let us emphasize here that it may well be that $\sigma_1 = 1$ even when K is ergodic. In such cases one may try to save the day by using the singular values of K^j where j is the smallest integer such that $\sigma_1(j) < 1$. This works well as long as j is relatively small. We will see below in Theorem 5.3 how to use all the singular values of K (or K^j) in the random walk case.

5.2 The Random Walk Case

Let us now return to the case of a left-invariant random walk driven by a probability measure p on a group G, i.e., the case when $K(x,y) = p(x^{-1}y)$ and $\pi = u$. In this case an important simplification occurs because, by left-invariance, the left-hand side of both (5.2) and (5.7) are independent of x. Averaging over $x \in G$ and using the fact that our eigenvectors are normalized in $L^2(G)$, we obtain the following.

Theorem 5.2. *Let p a probability measure on a finite group G. Assume that $p * \check{p} = \check{p} * p$, then we have*

$$d_2(p^{(n)}, u)^2 = \sum_{i \geq 1} |\beta_i|^{2n} \tag{5.8}$$

where β_i, $0 \leq i \leq |G|-1$ are the eigenvalues associated to $K(x,y) = p(x^{-1}y)$ as above. In particular, if $\beta_* = \max\{|\beta_i| : i = 1, \ldots, |G|-1\}$ denotes the second largest eigenvalue in modulus, we have

$$d_2(p^{(n)}, u)^2 \leq (|G|-1)\beta_*^{2n}. \tag{5.9}$$

Note that p and $\check{p}$ always commute on abelian groups. Sections 6 and 10 below discuss techniques leading to eigenvalues estimates.

Theorem 5.3. *Let p a probability measure on a finite group G. Then, for any integers n, m we have*

$$d_2(p^{(nm)}, u)^2 \leq \sum_{i \geq 1} \sigma_i(m)^{2n} \tag{5.10}$$

where $\sigma_i(m)$, $0 \leq i \leq |G|-1$ are the singular values associated to $K_m(x,y) = p^{(m)}(x^{-1}y)$ in non-increasing order. In particular, for each m, we have

$$d_2(p^{(nm)}, u)^2 \leq (|G|-1)\sigma_1(m)^{2n}. \tag{5.11}$$

Proof. Use (5.7) and the fact (see e.g., [91, Th. 3.3.14]) that, for all k, n, m,

$$\sum_0^k \sigma(nm)^2 \leq \sum_0^k \sigma_i(m)^{2n}.$$

□

It is worth restating (5.10) as follows.

Theorem 5.4. *Let p a probability measure on a finite group G and let q_m denote either $\check{q}^{(m)} * q^{(m)}$ or $q^{(m)} * \check{q}^{(m)}$. Then*

$$d_2(p^{(nm)}, u) \leq d_2(q_m^{(\lfloor n/2 \rfloor)}, u).$$

For applications of Theorem 5.4, see Section 10.3.

Let us point out that the fact that (5.8) and (5.10) do not involve eigenfunctions is what makes eigenvalue and comparison techniques (see Section 10) so powerful when applied to random walks on finite groups. For more general Markov chains, the presence of eigenfunctions in (5.2) and (5.7) make them hard to use and one often needs to rely on more sophisticated tools such as Nash and logarithmic Sobolev inequalities. See, e.g., [3, 47, 48, 124] and Martinelli's article in this volume.

5.3 Lower Bounds

This section discusses lower bounds in total variation. The simplest yet useful such lower bound follows from a direct counting argument: Suppose the probability p has a support of size at most r. Then $p^{(k)}$ is supported on at most r^k elements. If k is too small, not enough elements have possibly been visited to have a small variation distance with the uniform probability on G. Namely,

$$\|p^{(k)} - u\|_{\text{TV}} \geq 1 - r^k/|G| \tag{5.12}$$

which gives

$$T(G, p) \geq \frac{\log(|G|/2)}{\log r}.$$

Useful improvements on this bound can be obtain if one has further information concerning the group law, for instance if G is abelian or if many of the generators commutes. See, e.g., [56] and [19].

Generally, lower bounds on total variation are derived by using specific test sets or test functions. For instance, for random transposition and for transpose top and random on the symmetric group, looking at the number of fixed points yields sharp lower bounds in total variation, see [27, p. 43]. For random transvection on $SL_n(\mathbb{F}_q)$, the dimension of the space of fixed vectors can be used instead [86].

Eigenvalues and eigenfunctions can also be useful in proving lower bounds on $d_2(p^{(k)}, u)$ and, more surprisingly, on $\|p^{(k)} - u\|_{\text{TV}}$. Start with the following two simple observations.

Proposition 5.5. *Let p be a probability measure on a finite group G. Assume that β is an eigenvalue of p with multiplicity m. Then*

$$d_2(p^{(k)}, u)^2 \geq m|\beta|^{2k}, \quad 2\|p^{(k)} - u\|_{\text{TV}} \geq |\beta|^k.$$

Proof. Let V be the eigenspace of β, of dimension m. It is not hard to show that V contains a function ϕ, normalized by $\|\phi\|_2 = 1$ and such that $\phi(e) = \|\phi\|_\infty \geq \sqrt{m}$. See [20, p. 103]. Then $d_2(p^{(k)}, u) \geq |\langle p^{(k)} - u, \phi\rangle| = |\phi * \check{p}^{(k)}(e)| = |\beta|^k |\phi(e)| = |\beta|^k \sqrt{m}$. For the total variation lower bound, use the last expression in (2.8) with any β-eigenfunction as a test function. □

Note that it is not uncommon for random walks on groups to have eigenvalues with high multiplicity. Both of the inequalities in Proposition 5.5 are sharp as k tends to infinity when β is the second largest eigenvalue in modulus. However, the first inequality often gives good lower bound on the smallest k such that $d_2(p^{(k)}, u) \leq \varepsilon$ for fixed ε whereas the second inequality seldom does for the similar question in total variation (the walk on the hypercube of Theorem 8.7 illustrates this point). The following proposition can often be used to obtain improved total variation lower bounds. It is implicit in [27] and in [141]. See also [123, 126].

Proposition 5.6. *Let β be an eigenvalue of p. Let ϕ be an eigenfunction associated to β. Let B_k be such that*

$$\forall\, k, \quad \mathrm{Var}_{p^{(k)}}(\phi) \leq B_k^2. \tag{5.13}$$

Then $\|p^{(k)} - u\|_{\mathrm{TV}} \geq 1 - \tau$ for any $\tau \in (0,1)$ and any integer k such that

$$k \leq \frac{1}{-2\log|\beta|} \log \frac{\tau|\phi(e)|^2}{4(\|\phi\|_2^2 + B_k^2)}.$$

The difficulty in applying this proposition is twofold. First, one must choose a good eigenfunction ϕ maximizing the ratio $\phi(e)^2/(\|\phi\|_2^2 + B_k^2)$. Second, one must prove the necessary bound (5.13) with good B_k's (e.g., B_k uniformly bounded) and this turns out to be a rather non-trivial task. Indeed, it involves taking advantage of huge cancellations in $\mathrm{Var}_{p^{(k)}}(\phi) = p^{(k)}(|\phi|^2) - |p^{(k)}(\phi)|^2$. In this direction, the following immediate proposition is much more useful than it might appear at first sight.

Proposition 5.7. *Let β, ϕ be as in* Proposition 5.6. *Assume that there are eigenvalues α_i and associated eigenfunctions ψ_i, $i \in I$, relative to p such that*

$$|\phi|^2 = \sum_{i\in I} a_i \psi_i.$$

Then

$$\mathrm{Var}_{p^{(k)}}(\phi) = \sum_{i\in I} a_i \alpha_i^k \psi_i(e) - |\beta|^{2k}|\phi(e)|^2.$$

The reason this is useful is that, in some cases, expanding $|\phi|^2$ along eigenfunctions requires only a few eigenfunctions which, in some sense, are close to ϕ. To see how this works, consider the simple random walk on the hypercube $G = \mathbb{Z}_2^d$ equipped with its natural set of generators $(e_i)_1^d$ where e_i is the

d-tuple with all entries zero except the i-th equal to 1. See [27, pg. 28-29]. To avoid periodicity, set $e_0 = (0, \ldots, 0)$ and consider the measure p given by

$$p(x) = \begin{cases} 1/(d+1) & \text{if } x = e_i \text{ for some } i \in \{0, \ldots, d\} \\ 0 & \text{otherwise.} \end{cases} \tag{5.14}$$

Denote by x_i the coordinates of $x \in \mathbb{Z}_2^d$. Then $(-1)^{x_i} = 1 - 2x_i$ is an eigenfunction with eigenvalue $1 - 2/(d+1)$ for each $i \in \{1, \ldots, d\}$ and so is $\phi(x) = \sum_1^d (-1)^{x_i} = 2|x| - d$ where $|x| = \sum_1^d x_i$. Now

$$|\phi(x)|^2 = d + 2 \sum_{1 \le i < j \le n} (-1)^{x_i + x_j} = d\psi_0(x) + 2\psi_2(x)$$

where $\psi_0 \equiv 1$ and $\psi_2 = \sum_{1 \le i < j \le n} (-1)^{x_i + x_j}$ are eigenfunctions with respective eigenvalues 1 and $1 - 4/(d+1)$. Hence

$$\mathrm{Var}_{p^{(k)}}(\phi) = d + d(d-1)\left(1 - \frac{4}{d+1}\right)^k - d^2\left(1 - \frac{2}{d+1}\right)^{2k}.$$

By careful inspection, for any integer k, the right-hand side is less than d. Using this in Proposition 5.6 shows that, for the simple random walk on the hypercube, $\|p^{(k)} - u\|_{TV} \ge 1 - \tau$ for $k \le \frac{1}{4} d \log(\tau d)$. This is sharp since the simple random walk on the hypercube has a cut-off at time $t_d = \frac{1}{4} d \log d$. See Theorem 8.7 below.

The next theorem and its illustrative example are taken from [141]. See also [126]. Set

$$\nabla\phi(x) = \left(\frac{1}{2} \sum_y |f(x) - f(xy)|^2 p(y)\right)^{1/2}.$$

Theorem 5.8. *Let β, ϕ be as in Proposition* 5.6. *Then*

$$\mathrm{Var}_{p^{(k)}}(\phi) \le \frac{2\|\nabla\phi\|_\infty^2}{1 - |\beta|^2}. \tag{5.15}$$

Moreover $\|p^{(k)} - u\|_{\mathrm{TV}} \ge 1 - \tau$ for all $\tau \in (0,1)$ and all k such that

$$k \le \frac{1}{2\log|\beta|} \log\left(\frac{\tau(1 - |\beta|^2)|\phi(e)|^2}{4(2 + |\beta|)\|\nabla\phi\|_\infty^2}\right).$$

As an example, consider the random adjacent transposition measure p_{AT}, i.e., the uniform measure on $\{e, (1,2), \ldots, (n, n-1)\} \subset S_n$. To find some eigenfunctions, consider how one given card moves, say card 1. It essentially performs a ± 1 random walk on $\{1, \ldots, n\}$ with holding $1/2$ at the endpoints. For this random walk, $v(j) = \cos[\pi(j - 1/2)/n]$ is an eigenfunction associated with the eigenvalue $\cos \pi/n$. For $\ell \in \{1, \ldots, n\}$, let $\ell(x)$ be the position of card ℓ in

the permutation x and $v_\ell(x) = v(\ell(x))$. Then, each v_ℓ is an eigenfunction of p with eigenvalue $1 - (2/n)(1 - \cos \pi/n)$. This is actually the second largest eigenvalue, see [12]. Obviously, the function

$$\phi(x) = \sum_{\ell=1}^{n} v_\ell(e) v_\ell(x)$$

is an eigenfunction for the same eigenvalue. Moreover, $\|\nabla\phi\|_\infty^2 \le 2\pi^2\phi(e)/n^3$ and $\phi(e) = n(1+o(1))$. Hence for all $\tau \in (0,1)$ and $k \le (1-o(1))\pi^{-2}n^3 \log \tau n$, Theorem 5.8 gives $\|p_{\text{AT}} - u\|_{\text{TV}} \ge 1 - \tau$. This is quite sharp since it is known that $T(S_n, p_{\text{AT}}) \le Cn^3 \log n$. See Sections 4.1, 10 and the discussion in [141].

6 Eigenvalue Bounds Using Paths

This section develops techniques involving the geometric notion of paths. Left-invariant random walks on finite groups can be viewed as discrete versions of Brownian motions on compact Lie groups. It is well understood that certain aspects of the behavior of Brownian motion on a given manifold depend on the underlying Riemannian geometry and this has been a major area of research for many years. Many useful ideas and techniques have been developed in this context. They can be harvested without much difficulty and be brought to bear in the study of random walks on groups. This has produced great results in the study of random walks on infinite finitely generated groups. See [125, 139, 145]. It is also very useful for random walks on finite groups and, more generally, for finite Markov chains. For the development of these ideas for finite Markov chains, see [3, 51, 124, 131]. In the finite Markov chain literature, the use of path techniques is credited to Jerrum and Sinclair. See [131] for an excellent account of their ideas.

6.1 Cayley Graphs

Fix a finite group G and a finite generating set S which is symmetric, i.e., satisfies $\Sigma = \Sigma^{-1}$. The (left-invariant) *Cayley graph* (G, Σ) is the graph with vertex set G and edge set

$$E = \{(x, y) \in G \times G : \exists\, s \in \Sigma,\ y = xs\}.$$

The *simple random walk* on the Cayley graph (G, Σ) is the walk driven by the measure $p = (\#\Sigma)^{-1}\mathbf{1}_\Sigma$. It proceeds by picking uniformly at random a generator in Σ and multiplying by this generator on the right.

Define a *path* to be any finite sequence $\gamma = (x_0, \ldots, x_n)$ of elements of G such that each of the pair (x_i, x_{i+1}), $i = 0, \ldots, n-1$ belongs to E, i.e., such that $x_i^{-1}x_{i+1} \in \Sigma$. The integer n is called the length of the path γ and we set $|\gamma| = n$. Denote by $\mathcal{P}$ the set of all paths in (G, Σ).

Definition 6.1. *For any $x, y \in G$, set*

$$|x|_\Sigma = \min\{k : \exists\, s_1, \dots, s_k \in \Sigma, x = s_1 \dots s_k\},$$
$$d_\Sigma(x,y) = \min\{|\gamma| : \gamma \in \mathcal{P},\ x_0 = x,\ x_n = y\},$$
$$D_\Sigma = \max_{x,y\in G} d_\Sigma(x,y).$$

We call d_Σ the graph distance *and D_Σ the* diameter *of (G, Σ).*

In words, $|x|_\Sigma$ is the minimal number of elements $s_1, \dots, s_k$ of the generating set Σ needed to write x as a product $x = s_1 \dots s_k$, with the usual convention that the empty product equals the identity element e. Obviously the graph distance is left invariant and

$$d_\Sigma(x,y) = |x^{-1}y|_\Sigma, \quad D_\Sigma = \max_{x\in G} |x|_\Sigma.$$

The reference to Σ will be omitted when no confusion can possibly arise. Babai [8] gives an excellent survey on graphs having symmetries including Cayley graphs.

6.2 The Second Largest Eigenvalue

Let G be a finite group and p be a probability measure on G whose support generates G. We assume in this section that p is symmetric, i.e., $p = \check{p}$. Hence the associated operator on $L^2(G)$ is diagonalizable with real eigenvalue $1 = \beta_0 \geq \beta_1 \geq \cdots \geq \beta_{|G|-1}$ in non-increasing order and repeated according to multiplicity. We will focus here on bounding β_1 from above. The results developed below can also be useful for non symmetric measure thanks to the singular value technique of Theorem 5.3. See Section 10.3.

There are a number of different ways to associate to p an adapted geometric structure on G. For simplicity, we will consider only the following procedure. Pick a symmetric set of generators Σ contained in the support of p and consider the Cayley graph (G, Σ) as defined in Section 6.1. In particular, this Cayley graph induces a notion of path and a left-invariant distance on G. The simplest result concerning the random walk driven by p and involving the geometry of the Cayley graph (G, Σ) is the following. See, e.g., [2, 42].

Theorem 6.2. *Let (G, Σ) be a finite Cayley graph with diameter D. Let p be a probability measure such that $p = \check{p}$ and $\varepsilon = \min_\Sigma p > 0$. Then the second largest eigenvalue β_1 of p is bounded by $\beta_1 \leq 1 - \varepsilon/D^2$.*

This cannot be much improved in general as can be seen by looking at the simple random walk on $G = \mathbb{Z}_2^n \times \mathbb{Z}_{2a}$ with $a \gg n$. See [45]. The papers [10, 11, 97] describe a number of deep results giving diameter estimates for finite Cayley

graphs. These can be used together with Theorem 6.2 to obtain eigenvalue bounds.

Two significant improvements on Theorem 6.2 involve the following notation. Recall from Section 6.1 that $\mathcal{P}$ denotes the set of all paths in (G, Σ). For $s \in \Sigma$ and any path $\gamma = (x_0, \dots, x_n) \in \mathcal{P}$, set

$$N(s, \gamma) = \#\{i \in \{0, \dots, n-1\} : x_i^{-1} x_{i+1} = s\}. \tag{6.1}$$

In words, $N(s, \gamma)$ counts how many times the generator s appears along the path γ. Let $\mathcal{P}_{x,y}$ be the set of all finite paths joining x to y and $\mathcal{P}_x$ be the set of all finite paths starting at x. For each $x \in G$, pick a path $\gamma_x \in \mathcal{P}_{e,x}$ and set

$$\mathcal{P}_* = \{\gamma_x : x \in G\}.$$

Theorem 6.3 ([42]). *Referring to the notation introduced above, for any choice of $\mathcal{P}_*$, set*

$$A_* = \max_{s \in \Sigma} \left\{ \frac{1}{|G| p(s)} \sum_{\gamma \in \mathcal{P}_*} |\gamma| N(s, \gamma) \right\}.$$

Then $\beta_1 \le 1 - 1/A_*$.

This theorem is a corollary of Theorem 6.4 which is proved below. The notation A_* reminds us that this bound depends on the choice of paths made to construct the set $\mathcal{P}_*$. To obtain Theorem 6.2, define $\mathcal{P}_*$ by picking for each x a path from e to x having minimal length. Then bound $|\gamma_x|$ and $N_*(s, \gamma_x)$ from above by D, and bound $p(s)$ from below by ε.

Making arbitrary choices is not always a good idea. Define a *flow* to be a non-negative function Φ on $\mathcal{P}_e$ (the set of all paths starting at e) such that,

$$\forall\, x \in G, \quad \sum_{\gamma \in \mathcal{P}_{e,x}} \Phi(\gamma) = \frac{1}{|G|}.$$

For instance, for each x, let $\mathcal{G}_{e,x}$ be the set of all geodesic paths (paths of minimal length) in $\mathcal{P}_{e,x}$. The function

$$\Phi(\gamma) = \begin{cases} \frac{1}{\#\mathcal{G}_{e,x}|G|} & \text{if } \gamma \in \mathcal{G}_{e,x} \text{ for some } x \in G \\ 0 & \text{otherwise} \end{cases}$$

is a flow.

Theorem 6.4 ([49, 124]). *Let Φ be a flow and set*

$$A(\Phi) = \max_{s \in \Sigma} \left\{ \frac{1}{p(s)} \sum_{\gamma \in \mathcal{P}_e} |\gamma| N(s, \gamma) \Phi(\gamma) \right\}.$$

Then $\beta_1 \le 1 - 1/A(\Phi)$.

