A curiously slowly mixing Markov chain

Persi Diaconis
Department of Mathematics
Stanford University

Andrew Lin
Department of Mathematics
Stanford University

Arun Ram
Department of Mathematics
University of Melbourne

Abstract

We study a Markov chain with very different mixing rates depending on how mixing is measured. The chain is the "Burnside process on the hypercube C_2^n." Started at the all-zeros state, it mixes in a bounded number of steps, no matter how large n is, in ℓ^1 and in ℓ^2. And started at general x, it mixes in at most $\log n$ steps in ℓ^1. But, in ℓ^2, it takes $\frac{n}{\log n}$ steps for most starting x. An interesting connection to Schur–Weyl duality between $\mathfrak{sl}_2(\mathbb{C})$ and S_n further allows for analysis of the mixing time from arbitrary starting states.

Contents

1 Introduction

A mainstay of careful analysis on the mixing rates of Markov chains is

"bound ℓ^1 by ℓ^2 and use eigenvalues."

While this often works to give sharp rates of convergence in ℓ^1, even permitting proof of cutoff (a review is found in Section 2), it can be off if the mixing rates in ℓ^1 and ℓ^2 are of different orders. The present paper studies a natural basic example, the Burnside process on C_2^n, where we find the mixing rates in ℓ^1 and ℓ^2 to be exponentially different.

We begin by describing a general version of the Burnside process. Let $\mathfrak{X}$ be a finite set and G a finite group acting on $\mathfrak{X}$. This group action splits $\mathfrak{X}$ into orbits

$$\mathfrak{X} = \mathcal{O}_1 \cup \mathcal{O}_2 \cup \cdots \cup \mathcal{O}_Z,$$

where we also write $\mathcal{O}_x$ for the orbit containing x. The Burnside process gives a method of choosing an *orbit* uniformly at random. Examples reviewed in Section 2 show that this is an extremely efficient way to generate random integer partitions, Pólya trees, and many other objects of "Pólya theory." It proceeds by a Markov chain on $\mathfrak{X}$ run as follows:

- From $x \in \mathfrak{X}$, choose s uniformly from the set $G_x = \{s : x^s = x\}$.
- From $s \in G$, choose y uniformly from the set $\mathfrak{X}_s = \{y : y^s = y\}$.

The chance of moving from x to y in one step of the chain is

$$K(x,y) = \frac{1}{|G_x|} \sum_{s \in G_x \cap G_y} \frac{1}{|\mathfrak{X}_s|}.$$

As discussed in Section 2, this is an ergodic, reversible Markov chain on $\mathfrak{X}$ with stationary distribution

$$\pi(x) = \frac{1}{Z|\mathcal{O}_x|}, \quad Z \text{ the number of orbits.}$$

Thus, running the chain and simply recording the current orbit gives a Markov chain on orbits with a uniform stationary distribution.

While experiments show extremely rapid mixing of the Burnside process, this fact has been hard to prove in most settings. A first example, the focus of the present paper, takes $\mathfrak{X} = C_2^n$, the set of binary n-tuples, and $G = S_n$, the symmetric group acting by permuting coordinates. Then letting $|x|$ denote the number of ones in $x \in C_2^n$, the orbits can be enumerated as

$$\mathcal{O}_i = \{x : |x| = i\}, \quad 0 \le i \le n.$$

For this example, the two steps of the Burnside process are easy to carry out:

- Given $x \in C_2^n$, G_x is the set of permutations which permute the zeros and ones in x among themselves. Thus $G_x \cong S_i \times S_{n-i}$ with $i = |x|$, and it is easy to choose $s \in G_x$ uniformly.
- Given $s \in S_n$, $\mathfrak{X}_s$ is the set of binary n-tuples fixed by the permutation of coordinates. Thus we may write s as a product of disjoint cycles, label each cycle with independent fair 0/1 coin flips, and install those zeros and ones as the cycles indicate. It is thus easy to choose $y \in \mathfrak{X}_s$ uniformly.

A closed form expression for $K(x, y)$ in the binary case is in Proposition 3.1.