Proof. The proof is based on the elementary variational inequality (5.5) which reduces Theorem 6.4 to proving the Poincaré inequality

$$\forall\, f \in L^2(G), \;\; \mathrm{Var}_u(f) \le A(\Phi)\mathcal{E}(f,f). \tag{6.2}$$

Here we have

$$\mathrm{Var}_u(f) = \frac{1}{2|G|^2} \sum_{x,y\in G} |f(xy) - f(x)|^2 \tag{6.3}$$

and

$$\mathcal{E}(f,f) = \frac{1}{2|G|} \sum_{x,y\in G} |f(xy) - f(x)|^2 p(y). \tag{6.4}$$

The similarity between these two expressions is crucial to the argument below. For any path $\gamma = (y_0, \ldots, y_n)$ from e to y of length $|\gamma| = n$, set $\gamma_i = y_i^{-1} y_{i+1}$, $0 \le i \le n-1$ and write

$$f(xy) - f(x) = \sum_{i=0}^{n-1} (f(xy_{i+1}) - f(xy_i)) = \sum_{i=0}^{n-1} (f(xy_i\gamma_i) - f(xy_i)).$$

Squaring and using the Cauchy-Schwarz inequality, gives

$$|f(xy) - f(x)|^2 \le |\gamma| \sum_{i=0}^{n-1} |f(xy_i\gamma_i) - f(xy_i)|^2.$$

Summing over $x \in G$ yields

$$\begin{aligned}\sum_{x\in G} |f(xy) - f(x)|^2 &\le |\gamma| \sum_{i=0}^{n-1} \sum_{x\in G} |f(x\gamma_i) - f(x)|^2 \\ &\le |\gamma| \sum_{s\in\Sigma} \sum_{x\in G} N(s,\gamma) |f(xs) - f(x)|^2.\end{aligned}$$

Multiplying by $\Phi(\gamma)$, summing over all $\gamma \in \mathcal{P}_{e,y}$ and then averaging over all $y \in G$ yields

$$\mathrm{Var}(f) \le \frac{1}{2|G|} \sum_{s\in\Sigma} \sum_{x\in G} \sum_{\gamma\in\mathcal{P}_e} |\gamma| N(s,\gamma)\Phi(\gamma) |f(xs) - f(x)|^2.$$

Hence

$$\begin{aligned}\mathrm{Var}(f) &\le \frac{1}{2|G|} \sum_{s\in\Sigma} \sum_{x\in G} \left\{ \frac{1}{p(s)} \sum_{\gamma\in\mathcal{P}_e} |\gamma| N(s,\gamma)\Phi(\gamma) \right\} |f(xs) - f(x)|^2 p(s) \\ &\le \left(\max_{s\in\Sigma} \left\{ \frac{1}{p(s)} \sum_{\gamma\in\mathcal{P}_e} |\gamma| N(s,\gamma)\Phi(\gamma) \right\} \right) \mathcal{E}(f,f).\end{aligned}$$

This proves (6.2). □

The next result is a corollary of Theorem 6.4 and use paths chosen uniformly over all geodesic paths from e to y.

Theorem 6.5 ([49, 124]). *Referring to the setting of Theorem* 6.2*, assume that the automorphisms group of G is transitive on Σ. Then*

$$\beta_1 \le 1 - \frac{\varepsilon \# \Sigma}{D^2}.$$

Let us illustrate these results by looking at the random transposition walk on the symmetric group S_n defined at (4.1). Thus $p(e) = 1/n$, $p(\tau) = 2/n^2$ if τ is a transposition and $p(\tau) = 0$ otherwise. From representation theory (see Section 9.2), we know that $\beta_1 = 1 - 2/n$. Here Σ is the set of all transpositions. Any permutation can be written as a product of at most $n-1$ transpositions (i.e., the diameter is $D = n-1$). Thus Theorem 6.2 gives

$$\beta_1 \le 1 - \frac{2}{n^2(n-1)^2}.$$

When writing a permutation as a (minimal) product of transpositions, any given transposition is used at most once. Hence $N(s,\gamma)$ at (6.1) is bounded by 1. Using this in Theorem 6.3 immediately gives

$$\beta_1 \le 1 - \frac{2}{n^2(n-1)}.$$

A more careful use of the same theorem actually yields

$$\beta_1 \le 1 - \frac{2}{n(n-1)}.$$

Finally, as the transpositions form a conjugacy class, it is easy to check that Theorem 6.5 applies and yields again the last inequality.

6.3 The Lowest Eigenvalue

Let p be a symmetric probability on G and Σ be a finite symmetric generating set contained in the support of p. Loops of odd length in the Cayley graph (G, Σ) can be used to obtain lower bounds on the lowest eigenvalue

$$\beta_{\min} = \beta_{|G|-1}.$$

Denote by $\mathcal{L}$ the set of loops of odd length anchored at the identity in (G, Σ). A *loop flow* is a non-negative function Ψ such that

$$\sum_{\gamma \in \mathcal{L}} \Psi(\gamma) = 1.$$

As above, let $N(s,\gamma)$ be the number of occurrences of $s \in \Sigma$ in γ.

Theorem 6.6 ([51, 42, 45]). *Let Ψ be a loop flow and set*

$$B(\Psi) = \max_{s \in \Sigma} \left\{ \frac{1}{p(s)} \sum_{\gamma \in \mathcal{L}} |\gamma| N(s, \gamma) \Psi(\gamma) \right\}.$$

Then the smallest eigenvalue is bounded by $\beta_{|G|-1} \geq -1 + 2/B(\phi)$.

As a trivial application, assume that $p(e) > 0$ and that $e \in \Sigma$. Then we can consider the loop flow concentrated on the trivial loop of length 1, that is, $\gamma = (e, e)$. In this case $B(\Psi) = 1/p(e)$ and we obtain

$$\beta_{|G|-1} \geq -1 + 2p(e).$$

This applies for instance to the random transposition measure p defined at (4.1) and gives $\beta_{|G|-1} \geq -1 + 2/n$ (there is, in fact, equality in this case).

For an example where a non-trivial flow is useful, consider the Borel–Chéron shuffle of Section 3.1: remove a random packet and place it on top. This allows for many loops of length 3. Consider the loops $\gamma_{a,b}$, $2 < a \leq b \leq n$ and a odd, defined as follows. Remove the packet $(a, \ldots, b)$ and place it on top; remove the packet corresponding to the cards originally in position $(a+1)/2$ through $a-1$ and place it on top; remove the packet of the cards originally in positions 1 through $(a-1)/2$ and place it on top. The crucial observation is that, given one of these moves and its position in the loop, one can easily recover the two other moves of the loop. Using the flow uniformly supported on these loops in Theorem 6.6 gives $\beta_{\min} \geq -(26n+2)/(27n)$ for the Borel–Chéron shuffle on S_n.

The following result is a corollary of Theorem 6.6 and complements Theorem 6.5. The proof uses the uniform flow on all loops of minimal odd length.

Theorem 6.7. *Assume that the automorphism group of G is transitive on Σ. Then*

$$\beta_{|G|-1} \geq 1 - \frac{2\varepsilon \# \Sigma}{L^2}$$

where $\varepsilon = \min\{p(s) : s \in \Sigma\}$ and L is the minimal length of a loop of odd length in (G, Σ).

To illustrate this result, consider the alternating group A_n. In A_n, consider any fixed element $\sigma \neq e$ and its orbit Σ under the action of the symmetric group, that is, $\Sigma = \{\tau = \varrho \sigma \varrho^{-1},\ \varrho \in S_n\}$. In words, Σ is the conjugacy class of σ in S_n. One can show that, except when σ is the product of two transpositions with disjoint supports in A_4, the set Σ is a generating set of A_n. Moreover, in any such case, the Cayley graph (A_n, Σ) contains cycles of length three (for details, see, e.g., [121]). For instance, if $\sigma = c$ is a cycle of odd length, we have $c^{-1}, c^2 \in \Sigma$ and $c^{-1}c^{-1}c^2 = e$. If $\sigma = (i,j)(k,l)$ is the product of two disjoint transpositions, we have $[(i,j)(k,l)][(k,i)(j,l)][(k,j)(i,l)] = e$. Set

$$p_\Sigma(\tau) = \begin{cases} 1/|\Sigma| & \text{if } \tau \in \Sigma \\ 0 & \text{otherwise.} \end{cases}$$

By construction, the automorphism group of A_n acts transitively on Σ. Hence, for any Σ as above, Theorem 6.7 shows that the lowest eigenvalue of p_Σ is bounded by $\beta_{\min} \geq -1 + 2/9 = -7/9$.

6.4 Diameter Bounds, Isoperimetry and Expanders

The goal of this section is to describe the relation between eigenvalues of random walks, isoperimetric inequalities and the important notion of expanders.

Diameter bounds. Let (G, Σ) be a finite Cayley graph with diameter D (recall that, by hypothesis, Σ is symmetric). Let p be a probability with support contained in Σ. For $k = \lfloor D/2 \rfloor - 1$, the support of $p^{(k)}$ contains less than half the elements of G. Hence

$$D \leq 2(T(G, p) + 2). \tag{6.5}$$

This gives an elementary relation between the diameter of (G, Σ) and random walks. Theorem 6.2 shows how the diameter can be used to control the second largest eigenvalue of an associated walk. Interestingly enough, this relation can be reversed and eigenvalues can be used to obtain diameter bounds. The best known result is the following [22, 117] which in fact holds for general graphs.

Theorem 6.8. *Let Σ be a symmetric generating set of a finite group G of order $|G| = N$. Let β_i, $0 \leq i \leq N-1$, be the eigenvalues in non-increasing order of a random walk driven by a measure p whose support is contained in $\{e\} \cup \Sigma$ and set $\lambda_i = 1 - \beta_i$. Then the diameter D of (G, Σ) is bounded by*

$$D \leq 1 + \left\lfloor \frac{\cosh^{-1}(N-1)}{\cosh^{-1}\left(\frac{\lambda_1 + \lambda_{N-1}}{\lambda_{N-1} - \lambda_1}\right)} \right\rfloor \leq 1 + \left\lfloor \frac{\cosh^{-1}(N-1)}{\cosh^{-1}\left(\frac{2+\lambda_1}{2-\lambda_1}\right)} \right\rfloor .$$

It is useful to observe that if $N = |G|$ goes to infinity and λ_1 goes to zero the asymptotics of the right most bound is $(2\lambda_1)^{-1/2} \log |G|$. One can also verify that, assuming $\lambda_1 \leq 1$, the second upper bound easily gives

$$D \leq 3\lambda_1^{-1/2} \log |G|. \tag{6.6}$$

When λ_1 is relatively small, the elementary bound (6.5) often gives better results than Theorem 6.8. For instance, consider the symmetric group S_n generated by the set of all transpositions. Let $p = p_{\mathrm{RT}}$ be the random transposition measure defined at (4.1). The diameter of this Cayley graph is $n-1$, the spectral gap λ_1 of p_{RT} is $2/n$ and $T(S_n, p_{\mathrm{RT}}) \sim \frac{1}{2} n \log n$. Hence, both (6.5) and Theorem 6.8 are off but (6.5) is sharper. Theorem 6.8 is of most interest for families of graphs and random walks having a spectral gap bounded away from 0. Such graphs are called expanders and are discussed below.

Isoperimetry. Let (G, Σ) be a finite Cayley graph. Recall that the edge set E of (G, Σ) is $E = \{(x, y) : x, y \in G, x^{-1}y \in \Sigma\}$. As always, we denote by u the uniform probability measure on G. We also denote by u_E the uniform probability on E so that for a subset F of E, $u_E(F) = |F|/|\Sigma||G|$ where $|F|$ denotes the cardinality of F.

Given a set $A \in G$, define the boundary of A to be

$$\partial A = \{(x, y) \in G \times G : x \in A, y \in G \setminus A, x^{-1}y \in \Sigma\}.$$

The *isoperimetric constants* $I = I(G, \Sigma)$, $I' = I'(G, \Sigma)$ are defined by

$$I = \min_{\substack{A \subset G \\ 2|A| \leq |G|}} \frac{u_E(\partial A)}{u(A)}, \quad I' = \min_{A \subset G} \frac{u_E(\partial A)}{2(1 - u(A))u(A)}. \tag{6.7}$$

We have $I/2 \leq I' \leq I$. Note that, in terms of cardinalities, this reads

$$I = \min_{\substack{A \subset G \\ 2|A| \leq |G|}} \frac{|\partial A|}{|\Sigma||A|}, \quad I' = \min_{A \subset G} \frac{|G||\partial A|}{2|\Sigma|(|G| - |A|)|A|}.$$

The following gives equivalent definitions of I, I' in function terms. See, e.g., [124]. For a function f on G and $e = (x, y) \in E$, we set $df(e) = f(y) - f(x)$.

Lemma 6.9. *We have*

$$2I = \min_f \left\{ \frac{u_E(|df|)}{u(|f - m(f)|)} \right\}, \quad 2I' = \min_f \left\{ \frac{u_E(|df|)}{u(|f - u(f)|)} \right\}$$

where $m(f)$ denote an arbitrary median of f.

For sharp results concerning isoperimetry on the hypercube and further discussion, see [84, 96] and the references therein.

The next result relates I and I' to the spectral gap λ_1 of random walks closely related to the graph (G, Σ). This type of result has become known under the name of a Cheeger inequality. See, e.g., [98, 124, 131]. An interesting development is in [111]. For the original Chegeer inequality in Riemannian geometry, see, e.g., [20].

Theorem 6.10. *Let G be a Cayley graph and p be a symmetric probability measure on G with spectral gap $\lambda_1 = 1 - \beta_1$.*

- *Assume* $\text{supp}\,(p) \subset \Sigma$ *and set* $\eta = \max_\Sigma p$. *Then* $\lambda_1 \leq 2\eta|\Sigma|I'$.
- *Assume that* $\inf_\Sigma p = \varepsilon > 0$. *Then* $\varepsilon|\Sigma|I^2 \leq 2\lambda_1$.
- *In particular, if $p = p_\Sigma$ is the uniform probability on Σ, $I^2 \leq 2\lambda_1 \leq 4I'$.*

Slightly better results are known. For instance, [110, Theorem 4.2] gives $I^2 \leq \lambda_1(2 - \lambda_1)$. See also [111].

The isoperimetric constants I, I' can be bounded from below in terms of the diameter. See, e.g., [9] and [131]. Using the notation of Section 6, we have the following isoperimetric version of Theorems 6.4, 6.5.

Theorem 6.11. *Let (G, Σ) be a finite Cayley graph. Let Φ be a flow as in Theorem 6.4. Then $2I' \geq 1/a(\Phi)$ with*

$$a(\Phi) = \max_{s \in \Sigma} \left\{ |\Sigma| \sum_{\gamma \in \mathcal{P}_e} N(s, \gamma) \Phi(\gamma) \right\}.$$

In particular, $I \geq I' \geq 1/(2|\Sigma|D)$ where D is the diameter of (G, Σ). If we further assume that the automorphism group of G is transitive on Σ then $I \geq I' \geq 1/(2D)$.

Although the notion of isoperimetry is appealing, it is rarely the case that good spectral gap lower bounds are proved by using the relevant inequality in Theorem 6.10. See the discussion in [62]. In fact, isoperimetric constants are hard to compute or estimate precisely and spectral bounds are often useful to bound isoperimetric constants.

Let us end this short discussion of isoperimetric constants by looking at the symmetric group S_n equipped with the generating set of all transpositions. This Cayley graph has diameter $n-1$ and the automorphism group of S_n acts transitively on transpositions. Hence Theorem 6.11 gives $I' \geq (2(n-1))^{-1}$. The random transposition walk defined at (4.1) has spectral gap $\lambda_1 = 2/n$ (See Section 9.2). By Theorem 6.10, this implies $(n-1)^{-1} \leq I' \leq I \leq 2(n-1)^{-1/2}$. Using $A = \{\sigma \in S_n : \sigma(n) = n\}$ as a test set shows that $I \leq 2n^{-1}$, $I' \leq (n-1)^{-1}$. Thus $(n-1)^{-1} \leq I \leq 2n^{-1}$ and $I' = (n-1)^{-1}$.

Expanders. The notion of *expander* depends on a different definition of the boundary than the one given above. Namely, for any $A \subset G$, set

$$\delta A = \{x \in G : d(x, A) = 1\}$$

where d is the graph distance introduced in Section 6.1. Define the expansion constant $h = h(G, \Sigma)$ by

$$h = \min_{\substack{A \subset G \\ 2|A| \leq |G|}} \frac{|\delta A|}{|A|}.$$

By inspection, we have $I \leq h \leq |\Sigma| I$. A variant of Theorem 6.11 in [9] states that, for any Cayley graph, $h \geq 2/(2D+1)$.

Definition 6.12. *A finite Cayley graph (G, Σ) is an* (N, r, ε)-expander *if $|G| = n$, $|\Sigma| = r$ and $h(G, \Sigma) \geq \varepsilon$.*

A family $((G_n, \Sigma_n))$ of finite Cayley graphs is a family of expanders *if $|G_n|$ tends to ∞ and there exists $\varepsilon > 0$ such that $h(G_n, \Sigma_n) > \varepsilon$.*

Comparing I and h and using Theorem 6.10 yields the following relation between spectral gap estimates and the notion of expander.

Proposition 6.13. *Let $((G_n, \Sigma_n))$ be a family of finite Cayley graphs such that $|G_n|$ tends to ∞. Let p_n denote the uniform probability on Σ_n and let $\lambda_1(n)$ be the spectral gap associated to p_n.*

- *If there exists $\varepsilon > 0$ such that $\lambda_1(n) \geq \varepsilon$ for all n then (G_n, Σ_n) is a family of expanders.*
- *If there exists r such that $|\Sigma_n| \leq r$ for all n then (G_n, Σ_n) is a family of expanders if and only if there exists $\varepsilon > 0$ such that $\lambda_1(n) \geq \varepsilon$ for all n.*

Theorem 9.8 in Section 9.4 gives a remarkable application of Proposition 6.13.

In the other direction, Proposition 6.13 shows that the symmetric groups S_n equipped with the generating sets $\Sigma_n = \{\tau, c, c^{-1}\}$ where τ is the transposition $(1, 2)$ and c the cycle $(1, 2, \ldots, n)$ do not form a family of expanders. Indeed, the diameter D of (S_n, Σ_n) is of order n^2 whereas Proposition 6.13 and Theorem 6.8 shows that any expander graph on S_n has diameter of order $n \log n$ at most. In fact, the present Cayley graph has λ of order $1/n^3$. See Section 10.

Recall that a finitely generated group Γ has property (T) (i.e., Kazhdan property (T)) if there exists a finite set $K \subset \Gamma$ and $\varepsilon > 0$ such that, for every non-trivial irreducible unitary representation (V, ϱ) of Γ and every unitary vector $v \in V$, $\|\varrho(x)v - v\| \geq \varepsilon$ for some $x \in K$. One shows that if this holds for one finite set K then it holds for any finite generating set Σ (with different $\varepsilon > 0$). See [98] for an excellent exposition and references concerning property (T). The groups $SL_n(\mathbb{Z})$, $n \geq 3$, have property (T). Non-compact solvable groups, free groups and $SL_2(\mathbb{Z})$ do not have property (T). Margulis [108] produced the first explicit examples of families of expanders by using property (T) to obtain infinite families of graphs with bounded degree and spectral gap bounded from below. See also [101] and [115, 146] for recent advances concerning property (T).

Theorem 6.14 ([98]). *Let Γ be a finitely generated infinite group. Let H_n be a family of normal finite index subgroups of Γ. Set $G_n = \Gamma/H_n$ and assume that $|G_n|$ tends to infinity. Let Σ be a symmetric generating set of Γ and $\Sigma_n \subset G_n$ be the projection of Σ.*

- *Assume that Γ has property (T). Then $((G_n, \Sigma_n))$ is a family of expanders.*
- *Assume that Γ is solvable. Then $((G_n, \Sigma_n))$ is not a family of expanders.*

The condition that the subgroups H_n are normal is not essential. It is added here simply to have Cayley graphs as quotients. For a proof, see [98, Prop. 3.3.1, 3.3.7]. The following simple result describes what happens for random walks on expanders. See, e.g., [46, 115].

Theorem 6.15. *Fix $r > 0$. Let (G_n, Σ_n) be a family of expanders with Σ_n containing the identity. For each n, let p_n be a probability measure on G_n such that $\inf_{\Sigma_n} p_n \geq 1/r$, $|\mathrm{supp}\ (p_n)| \leq r$. Then there are constants $C, c > 0$ such that*

$$c \log |G_n| \leq T(G_n, p_n) \leq C \log |G_n|.$$

Moreover, the family (G_n, p_n) has a precut-off at time $\log |G_n|$.

Proof. For the upper bound, use (5.9) and the fact that the hypotheses and Proposition 6.13 imply $\beta_{n,*} \leq 1 - \varepsilon$. For the lower bound, use (5.12). □

The next theorem due to Alon and Roichman [6] says that most Cayley graphs (G, Σ) with $|\Sigma| \gg \log |G|$ are expanders.

Theorem 6.16. *For every $\varepsilon > 0$ there exists $c(\varepsilon) > 0$ such that, if G is a group of order n, $t \geq c(\varepsilon) \log n$, T is a uniformly chosen t-subset of G and $\Sigma = T \cup T^{-1}$ then the Cayley graph (G, Σ) is an $(|G|, |\Sigma|, \varepsilon)$-expander with probability $1 - o(1)$ when n tends to infinity.*

Next, we describe some explicit examples of expanders. In $SL_n(\mathbb{Z})$, consider the matrices

$$A_n = \begin{pmatrix} 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}, \quad B_n = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ j & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{pmatrix}$$

where $j = (-1)^{n+1}$. These generates $SL_n(\mathbb{Z})$.

Theorem 6.17 ([98]). *Fix $n \geq 2$. consider the symmetric generating set $\Sigma_n = \{A_n^{\pm 1}, B_n^{\pm 1}\}$ of $SL_n(\mathbb{Z}_q)$ where q is prime. Let p_n denote the uniform probability on $\{I_n, A_n^{\pm 1}, B_n^{\pm 1}\}$. Then $((SL_n(\mathbb{Z}_q), \Sigma_n))$ is a family of expanders. In particular, for fixed n and varying prime q, $((SL_n(\mathbb{Z}_q), p_n))$ has a precut-off at time $\log q$.*

The proof differs depending on whether $n = 2$ or $n > 2$ because, as mentioned earlier, $SL_2(\mathbb{Z})$ does not have property (T). See [98, 99].

We close our discussion of expanders by stating a small selection of open problems. See [98, 99] for more.

Problem 6.18. Can one find generating subsets Σ_n of the symmetric groups S_n of bounded size $|\Sigma_n| \leq r$ such that (S_n, Σ_n) form a family of expanders?

In [100, Section 5], Lubotzky and Pak notice that this problem is related to another open problem, namely, to whether or not the automorphism group of a regular tree of degree at least 4 has property (T). One can also state Problem 6.18 with the symmetric groups replaced by an infinite family of simple finite groups.

Problem 6.19. Can one find a family of finite groups G_n and generating sets Σ_n^1, Σ_n^2 of bounded size $|\Sigma_n^i| \leq r$ such that $((G_n, \Sigma_n^1))$ is a family of expanders but $((G_n, \Sigma_n^2))$ is not?

If Problem 6.18 has a positive answer then the same is true for Problem 6.19 since $((S_n, \Sigma_n))$ with $\Sigma_n = \{(1, 2), (1, \ldots, n)^{\pm 1}\}$ is not a family of expanders (see, e.g., [115]).

Problem 6.20. Fix r and let Σ_p denote an arbitrary generating set of $SL_2(\mathbb{Z}_p)$, p prime, with $|\Sigma_p| \leq r$. Is $((SL_2(\mathbb{Z}_p), \Sigma_p))$ always a family of expanders?

With respect to this last problem, set

$$\Sigma_p^i = \left\{ \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}^{\pm 1}, \begin{pmatrix} 1 & 0 \\ i & 1 \end{pmatrix}^{\pm 1} \right\}.$$

Then $((SL_2(\mathbb{Z}_p), \Sigma_p^i))$ is a family of expanders if $i = 1, 2$ but it is not known if the same result holds for $i = 3$. See [63, 98, 99].

Problem 6.21. Let $((G_n, \Sigma_n))$ be a family of expanders. Under the assumptions and notation of Theorem 6.15, does the family $((G_n, p_n))$ admit a cut-off?

For further information on these problems, see [63, 64, 65, 98, 99].

Ramanujan graphs. Alon and Bopanna (see, e.g., [74, 98, 99, 127, 136]) observed that any infinite family of finite Cayley graphs $((G_n, \Sigma_n))$ with $|\Sigma_n| = r$ for all n (more generally, r-regular graphs) satisfies

$$\liminf_{n \to \infty} \beta_1(G_n, p_n) \geq \frac{2\sqrt{r-1}}{r}.$$

where p_n denotes the uniform probability on Σ_n.

Definition 6.22. *A Cayley graph (G, Σ) is* Ramanujan *if*

$$\beta_1(G, p_\Sigma) \leq \frac{2\sqrt{r-1}}{r}$$

where p_Σ denotes the uniform probability on Σ and $r = |\Sigma|$.

Examples of Ramanujan Cayley graphs with $G = PGL_2(\mathbb{Z}_q)$ are given in [127]. See also [25, 98, 101, 136]. For fixed r, asymptotically as the cardinality goes to infinity, Ramanujan graphs are graphs whose second largest eigenvalue is as small as possible. By Proposition 6.13, they are expanders, in fact very good expanders, and have many other remarkable properties. After taking care of possible periodicity problems, the simple random walks on any infinite family of Ramanujan Cayley graphs $((G_n, \Sigma_n))$ have a precut-off at time $\log |G_n|$.

Infinite families of Ramanujan graphs are hard to find and most (if not all) known examples are obtained by applying rather deep number theoretic results. See [98, 127]. In particular, the construction of expanders as in Theorem 6.14 cannot work for Ramanujan graphs [71, 98, 99].

Theorem 6.23. *Let Γ be a finitely generated infinite group. Let H_n be a family of normal finite index subgroups of Γ. Set $G_n = \Gamma/H_n$ and assume that $|G_n|$ tends to infinity. Let Σ be a symmetric generating set of Γ such that the graph (Γ, Σ) is not a tree. Let $\Sigma_n \subset G_n$ be the projection of Σ. Then at most finitely many (G_n, Σ_n) are Ramanujan.*

As in Theorem 6.14, the condition that the subgroups H_n are normal is not essential.

7 Results Involving Volume Growth Conditions

On a finite group G, consider a symmetric probability p whose support generates G. Fix a symmetric generating set Σ contained in the support of p and consider the Cayley graph (G, Σ) as in Section 6.1.

Definition 7.1. *Referring to the notation of Section* 6.1*, set*

$$V(n) = V_\Sigma(n) = \#\{x \in G : |x|_\Sigma \le n\}.$$

The function V_Σ is called the volume growth function *of (G, Σ).*

Sections 7.1 and 7.2 below describe results that involve the volume growth function V and apply to walks based on a bounded number of generators. Examples include nilpotent groups with small class and bounded number of generators. Section 7.3 presents contrasting but related results for some families of nilpotent groups with growing class and/or number of generators.

7.1 Moderate Growth

This section gives a large class of finite groups which carry natural random walks whose behavior is similar to that of the simple random walk on the finite circle group $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$. More precisely, on $\mathbb{Z}_n$, consider the random walk which goes left, right or stays put, each with probability 1/3. For this walk, the spectral gap $\lambda_1 = 1 - \beta_1$ is of order $1/n^2$ and there are continuous positive decreasing functions f, g tending to 0 at infinity such that

$$f(k) \le \|p^{(kn^2)} - u\|_{\mathrm{TV}} \le g(k).$$

Thus, there is no cut-off phenomenon in this case: a number of steps equal to a large multiple of $1/\lambda_1$ suffices to reach approximate equilibrium whereas a small multiple of $1/\lambda_1$ does not suffice.

We start with the following definition.

Definition 7.2 ([44, 47]). *Fix $A, \nu > 0$. We say that a Cayley graph (G, Σ) has* (A, ν)-moderate growth *if its volume growth function satisfies*

$$V(k) \ge \frac{|G|}{A} \left(\frac{k}{D}\right)^\nu$$

for all integers $k \le D$ where D is the diameter of (G, Σ).

Let us illustrate this definition by some examples.

– The circle group $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ with $\Sigma = \{0, \pm 1\}$ has $V(k) = 2k + 1$. Here $|G| = n$, $D = \lfloor n/2 \rfloor$. Thus the circle group has moderate growth with $A = 3/2$ and $\nu = 1$.

- The group $\mathbb{Z}_n$ with $\Sigma = \{0, \pm 1, \pm m\}$ with $m \leq n$ has diameter D of order $\max\{n/m, m\}$. The Cayley graph $(\mathbb{Z}_n, \Sigma)$ has moderate growth with $A = 5$ and $\nu = 2$ although this is not entirely obvious to see.
- Consider the group $\mathbb{Z}_n^d$ with $\Sigma = \{0, \pm e_i\}$ where e_i denotes the element with all coordinates 0 except the i-th which equals 1. This Cayley graph has diameter $D = d\lfloor n/2 \rfloor$. For fixed d, there exists a constant A_d such that $(\mathbb{Z}_n^d, \Sigma)$ has (A_d, d)-moderate growth for all n.
- For any odd prime p, consider the affine group A_p which is the set of all pairs $(a, b) \in \mathbb{Z}_p^* \times \mathbb{Z}_p$ with multiplication given by $(a, b)(a', b') = (aa', a'b + b')$. Let α be a generator of $\mathbb{Z}_p^*$, β a generator of $\mathbb{Z}_p$, and set $\Sigma = \{(1, 0), (\alpha, 0), (\alpha^{-1}, 0), (1, \beta), (1, -\beta)\}$. This group has diameter D of order p and it has $(6, 2)$-moderate growth.
- Let $U_3(n)$ be the Heisenberg group mod n, i.e., the group of all 3 by 3 upper diagonal matrices with 1 on the diagonal and integer coefficients mod n. Let I denote the identity matrix in $U_3(n)$. Let $E_{i,j}$ be the matrix in U_3 whose non-diagonal entries are all 0 except the (i, j) entry which is 1. Then $\Sigma = \{I, \pm E_{1,2}, \pm E_{2,3}\}$ is a generating set of $U_3(n)$. The Cayley graph $(U_3(n), \Sigma)$ has diameter of order n and $(48, 3)$-moderate growth.

The next theorem gives sharp bounds under the assumption of moderate growth.

Theorem 7.3 ([44, 47]). *Let (G, Σ) be a finite Cayley graph with diameter D and such that $e \in \Sigma$. Let p be a probability measure on G supported on Σ. For any positive numbers A, d, ε, there exists six positive constants $c_i = c_i(A, d, \varepsilon)$, $1 \leq i \leq 6$, such that if (G, Σ) has (A, d)-moderate growth and p satisfies $\inf_\Sigma p \geq \varepsilon$ then we have*

$$\forall\, k \in \mathbb{N}, \quad a_1 \mathrm{e}^{-a_2 k/D^2} \leq \|p^{(k)} - u\|_{TV} \leq a_3 \mathrm{e}^{-a_4 k/D^2}$$

and

$$\forall\, k \geq D^2, \quad d_2(p^{(k)}, u) \leq a_5 \mathrm{e}^{-a_6 k/D^2}.$$

The condition $\inf_\Sigma p \geq \varepsilon$ has two different consequences. On the one hand, it forces p to be, in some sense, adapted to the underlying graph structure. On the other hand, it implies a uniform control over the size of the generating set Σ since we have $1 \geq p(\Sigma) \geq \varepsilon |\Sigma|$.

Moderate growth was first introduced in [44]. It is related to the following notion of doubling growth which has been used in many different contexts.

Definition 7.4. *Fix $A > 0$. We say that a Cayley graph (G, Σ) has* A-doubling growth *if its volume growth function satisfies*

$$\forall k \in \mathbb{N}, \quad V(2k) \leq A\, V(k).$$

Doubling growth provides a useful way to obtain examples of groups with moderate growth thanks to the following two propositions. The first is elementary.

Proposition 7.5. *If the Cayley graph (G, Σ) has A-doubling growth, then it has (A, d)-moderate growth with $d = \log_2 A$.*

Let us observe that the notion of doubling growth make sense for infinite Cayley graphs.

Proposition 7.6. *Let (Γ, Σ) be an infinite Cayley graph and assume that (Γ, Σ) has A-doubling growth. Then, for any quotient group $G = \Gamma/N$, N normal in Γ, the Cayley graph (G, Σ_G) where Σ_G is the canonical projection of Σ in G has A^2-doubling growth.*

We illustrate this with two examples. First, consider $\mathbb{Z}_n$ with generating set $\Sigma = \{0, \pm 1, \pm m\}$, $m < n$. We can view this Cayley graph as a quotient of the square grid, i.e., the natural graph on $\mathbb{Z}^2$. Indeed, one can check that there is a unique surjective group homomorphism π from $\mathbb{Z}^2$ to $\mathbb{Z}_n$ such that $\pi((1,0)) = 1$, $\pi((0,1)) = m$ (this is because $\mathbb{Z}^2$ is the free abelian group on two generators). Proposition 7.6 applies and easily shows that $(\mathbb{Z}_n, \Sigma)$ is 5-doubling. As a second example, consider the Heisenberg group $U_3(n)$ with its natural generating set $\Sigma = \{I, E_{1,2}, E_{2,3}\}$ as defined above after Definition 7.2. This is a quotient (simply take all coordinates mod n) of the infinite discrete Heisenberg group U_3, i.e., the group of all 3 by 3 upper-triangular matrices with entries in $\mathbb{Z}$ and 1 on the diagonal. It is well known (see e.g., [82, Pro. VII.22]) that the volume growth function of this group satisfies $c_1 n^4 \le V(n) \le c_2 n^4$. Hence $(U_3(n), \Sigma)$ has A-doubling growth with $A = c_2 c_1^{-1} 3^4$.

The next result is derived from a deep theorem of Gromov [75].

Theorem 7.7. *Given two positive reals C, d, there is a constant $A = A(C, d)$ such that any finite Cayley graph (G, Σ) satisfying $V(n) \le C n^d$ for all integers n has A-doubling growth.*

In contrast to all the other results presented in this survey, there is no known explicit control of A as a function of C, d.

Doubling growth is a stronger assumption than moderate growth. Under the latter condition one can complement Theorem 7.3 with the following result.

Theorem 7.8 ([44, 46]). *Let (G, Σ) be a finite Cayley graph with diameter D and such that $e \in \Sigma$. Let p be a symmetric probability measure on G supported on Σ. For any positive numbers A, ε, there exist four positive constants $c_i = c_i(A, \varepsilon)$, $1 \le i \le 4$, such that if (G, Σ) has A-doubling growth and p satisfies $\inf_\Sigma p \ge \varepsilon$ then we have*

$$\forall\, k \in \mathbb{N},\ \ \frac{a_1 |G|}{V(k^{1/2})} \mathrm{e}^{-a_2 k/D^2} \le d_2(p^{(k)}, u) \le \frac{a_3 |G|}{V(k^{1/2})} \mathrm{e}^{-a_4 k/D^2}.$$

The same upper bound holds for any non-symmetric measure that charges e and a generating set Σ (which can be non-symmetric). See Theorem 10.8.

Thus doubling growth gives a very satisfactory control over the behavior of random walks adapted to the underlying graph structure. The next section describes a large class of examples with doubling growth.

7.2 Nilpotent Groups

In a group G, let $[x, y] = x^{-1}y^{-1}xy$ denote the *commutator* of $x, y \in G$. For $A, B \subset G$, let $[A, B]$ denote the group generated by all the commutators $[a, b]$, $a \in A, b \in B$. The *lower central series* of a group G is the non-increasing sequence of subgroups G_k of G defined inductively by $G_1 = G$ and $G_k = [G_{k-1}, G]$. A group (finite or not) is *nilpotent of class* c if $G_c \neq \{e\}$ and $G_{c+1} = \{e\}$. See [79, 78, 135]. Abelian groups are nilpotent of class 1. The group $U_m(n)$ of all m by m upper-triangular matrices with 1 on the diagonal is nilpotent of class $m - 1$.

Doubling growth for nilpotent groups. The next statement shows that nilpotent groups give many infinite families of Cayley graphs having A-doubling growth. See [82, p. 201] and [44].

Theorem 7.9. *Given any two integers c, s, there exists a constant $A = A(c, s)$ such that any Cayley graph (G, Σ) with G nilpotent of class at most c and Σ of cardinality at most s has A-doubling growth.*

The constant $A(c, s)$ can be made explicit, see [44]. Of course, this result brings Theorem 7.3 and 7.8 to bear. For concrete examples, consider the group $U_m(n)$ of all m by m upper-triangular matrices with 1 on the diagonal and entries in $\mathbb{Z}_n$. We noticed earlier that this group is nilpotent of class $m - 1$. Let $E_{i,j} \in U_m(n)$ be the matrix with zero non-diagonal entries except the (i, j)-th which is 1. The set $\Sigma = \{I, E_{1,2}^{\pm 1}, \ldots, E_{m-1,m}^{\pm 1}\}$ generates $U_m(n)$. Let p_Σ be the uniform probability measure on Σ. For each fixed integer m, Theorem 7.9 applies uniformly to $U_m(n)$, $n = 2, 3, \ldots$. As $(U_m(n), \Sigma)$ has diameter of order n this shows that, given m, there are positive constants a_i such that, uniformly over all integers n, k, the measure p_Σ on $U_m(n)$ satisfies

$$a_1 \mathrm{e}^{-a_2 k/n^2} \leq \|p_\Sigma^{(k)} - u\|_{\mathrm{TV}} \leq a_3 \mathrm{e}^{-a_4 k/n^2}.$$

***p*-groups and Frattini walks.** Let p be a prime. A p-group is a group of order a power of p. Any group of order p^a is nilpotent of class at most $a - 1$ and contains generating sets of size less than or equal to a. In fact, in a group of order p^a, the minimal generating sets (i.e., sets that contains no generating proper subsets) all have the same size and can be described in terms of the *Frattini subgroup* which is defined as the intersection of all subgroups of order p^{a-1}. By a theorem of Burnside, the quotient of any p-group G by its Frattini subgroup is a vector space over $\mathbb{Z}_p$ whose dimension is the size of any minimal generating set and is called the *Frattini rank* of G. For instance, the group $U_m(p)$ has order p^a with $a = \binom{m}{2}$ and the matrices $E_{i,i+1}$, $1 \leq i \leq m - 1$

form a minimal set of generators. Hence $U_m(p)$ has Frattini rank $m-1$. See [79, 78, 135]. The following theorem describes how the results of the previous two sections apply to this very natural class of examples we call Frattini walks. Recall that the exponent of a group G is the smallest n such that $g^n = e$ for all $g \in G$.

Theorem 7.10 ([44, 45]). *Fix an integer c. Then there exists four positive constants $a_i = a_i(c)$ such that, for any p-group G of nilpotency class and Frattini rank at most c, for any minimal set F of generators of G, we have*

$$a_1 \mathrm{e}^{-a_2 k/p^{2\omega}} \le \|q_F^{(k)} - u\|_{TV} \le a_3 \mathrm{e}^{-a_4 k/p^{2\omega}}$$

where q_F denotes the uniform probability measure on $\{e\} \cup F \cup F^{-1}$ and p^ω is the exponent of $G/[G,G]$.

The proof consists in applying Theorems 7.9, 7.3 and showing that the diameter of (G, Σ) is of order p^ω, uniformly over the class of group considered here. Note that, for any fixed a, Theorem 7.10 applies uniformly to all groups of order p^a and their minimal sets of generators since such groups have nilpotency class and Frattini rank bounded by a. Also, the conclusion of Theorem 7.10 holds true if we replace the probability q_F by any symmetric probability q such that $\inf\{q(s) : s \in \{e\} \cup F\} \ge \varepsilon$ for some fixed $\varepsilon > 0$ and $\mathrm{supp}(q) \subset (\{e\} \cup F \cup F^{-1})^m$ for some fixed m. Theorem 10.9 extends the result to non-symmetric walks.

7.3 Nilpotent Groups with many Generators

The results described in the previous sections give a rather complete description of the behavior of simple random walks on Cayley graphs of finite nilpotent groups when the nilpotency class and the number of generators stay bounded. There are however many interesting examples where one or both of these conditions are violated. The simplest such example is the hypercube $\mathbb{Z}_2^d$ as d varies. In this case, the class is 1 but the minimal number of generators is d. Of course, this walk is well understood. If we denote by $e_1, \dots, e_d$ the natural generators of $\mathbb{Z}_2^d$ and take p to be the uniform probability on $\{e, e_1, \dots, e_d\}$, then the walk driven by p has a cut-off at time $t_n = \frac{1}{4} d \log d$. See Theorem 8.2. It seems very likely that the walks described below present a similar cut-off phenomenon. However, even the existence of a precut-off in the sense of Definition 3.8 is an open problem for these walks. The results presented in this section are taken from Stong's work [132, 133, 134]. They are all based on similar basic ideas introduced by Stong: using the the action of large abelian subgroups and eigenvalue bounds for *twisted graphs*, i.e., weighted graphs whose weights can be complex numbers. These techniques lead to sharp bounds on the second largest eigenvalue β_1 in interesting hard problems. Together with easier bounds on the smallest eigenvalue $\beta_{\min} = \beta_{|G|-1}$, this brings to bear the simple eigenvalue bound (5.9), that is,

$$2\|p^k - u\|_{\text{TV}} \le d_2(p^{(k)}, u) \le \sqrt{|G| - 1}\ \beta_*^k \tag{7.1}$$

where $\beta_* = \max\{\beta_1, -\beta_{\min}\}$ as in (5.9).

Random walk on $U_m(q)$ as m and q vary. Let q be an odd prime and recall that $U_m(q)$ denotes the group of all m by m upper-triangular matrices with coefficients mod q and 1 on the diagonal. This group is generated by the matrix $E_{i,i+1}$, $1 \le i \le m-1$, where $E_{i,j}$ has all its non-diagonal entries 0 except the (i,j) entry which is 1. We set $\Sigma = \{E_{1,2}^{\pm 1}, \ldots, E_{m-1,m}^{\pm 1}\}$ and denote by p the uniform probability on Σ. It is easy to apply Theorem 6.6 using a flow equidistributed on the $2(m-1)$ loops of odd length q defined by $E_{i,i+1}^{\pm j}$, $j = 0, 1, \ldots, q$. This gives $\beta_{\min} \ge -1 + 2/q^2$.