For the binary case, a first analysis by Jerrum [Jer93] showed that order $\sqrt{n}$ steps suffice for ℓ^1 mixing. This was improved by Aldous [AF02], who showed that $\log n$ steps suffice. More precisely, for any $x \in C_2^n$, the *total variation distance* satisfies

$$||K_x^\ell - \pi||_{\text{TV}} \leq n \left(\frac{1}{2}\right)^\ell \leq \left(\frac{1}{2}\right)^c \text{ for } \ell = \log_2 n + c. \tag{1.1}$$

Of course, the starting state can matter. In [DZ21], it is shown that starting at the all-zeros state $\underline{0}$, just a bounded number of steps suffice:

$$\frac{1}{4}\left(\frac{1}{4}\right)^\ell \leq ||K_{\underline{0}}^\ell - \pi||_{\text{TV}} \leq 4\left(\frac{1}{4}\right)^\ell. \tag{1.2}$$

This result is proved by the "bound ℓ^1 by ℓ^2" approach. Here, the ℓ^2 or *chi-square distance* after ℓ steps is

$$\chi_x^2(\ell) = \sum_y \frac{(K^\ell(x,y) - \pi(y))^2}{\pi(y)} = \left|\left|\frac{K_x^\ell}{\pi} - 1\right|\right|_2^2.$$

As illustrated in Eq. (1.2), often the bound

$$4||K_x^\ell - \pi||_{\text{TV}}^2 \leq \chi_x^2(\ell)$$

is fairly sharp (Section 2 provides bounds in the other direction). This requires the ℓ^1 and ℓ^2 mixing times to be of the same order.

Our first main result shows that for the binary Burnside process, the ℓ^1 and ℓ^2 mixing times can have very different orders. For most starting states, order $\frac{n}{\log n}$ steps are required to make $\chi_x^2(\ell)$ small, which is exponentially slower than Eq. (1.1). To state the result, define the average chi-square distance as

$$\chi_{\text{avg}}^2(\ell) = \sum_x \pi(x)\chi_x^2(\ell).$$

Theorem 1.1. *For the binary Burnside process on C_2^n, we have the following:*

1. $\chi_{avg}^2(\ell) = \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{2k} \left(\frac{1}{2^{4k}}\binom{2k}{k}^2\right)^{2\ell}$.
2. *For $\ell \leq \frac{0.1n}{\log n}$, $\chi_{avg}^2(\ell) \to \infty$ as $n \to \infty$. In particular, if $x^{(n)} \in C_2^n$ is the state with $\lfloor \frac{n}{2} \rfloor$ zeros followed by $\lceil \frac{n}{2} \rceil$ ones, we have $\chi_{x^{(n)}}^2(\ell) \to \infty$ for such ℓ.*
3. *For $\ell \geq \frac{10n}{\log n}$, $\chi_{avg}^2(\ell) \to 0$. In fact, more is true: for such ℓ, we have $\chi_{x^{(n)}}^2(\ell) \to 0$ for all $x^{(n)} \in C_2^n$.*

In the process of proving Theorem 1.1, we will also prove the following two refinements of the result:

- Proposition 5.1 shows that order $\frac{n}{\log n}$ steps are in fact required for "almost all" starting states x, with exceptions only occurring if the fraction of either zeros or ones in x approaches 0.
- Theorem 5.2 describes the leading constant at which ℓ^2 cutoff occurs: we have for any $\varepsilon > 0$ that

$$\begin{cases} \chi^2_{\text{avg}}(\ell) \to \infty \\ \chi^2_{\text{avg}}(\ell) \to 0 \end{cases} \quad \text{if} \quad \begin{cases} \ell \leq (1-\varepsilon)\frac{\log 2}{2}\frac{n}{\log n} \\ \ell \geq (1+\varepsilon)\frac{\log 2}{2}\frac{n}{\log n} \end{cases}.$$

The proofs of these three results rely on an explicit diagonalization, which we describe now:

Theorem 1.2. *Let $K(x,y)$ be the transition matrix of the binary Burnside process on C_2^n. We have the following:*

1. *The eigenvalues of K are 0 and*

$$\beta_k = \frac{1}{2^{4k}}\binom{2k}{k}^2, \quad 0 \leq k \leq \left\lfloor \frac{n}{2} \right\rfloor.$$

2. *The eigenvalue multiplicity of 0 is 2^{n-1}. The eigenvalue multiplicity of β_k is $\binom{n}{2k}$ for all $0 \leq k \leq \lfloor \frac{n}{2} \rfloor$.*
3. *For $0 \leq k \leq \lfloor \frac{n}{2} \rfloor$, a basis of eigenvectors for the β_k-eigenspace is*

$$\left\{ f_S(x) = (-1)^{|x_S|}\binom{2k}{|x_S|} : |S| \subset [n], |S| = 2k \right\},$$

where $|x_S|$ denotes the number of ones of $x \in C_2^n$ among the coordinate set S.