Theorem 7.11 ([132]). *Referring to the walk driven by p on $U_m(q)$ as defined above, there are two constants $c_1, c_2 > 0$ such that for any integer m and any odd prime q, we have*

$$1 - \frac{c_1}{mq^2} \le \beta_1 \le 1 - \frac{c_2}{mq^2}.$$

Ellenberg [56] proved that there are two constants $a_1, a_2 > 0$ such that the diameter D of $(U_m(q), \Sigma)$ satisfies

$$a_1(mq + m^2 \log q) \le D \le a_2(mq + m^2 \log q).$$

Thus the upper bound in Theorem 7.11 is a substantial improvement upon the bound of Theorem 6.2.

As $U_m(q)$ has order $q^{m(m-1)/2}$, the bound (7.1) shows that k of order $m^3 q^2 \log q$ suffices for $p^{(k)}$ to be close to the uniform distribution on $U_m(q)$. For a lower bound, it is not hard to see that $p^{(k)}$ is far from the uniform distribution for $k < \max\{n^2, q^2 n\}$. It would be nice to have a better lower bound.

The Burnside group $B(3, r)$. Around 1900, Burnside asked whether or not a finitely generated group G all of whose elements have finite order must be finite. Golod and Shafarevich proved that the answer is no. Another version of this problem is as follows: Given n, is any finitely generated group of exponent n a finite group? This can be phrased in terms of the Burnside groups $B(n, r)$. By definition, the group $B(n, r)$ is the free group of exponent n with r generators. This means that any group with exponent n and r generators is a quotient of $B(n, r)$. The group $B(n, r)$ can be constructed from the free group F_r on r generators by taking the quotient by the normal subgroup generated by $\{g^n : g \in F_r\}$. It turns out that for all n large enough, $B(n, r)$ is infinite. However $B(n, r)$ is finite for $n = 2, 3, 4, 6$. At this writing, it is not known if $B(5, r)$ is finite or not. See [78, Chapter 18] and also [82, p. 224] for a short discussion and further references. When $B(n, r)$ is infinite, the solution of the restricted Burnside problem due

to Zelmanov asserts that there is a *finite* group $\widetilde{B}(n,r)$ which covers all finite groups generated by r elements and of exponent n. Studying natural random walks on these groups is a tempting but probably extremely hard problem.

For $n=2$, $B(2,r)=\mathbb{Z}_2^r$. The group $B(3,r)$ has order $M=3^{N(r)}$ where $N(r)=r+\binom{r}{2}+\binom{r}{3}$ and its structure is described in [78, p. 322]. In particular, it is nilpotent of class 2 and $B(3,r)/[B(3,r),B(3,r)]=\mathbb{Z}_3^r$.

Theorem 7.12 ([133]). *Consider the Burnside group $B(3,r)$ and let p denote the uniform probability on the r canonical generators and their inverses. Then*

$$1-\frac{3}{2r}\le\beta_1\le 1-\frac{1}{8r}.$$

For the walk in Theorem 7.12, Theorem 6.7 easily gives the lower bound $\beta_{\min}\ge -7/9$. Indeed, by definition of $B(3,r)$, the group of automorphism acts transitively on the generators and any generator gives an obvious loop of length 3. Inequality (7.1) shows that $p^{(k)}$ is close to the uniform distribution on $B(3,r)$ for k of order r^4. The elementary lower bound (5.12) gives that $p^{(k)}$ is not close to the uniform distribution if k is of order $r^3/\log r$.

Polynomials under composition. Let n be an integer and q an odd prime. Let $P_{n,q}$ be the group of all polynomials $\alpha_1 x+\cdots+\alpha_n x^n \mod x^{n+1}$ with $\alpha_1\in\mathbb{Z}_q^*$, $\alpha_2,\dots,\alpha_n\in\mathbb{Z}_q$. The group law is composition. Let α be a generator of $\mathbb{Z}_q^*$. Then $\Sigma=\{x,\alpha^{\pm1}x,(x+x^2)^{\pm1},\dots,(x+x^n)^{\pm1}\}$ is a symmetric generating set. This group is not nilpotent but it contains a large normal nilpotent subgroup, namely, the group $P_{n,q}^1$ of polynomials in $P_{n,q}$ with $\alpha_1=1$. This subgroup has order q^{n-1}. It is proved in [44] that for fixed n, $P_{n,q}$ has A-moderate growth uniformly over the prime q and diameter of order q. Hence, Theorem 7.3 shows that the simple random walk on $(P_{n,q},\Sigma)$ is close to stationarity after order q^2 steps. In [134], Stong is able to compute exactly the second largest eigenvalue of this walk.

Theorem 7.13. *For the simple random walk on the Cayley graph $(P_{n,q},\Sigma)$ defined above, the second largest eigenvalue is*

$$\beta_1=1-\frac{2}{2n+1}\left(1-\cos\frac{2\pi}{q-1}\right).$$

The value given above is slightly different than that found in [134] because we have included the identity element x in Σ to have the easy lower bound $\beta_{\min}\ge -1+2/(2N+1)$ at our disposal. Note that the spectral gap $\lambda_1=1-\beta_1$ is of order $1/(q^2n)$ and that (7.1) shows that order $q^2n^2\log q$ steps suffices to be close to stationarity.

The group $P_{n,q}^1$ is generated by two elements, e.g., $x+x^2$ and $x+x^3$. It is an interesting open problem to study the random walks on $P_{n,q}^1$ and $P_{n,q}$ associated with such small sets of generators.

8 Representation Theory for Finite Groups

Representation theory was first developed as a diagonalization tool. As such, it applies to all convolution operators. On abelian groups, it provides a powerful technique to study random walks as witnessed for instance by the classical proof of the central limit theorem on $\mathbb{R}$. Early references discussing applications to random walks on finite groups are [70, 81] but the first serious application of the representation theory of a non-abelian group to a random walk seems to be in [50] which studies the random transposition walk on the symmetric group. See also [59]. Useful references are [27, 28, 98, 136].

8.1 The General Set-up

A (finite dimensional) *representation* of a group G is a group homomorphism ϱ from G to the group $GL(V)$ of all linear invertible maps of a (finite dimensional) vector space V over the complex numbers. The dimension of V will be denoted by d_ϱ and is called the dimension of the representation. Here, we will consider only finite groups and finite dimensional representations. There always exists on V a Hermitian structure $\langle \cdot, \cdot \rangle$ for which each $\varrho(s)$ is a unitary operator and we always assume that V is equipped with such a structure. The trivial representation of G is (ϱ, V) where $V = \mathbb{C}$ and $\varrho(s)(z) = z$ for all $s \in G$ and $z \in \mathbb{C}$.

The *left regular representation* $\varrho : s \mapsto \varrho(s)$ on $L^2(G)$ is defined by $\varrho(s)f(x) = f(s^{-1}x)$ for all $f \in L^2(G)$. A representation is *irreducible* if any linear subspace W which is invariant by ϱ, i.e., such that $\varrho(s)W \subset W$ for all $s \in G$ is trivial, i.e., is equal to either $\{0\}$ or V. Irreducible representations are the basic building blocks of Fourier analysis. For instance, if the group G is abelian, all the unitary operators $\varrho(s)$, $s \in G$, commute. Thus they can all be diagonalized in the same basis. It follows that any irreducible representation must be 1-dimensional. When the group is not abelian, irreducible representations are typically of dimension greater than 1. Two representations $(\varrho_1, V_1), (\varrho_2, V_2)$ of a group G are *equivalent* if there exists a unitary map $T : V_1 \to V_2$ such that $\varrho_2(s) \circ T = T \circ \varrho_1(s)$. Constructing and classifying irreducible representations up to equivalence is the basic goal of representation theory. We denote by $\widehat{G}$ the set of equivalence classes of irreducible representations of G. For instance, when G is a finite abelian group, one can show that $\widehat{G}$ admits a natural group structure and is isomorphic to G itself.

The famous Shur's lemma implies the following fundamental orthogonality relations. Let (ϱ, V) be an irreducible representation which is not equal to the trivial representation. Let $(e_i)_{1 \le i \le d_\varrho}$ be a Hermitian basis of V and set $\varrho_{i,j}(s) = \langle \varrho(s)e_i, e_j \rangle$. The functions $\varrho_{i,j}$ are called the matrix coefficients of ϱ. For any (i, j) and (k, ℓ) in $\{1, \ldots, d_\varrho\}^2$, the functions $\varrho_{i,j}$ and $\varrho_{k,\ell}$ satisfy

$$\sum_{s \in G} \varrho_{i,j}(s)\overline{\varrho_{k,\ell}(s)} = \frac{|G|}{d_\varrho}\delta_{(i,j),(k,\ell)}.$$

Moreover, for any two inequivalent irreducible representations (ϱ^1, V_1), (ϱ^2, V_2), we have

$$\sum_{s\in G} \varrho^1_{i,j}(s)\overline{\varrho^2_{k,\ell}(s)} = 0$$

for any $1 \le i, j \le d_{\varrho^1}$ and $1 \le k, \ell \le d_{\varrho^2}$. Finally, analyzing the left regular representation, one shows that each irreducible representation ϱ occurs in the left regular representation exactly as many times as its dimension d_ϱ. It follows that

$$|G| = \sum_{\varrho\in\widehat{G}} d_\varrho^2$$

and that the normalized matrix coefficients $d_\varrho^{-1/2}\varrho_{i,j}$, $1 \le i, j \le d_\varrho$, $\varrho \in \widehat{G}$, form an orthonormal basis of $L^2(G)$.

Let p be a measure (a function) on G. Set, for any representation ϱ,

$$\widehat{p}(\varrho) = \sum_{s\in G} p(s)\varrho(s).$$

The linear operator $\widehat{p}(\varrho)$ is called the Fourier transform of p at ϱ. If p, q are two measures, then

$$\widehat{p * q}(\varrho) = \widehat{p}(\varrho)\widehat{q}(\varrho).$$

Hence the Fourier transform turns the convolution product $p * q$ into the product $\widehat{p}(\varrho)\widehat{q}(\varrho)$ of two unitary operators (i.e., the product of matrices once a basis has been chosen in V). In general, one mostly computes the Fourier transform at irreducible representations. For instance, for the uniform measure $u(s) = 1/|G|$, the orthogonality relations recalled above imply that

$$\hat{u}(\varrho) = \begin{cases} 1 & \text{if } \varrho = \mathbf{1} \text{ is the trivial representation} \\ 0 & \text{otherwise.} \end{cases} \tag{8.1}$$

There are straightforward analogs of the Fourier inversion and Plancherel formula which read

$$p(s) = \frac{1}{|G|}\sum_{\varrho\in\widehat{G}} d_\varrho \ \mathrm{tr}[\hat{p}(\varrho)\varrho(s^{-1})],$$

$$\sum_{s\in G} p(s^{-1})q(s) = \frac{1}{|G|}\sum_{\varrho\in\widehat{G}} d_\varrho \ \mathrm{tr}[\hat{p}(\varrho)\hat{q}(\varrho)]$$

where $|G|$ is the cardinality of G. Since $\varrho(s^{-1}) = \varrho(s)^{-1} = \varrho(s)^\dagger$ where $\dagger$ stands for "conjugate-transpose", we have

$$\sum_{s\in G} |p(s)|^2 = \frac{1}{|G|}\sum_{\varrho\in\widehat{G}} d_\varrho \ \mathrm{tr}[\hat{p}(\varrho)\hat{p}(\varrho)^\dagger] \tag{8.2}$$

which is the most important formula for our purpose. Behind this formula is the decomposition of the left regular representation into irreducible components and the fact that each irreducible representation $\varrho \in \widehat{G}$ appears with multiplicity equal to its dimension d_ϱ.

The following lemma follows from (8.1) and (8.2).

Theorem 8.1. *Let p be a probability measure on the finite group G and u the uniform distribution on G. Then, for any integer k,*

$$|G| \sum_{s\in G} |p^{(k)}(s) - u(s)|^2 = \sum_{\varrho\in\widehat{G}^*} d_\varrho \; tr[\hat{p}(\varrho)^k (\hat{p}(\varrho)^k)^\dagger]$$

where $\widehat{G}^* = \widehat{G} \setminus \{\mathbf{1}\}$.

In principle, the meaning of this lemma for random walks on finite groups is clear. Using representation theory, one can compute (or estimate) the square of the L^2-distance

$$d_2(p^{(k)}, u) = |G| \sum_{s\in G} |p^{(k)}(s) - u(s)|^2$$

whenever one can compute (or estimate)

$$\sum_{\varrho\in\widehat{G}^*} d_\varrho \; \mathrm{tr}[\hat{p}(\varrho)^k (\hat{p}(\varrho)^k)^\dagger].$$

This requires having formula for the dimensions d_ϱ of all irreducible representations and being able to compute the powers of the matrices $\hat{p}(\varrho)$. Once these preliminary tasks have been tackled, one still has to sum over all irreducible representations.

8.2 Abelian Examples

Let G be a finite abelian group and p a probability measure on G. Viewed as a convolution operator acting on $L^2(G)$, p has adjoint $\check{p}$. As G is abelian, the convolution product is commutative. It follows that p is normal, hence diagonalizable. As all the irreducible representations are one dimensional, each gives rise to exactly one matrix coefficient called the character χ of the representation. The characters form an orthonormal basis of $L^2(G)$ and they also form a group, the dual group $\widehat{G}$, isomorphic to G. The Fourier transform $\hat{p}$ at the character χ (i.e., at the representation with character χ) is given by

$$\hat{p}(\chi) = \sum_{s\in G} p(s)\chi(s).$$

The collection $(\hat{p}(\chi))_\chi$ indexed by the characters, is exactly the spectrum of p viewed as a convolution operator. In this case, the formula of Theorem 8.1 gives

$$d_2(p^{(k)}, u)^2 = |G| \sum_{s\in G} |p^{(k)} - u(s)|^2 = \sum_{\chi\in\widehat{G}^*} |p(\chi)|^{2k}. \tag{8.3}$$

The simple random walk on $\mathbb{Z}_n$. Consider the group $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{0, 1, \dots, n-1\}$. In this case, the characters are the functions

$$\chi_\ell(x) = \mathrm{e}^{-2\mathrm{i}\pi\ell x/n}, \ \ \ell = 0, \dots, n-1.$$

Let $p(+1) = p(-1) = 1/2$. Then $\hat{p}(\chi_\ell) = \cos(2\pi\ell/n)$. Hence,

$$d_2(p^{(k)}, u) = \left(\sum_{\ell=1}^{n-1} |\cos(2\pi\ell/n)|^{2k} \right)^{1/2}.$$

If n is even, for $\ell = n/2$, we get $\cos\pi = -1$ as an eigenvalue. Indeed, the chain is periodic of period 2 in this case. As a typical careful application of eigenvalue techniques, we state the following result.

Theorem 8.2. *There exist two constants $0 < c_1 \le C_1 < \infty$ such that, for all odd integers $n = 2m+1$ and all integers k, we have*

$$2|\cos(\pi/n)|^{2k} \left(1 + \frac{c_1 n}{\sqrt{k}}\right) \le d_2(p^{(k)}, u)^2 \le 2|\cos(\pi/n)|^{2k} \left(1 + \frac{C_1 n}{\sqrt{k}}\right).$$

Proof. Assume that $n = 2m+1$ is odd. Using the symmetries of cos, we get

$$d_2(p^{(k)}, u)^2 = 2 \sum_{\ell=1}^{m} |\cos(\pi\ell/n)|^{2k}.$$

Calculus gives

$$\log \frac{\cos t}{\cos s} \left\} \begin{array}{ll} \le -\frac{1}{2}(t^2 - s^2) & \text{for } 0 < s < t < \pi/2 \\ \ge -\frac{2}{\pi}(t^2 - s^2) & \text{for } 0 < s < t < \pi/4. \end{array} \right.$$

Hence

$$\begin{aligned} d_2(p^{(k)}, u)^2 &\ge 2|\cos(\pi/n)|^{2k} \left(\sum_{\ell=1}^{m/2} \mathrm{e}^{-4\pi(\ell^2-1)k/n^2} \right) \\ &\ge 2|\cos(\pi/n)|^{2k} \left(1 + c_1\sqrt{n^2/k}\right) \end{aligned}$$

where $c_1 = \mathrm{e}^{-8\pi}$. For an almost matching upper bound, write

$$\begin{aligned} \sum_{\ell=1}^{m} \mathrm{e}^{-2\pi^2(\ell^2-1)k/n^2} &\le 1 + \sum_{\ell=1}^{\infty} \mathrm{e}^{-2\pi^2\ell^2 k/n^2} \le 1 + \int_0^\infty \mathrm{e}^{-2\pi^2 t^2 k/n^2} dt \\ &= 1 + C_1\sqrt{n^2/k}. \end{aligned}$$

with $C_1 = 1/\sqrt{8\pi}$. Hence $d_2(p^{(k)}, u)^2 \le 2|\cos(\pi/n)|^{2k} \left(1 + C_1\sqrt{n^2/k}\right)$. □

Other random walks on $\mathbb{Z}_n$. Let $a, b \in \mathbb{Z}_n$ and let $p_{a,b}$ be the uniform probability measure on $\{a, b\}$, i.e., $p(a) = p(b) = 1/2$. Thus the measure p of the previous example is $p_{-1,1}$ in this notation. Let us look at $p_{0,1}$. The associated random walk is not reversible but it is ergodic for all n. Here the eigenvalues are $\frac{1}{2}(1 + e^{2i\pi\ell/n})$. As $|1 + e^{2i\pi\ell/n}|^2 = |\cos(\pi\ell/n)|^2$, we get

$$d_2(p_{0,1}^{(k)}, u)^2 = \sum_1^{n-1} |\cos(\pi\ell/n)|^2.$$

Now, if n is odd, one easily checks that

$$\sum_1^{n-1} |\cos(\pi\ell/n)|^2 = \sum_1^{n-1} |\cos(2\pi\ell/n)|^2.$$

This shows that, for all odd n and all k, $d_2(p_{-1,1}^{(k)}, u) = d_2(p_{0,1}^{(k)}, u)$. The following result generalizes this observation.

Theorem 8.3. *Let $a, b \in \mathbb{Z}_n$. Then the random walk driven by the uniform probability measure $p_{a,b}$ on $\{a, b\}$ is ergodic if and only if*

$$b - a \text{ and } n \text{ are relatively prime.} \tag{8.4}$$

For any $s \in [1, \infty]$, any a, b satisfying (8.4) *and any integer k, we have*

$$d_s(p_{a,b}^{(k)}, u) = d_s(p_{0,1}^{(k)}, u).$$

Moreover, there are constants c, C such that for any a, b satisfying (8.4) *and any integer k, we have*

$$2|\cos(\pi/n)|^{2k}\left(1 + \frac{c_1 n}{\sqrt{k}}\right) \le d_2(p_{a,b}^{(k)}, u)^2 \le 2|\cos(\pi/n)|^{2k}\left(1 + \frac{C_1 n}{\sqrt{k}}\right).$$

Proof. The first assertion follows for instance from Proposition 2.3. Given that (8.4) holds, there is an invertible affine transformation $\phi : x \mapsto uz + v$ such that $\phi(a) = 0$, $\phi(b) = 1$. Hence, as functions on $\mathbb{Z}_n$, $p_{a,b} = p_{0,1} \circ \phi$. Moreover, because ϕ is affine, for any two probabilities p, q, $[p \circ \phi] * [q \circ \phi](x) = p * q(\phi(x) + v)$. Hence, $p_{a,b}^{(k)}(x) = p_{0,1}^{(k)}(\phi(x) + (k-1)v)$. As $z \mapsto \phi(z) + (k-1)v$ is a bijection, we have $d_s(p_{a,b}^{(k)}, u) = d_s(p_{0,1}^{(k)}, u)$. The last assertion is obtained as in Theorem 8.2. □

We now consider what happens when $p = p_\Sigma$ is uniform on a subset Σ of $\mathbb{Z}_n$ having $m > 2$ elements where m is fixed. Theorems 7.3, 7.8 and 7.9 apply in this case and show that if Σ is symmetric, and $0 \in \Sigma$ then $c(m)D^2 \le T(Z_n, p_\Sigma) \le C(m)D^2$ where D is the diameter of the associated Cayley graph (the condition that Σ be symmetric and contains 0 can be removed and replaced by the condition that $\Sigma\Sigma^{-1}$ generates). For instance,

it is not hard to use this to show that, for any fixed m the walk driven by the uniform measure p_{Σ_m} on $\Sigma_m = \{0, \pm 1, \pm\lfloor n^{1/m}\rfloor, \dots, \pm\lfloor n^{(m-1)/m}\rfloor\}$ satisfies $c(m)n^{2/m} \le T(\mathbb{Z}_n, \Sigma_k) \le C(m)n^{2/m}$ (the same is true for the non-symmetric version of Σ_m, i.e., $\Sigma'_m = \{\{0, 1, \lfloor n^{1/m}\rfloor, \dots, \lfloor n^{(m-1)/m}\rfloor\}$).

The works [24, 72, 87] contain interesting complementary results derived through a careful use of representation theory in the spirit of this section.

Theorem 8.4 ([72], see also [87]). *Let p be any probability measure on $\mathbb{Z}_n$. Assume that the support of p is of size $m+1 > 2$. There exist $c = c(m)$ and $N = N(m)$ such that, for $k < cn^{2/m}$ and for all $n > N$, we have $\|p^{(k)} - u\|_{\text{TV}} \ge 1/4$.*

Call a subset $\{a_0, \dots, a_m\} \subset \mathbb{Z}_m$ aperiodic if the greater common divisor of $a_1 - a_0, \dots, a_m - a_0$ and n is 1. Let u_Σ denote the uniform probability on Σ.

Theorem 8.5 ([24]). *Fix $m \ge 2$. Let Σ be chosen uniformly at random from all aperiodic $m+1$-subsets of $\mathbb{Z}_n$. Let $\psi(n)$ be any function increasing to infinity and assume that $k_n \ge \psi(n)n^{2/m}$. Then*

$$\mathbf{E}(\|u_\Sigma^{(k_n)} - u\|_{\text{TV}}) \to 0 \text{ as } n \to \infty$$

where the expectation is relative to the choice of the set Σ.

When n is prime this can be improved as follows.

Theorem 8.6 ([87]). *Fix $m \ge 2$ and assume that n is a prime. Let Σ be chosen uniformly at random from all $m+1$-subsets of $\mathbb{Z}_n$. Given $\varepsilon > 0$, there exist $c = c(m, \varepsilon)$ and $N = N(m, \varepsilon)$ such that, for all $n > N$ and $k > cn^{2/m}$, we have $\mathbf{E}(\|u_\Sigma^{(k)} - u\|_{\text{TV}}) < \varepsilon$.*

The simple random walk on the hypercube. Let $G = \mathbb{Z}_2^d$ be the hypercube and consider the simple random walk driven by the measure p at (5.14), i.e., the uniform measure on $\{e_0, e_1, \dots, e_d\}$ where $e_0 = (0, \dots, 0)$ and e_i, $1 \le i \le d$ are the natural basis vectors of $\mathbb{Z}_2^d$.

The characters of G, indexed by $\widehat{G} = G$ are given by $\chi_y(x) = (-1)^{x.y}$ where $x.y = \sum_1^d x_i y_i$. Hence, p has eigenvalues $\hat{p}(\chi_y) = 1 - 2|y|/(d+1)$ where $|y| = \sum_1^d y_i$. Now (8.3) becomes

$$d_2(p^{(k)}, u)^2 = \sum_1^d \binom{d}{j} \left(1 - \frac{2j}{d+1}\right)^{2k}.$$

For $k = \frac{1}{4}(d+1)[\log d + c]$ with $c > 0$, this yields (see [27, p. 28])

$$2\|p^{(k)} - u\|_{\text{TV}} \le d_2(p^{(k)}, u)^2 \le 2\left(\mathrm{e}^{\mathrm{e}^{-c}} - 1\right).$$

Together with the lower bound in total variation of Section 5.3, this proves that the simple random walk on the hypercube has a cut-off at time $t_d = \frac{1}{4}d\log d$. By a more direct method, Diaconis, Graham and Morrison prove the following complementary results.

Theorem 8.7 ([35]). *Referring to the above walk on the hypercube $\mathbb{Z}_2^d$, for any $k = \frac{1}{4}(d+1)[\log d + c]$, $c \in \mathbb{R}$*

$$\|p^{(k)} - u\|_{\text{TV}} = 1 - 2\Phi\left(-\frac{e^{-2c}}{4}\right) + o(1)$$

where

$$\Phi(t) = \frac{1}{2\pi}\int_{-\infty}^{t} e^{-s^2/2} ds.$$

Note that the automorphism group of $\mathbb{Z}_2^d$ acts transitively on the set of all d-tuples that generate $\mathbb{Z}_2^d$ which means that all generating d-tuples are equivalent from our viewpoint.

Other walks on the hypercube. The papers [73, 140] consider what typically happens for walks on the hypercube driven by the uniform measure u_Σ on a generating set Σ with $n > d$ elements. In particular, [140] proves the following result. Set

$$H(x) = x \log_2 x^{-1} + (1-x)\log_2(1-x)^{-1}.$$

This function is increasing from $H(0) = 0$ to $H(1/2) = 1$. Let H^{-1} be the inverse function from $[0,1]$ to $[0,1/2]$ and set

$$T(d,n) = \frac{n}{2}\log\frac{1}{1-2H^{-1}(d/n)}.$$

Theorem 8.8 ([140]). *Assume that the random walk driven by the uniform probability u_Σ on the set Σ of n elements in $\mathbb{Z}_2^d$ is ergodic. For any $\varepsilon > 0$, for all d large enough and $n > d$, we have:*

- *For any set Σ, if $k \le (1-\varepsilon)T(d,n)$ then $\|u_\Sigma^{(k)} - u\|_{\text{TV}} > 1 - \varepsilon$.*
- *For most sets Σ, if $k \ge (1+\varepsilon)T(d,n)$ then $\|u_\Sigma^{(k)} - u\|_{\text{TV}} < \varepsilon$.*

Thus the lower bound holds for all choices of Σ whereas the upper bounds holds only with probability $1-\varepsilon$ when the set Σ is chosen at random. Also, when n is significantly larger than d, the walk is ergodic for most choices of Σ. The function $T(d,n)$ has the following behavior (see [140]):

$$T(d,n) \sim \frac{d}{4}\log\frac{d}{n-d} \qquad \text{if } n - d = o(d)$$

$$T(d,n) \sim \frac{d}{\log_2(n/d)} \qquad \text{if } d/n = o(1).$$

When n is linear in d then $T(d,n)$ is also linear in d. For instance, $T(d,2d) \sim ad$ with $0.24 < a < 0.25$. This leads to the following open question.

Problem 8.9. Find an explicit set of $2d$ elements in $\mathbb{Z}_2^d$ whose associated walk reaches approximate stationarity after order d steps.

The arguments in [140] do not use characters or eigenvalues directly. In fact, Wilson observes in [140] that for n linear in d the walk driven by u_Σ typically reaches stationarity strictly faster in total variation than in the d_2 distance for which we have the equality (5.8).

Wilson's result for random subsets contrasts with what is known for explicit sets. Uyemura-Reyes [138] studies the walk on the hypercube driven by

$$p(x) = \begin{cases} 1/(2d) & \text{if } x = (0, \dots,) \text{ or } (1, \dots, 1) \\ 1/d^2 & \text{if } x = \sum_{\ell=i}^{i+j} e_\ell, 1 \le i \le d, 1 \le j < d \\ 0 & \text{otherwise} \end{cases}$$

where, in the second line, $i + j$ is understood mod d. For reasons explained in [138], this is called the random spatula walk. It is proved in [138] that this walk has a cut-off at time $t_n = \frac{1}{8} d \log d$.

The simple random walk on $\mathbb{Z}_n^d$. In $\mathbb{Z}_n^d$, let $e = (0, \dots, 0)$ and e_i have a single non-zero coordinate, the i-th, equal to 1. Let n be odd and p be the uniform measure on $\{\pm e_i : 0 \le i \le d\}$. It is noteworthy that obtaining good uniform bounds over the two parameters n and d for this walk is not entirely trivial. The eigenvalues are easy to write down. They are

$$\alpha_\ell = \frac{1}{d} \left(\sum_1^d \cos(2\pi \ell_i / n) \right)$$

with $\ell = (\ell_1, \dots, \ell_d) \in \{0, \dots, n-1\}^d$. But bounding $d_2(p^{(k)}, u)^2 = \sum_{\ell \neq 0} \alpha_\ell^{2k}$ is not an easy task. One way to solve this difficulty is to use the associated continuous-time measure H_t defined at (2.10) and Theorem 5.1. This technique works for problems having a product structure similar to the present example. See [42, Section 5]. The reason this is useful is because H_t turns out to be a product measure. Namely, if $x = (x_1, \dots, x_d)$,

$$H_t(x) = \prod_1^d H_{1,t/d}(x_i)$$

where $H_{1,t}$ corresponds to the random walk on $\mathbb{Z}_n$ driven by the measure $p_1(\pm 1) = 1/2$. It follows that (u_1 denotes the uniform measure on $\mathbb{Z}_n$)

$$d_2(H_t, u)^2 = \left(1 + d_2(H_{1,t/d}, u_1)^2\right)^d - 1.$$

It is not hard to obtain good upper and lower bounds for

$$d_2(H_{1,t}, u_1)^2 = \sum_{j=1}^{n-1} \mathrm{e}^{-2t[1-\cos(2\pi j/n)]}.$$

Namely, setting $\lambda(n) = 1 - \cos(2\pi/n)$ we have

$$\left(1+\frac{cn}{\sqrt{t}}\right)\mathrm{e}^{-2t\lambda(n)} \le d_2(H_{1,t},u_1)^2 \le \left(1+\frac{Cn}{\sqrt{t}}\right)\mathrm{e}^{-2t\lambda(n)}.$$

This analysis, the elementary inequalities

$$\forall\, x>0,\ d\in\mathbb{N},\ \ dx(1+x/2)^{d-1} \le (1+x)^d - 1 \le dx(1+x)^{d-1},$$

and Theorem 5.1 yield the following result.

Theorem 8.10. *There are constants $c, C \in (0,\infty)$ such that, for the simple random walk on $\mathbb{Z}_n^d$, we have*

$$F_{n,d}(c,t) \le d_2(H_t,u)^2 \le F_{n,d}(C,t)$$

with $\lambda(n) = 1-\cos(2\pi/n)$ and

$$F_{n,d}(a,t) = d\left(1+a\sqrt{\frac{dn^2}{t}}\right)\left(1+\left(1+a\sqrt{\frac{dn^2}{t}}\right)\mathrm{e}^{-2t\lambda(n)/d}\right)^{d-1}\mathrm{e}^{-2t\lambda(n)/d}$$

Moreover, there exists a constant C_1 such that, if n is an odd integer, d is large enough, and

$$k > 1+\frac{d\log d}{2\lambda(n)}+\frac{d\theta}{2\lambda(n)}$$

with $\theta>0$, then

$$2\|p^{(k)}-u\|_{\mathrm{TV}} \le d_2(p^{(k)},u) \le C_1\mathrm{e}^{-\theta}.$$

Finally, for any $\tau > 6/d$, we have $\|p^{(k)}-u\|_{\mathrm{TV}} \ge 1-\tau$ if

$$k < \frac{\log(d\tau/6)}{-2\log(1-\lambda(n)/d)}.$$

Note that the discrete time upper bound uses the fact that when n is odd, the lowest eigenvalue is $\cos(\pi/n)$ whose absolute value is much smaller than $1-\lambda(n)/d$ for d large enough ($d\ge 8$ suffices). Theorem 8.10 proves a cut-off at time $(d/2\lambda(n))\log d$ as long as d tends to infinity (n can be fixed or can tend to infinity).

8.3 Random Random Walks

In the spirit of Theorem 8.8, consider a group G, an integer m, and pick uniformly at random an m-set $\Sigma = \{g_1,\dots,g_m\}$. Consider the random walk on G driven by the uniform probability measure u_Σ. What is the "typical" behavior of such a walk? Let $\mathbf{E}$ denote the expectation relative to the random choice of Σ. What can be said about $\mathbf{E}\left(\|u_\Sigma^{(k)}-u\|_{\mathrm{TV}}\right)$? To obtain some meaningful answers, we consider this problem for families of groups (G_n) where the size of G_n grows to infinity with n as in the following open problem. Recall that a classical result [52] asserts that the probability that a random pair of elements of the alternating group A_n generates A_n tends to 1 as n tends to infinity.

Problem 8.11. What is the typical behavior of the random walk driven by u_Σ when Σ is a random pair (more generally a random m-set) in A_n and n tends to infinity?

This is a wide open question. However, interesting results have been obtained in the case where $m = m(G)$ is allowed to grow with the order $|G|$ of G and this growth is fast enough.

Large random sets. In his unpublished thesis [53], C. Dou proves the following result using Theorem 8.1 and some combinatorics.

Theorem 8.12. *Let G be a finite group of order $|G|$. Let Σ be an m-element set chosen uniformly at random from G. Then*

$$\mathbf{E}\left(\|u_\Sigma^{(k)} - u\|_{TV}\right) \le \frac{1}{2}\left(\frac{(2k)^{2k}|G|}{m^k}\right)^{1/2}.$$

To illustrate this result, fix an integer s and take $m \ge |G|^{1/s}$ and $k = s+1$. Then the right-hand side is $\frac{1}{2}[2(s+1)]^{2(s+1)}|G|^{-1/s}$ which tends to 0 as $|G|$ tends to ∞. For instance, most random walks based on sets of size $\sqrt{|G|}$ reach approximate stationarity in 3 steps. As a second example, consider sets of fixed size $m \ge a(\log|G|)^{2s}$ with $a > 4$ and $s > 1$. Then, there exists $\delta > 0$ such that for $k = (\log|G|)^s$ we have

$$\mathbf{E}\left(\|u_\Sigma^{(k)} - u\|_{\mathrm{TV}}\right) \le \exp(-\delta(\log|G|)^s).$$

In [54], the approach of [53] is developed further to obtain the following.

Theorem 8.13 ([54]). *Let $m = \lfloor(\log|G|)^s\rfloor$ for some fixed $s > 1$. Let $\varepsilon > 0$ be given. Let Σ be a m-element set chosen uniformly at random in a finite group G. Then for*

$$k > \frac{s}{s-1}\frac{\log|G|}{\log m}(1+\varepsilon)$$

we have that $\mathbf{E}\left(\|u_\Sigma^{(k)} - u\|_{TV}\right)$ tends to 0 as $|G|$ tends to infinity.

This result cannot be improved as shown by an earlier result of Hildebrand [87] concerning abelian finite groups. See [54] for a slightly more general result.

Theorem 8.14 ([87]). *Let $\varepsilon > 0$ be given. Let G be a finite abelian group. Let $m = \lfloor(\log|G|)^s\rfloor$ for some fixed $s > 1$. Let Σ be a m-element set chosen uniformly at random in a finite abelian group G. Then for*

$$k < \frac{s}{s-1}\frac{\log|G|}{\log m}(1-\varepsilon)$$

we have that $\mathbf{E}\left(\|u_\Sigma^{(k)} - u\|_{TV}\right)$ tends to 1 as $|G|$ tends to infinity.

For further results in this direction, see [88, 89, 113, 120].

9 Central Measures and Bi-invariant Walks

9.1 Characters and Bi-invariance

When the group G is not abelian, e.g., $G = S_n$, the formula of Theorem 8.1 is often quite hard to use in practice, even when $p = \check{p}$ is symmetric. Indeed, $p(x^{-1}y)$ defines a $|G| \times |G|$ matrix whose eigenvalues we would like to find. What Theorem 8.1 does is to decompose this into $|\widehat{G}|$ smaller problems, one for each irreducible representation ϱ. The matrix $\hat{p}(\varrho)$ has size $d_\varrho \times d_\varrho$. This is very useful if d_ϱ is small. Unfortunately, irreducible representations of non-abelian finite groups tend to have large dimensions. For instance, for the symmetric group S_n, it is known that the typical dimension of a representation is $\sqrt{n!}$. Because of this, Theorem 8.1 is useful mostly in cases where p has further symmetries. The typical case is when p is a *central probability*, that is, it satisfies

$$\forall\, x, y \in G, \quad p(y^{-1}xy) = p(x). \tag{9.1}$$

Functions (probabilities) with this property are also called *class functions* since they are exactly the functions which are constant on conjugacy classes. Indeed, by definition, the conjugacy classes are exactly the classes of elements of G for the equivalence relation defined by $x \sim y$ iff $x = z^{-1}xz$ for some $z \in G$. When p is central, the associated Markov chain is not only left- but also right-invariant, that is, satisfies

$$\mathbb{P}_e(X_n = y) = \mathbb{P}_x(X_n = xy) = \mathbb{P}_x(X_n = yx)$$

for all $x, y \in G$. Such random walks are called *bi-invariant* random walks.

To each representation ϱ of G, one associates its *character*

$$\chi_\varrho(x) = \mathrm{tr}(\varrho(x)) = \sum_1^{d_\varrho} \varrho_{i,i}(x).$$

These functions are all central functions and $\chi_\varrho(s^{-1}) = \overline{\chi_\varrho(s)}$. Moreover $|\chi_\varrho(s)|$ is maximum at $s = e$ where $\chi_\varrho(e) = d_\varrho$. From the orthogonality relations it follows immediately that the characters of all irreducible representations form an orthonormal family in $L^2(G)$. Moreover, if p is any central measure (function) and ϱ is an irreducible representation, then

$$\hat{p}(\varrho) = \lambda_\varrho(p) I_{d_\varrho}, \quad \lambda_\varrho(p) = \frac{1}{d_\varrho} \sum_{s \in G} p(s)\chi_\varrho(s)$$

where I_{d_ϱ} is the $d_\varrho \times d_\varrho$ identity matrix. See, e.g., [27, 28, 59]. It follows that the *irreducible characters*, i.e., the characters associated with irreducible representations, form a basis of the subspace of all central functions in $L^2(G)$. Hence the number of irreducible representations up to equivalence, i.e., $|\widehat{G}|$, equals the number of conjugacy classes in G. This leads to the following general result. See, e.g., [27, 59].

Theorem 9.1. *Let $C_1, \dots, C_m$ be conjugacy classes in G with representatives $c_1, \dots c_m$. Assume that p is a central probability measure supported on $\cup_1^m C_i$. Then*

$$d_2(p^{(k)}, u)^2 = \sum_{\varrho \in \widehat{G}} d_\varrho^2 \left(\sum_1^m p(C_i) \frac{\chi_\varrho(c_i)}{\chi_\varrho(e)} \right)^{2k}. \tag{9.2}$$

Representation and character theory of finite groups is an important and well studied subject and there is sometimes enough information on characters available in the literature to make this theorem applicable. What is needed are manageable formulas or estimates for the dimensions d_ϱ of all irreducible representations and for the character ratios $\chi(c_i)/\chi(e)$.

Even when such data is available, estimating the sum on the left-hand side of (9.2) can still be quite a challenge. Indeed, this is a huge sum and it is often not clear at all how to identify the dominant terms.

9.2 Random Transposition on the Symmetric Group

Representation theory of the symmetric group. We will illustrate Theorem 9.1 by examples of bi-invariant walks on the symmetric group S_n. See [27] for a detailed treatment and [31] for a survey of further developments. The irreducible representations of the symmetric group are indexed by the set of all partitions λ of n where a partition $\lambda = (\lambda_1, \dots, \lambda_r)$ has $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and $\sum_1^r \lambda_i = n$. It is useful to picture the partition $\lambda = (\lambda_1, \dots, \lambda_r)$ as a diagram made of r rows of square boxes, the i-th row having λ_i boxes. The rows are justified on the left. See [27, 59] for pointers to the literature concerning the representation theory on the symmetric group. For instance, for $n = 10$ the partition $\lambda = (5, 4, 1)$ is pictured in Figure 1.

Denote by d_λ the dimension of the irreducible representation ϱ_λ indexed by λ. Then d_λ equals the number of ways of placing the numbers $1, 2, \dots, n$ into the diagram of λ such that the entries in each row and column are increasing. This is by no mean an easy number to compute or estimate.

The partition $\lambda = (n)$ corresponds to the trivial representation, (dimension 1). The partition $(1, 1, \dots, 1)$ corresponds to the sign representation (dimension 1). The partition $(n-1, 1)$ corresponds to the representation $\varrho_{(n-1,1)}$ of S_n on $V = \{(z_1, \dots, z_n) \in \mathbb{C}^n : \sum z_i = 0\}$ where $\varrho_{(n-1,1)}(\sigma)$ is represented

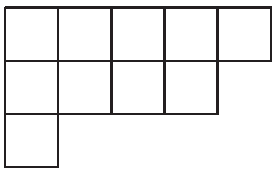

Fig. 1. $\lambda = (5, 4, 1)$

in the canonical basis of $\mathbb{C}^n$ by the matrix with coefficients $m_{i,j} = \delta_{i,\sigma(j)}$. This representation $\varrho_{(n-1,1)}$ has dimension $d_\lambda = n-1$ (the only free choice is the number between 2 and n which goes in the unique box on the second row of the diagram).

The next necessary ingredient in applying Theorem 9.1 are formulas for character values. Such formulas were given by Frobenius but they become unwieldy for conjugacy classes with a complex cycle structure. Which character values are needed depend on exactly which random walk is considered. The simplest case concerns the walk called *random transposition*.