Remark. *Alas, the eigenvectors f_S are not orthogonal. We describe formulas in Section 4 for the inner products $\langle f_S, f_T \rangle$ for any subsets S, T, but the usual route of expressing $\chi_x^2(\ell)$ requires an orthonormal eigenbasis, meaning that we would need to find appropriate linear combinations of the f_Ss. That is exactly what is done in the main result of Section 6.*

We conclude with a summary of the rest of the paper. Section 2 below gives background on some required analytic tools (Section 2.1), as well as some additional uses for the eigenvectors (Section 2.2). It also gives a survey of examples where ℓ^1 and ℓ^2 rates are the same and different (Section 2.3) and a brief review of the Burnside process (Section 2.4).

Properties of the transition matrix K are developed in Section 3. We point to a curious feature: for any subset $S \subseteq [n]$, the chain on C_2^n lumped to S is *precisely* the Burnside process on $C_2^{|S|}$. Theorem 1.2 is then proved in Section 4, and Theorem 1.1 and its refinements are proved in Section 5.

The binary Burnside process has all kinds of symmetries. Section 6 shows that because the matrix $K(x,y)$ commutes with the natural action of $\mathfrak{sl}_2(\mathbb{C})$ on $\mathbb{C}[C_2^n]$, Schur–Weyl duality provides a decomposition of orthogonal eigenvectors for the function space. Theorem 6.2 explicitly constructs this complete set of eigenvectors

$$\left\{ f_Q^{m,\ell} : m \in \{0,1,\ldots,\lfloor n/2 \rfloor\},\ \ell \in \{0,1,\ldots,n-2m\},\ Q \text{ a standard Young tableau of shape } (n-m,m) \right\}$$

using certain linear combinations of the f_S eigenvectors above. With these eigenvectors, we are able to determine sharp rates of convergence for some particular starting states: specifically, Theorem 6.7 shows that starting from a state x with one coordinate 1 and all other coordinates 0, we have

$$5\left(\frac{1}{4}\right)^{2s} \leq \chi_x^2(s) \leq 270\left(\frac{1}{4}\right)^{2s},$$

so that the Markov chain indeed mixes much more quickly (in a constant number of steps) from these special states.

Finally, Section 7 compares some of the properties of the binary Burnside process, including eigenvalue multiplicities and duality, to other chains in which these properties manifest. It then shows that some of the magical properties of the binary Burnside chain also hold for the chain on C_k^n for $k \geq 2$.

Acknowledgements

This paper is dedicated to our friend Svante Janson on the occasion of his 70th birthday. Over the years, we have tried (and failed) to interest him in mixing questions. We sure could use his help on this problem.

We also thank Evita Nestoridi, Allan Sly, Amol Aggarwal, Yuval Peres, Michael Howes, Jonathan Hermon, Lucas Teyssier, Laurent Bartholdi, Christoph Koutschan, and Richard Stanley for helpful discussions and ideas.

P.D. was partially supported by the NSF under Grant No. DGE-1954042, and A.L. was partially supported by the NSF under Grant No. DGE-2146755.

2 Background

This section contains needed background and a literature review. Section 2.1 reviews the analytic background for bounding ℓ^1 and ℓ^2 distances using eigenvalues and eigenvectors, and Section 2.2 provides some further applications of those eigenvectors specialized to our chain. Section 2.3 gives examples of Markov chains where both distances are well-understood enough to give useful comparisons. Finally, Section 2.4 discusses previous literature on the Burnside process.

2.1 Analytic background

An exceptional text for mixing time results is the book by Levin and Peres [LP17]. Chapter 12 of their book contains the basics for bounding ℓ^1 and ℓ^2 distances using eigenvalues. The comprehensive text of Saloff-Coste [SC97] develops analytic tools more deeply.