Random transposition. Consider n cards laid out on a table in a row. Let the right and left hands each pick a card uniformly and independently and switch the positions of the cards (if both hands pick the same card, the row of card stays unchanged). This description gives the random transposition measure p_{RT} on S_n defined at (4.1). Since $\{e\}$ and $T = \{\tau_{i,j} : 1 \le i < j \le n\}$ are conjugacy classes, Theorem 9.1 applies. Now, we need the character values $\chi_\lambda(e) = d_\lambda$ and $\chi_\lambda(t)$ where t is any fixed transposition. Frobenius' formula gives

$$\frac{\chi_\lambda(t)}{\chi_\lambda(e)} = \frac{1}{n(n-1)} \sum_j \left(\lambda_j^2 - (2j-1)\lambda_j\right)$$

from which it follows that the eigenvalues of this walk are

$$\begin{aligned} p_{\text{RT}}(e) + p_{\text{RT}}(T)\frac{\chi_\lambda(t)}{\chi_\lambda(e)} &= \frac{1}{n} + \frac{n-1}{n}\frac{\chi_\lambda(t)}{\chi_\lambda(e)} \\ &= \frac{1}{n} + \frac{1}{n^2}\sum_j \left(\lambda_j^2 - (2j-1)\lambda_j\right) \end{aligned}$$

with multiplicity d_λ^2. With some work, one shows that the second largest eigenvalue is $1-2/n$ with multiplicity $(n-1)^2$, attained for $\lambda = (n-1,1)$. The lowest eigenvalue is $-1+2/n$ with multiplicity 1, attained for $\lambda = (1,1,\ldots,1)$.

Using the above data and estimates on d_λ, Diaconis and Shahshahani obtained in 1981 the following theorem which gives first precise result about the convergence of a complex finite Markov chain.

Theorem 9.2 ([50]). *For the random transposition walk on the symmetric group S_n, there exists a constant A such that, for all n and $c > 0$ for which $k = \frac{1}{2}n(\log n + c)$ is an integer, we have*

$$2\|p_{\text{RT}}^{(k)} - u\|_{\text{TV}} \le d_2(p_{\text{RT}}^{(k)}, u) \le A\mathrm{e}^{-c}.$$

Moreover, there exist a function f with limit 0 at ∞ such that for all $n > 5$ and all $c > 0$ for which $k = \frac{1}{2}n(\log n - c)$ is an integer,

$$\|p_{\text{RT}}^{(k)} - u\|_{\text{TV}} \ge 1 - 12\left(\mathrm{e}^{-c} + n^{-1}\log n\right).$$

This theorem proves that (S_n, p_{RT}) has a total variation cut-off and a L^2-cut-off, both a time $\frac{1}{2}n\log n$. Let us comment further on the lower bound. It can be proved ([27, p. 44]) by using Propositions 5.6, 5.7, the fact that

$$\chi^2_{(n-1,1)} = \chi_{(n)} + \chi_{(n-1,1)} + \chi_{(n-2,2)} + \chi_{(n-2,1,1)},$$

and the values of the corresponding eigenvalues and dimensions. This formula giving $\chi^2_{(n-1,1)}$ is a classical result in representation theory. It corresponds to the decomposition into irreducible components of the tensor product $\varrho_{(n-1,1)} \otimes \varrho_{(n-1,1)}$. Another proof using classical probability estimates can be obtained by adapting the argument of [27, p. 43].

9.3 Walks Based on Conjugacy Classes of the Symmetric Group

A conjecture. In principle, it is possible to use character bounds to study any random walk on the symmetric group whose driving measure is central. However, the computational difficulty increases rapidly with the complexity of the conjugacy classes involved. To state some results and conjectures, recall that any conjugacy class C on S_n can be described by the common disjoint cycle structure of its elements. Thus $C = (2)$ means C is the class of all transpositions, $C = (5, 3, 3, 2, 2, 2, 2)$ means C is the class of all permutations that can be written as a product of one 5-cycle, two 3-cycles and four 2-cycles where the supports of those cycles are pairwise disjoint. It is known (and not hard to prove) that any odd conjugacy class (i.e., whose elements have sign -1) generates the symmetric group. However the walk associated to the uniform measure on an odd conjugacy class is always periodic of period 2. To cure this parity problem consider, for any odd conjugacy class C on S_n the probability measure p_C defined by

$$p_C(\theta) = \begin{cases} 1/2 & \text{if } \theta = e \\ 1/[2\#C] & \text{if } \theta \in C \\ 0 & \text{otherwise.} \end{cases}$$

This is sometimes referred to as a *lazy random walk* because, on average, it moves only every other steps, see, e.g., [88, 89]. Thus, the walk driven by $p_{(2)}$ is similar to the random transposition walk except that it stay put with probability $1/2$ instead of $2/n$. One can show that Theorem 9.2 applies to the walk generated by $p_{(2)}$ if $k = \frac{1}{2}n(\log n \pm c)$ is changed to $k = n(\log n \pm c)$.

For $C = (c_1, c_2, \ldots, c_\ell)$, set $|C| = \sum_1^\ell c_i$. Note that $|C|$ is the size of the support of any permutation in C, i.e., n minus the number of fixed points. With this notation one can make the following conjecture.

Conjecture 9.3. There exists a constant A such that, for all n, all odd conjugacy classes C with $|C| \ll n$, and all $c > 0$ for which $k = (2n/|C|)(\log n + c)$ is an integer, we have

$$2\|p_C^{(k)} - u\|_{\text{TV}} \le d_2(p_C^{(k)}, u) \le A\mathrm{e}^{-c}.$$

Moreover, there exist two functions f_C^1, f_C^2 with limit 0 at ∞ such that for all n and all $c > 0$ for which $k = (2n/|C|)(\log n\ - c)$ is an integer,

$$\|p_C^{(k)} - u\|_{\text{TV}} \ge 1 - f_C^1(c) - f_C^2(n).$$

Any even conjugacy class C of S_n generates the alternating group A_n (except for $n = 4$) and one can consider the random walk on A_n driven by the uniform measure on C. Denote by $\widetilde{p}_C$ the uniform measure on the conjugacy class C viewed as a subset of A_n. For $\widetilde{p}_C$ it is conjectured that the statement of Conjecture 9.3 holds with $k = (n/|C|)(\log n\ + c)$ instead of $k = (2n/|C|)(\log n\ + c)$.

Conjecture 9.3 can be interpreted in various ways depending of what is meant by $|C| \ll n$. It is open even for fixed $|C|$ such as $|C| = 20$ and n tending to infinity. The strongest reasonable interpretation is $|C| \le (1 - \varepsilon)n$, for some fixed $\varepsilon > 0$. What is known at this writing is described in the next section.

Small conjugacy classes. For $|C| \le 6$ and n tending to infinity, Conjecture 9.3 (and its even conjugacy class version on A_n) is proved in [121, 122]. Moreover, [121, 122] shows that the lower bound holds true for all C such that $|C| < n/(1 + \log n)$ (some of the computations in the proof given in [121, 122] are incorrect but these errors can easily be fixed).

To give an idea of the difficulties that arise in adapting the method used for random transposition, we give below some explicit character values. The source is [93] and [121, 122]. For any partition $\lambda = (\lambda_1, \ldots, \lambda_r)$ and $\ell = 1, 2, \ldots$, set

$$M_{2\ell,\lambda} = \sum_{j=1}^{r} \left[(\lambda_j - j)^\ell(\lambda_j - j + 1)^\ell - j^\ell(j-1)^\ell\right]$$

$$M_{2\ell+1,\lambda} = \sum_{j=1}^{r} \left[(\lambda_j - j)^\ell(\lambda_j - j + 1)^\ell(2\lambda_j - 2j + 1) + j^\ell(j-1)^\ell(2j-1)\right].$$

For a conjugacy class C, set $r_\lambda(C) = \chi_\lambda(c)/\chi_\lambda(e)$ where c is any element of C. These character ratios are the building blocks needed to apply formula (9.2). For the conjugacy classes $(4), (2,2)$ and (6), one has:

$$r_\lambda((4)) = \frac{(n-4)!}{n!}\left(M_{4,\lambda} - 2(2n-3)M_{2,\lambda}\right)$$

$$r_\lambda((2,2)) = \frac{(n-4)!}{n!}\left(M_{2,\lambda}^2 - 2M_{3,\lambda} + 4n(n-1)\right)$$

$$r_\lambda((6)) = \frac{(n-6)!}{n!}\left(M_{6,\lambda} - (6n-37)M_{4,\lambda} - 3M_{2,\lambda}M_{3,\lambda} + 6(3n^2 - 19n + 20)M_{2,\lambda}\right).$$

A weak form of the conjectures stated in the previous section is proved by Roichman in [119] where interesting uniform bounds for the character ratios $r_\lambda(C)$ are also derived.

Theorem 9.4 ([119]). *Fix $\eta, \varepsilon \in (0,1)$. Then there are constants $a, A, N \in (0,\infty)$ such that for any $n \geq N$, any odd conjugacy class C with $|C| \leq (1-\eta)n$, we have*

$$2\|p^{(k)} - u\|_{\mathrm{TV}} \leq d_2(p_C^{(k)}, u) \leq \varepsilon \quad \textit{for all} \quad k \geq \frac{An}{|C|}\log n$$

whereas

$$\|p_C^{(k)} - u\|_{\mathrm{TV}} \geq \varepsilon \quad \textit{for all} \quad k \leq \frac{an}{|C|}\log n.$$

The same result holds on A_n for even conjugacy classes.

This theorem of Roichman proves the existence of a precut-off at time $(n/|C|)\log n$ for (S_n, p_C) when $|C| \leq (1-\eta)n$.

Large conjugacy classes. In his thesis [102], Lulov considers the walks driven by the uniform measure on the conjugacy classes $C_r = (n/r, \ldots, n/r)$, where r divides n. These are huge conjugacy classes. Consider the case where C_r is even and the walk is restricted to A_n. Obviously, $\widetilde{p}_{C_r}$ is not close to the uniform distribution on A_n. However, Lulov uses character ratios estimates to show that $\widetilde{p}_{C_r}^{(k)}$ is close to uniform on A_n for $k = 3$ if $r = 2$ and for $k = 2$ if $r \geq 3$. In [103] the authors conjecture that, for conjugacy classes with no fixed points, it always takes either 2 or 3 steps to reach approximate stationarity. They also prove the following Theorem by deriving sufficiently good character ratio estimates.

Theorem 9.5 ([103]). *Let C_n be an even conjugacy class in S_n with a single cycle, i.e., $C_n = (r_n)$ and assume that $|C_n| = r_n > n/2$ and $n - r_n$ tends to infinity. Then the sequence $(A_n, \widetilde{p}_{C_n})$ presents a cut-off at time*

$$t_n = \frac{\log n}{\log[n/(n-r_n)]}.$$

For the lower bound, [103] refers to [119]. The lower bound in [119] is based on Propositions 5.6 and 5.7. The proof in [119] needs to be adapted properly in order to prove the lower bound stated in Theorem 9.5.

The authors of [103] conjecture that the conclusion of Theorem 9.5 is valid for all sequences C_n of even conjugacy classes whose number of fixed points $n - |C_n|$ is $o(n)$ and tends to infinity.

Other walks related to random transposition. Imagine a deck of cards where each card, in addition to its face value, has an orientation (or spin), say up or down (think of the faces of the cards being up or down in the deck, or of the back of each card being marked by an arrow that can be up or down). A natural generalization of random transposition is as follows. Pick a pair of positions uniformly at random in the deck. Transpose the cards in these positions and, at the same time, uniformly pick an orientation for these cards. This is a random walk on the wreath product $\mathbb{Z}_2 \wr S_n = (\mathbb{Z}_2)^n \rtimes S_n$ where the action of S_n is by permutation of the coordinates in $\mathbb{Z}_2^n$. The above description generalizes straightforwardly to the case where $\mathbb{Z}_2$ is replace by an arbitrary finite group H. For instance, taking $H = S_m$, we can think of the corresponding walk as mixing up n decks of m cards. Here cards of different decks are never mixed together. What is mixed up is the relative order of the decks and the cards in each individual deck. Schoolfield [128, 129] studies such walks and some variants using character theory. He finds that $ae^{-c} \leq d_2(p^{(k)}, u) \leq Ae^{-c}$ if $k = \frac{1}{2}n\log(n\sqrt{|G|}) + c$, $c > 0$. Using a stopping time argument as in Theorem 4.6, he also proves a cut-off in total variation at tine $t_n = \frac{1}{2}\log n$. Hence, if G depends on n and $|G|$ grows fast enough with n then stationarity is reached at different times in total variation and in L^2. See also [58].

9.4 Finite Classical Groups

Together with the symmetric and alternating groups, one of the most natural families of finite groups is formed by the classical groups over finite fields. These are groups of matrices resembling the classical real compact Lie groups. Representation and character theory of these groups are an important domain of research from several viewpoints but what is known is much less complete than for the symmetric groups. Many of these groups contains some relatively small conjugacy classes (or union of conjugacy classes), resembling the class of all transpostions in S_n, which generates the whole group. This leads to interesting random walks that can, in principle, be studied by using Theorem 9.1, i.e., character theory. We describe below some of the known results in this direction.

Random transvection in $SL_n(\mathbb{F}_q)$. $SL_n(\mathbb{F}_q)$ is the group of $n \times n$ matrices with determinant 1 over the finite field $\mathbb{F}_q$ with q elements (hence $q = p^n$ for some prime p). By definition, a *transvection* is an element in $SL_n(\mathbb{F}_q)$ which is not the identity and fixes all the points of a hyperplane in $\mathbb{F}_q^n$, the n dimensional vector space over $\mathbb{F}_q$. The transvections generate $SL_n(\mathbb{F}_q)$ and form a conjugacy class when $n > 2$. Good examples of transvections are the elementary matrices $I + aE_{i,j}$, $a \in \mathbb{F}_q \setminus \{0\}$, $i \neq j$, where I is the $n \times n$ identity matrix, and the matrix $E_{i,j}$ has a unique non-zero entry equal to 1 in the (i, j)-th position. A general transvection has the form $I + uv^t$ where u, v are two arbitrary non-zero vectors in $\mathbb{F}_q^n$ with $u^t v = 0$ (an element u of $\mathbb{F}_q^n$ is

a column vector and u^t is its transpose). Moreover, $uv^t = u_0 v_0^t$ if and only if $u = au_0$, $v = a^{-1}v_0$ for some $a \in \mathbb{F}_q \setminus \{0\}$. Thus picking u, v independently and uniformly in $\mathbb{F}_q^n \setminus \{0\}$ gives a uniformly distributed transvection $I + u^t v$. We denote by p the uniform measure on the set of all transvections and call the corresponding random walk the random transvection walk. This walk is studied by Hildebrand in [86] who proves the following remarkable result.

Theorem 9.6 ([86]). *For the random transvection measure p on $SL_n(\mathbb{F}_q)$ defined above, there are two positive constants A, N such that, for all $q \geq 2$, $n \geq N$ and $k = n + m$ with $m = 1, 2, \ldots$, we have*

$$d_2(p^{(m)}, u) \leq A\ q^{-m}.$$

Moreover, for all q and all integers n, m with $k = n - m > 0$ and $m \geq 3$, we have

$$\|p^{(k)} - u\|_{\mathrm{TV}} \geq 1 - 4q^{1-m}.$$

The upper bound uses (9.2) and a formula for character ratios that Hildebrand obtains from results in McDonald's book [109]. The task is significantly harder than for random transposition on S_n. The lower bound follows from a relatively simple argument concerning the dimension of the space of fixed vectors by a product of m transvections. Hildebrand's results demonstrate that the random transvection walk presents a very sharp cut-off: for random transvection on $SL_n(\mathbb{F}_q)$, it takes at least $n-6$ steps to reduce the total variation distance from 1 to 0.9. After that, a fixed number of steps suffices to drop the variation distance to, say 0.1.

Small conjugacy classes on finite classical groups. In a remarkable work [67, 68, 69], David Gluck studies in a unified and uniform way a large class of random walks on the finite classical groups. The results that Gluck obtains are somewhat less precise than Hildebrand's Theorem 9.6 but they have the same flavor: for any random walk whose driving measure is central, that is, constant on conjugacy classes and supported on *small* conjugacy classes, convergence to the uniform distribution occurs after order k steps where k is the rank of the underlying finite classical group. For instance, $SL_n(\mathbb{F}_q)$ has rank $n-1$ and it follows from Gluck's results that the random transvection walk studied by Hildebrand reaches approximate stationarity after order n steps.

Technically, the results obtained by Gluck are by no means simple generalizations of the previous results of Diaconis–Shahshahani and Hildebrand. The exact character formulas used by both Diaconis–Shahshahani and Hildebrand do not seem to be available for the problems treated by Gluck. Even if they were, it would be an immense task to obtain Gluck's results through a case by case analysis. A massive amount of (very advanced) algebra is at work behind Gluck's approach. To avoid technicalities, we present below two specific examples that falls into Gluck's theory: random symplectic transvection and random unitary transvection. A friendly reference for basic facts and notation

concerning these examples is [76]. Let $\mathbb{F}_q$ be a finite field with q elements and consider the vector space $\mathbb{F}_q^n$. For simplicity, we assume that $n, q \geq 4$ and q odd.

Assume that $n = 2m$ and fix a non-degenerate alternating form B (the choice of the form is irrelevant). A *symplectic transformation* is any invertible linear transformations of F_q^n that preserve B and $Sp_n(\mathbb{F}_q) \subset SL_n(\mathbb{F}_q)$ is the group of all symplectic transformations. The group $Sp_n(\mathbb{F}_q)$ satisfies $Sp_n(\mathbb{F}_q)' = Sp_n(\mathbb{F}_q)$. It has order

$$|Sp_n(\mathbb{F}_q)| = q^{m^2} \prod_{i=1}^{m} (q^{2i} - 1), \quad n = 2m.$$

To define $SU_n(\mathbb{F}_q)$, assume that $\mathbb{F}_q$ admits an automorphism α such that $\alpha^2 = 1$ (this implies that $q = q_0^2$ for some prime power q_0). Fix a Hermitian form B (relative to α)). Again, because we work on finite fields, the precise choice of B is irrelevant. The special unitary group $SU_n(\mathbb{F}_q^n)$ is the group of all invertible linear transformations with determinant 1 which preserve the Hermitian form B. The group $SU_n(\mathbb{F}_q)$ satisfies $SU_n(\mathbb{F}_q)' = SU_n(\mathbb{F}_q)$. It has order

$$|SU_n(\mathbb{F}_q)| = q^{n(n-1)} \prod_{j=1}^{n} (q^{j/2} - (-1)^j).$$

A *symplectic transvection* (resp. unitary transvection) is a transvection that preserve the Hermitian (resp. unitary) form B. Symplectic (resp. unitary) transvections are exactly the linear transformations of the form

$$\tau_{u,a} : v \mapsto v + aB(v, u)u$$

where $u \in \mathbb{F}_q^n \setminus \{0\}$ is a non-zero vector and $a \in \mathbb{F}^*$ is a non-zero scalar (resp. $u \in \mathbb{F}_q^n \setminus \{0\}$, $B(u, u) = 0$, and $a \in \mathbb{F}^*$, $a = -\alpha(a)$). Both the symplectic groups and the special unitary groups are generated by transvections.

Note that $\tau_{u,a} = \tau_{u_0,a_0}$ if and only if there exists $b \in \mathbb{F}^*$ such that $u = bu_0$, $a = b^{-1}a_0$. Thus we can pick a symplectic (resp. unitary) transformation uniformly at random by picking uniformly at random $u \in \mathbb{F}_q \setminus \{0\}$ and $a \in \mathbb{F}^*$ (resp. $u \in \mathbb{F}_q \setminus \{0\}$ satisfying $B(u, u) = 0$ and $a \in \mathbb{F}^*$ satisfying $a = -\alpha(a)$).

For any symplectic (resp. unitary) transformation σ, and any symplectic (resp. unitary) transvection $\tau_{u,a}$, we have $\sigma\tau_{u,a}\sigma^{-1} = \tau_{\sigma(u),a}$. This shows that the set T of all symplectic (resp. unitary) transvections is a union of conjugacy classes (it is not, in general, a single conjugacy class). Gluck's results in [68, Th. 42 and Cor. 64] specialize to the present examples as follows.

Theorem 9.7 ([68]). *Let p denote the uniform measure on symplectic or unitary transvections in $Sp_n(\mathbb{F}_q)$ or in $SU_n(\mathbb{F}_q)$, respectively. Assume that q is odd and n is large enough. Then there exists N such that for $k = N(n + c)$ with $c > 0$, we have*

$$d_2(p^{(k)}, u) \leq q^{-n/4-2c}.$$

One of the typical character ratio estimates obtained by Gluck [67] says that there exist $a \in (0,1)$ and $M > 0$ such that for every finite simple group of Lie type G_q over the finite field with q elements, for every non-central element $g \in G_q$, and for every irreducible character χ of $G(q)$,

$$|\chi(g)/\chi(e)| \le \min\{a, Mq^{-1/2}\}.$$

This is not enough to prove Theorem 9.7 for which the refinements obtained in [68] are needed but, as noted in [99], it gives the following result.

Theorem 9.8. *Let G_{q_n} be a family of finite groups of Lie type of order growing to infinity. Let C_n be a non-central conjugacy class in G_{q_n} and $\Sigma_n = C_n \cup C_n^{-1}$. Then the Cayley graphs (G_{q_n}, Σ_n) form a family of expanders.*

9.5 Fourier Analysis for Non-central Measures

The extent to which Fourier analysis fails to provide useful results for random walks that are not bi-invariant (i.e., driven by non-central measures) is somewhat surprising. Still, there are cases in which the analysis of Sections 9.1 and 9.2 can be extended but few have been worked out in detail. A typical example is the *transpose top and random* shuffle. On S_n, consider the measure

$$p_\star(\tau) \begin{cases} 1/n & \text{if } \tau = (1,i), \quad i = 1,\dots,n \\ 0 & \text{otherwise,} \end{cases} \tag{9.3}$$

where $(1,1)$ is the identity and $(1,i)$, $i \neq 1$, is transpose 1 and i. This measure is not central (see (9.1)) but it is invariant by $\tau \mapsto \theta\tau\theta^{-1}$, $\theta \in S_{n-1}$ where S_{n-1} is understood as the subgroup of S_n of those permutations that fix 1. Because of this property, for any irreducible representation ϱ of S_n, the matrix $\hat{p}_\star(\varrho)$ has a relatively small number of distinct eigenvalues and manageable formulas for the eigenvalues and their multiplicity can be obtained. See [27, 28, 59]. Using this spectral information and (5.8) gives the upper bound in the following theorem. The lower bound can be obtained by adapting the argument used for random transposition in [27, p.43].

Theorem 9.9. *For transpose top and random, i.e., the walk on S_n driven by $p_\star$, there exists a constant A such that, for all n and $c > 0$ for which $k = n(\log n + c)$ is an integer, we have*

$$2\|p_\star^{(k)} - u\|_{\text{TV}} \le d_2(p_\star^{(k)}, u) \le A\mathrm{e}^{-c}.$$

Moreover, there are two functions f_1, f_2 with limit 0 at ∞ such that for all n and all $c > 0$ for which $k = n(\log n - c)$ is an integer,

$$\|p_\star^{(k)} - u\|_{\text{TV}} \ge 1 - f_1(c) - f_2(n).$$

10 Comparison Techniques

The path technique used in Section 6 to bound the spectral gap generalizes in a very useful way to yield comparison inequalities between the Dirichlet form of different random walks. Such inequalities are important because they lead to a full comparison of the higher part of the spectrum of the two walks as sated in the next result.

10.1 The min-max Characterization of Eigenvalues

Dirichlet form comparison leads to spectrum comparison by a simple application of the Courant–Fisher min-max characterization of the ordered eigenvalues $q_0 \le q_1 \le \dots$ of a self-adjoint linear operator Q on a Hilbert space $(V, \langle \cdot, \cdot \rangle)$ (here, finite dimensional and real). See, e.g., [90, 4.2.11].

Theorem 10.1 ([42]). *Let $p, \widetilde{p}$ be two symmetric probability measures on a finite group G with respective Dirichlet forms $\mathcal{E}, \widetilde{\mathcal{E}}$ and respective eigenvalues, in non-increasing order $\beta_i, \widetilde{\beta}_i$. Assume that there is a constant A such that $\widetilde{\mathcal{E}} \le A\mathcal{E}$. Then, for all $i = 0, 1, \dots, |G|-1$, $\beta_i \le 1 - A^{-1}\left(1 - \widetilde{\beta}_i\right)$. In particular, for the continuous-time random walks associated to p and $\widetilde{p}$ as in* (2.10)*, we have*

$$d_2(H_t, u) \le d_2(\widetilde{H}_{t/A}, u). \tag{10.1}$$

The inequality $\widetilde{\mathcal{E}} \le A\mathcal{E}$ does not provide good control on the small positive eigenvalues and the negative eigenvalues of p. Thus there is no clean statement in discrete time analogous to (10.1). However, there are various ways to cope with this difficulty. Often, negative and small positive eigenvalues do not play a crucial role in bounding $d_2(p^{(k)}, u)$. In particular, (10.1) and Theorem 5.1 give the following useful result.

Theorem 10.2 ([42]). *Referring to the notation of Theorem* 10.1*, assume that there is a constant $A > 0$ such that $\widetilde{\mathcal{E}} \le A\mathcal{E}$. Then*

$$d_2(p^{(k)}, u)^2 \le \beta_-^{2k_1}(1 + d_2(\widetilde{H}_{k_2/A}, u)^2) + d_2(\widetilde{H}_{k/A}, u)^2$$

and

$$\begin{aligned} d_2(p^{(k)}, u)^2 \le \beta_-^{2k_1}(1 + |G|\mathrm{e}^{-k_2/2A} + d_2(\widetilde{p}^{(\lfloor k_2/2A \rfloor)}, u)^2) \\ + |G|\mathrm{e}^{-k/2A} + d_2(\widetilde{p}^{(\lfloor k/2A \rfloor)}, u)^2 \end{aligned}$$

where $k = k_1 + k_2 + 1$ and $\beta_- = \max\{0, -\beta_{|G|-1}\}$.

For best results, one should use the first inequality stated in this theorem since an extra factor of 2 is lost in bounding $d_2(\widetilde{H}_t, u)$ in terms of $d_2(\widetilde{p}^{(k)}, u)$. To use Theorems 10.1, 10.2, one needs a measure $\widetilde{p}$ that can be analyzed in terms of the L^2-distance d_2. A general scheme that has proved very successful is to start with a central measure $\widetilde{p}$ for which representation theory can be used as in Theorem 9.1. Then Theorems 10.1, 10.2 can be used to obtain results for other walks.

10.2 Comparing Dirichlet Forms Using Paths

We now present some comparison inequalities between Dirichlet forms taken mostly from [42, 49]. The proofs are similar to the proof of Theorem 6.4 given in Section 6.2. Fix two probability measures p and $\widetilde{p}$ on G. Think of p as driving the unknown walk we wish to study whereas we already have some information on the walk driven by $\widetilde{p}$. Fix a symmetric generating set Σ contained in the support of p. We will use the notation introduced in Section 6. Given a subset T of G, pick a path γ_x from e to x in the Cayley graph (G, Σ) and set $\mathcal{P}_*(T) = \{\gamma_x : x \in T\}$.

Theorem 10.3 ([42, 45, 49]). *Let T denote the support of $\widetilde{p}$. Referring to the setting and notation introduced above, we have $\widetilde{\mathcal{E}} \le A_* \mathcal{E}$ where*

$$A_* = \max_{s \in \Sigma} \left\{ \frac{1}{p(s)} \sum_{\gamma \in \mathcal{P}_*(T)} |\gamma| N(s, \gamma) \widetilde{p}(\gamma) \right\}$$

with $\widetilde{p}(\gamma) = \widetilde{p}(x)$ if $\gamma = \gamma_x \in \mathcal{P}_*(T)$.

The following result concerns the walks based on fixed subsets of transpositions and is obtained by comparison with random transposition [42]. Let $\mathbf{G} = (V, E)$ be a graph with vertex set $V = \{1, \dots, n\}$ and symmetric edge set $E \subset V \times V$ containing no loops ($(i, i) \notin E$ and $(i, j) \in E$ if and only if $(j, i) \in E$). Consider the walk on the symmetric group driven by the measure

$$p_{\mathbf{G}}(\tau) = \begin{cases} 1/n & \text{if } \tau = e \\ 2(n-1)/|E|n & \text{if } \tau = (i, j) \text{ with } (i, j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

Thus this walk is based on those transpositions which corresponds to neighbors in $\mathbf{G}$. It is irreducible if and only if the graph is connected. If $\mathbf{G}$ is the complete graph then $p_{\mathbf{G}} = p_{\mathrm{RT}}$ is the random transposition measure defined at (4.1). If $\mathbf{G}$ is the line graph $1 - 2 - \cdots - n$ then $p_{\mathbf{G}} = p_{\mathrm{AT}}$ is the adjacent transposition measure. If $\mathbf{G}$ is the star graph with center 1 then $p_{\mathbf{G}} = p_\star$ is the transpose top and random measure defined at (9.3). These walks were introduced in [42]. They are also considered in [80]. To state a general result, for each $x, y \in V$, pick paths $\mu_{x,y}$ from x to y in $\mathbf{G}$ of length (i.e number of edges) $|\mu_{x,y}|$ and set

$$\Delta = \max_{e \in E} \sum_{\substack{(x,y) \in V \times V \\ e \in \mu_{x,y}}} |\mu_{x,y}|.$$

The quantity Δ depends on both the length of the paths and the number of bottlenecks in the family $\{\mu_{x,y} : x, y \in V\}$ (see, e.g., [51, 57, 42, 43]).

Theorem 10.4 ([42]). *Referring to the notation introduced above, there exists a constant A such that for $k > (4(n-1)^{-1}|E|\Delta + n)(\log n + c)$, $c > 0$, we have*

$$2\|p_{\mathbf{G}}^{(k)} - u\|_{\mathrm{TV}} \le d_2(p_{\mathbf{G}}^{(k)}, u) \le A\mathrm{e}^{-c}.$$

For the star graph and the line graph this theorem gives upper bounds on $T(S_n, p_\star)$, $T(S_n, p_{\text{AT}})$ that are of order $n \log n$ and $n^3 \log n$ respectively. Both capture the right order of magnitude. If $\mathbf{G}$ is a two dimensional finite square grid with side size $\sqrt{n}$, the theorem gives $T(S_n, p_{\mathbf{G}}) \leq Cn^2 \log n$. A matching lower bound is proved in [141]. The bound of Theorem 10.4 is probably not sharp in general. For instance, assume $n = 2^d$ and let $\mathbf{G}$ be the hypercube. In this case, Theorem 10.4 gives $T(S_n, p_{\mathbf{G}}) \leq Cn(\log n)^3$. Wilson [141] proves $T(S_n, p_{\mathbf{G}}) \geq cn(\log n)^2$ which is probably sharp.

An interesting example is obtained for $E = \{(i, j) : |i - j| \leq \ell\}$ with $1 \leq \ell \leq n$. We call the associated walk the *ℓ-adjacent transposition* walk and denote by $p_{\ell\text{-AT}}$ the corresponding measure. For $\ell = 1$, this is the adjacent transposition walk. For $\ell = n$, we get random transposition. Durrett [55] uses Theorem 10.4 and Theorem 5.8 to show that there are constants $C, c > 0$ such that $c(n^3/\ell^2) \log n \leq T(S_n, p_{\ell\text{-AT}}) \leq Cn^3/\ell^2) \log n$ (in fact, the walk considered in [55] is slightly different but the same analysis applies).

Next we describe other examples where comparison with random transposition gives good results.

- The crude overhand shuffle and the Borel–Chéron shuffle of Section 3.1. In both cases, comparing with random transposition, the constant A_* in Theorem 10.3 stays bounded, uniformly in n. This shows that order $n \log n$ such shuffles suffice to mix up n cards. Details and matching lower bounds can be found in [42].
- Random insertions. For $i < j$, the insertion $c_{i,j}$ is the cycle $(j, j-1, \ldots, j-i+1, i)$ and $c_{j,i} = c_{i,j}^{-1}$. The random insertion measure p_{RI} is given by $p_{\text{RI}}(e) = 1/n$, $p(c_{i,j}) = 1/n^2$ for $i \neq j$. The mixing time $T(S_n, p_{\text{RI}})$ is of order $n \log n$. See [42, 45] where other insertion walks are also considered.
- Random reversal. A reversal is a transposition that takes a packet and puts it back in reverse order. Thus for $i < j$, $r_{i,j} = (i, j)(i-1, j-1) \ldots (\lfloor (j-i)/2 \rfloor)(\lceil (j-i)/2 \rceil)$ is the reversal corresponding to the i to j packet. The random reversal measure is p_{RR} given by $p_{\text{RR}}(e) = 1/n$, $p_{\text{RR}}(r_{i,j}) = 2/n^2$. The ℓ-reversal measure $p_{\ell\text{-RR}}$ has $p_{\ell\text{-RR}}(e) = 1/n$ and $p_{\ell\text{-RR}}(r_{i,j}) = 1/\ell(n - \ell/2 - 1)$ if $i < j$ with $j - i \leq \ell$. Durrett [55] shows that there exists $C, c > 0$ such that $c(n^3/\ell^3) \log n \leq T(S_n, p_{\ell\text{-RR}}) \leq C(n^3/\ell^2) \log n$. The upper bound is by comparison with random transposition. The lower bound uses Theorem 5.8. The walk "reverse top to random" is studied in [42]. It has a precut-off at time $n \log n$.
- A slow shuffle. Let p be uniformly supported on $\Sigma = \{e, \tau, c, c^{-1}\}$ where τ is the transposition $(1, 2)$ and c is the long cycle $c = (1, 2, \ldots, n)$. It is easy to write any transposition using τ, c, c^{-1}. In this case the constant A_* is of order n^2 and this proves that there is a constant C such that $T(S_n, p) \leq Cn^3 \log n$, see [42]. A matching lower bound is proved in [142]. Hence this walk has a precut-off at time $n^3 \log n$.
- A fast shuffle. This example is taken from [10] and [42]. For any even integer n, let S_n act by permutation on the n-set $\mathbb{Z}_{n-1} \cup \{\infty\}$. Let π_i :

$x \mapsto 2x + i, \mod n-1$, $i = 0, 1$, and $\pi_2 = (0, \infty)$, i.e., transpose 0 and ∞. Let p be the uniform probability on $\Sigma = \{e, \pi_0^{\pm 1}, \pi_1^{\pm 1}, \pi_2\}$. The diameter of (S_n, Σ) is of order $n \log n$ (by an obvious counting argument, this is optimal for a bounded number of generators). Moreover, comparison with random transposition gives $T(S_n, p) \le Cn(\log n)^3$, see [42]. It is an open problem to find a bounded number of generators in S_n such that the mixing time of the associated walk is of order $n \log n$.

We now give a slightly more sophisticated version of Theorem 10.3 using the notion of $\widetilde{p}$-flow. Let $\mathcal{P}_e, \mathcal{P}_{e,x}$ be as defined in Section 6.2. A $\widetilde{p}$-flow is a non-negative function Φ on $\mathcal{P}_e$ such that

$$\sum_{\gamma \in \mathcal{P}_{e,x}} \Phi(\gamma) = \widetilde{p}(x).$$

Theorem 10.5 ([45]). *Referring to the setting and notation introduced above, let Φ be $\widetilde{p}$-flow. Then $\widetilde{\mathcal{E}} \le A(\Phi)\mathcal{E}$ where*

$$A(\phi) = \max_{s \in \Sigma} \left\{ \frac{1}{p(s)} \sum_{\gamma \in \mathcal{P}} |\gamma| N(s, \gamma) \Phi(\gamma) \right\}.$$

As a corollary, we obtain the following result.

Theorem 10.6. *Assume that there is a subgroup H of the automorphism group of G which is transitive on Σ and such that $\widetilde{p}(hx) = \widetilde{p}(x)$ for all $x \in G$ and $h \in H$. Set $\varepsilon = \min\{p(s) : s \in \Sigma\}$. Then $\widetilde{\mathcal{E}} \le A\mathcal{E}$ where*

$$A = \frac{1}{\varepsilon \# \Sigma} \sum_{x \in G} |x|^2 \widetilde{p}(x).$$

Proof. Consider the set $\mathcal{G}_{e,x}$ of all geodesic paths from e to x in (G, Σ) and set

$$\Phi(\gamma) = \begin{cases} (\#\mathcal{G}_{e,x})^{-1} \widetilde{p}(x) & \text{if } \gamma \in \mathcal{G}_{e,x} \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that this defines a $\widetilde{p}$-flow. Moreover, since each $\gamma \in \mathcal{G}_{e,x}$ has length $|\gamma| = |x|$, the constant $A(\phi)$ of Theorem 10.5 is bounded by

$$\begin{aligned} A(\Phi) &= \max_{s \in \Sigma} \left\{ \frac{1}{p(s)} \sum_{x \in G} |x| \sum_{\gamma \in \mathcal{G}_{e,x}} N(s, \gamma) \frac{\widetilde{p}(x)}{\#\mathcal{G}_{e,x}} \right\} \\ &\le \varepsilon^{-1} \max_{s \in \Sigma} \left\{ \sum_{x \in G} |x| \sum_{\gamma \in \mathcal{G}_{e,x}} N(s, \gamma) \frac{\widetilde{p}(x)}{\#\mathcal{G}_{e,x}} \right\}. \end{aligned}$$

By assumption, the quantity inside the parentheses is independent of s. Averaging over $s \in \Sigma$ yields the desired bound. □

As an application of Theorem 10.6, we state the following result for which the construction of the paths is rather involved. See [49] and the references cited therein. On $SL_n(\mathbb{Z}_m)$, m prime, let p be the uniform measure on the the set $\Sigma = \{E_{i,j} : 0 \le i, j \le n\}$ where $E_{i,j}$ denotes the elementary matrix with 1's along the diagonal, a 1 in position (i, j) and 0's elsewhere. Let $\widetilde{p}$ be the random transvection measure of Theorem 9.6.

Theorem 10.7 ([49]). *Referring to the notation introduced above, there exists a constant C such that, for any integer n and prime number m,*

$$\widetilde{\mathcal{E}} \le C[n \log m]^2 \mathcal{E}.$$

In particular, the second largest eigenvalue β_1 of p is bounded by

$$\beta_1 \le 1 - \frac{1}{2C[n \log m]^2}$$

for all integers n, m large enough, m prime.

10.3 Comparison for Non-symmetric Walks

This section applies Dirichlet form comparison and Theorem 5.4 to study non-symmetric examples.

Let us start with two examples on the symmetric group S_n. Let $\tau = (1, 2)$, $c = (1, 2, \ldots, n)$, $c' = (1, 2, \ldots, n-1)$ and consider the probabilities p_1, p_2 defined by

$$p_1(\tau) = p_1(c) = 1/2, \quad p_2(c) = p_2(c') = 1/2.$$

These are essentially the probabilities corresponding to the slow shuffles discussed at the end of Section 4.1.

As the walk driven by p_1 is periodic if n is even, we assume that n is odd. It is easy to see (see [45]) that the second largest singular value $\sigma_1(1) = \sigma_1$ of p_1 is 1 but that the support of $q = p_1^{(2)} * \check{p}_1^{(2)}$ generates S_n so that $\sigma_1(2) < 1$. Comparison between q and random transposition, together with Theorem 5.4, gives $T(S_n, p_1) \le Cn^3 \log n$. A matching lower bounds is given in [142].

Surprisingly, this argument does not work for the walk driven by p_2. Indeed, the support of $p_2^{(j)} * \check{p}_2^{(j)}$ does not generate S_n unless $j \ge n$ and it is not clear how to study the walk driven by $p_2^{(n)} * \check{p}_2^{(n)}$ using comparison. See [45]. A coupling argument gives $T(S_n, p_2) \le Cn^3 \log n$, [85]. A matching lower bounds is given in [142].

The next result shows that non-symmetric walks with significant holding probability can always be controlled by additive symmetrization.

Theorem 10.8. *Let p be a probability measure on a finite group G. let $q_+ = \frac{1}{2}(p + \check{p})$ be the additive symmetrization of p and assume that $p(e) = \varepsilon > 0$. Then*

$$d^2(p^{(2k)}, u)^2 \le d_2(Q^+_{\varepsilon k}, u)^2 \le |G| \mathrm{e}^{-\varepsilon k} + d_2(q_+^{(\lfloor \varepsilon k/2 \rfloor)}, u)^2.$$

Proof. By assumption $q = p * \check{p} \geq \varepsilon q_+$ leading to an immediate comparison of the associated Dirichlet forms. For the continuous-time probabilities Q_t, Q_t^+ associated respectively to q, q_+ by (2.10), Theorem 10.1 gives

$$d_2(Q_t, u) \leq d_2(Q_{\varepsilon t}^+, u).$$

As q has non-negative eigenvalues, Theorem 5.1 gives $d_2(q^{(k)}, u) \leq d_2(Q_k, u)$. Also, by Theorem 5.4, we have $d_2(p^{(2k)}, u) \leq d_2(q^{(k)}, u)$. Hence,

$$d_2(p^{(2k)}, u) \leq d_2(Q_{\varepsilon k}^+, u).$$

Using Theorem 5.1 again finishes the proof. □

As a typical application, we consider the Frattini walks on p-groups of Section 7.2.

Theorem 10.9. *Fix an integer c. Then there are positive constants $a_i = a_i(c)$, $i = 1, 2$, such that for any p-group G of nilpotency class and Frattini rank at most c, for any minimal set F of generators of G, we have*

$$\|q_F^{(k)} - u\|_{TV} \leq a_3 \mathrm{e}^{-a_4 k / p^{2\omega}}$$

where q_F denotes the uniform probability measure on $\{e\} \cup F$ and p^ω is the exponent of $G/[G, G]$.