Let $\mathfrak{X}$ be a finite set and $K(x, y)$ a Markov transition matrix with state space $\mathfrak{X}$ and stationary distribution $\pi(x)$. Throughout this section, assume that (K, π) is ergodic and reversible. For any $0 < p < \infty$, let $\ell^p(\pi) = \{f : \mathfrak{X} \to \mathbb{R}\}$ denote the function space with norm

$$||f||_p^p = \sum_x |f(x)|^p \pi(x).$$

Reversibility implies that $Kf(x) = \sum_y K(x, y)f(y)$ is self-adjoint as a map $\ell^2 \to \ell^2$; that is, for any functions f, g, we have $\langle Kf, g\rangle = \langle f, Kg\rangle$, where $\langle f, g\rangle = \sum_x f(x)g(x)\pi(x)$. Thus, the spectral theorem shows that K has an orthonormal set of eigenvectors and corresponding eigenvalues f_i, β_i with $Kf_i(x) = \beta_i f_i(x)$ for all i. As usual, we reorder so that $1 = \beta_0 > \beta_1 \geq \beta_2 \geq \cdots \geq \beta_{|\mathfrak{X}|-1} > -1$. Started at a state x, the ℓ^1 or *total variation* distance

$$||K_x^\ell - \pi||_{\mathrm{TV}} = \frac{1}{2}\sum_y |K^\ell(x,y) - \pi(y)|$$

and the ℓ^2 or *chi-square* distance

$$\chi_x^2(\ell) = \sum_y \frac{(K^\ell(x,y) - \pi(y))^2}{\pi(y)} = \left|\left|\frac{K_x^\ell}{\pi} - 1\right|\right|_2^2$$

can be bounded as

$$4||K_x^\ell - \pi||_{\mathrm{TV}}^2 \leq \chi_x^2(\ell) = \sum_{i=1}^{|\mathfrak{X}|-1} f_i^2(x)\beta_i^{2\ell} \leq \frac{1}{\pi(x)}\beta_*^{2\ell}, \tag{2.1}$$

where $\beta_* = \max(|\beta_1|, |\beta_{|\mathfrak{X}|-1}|)$ is the second absolute eigenvalue.

If f is an eigenfunction for K with eigenvalue $\beta \neq 1$,

$$\begin{aligned}|\beta^\ell f(x)| = \left|K^\ell f(x)\right| &= \left|\sum_y K^\ell(x,y)f(y)\right| \\ &= \left|\sum_y K^\ell(x,y)f(y) - \pi(y)f(y)\right| \\ &\leq \sum_y |K^\ell(x,y) - \pi(y)|f^*,\end{aligned}$$

where $f^* = \max_y |f(y)|$. Choosing x^* so that $|f(x^*)| = f^*$ then yields

$$|\beta^\ell| \leq \sum_y |K^\ell(x,y) - \pi(y)| = 2||K_{x^*}^\ell - \pi||_{\mathrm{TV}}.$$

Combining this with $\sum_{i=0}^{|\mathfrak{X}|-1} f_i^2(x) = \frac{1}{\pi(x)}$ and Eq. (2.1) yields the following:

Proposition 2.1. *Let (K, π) be a reversible Markov chain with second absolute eigenvalue β_*, and let f be an eigenvector for β_*. Suppose $x_* \in \mathfrak{X}$ satisfies $|f(x^*)| = \max_x |f(x)|$. Then*

$$4||K_{x^*}^\ell - \pi||_{TV}^2 \leq \chi_{x^*}^2(\ell) \leq \frac{1}{\pi(x^*)}||K_{x^*}^{2\ell} - \pi||_{TV}.$$

This is one of several results showing roughly that if $||K^\ell - \pi||_{\mathrm{TV}}$ is close to zero in order ℓ^* steps, then $\chi^2(\ell)$ is close to zero in order $\ell^* + \log \pi(x^*)$ steps. (See [LP17, Chapter 12].) Here is an example application to the binary Burnside process:

Corollary 2.2. *For the binary Burnside process on C_2^n and any $x \in C_2^n$,*

$$\chi_x^2(\ell) \leq \frac{1}{\pi(x)}||K_x^{2\ell} - \pi||_{TV}.$$

Proof. By Theorem 1.2, the second absolute eigenvalue is $\beta_1 = \frac{1}{4}$. Furthermore, for each $1 \le i < j \le n$, we have the corresponding β_1-eigenfunction

$$f_{\{i,j\}}(x) = (-1)^{x_i+x_j}\binom{2}{x_i+x_j} = \begin{cases} 1 & \text{if } (x_i,x_j) = (0,0) \text{ or } (1,1), \\ -2 & \text{if } (x_i,x_j) = (0,1) \text{ or } (1,0). \end{cases} \tag{2.2}$$

Thus if x is not the all-zeros $\underline{0}$ or all-ones $\underline{1}$ state, there is at least one choice of i, j with $|f_{\{i,j\}}(x)| = 2$ and the result follows. And if $x = \underline{0}$ or $\underline{1}$, then the sum of all $\binom{n}{2}$ $f_{\{i,j\}}$s is a β_1-eigenfunction achieving its maximum magnitude at $x = \underline{0}$ and $\underline{1}$, so again we may apply the previous result. ☐

Example 2.3. *Using Eq. (1.1) to bound the right-hand side of the corollary, and noting that for any x,*

$$\frac{1}{\pi(x)} = (n+1)\binom{n}{|x|} \le 2n^{1/2}2^n,$$

we have that for any starting state,

$$\chi_x^2(\ell) \le 2n^{3/2}2^n\left(\frac{1}{2}\right)^\ell.$$

This shows that $n + c\log n$ steps suffice for ℓ^2 mixing for every starting state x, and Theorem 1.1 shows that this is indeed close to best possible.

On the other hand, the bound is sometimes much tighter. If $|x|$ or $n - |x|$ is of constant size (meaning that all but a constant number of coordinates in the starting state are all 0s or all 1s), then the corollary states that $\chi_x^2(\ell)$ is at most $\left(\frac{1}{2}\right)^\ell$ times a polynomial in n, and therefore just $c\log n$ steps will suffice for such states.

Remark. *It is of course important to point out that there are many other approaches for proving rates of convergence. For example, Aldous' bound (Eq. (1.1)) is proved by coupling, and strong stationary times have also proved useful; see [LP17, Chapters 5 and 6] for definitions and references. The direct use of Cauchy-Schwarz to bound ℓ^1 by ℓ^2 can also be refined, as is done in [Tey20] and [Nes24] in the analysis of limit profiles of reversible Markov chains. Other tools include functional inequalities such as the Nash [DSC96b], Harnack [MSC23], and log-Sobolev [DSC96a] inequalities, as well as the emerging work on spectral independence [SV23].*

Despite all of this, "bound ℓ^1 by ℓ^2 and use eigenvalues" is still basic and useful. In particular, it must be mentioned that one of the advantages of ℓ^2 bounds is the availability of comparison theory [DSC93]. If one has a rate in ℓ^2, then it is often possible to get good rates of closely related chains, such as perturbations of the kernel or more drastic variations. (For example, the walk on permutations generated by a single transposition and a single n-cycle gets sharp bounds via comparison with random transpositions.) This kind of robustness does not seem to be available for other methods of proof.

2.2 Some uses for the eigenvectors

The bounds for $\chi_x^2(\ell)$ above show an example calculation where "the eigenvectors can be used for something." When studying convergence of a Markov chain, it can also be informative to see how certain key features of that chain converge. We show now in several examples how the explicit form of our eigenvectors f_S can be useful for such questions. (And we also note that the algebraic approach of Section 6 relies heavily on the expressions for the f_Ss as well.)

One natural statistic on binary strings is the number of *alternations* $T(x)$ (that is, the count of adjacent differing coordinates); for example, $T(011011) = 3$. The celebrated work of Tversky and Kahneman