Proof. Use Theorem 10.8 and Theorem 7.10. □

References

1. Aldous, D. (1983): Random walks on finite groups and rapidly mixing Markov chains. In Séminaire de Probabilités, XVII, Lec. Notes in Math. **986**, Springer, Berlin.
2. Aldous, D. (1987): On the Markov-chain simulation method for uniform combinatorial simulation and simulated annealing. Prob. Eng. Info. Sci. **1**, 33–46.
3. Aldous, D., Fill, J.A. (1995) Preliminary version of a book on finite Markov chains. http://www.stat.berkeley.edu/users/aldous
4. Aldous, D., Diaconis, P. (1986): Shuffling cards and stopping times. Amer. Math. Monthly **93**, 333–348
5. Aldous, D., Diaconis, P. (1987): Strong uniform times and finite random walks. Adv. Appl. Math. **8**, 69–97.
6. Alon, N., Roichman, Y. (1994): Random Cayley graphs and expanders. Random Struct. and Alg. **5**, 271–284.
7. Astashkevich, A., Pak, I. (2001): Random walks on nilpotent groups. Preprint.
8. Babai, L. (1995): Automorphism groups, isomorphism, reconstruction. Handbook of combinatorics, Vol. 1, 2, 1447–1540, Elsevier.
9. Babai, L., Szegedy, M. (1992): Local expansion of symmetrical graphs. Combin. Probab. Comput. **1**, 1–11.

10. Babai, L., Hetyii, G., Kantor, W., Lubotzky, A., Seress, A. (1990): On the diameter of finite groups. 31 IEEE Symp. on Found. of Comp. Sci. (FOCS 1990) 857–865.
11. Babai, L., Kantor, W., Lubotzky, A. (1992): Small diameter Cayley graphs for finite simple groups. European J. Comb. **10**, 507–522.
12. Bacher, R. (1994): Valeur propre minimale du laplacien de Coxeter pour le groupe symétrique. J. Algebra **167**, 460–472.
13. Bayer, D., Diaconis, P. (1986): Trailing the dovetail shuffle to its lair. Ann. Appl. Probab. **2**, 294–313.
14. Billera, L., Brown, K., Diaconis, P. (1999): Random walks and plane arrangements in three dimensions. Amer. Math. Monthly 106, 502–524.
15. Borel, E., Chéron, A. (1940): Théorie Mathématique du Bridge à la Portée de Tous, Gauthier-Villars, Paris.
16. Brown, K. (2000): Semigroups, rings, and Markov chains. J. Theoret. Probab. **13**, 871–938.
17. Brown, K., Diaconis, P. (1998): Random walks and hyperplane arrangements. Ann. Probab. **26**, 1813–1854.
18. Burdzy, K., Kendall, W. (2000): Efficient Markovian couplings: examples and counterexamples. Ann. Appl. Probab. **10**, 362–409.
19. Cartier, P., Foata, D. (1969): Problèmes Combinatoires de Commutation et Réarrangements. Lec. Notes. Math. **85**, Springer.
20. Chavel, I. (1984): Eigenvalues in Riemannian Geometry. Academic Press.
21. Coppersmith, D., Pak, I. (2000): Random walk on upper triangular matrices mixes rapidly. Probab. Theory Related Fields **117**, 407–417.
22. Chung, F., Faber, V., Manteuffel, T. (1994): An upper bound on the diameter of a graph from eigenvalues associated with its Laplacian. SIAM J. Discrete Math. **7**, 443–457.
23. Dai, J. (1998): Some results concerning random walk on finite groups. Statist. Probab. Lett. **37**, 15–17.
24. Dai, J., Hildebrand, M. (1997): Random random walks on the integers mod n. Statist. Probab. Lett. **35**, 371–379.
25. Davidoff, G., Sarnak, P. (2003): Elementary Number Theory, Group Theory and Ramanujan Graphs. Cambridge University Press.
26. Diaconis, P. (1982): Applications of non-commutative Fourier analysis to probability problems. Lec. Notes in Math. **1362**, 51–100, Springer.
27. Diaconis, P. (1988): Group representations in probability and statistics. Institute of Mathematical Statistics Lecture Notes-Monograph Series, **11**. Hayward, CA.
28. Diaconis, P. (1991): Finite Fourier methods: Access to tools. Proc. Symp. Appl. Math. **44**, 171–194.
29. Diaconis, P. (1998): From shuffling cards to walking around the building: an introduction to modern Markov chain theory. Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998). Doc. Math., 187–204.
30. Diaconis, P. (2000): The cut-off phenomenon in finite Markov chains. Proc. Natl. Acad. Sci. USA **93**, 1659–1664.
31. Diaconis, P. (2003): Random walks on groups: characters and geometry. Groups St. Andrews, Neuman, P. et al (eds).
32. Diaconis, P. (2003): Mathematical developments from the analysis of riffle shuffling. In: M. Liebeck (ed), Proc. Durham conference on groups.

33. Diaconis, P., Fill, J.A. (1990): Srong stationary times via a new form of duality. Ann. Probab. **18**, 1483–1522.
34. Diaconis, P., Fill, J.A., Pitman, J. (1992): Analysis of top to random shuffles. Combin. Probab. Comput. **1**, 135–155.
35. Diaconis, P., Graham, R., Morrison, J. (1990): Asymptotic analysis of a random walk on a hypercube with many dimensions. Random Struct. and Alg. **1**, 51–72.
36. Diaconis, P., Hanlon, P. (1992): Eigen-analysis for some examples of the Metropolis algorithm. Contemp. Math. **138**, 99–117.
37. Diaconis, P., Holmes, S. (2001): Analysis of a card mixing scheme, unpublished report.
38. Diaconis, P., Holmes, S. (2002): Random walks on trees and matchings. Electron. J. Probab. **7**, 17 pp. (electronic).
39. Diaconis, P., Holmes, S., Neals, B. (2000): Analysis of a nonreversible Markov chain sampler. Ann. Appl. Probab. **10**, 726–752.
40. Diaconis, P., McGrath, M., Pitman, J. (1995): Riffle shuffles, cycles, and descents. Combinatorica **15**, 11–29.
41. Diaconis, P., Ram, A. (2000): Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. Mich. Math. jour. **48**, 157–190.
42. Diaconis, P., Saloff-Coste, L. (1993): Comparison techniques for random walk on finite groups. Ann. Probab. **21**, 2131–2156.
43. Diaconis, P., Saloff-Coste, L. (1993): Comparison techniques for reversible Markov chains. Ann. Probab. **3**, 696–730.
44. Diaconis, P., Saloff-Coste, L. (1994): Moderate growth and random walk on finite groups. GAFA, **4**, 1–36.
45. Diaconis, P., Saloff-Coste, L. (1995): Random walks on finite groups: a survey of analytic techniques. In Probability measures on groups and related structures XI (Oberwolfach, 1994), 44–75. World Scientific.
46. Diaconis, P., Saloff-Coste, L. (1995): An application of Harnack inequalities to random walk on nilpotent quotients. J. Fourier Anal. Appl. Proceedings of the Conference in Honor of J.P. Kahane. 190–207.
47. Diaconis, P., Saloff-Coste, L. (1996): Nash inequalities for finite Markov chains. J. Theoret. Probab. **9**, 459–510.
48. Diaconis, P., Saloff-Coste, L. (1996): Logarithmic Sobolev inequalities for finite Markov chains. Ann. Appl. Probab. **6**, 695–750.
49. Diaconis, P., Saloff-Coste, L. (1996): Walks on generating sets of abelian groups. Probab. Theory Related Fields **105**, 393–421.
50. Diaconis, P., Shahshahani, M. (1981): Generating a random permutation with random transpositions. Z. Wahrsch. Verw. Geb. **57**, 159–179.
51. Diaconis, P., Stroock, D. (1991): Geometric bounds for eigenvalues of Markov chains. Ann. Appl. Probab. **1**, 36–61.
52. Dixon, J. (1969): The probability of generating the symmetric group. Math. Z. **110**, 199–205.
53. Dou C. (1992): Studies of random walks on groups and random graphs. Ph.D. Dissertation, Dept. of Math., Massachusetts Institute of Technology.
54. Dou, C., Hildebrand, M. (1996): Enumeration and random walks on finite groups. Ann. Probab. **24** 987–1000.
55. Durrett, R. (2003): Shuffling Chromosomes. J. Theoret. Probab. (to appear)
56. Ellenberg, J. (1993) A sharp diameter bound for upper triangular matrices. Senior honors thesis, Dept. Math. Harvard University.

57. Fill, J.A. (1991): Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains with an application to the exclusion processes. Ann. Appl. Probab. **1**, 62–87.
58. Fill, J.A., Schoolfield, C. (2001): Mixing times for Markov chains on wreath products and related homogeneous spaces. Electron. J. Probab. **6**, 22p.
59. Flatto, L., Odlyzko, A., Wales, D. (1985): Random shuffles and group representations. Ann. Probab. **13**, 151–178.
60. Fulman, J. (2000): Semisimple orbits of Lie algebra and card shuffling measures on Coxeter groups, J. Algebra **224**, 151–165.
61. Fulman, J. (2000): Application of the Brauer complex: card shuffling, permutation statistics, and dynamical systems, J. Algebra **243**, 96–122.
62. Fulman, J. Wilmer, E. (1999): Comparing eigenvalue bounds for Markov chains: when does Poincaré beat Cheeger. Ann. Appl. Probab. **9**, 1–13.
63. Gamburd, A. (2002): On the spectral gap for infinite index "congruence" subgroups of $SL_2(\mathbf{Z})$. Israel J. Math. **127**, 157–2000
64. Gamburd, A. (2003): Expander graphs, random matrices and quantum chaos. In: Kaimanovich, V. et al eds., Random walks and Geometry (Vienna, 2001), de Gruyter.
65. Gamburd, A., Pak, I. (2001): Expansion of product replacement graphs. Preprint.
66. Gilbert, E. (1955): Theory of Shuffling. Technical Memorandum, Bell Laboratories.
67. Gluck, D. (1995): Sharper character value estimates for groups of Lie type. J. Algebra **174**, 229–266.
68. Gluck, D. (1997): Characters and random walks on finite classical groups. Adv. Math. **129**, 46–72.
69. Gluck, D. (1999): First hitting time for some random walks on finite groups. J. Theoret. Probab. **12**, 739–755.
70. Good, I. (1951): Random motion on a finite Abelian group. Proc. CambridgePhil. Soc. **47**, 756–762.
71. Greenberg, Y. (1995): Ph.D. Thesis, Hebrew University, Jerusalem.
72. Greenhalgh, A. (1987): Random walks on groups with subgroup invariance properties. Ph.D. Thesis, Dept. of Math., Stanford University.
73. Greenhalgh, A (1997). A model for random random-walks on finite groups. Combin. Probab. Comput. **6**, 49–56.
74. Grigorchuck, R., Żuk, A. (1999): On the asymptotic spectrum of random walks on infinite families of graphs. In: Picardello and Woess, eds., Random walks and discrete potential theory (Cortona, 1997), 188–204, Sympos. Math., XXXIX, Cambridge Univ. Press
75. Gromov, M. (1981): Groups of polynomial growth and expanding maps. Publ. Math. I.H.E.S. **53**, 53–81.
76. Grove, L. (2001): Classical Groups and Geometric Algebra. Graduate Studies in Mathematics **39**, American Math. Soc.
77. Häggström, O., Jonasson, J. (1997): Rates of convergence for lamplighter processes. Stochastic Process. Appl. **67**, 227–249.
78. Hall, M. (1976): The theory of groups, sec. ed., Chelsea, New York.
79. Hall, P. (1957): Nilpotent groups. In Collected Works of Philip Hall, Oxford University press, 417–462.
80. Handjani, S., Jungreis, D. (1996): Rate of convergence for shuffling cards by transpositions. J. Theoret. Probab. **9**, 983–993.

81. Hannan, E.J. (1965) Group representation and applied probability. J. Appl. Probab. **2** 1–68.
82. de la Harpe, P. (2000): Topics in Geometric Group Theory. Chicago Lectures in Mathematics, Chicago University Press.
83. de la Harpe, P., Valette, A. (1989): La propriété (T) de Kazhdan pour les groupes localement compacts. Astérisque **175**, SMF.
84. Harper, L. (2003) Global Methods for Combinatorial Isoperimetric Problems, monograph to be published by Cambridge University Press.
85. Hildebrand, M. (1990): Rates of convergence of some random processes on finite groups. Ph. D thesis, Department of Mathematics, Harvard University.
86. Hildebrand, M. (1992): Generating random elements in $SL_n(F_q)$ by random transvections. J. Alg. Combinatorics **1**, 133–150.
87. Hildebrand, M. (1994): Random walks supported on random points of $\mathbb{Z}/n\mathbb{Z}$. Probab. Theory Related Fields **100**, 191–203.
88. Hildebrand, M. (2001): Random lazy random walks on arbitrary finite groups. J. Theoret. probab. **14**, 1019–1034.
89. Hildebrand, M. (2002): A note on various holding probabilities for random lazy random walks on finite groups. Statist. Probab. Lett. **56**, 199–206.
90. Horn, R., Johnson, C. (1985): Matrix analysis. Cambridge University Press.
91. Horn, R., Johnson, C. (1991): Topics in matrix analysis. Cambridge University Press.
92. Hostinsky, M. (1931): Méthodes générales du calcul des probabilités. Gauthier-Villars, Paris.
93. Ingram, R.E. (1950): Some characters of the symmetric group. Proc. Amer. Math. Soc. **1**, 358–369.
94. Jerrum, M. (1998): Mathematical foundations of the Markov chain Monte Carlo method. In Probabilistic methods for algorithmic discrete mathematics Algorithms Combin. **16**, 116–165.
95. Kosambi, D., Rao, U.V.R. (1958) The efficiency of randomization by card shuffling. J. R. Statist. Soc. A **128**, 223–233.
96. Leader, I. (1991): Discrete isoperimetric inequalities. In Probabilistic combinatorics and its applications (San Francisco, CA, 1991). Proc. Sympos. Appl. Math. **44**, 57–80. Amer. Math. Soc.
97. Liebeck, M., Shalev, A. (2001): Diameters of finite simple groups: sharp bounds and applications. Ann. of Math. **154**, 383–406.
98. Lubotzky, A. (1994): Discrete Groups, expanding graphs and invariant measures. Birkhäuser.
99. Lubotzky, A. (1995): Cayley graphs: Eigenvalues, Expanders and Random Walks. Surveys in combinatorics, 155–189, London Math. Soc. Lecture Note Ser., **218**, Cambridge Univ. Press.
100. Lubotzky, A., Pak, I. (2000): The product replacement algorithm and Kazhdan's property (T). J. Amer. Math. Soc. **14**, 347–363.
101. Lubotzky, A., Phillips, R., Sarnak, P. (1988): Ramanujan graphs. Combinatorica, **8**, 261–277.
102. Lulov, N. (1996): Random walks on the symmetric group generated by conjugacy classes. Ph.D. Thesis, Harvard University.
103. Lulov, N., Pak, I. (2002): Rapidly mixing random walks and bounds on characters of the symmetric group. Preprint.
104. Markov, A. (1906): Extension of the law of large numbers to dependent events, Bull. Soc. Math. Kazan **2**, 155–156.

105. Matthews, P. (1987): Mixing rates for a random walk on the cube. SIAM J. Algebraic Discrete Methods **8**, no. 4, 746–752.
106. Matthews, P. (1988): A strong uniform time for random transpositions. J. Theoret. Probab. **1**, 411–423.
107. Matthews, P. (1992): Strong statinary times and eigenvalues. J. Appl. Probab. **29**, 228–233.
108. Margulis, G. (1975): Explicit constructions of concentrators. Prob. of Inform. Transm. **10**, 325–332.
109. McDonald, I. (1979): Symmetric functions and Hall polynomials. Clarendon Press, Oxford.
110. Mohar, B. (1989): Isoperimetric numbers of graphs. J. Combin. Theory **47**, 274–291.
111. Morris, B., Peres, Y. (2002): Evolving sets and mixing. Preprint.
112. Pak, I. (1997): Random walks on groups: strong uniform time approach. Ph.D. Thesis, Department of Math. Harvard University.
113. Pak, I. (1999): Random walks on finite groups with few random generators. Electron. J. Probab. **4**, 1–11.
114. Pak, I. (2000): Two random walks on upper triangular matrices. J. Theoret. Probab. **13**, 1083–1100.
115. Pak, I, Żuk, A. (2002): On Kazhdan constants and mixing of random walks. Int. Math. Res. Not. 2002, no. 36, 1891–1905.
116. Pemantle, R. (1989): An analysis of the overhand shuffle. J. Theoret. Probab. **2**, 37–50.
117. Quenell, G. (1994): Spectral diameter estimates for k-regular graphs. Adv. Math. **106**, 122–148.
118. Reeds, J. (1981): Theory of riffle shuffling. Unpublished manuscript.
119. Roichman, Y. (1996): Upper bound on the characters of the symmetric groups. Invent. Math. **125**, 451–485.
120. Roichman, Y. (1996): On random random walks. Ann. Probab. **24**, 1001–1011.
121. Roussel, S. (1999): Marches aléatoires sur le groupe symétrique. Thèse de Doctorat, Toulouse.
122. Roussel, S. (2000): Phénomène de cutoff pour certaines marches aléatoires sur le groupe symétrique. Colloquium Math. **86**, 111–135.
123. Saloff-Coste, L. (1994): Precise estimates on the rate at which certain diffusions tend to equilibrium. Math. Zeit. **217**, 641–677.
124. Saloff-Coste, L. (1997): Lectures on finite Markov Chains. In Lectures in Probability and Statistics, Lect. Notes in Math. **1665**, Springer.
125. Saloff-Coste, L. (2001): Probability on groups: random walks and invariant diffusions. Notices Amer. Math. Soc. **48**, 968–977.
126. Saloff-Coste, L. (2003): Lower bounds in total variation for finite Markov chains: Wilson's lemma. In: Kaimanovich, V. et al eds., Random walks and Geometry (Vienna, 2001), de Gruyter.
127. Sarnak, P. (1990): Some applications of Modular Forms. Cambridge Tracts in Mathematics **99**, Cambridge University Press.
128. Schoolfield, C. (1998): Random walks on wreath products of groups and Markov chains on related homogeneous spaces. Ph.D. dissertation, Department of Mathematical Sciences, The John Hopkins University.
129. Schoolfield, C. (2002): Random walks on wreath products of groups. J. Theoret. Probab. **15**, 667–693.

130. Shalev, A. (2000): Asymptotic group theory. Notices Amer. Soc. **48** 383–389.
131. Sinclair, A. (1993): Algorithms for random generation and counting: a Markov chain approach. Birkhäuser, Boston.
132. Stong, R. (1995): Random walks on the group of upper triangular matrices. Ann. Probab. **23**, 1939–1949.
133. Stong, R. (1995): Eigenvalues of the natural random walk on the Burnside group $B(3, n)$. Ann. Probab. **23**, 1950-1960.
134. Stong, R. (1995): Eigenvalues of random walks on groups. Ann. Probab. **23**, 1961–1981.
135. Suzuki, M. (1982,1986): Group theory I,II. Springer, New York.
136. Terras, A. (1999): Fourier Analysis on Finite Groups and Applications. London Math. Soc. Student Texts **43**, Cambridge University Press.
137. Thorpe, E. (1973): Nonrandom shuffling with applications to the game of Faro. J.A.S.A. **68**, 842–847.
138. Uyemura-Reyes, J-C. (2002): Random walk, semidirect products, and card shuffling. Ph.D. dissertation, Department of Mathematics, Stanford University.
139. Varopulos, N. Saloff-Coste, L., Coulhon, T. (1992): Analysis and Geometry on Groups. Cambridge Tracts in Mathematics **100**, Cambridge University Press.
140. Wilson, D. (1997): Random random walks on $\mathbb{Z}_2^d$. Probab. Theory Related Fields **108**, 441–457.
141. Wilson, D. (2001): Mixing times of lozenge tiling and card shuffling Markov chains. To appear in Ann. Appl. Probab. arXiv:math.PR/0102193 26 Feb 2001.
142. Wilson, D. (2002): Mixing time of the Rudvalis shuffle. Preprint.
143. Woess, W. (1980): Aperiodische Wahrscheinlichkeitsmasse auf topologischen Gruppen. Mh. Math. **90**, 339–345.
144. Woess, W. (1983): Périodicité de mesures de probabilité sur les groupes topologiques. In Marches Aléatoires et Processus Stochastiques sur le Groupe de Lie. Inst. Élie Cartan, **7**, 170–180. Univ. Nancy.
145. Woess, W. (2000): Random walks on infinite graphs and groups. Cambridge Tracts in Mathematics **138**. Cambridge University Press.
146. Żuk, A. (2002): On property (T) for discrete groups. In Rigidity in dynamics and geometry (Cambridge, 2000), 473–482, Springer, Berlin.