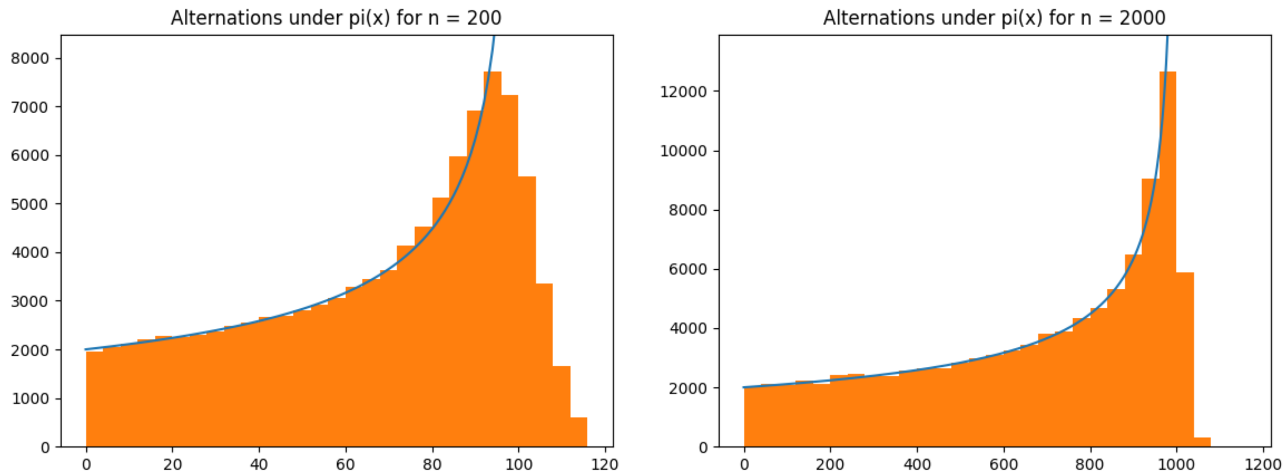

Figure 1: Alternation count histograms for 100000 binary strings sampled under $\pi_n(x)$ for $n = 200$ and $n = 2000$. The smooth curve corresponds to the limiting density $\frac{1}{\sqrt{1-2x}}$ for a random variable distributed as $2U(1-U)$ for U uniform.

[TK74] on misperceptions of chance applies to the way we view the "hot hand" in basketball games [GVT85], the effect of weather on arthritis pain [RT96], and many other examples. All of these misperceptions happen because most people think that a random sequence should have very many alternations and no long runs of zeros or ones.

Under a fair coin-tossing model for a uniformly random binary sequence of length n, all alternations are independent and thus $T(x)$ has mean $\frac{n-1}{2}$ with standard deviation $\frac{\sqrt{n-1}}{2}$. In particular, when n is large, $T(x)$ normalized by this mean and standard deviation has an approximately normal distribution. If we instead observe a coin-tossing sequence where the probability of heads p is unknown and has uniform prior on $[0, 1]$, then the resulting sequence is exactly distributed as $\pi(x)$ for the binary Burnside process (see the second half of Section 7.1 for some further discussion on a related Gibbs sampler chain). Alternations at different coordinates are now no longer independent, and the following discussion shows that $T(x)$ under $\pi(x)$ has a completely different behavior to that in the uniform case.

Proposition 2.4. *Let T_n be the random variable $T(x)$ under π_n on C_2^n. Then $\frac{T_n}{n-1}$ converges in distribution to $2U(1-U)$, where U is uniform on $[0, 1]$.*

A histogram of sampled alternation counts (along with the limiting density) is shown in Fig. 1. Notice in particular that even though $T(x)$ can be as large as $n-1$, it typically takes values below $\frac{n}{2}$; the following proof provides an explanation for this.

Proof. The measure π_n on C_2^n restricted to a subset of $m < n$ coordinates is exactly the measure π_m on C_2^m; this can be shown by direct calculation or by interpreting binary strings as draws from Pólya's urn. (In fact, more is true: Proposition 3.3 shows that a similar statement also holds for the binary Burnside process.) Thus, by consistency of finite-dimensional distributions, we may extend the π_ns to a measure on infinite binary sequences. Recall that each π_n is uniform on the number of ones and that the coordinates are exchangeable under π_n. So by de Finetti's theorem, this limiting infinite measure can be described as "pick p uniformly on $[0, 1]$ and flip a p-coin independently for each coordinate."

Thus if we condition on the tail σ-field and let p be the conditional probability that $x_i = 1$, then the x_is form an iid string of p-coin-tosses. Letting $Y_i = 1\{x_i \neq x_{i+1}\}$ (and still conditioning on the tail), the Y_is are each Bernoulli with parameter $2p(1-p)$ and form a 1-dependent sequence of random variables; in particular we have that $\frac{T_n}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1} Y_i$ converges in probability to $2p(1-p)$ by the weak law of large numbers, uniformly in p. Thus unconditionally we must have $\frac{T_n}{n-1}$ converge to $2U(1-U)$ for U uniform, as desired. □

In particular, while $\frac{T_n}{n-1}$ converges in probability to $\frac{1}{2}$ under the uniform measure on binary strings, it instead converges to some nondegenerate random variable with mean $\frac{1}{3}$ under π. We now show convergence of $\mathbb{E}[T(x)]$ to its average using our β_1-eigenfunctions:

Example 2.5. *Consider again the binary Burnside process on C_2^n. We have (using the formula Eq. (2.2)) that $1\{x_i \neq x_{i+1}\} = \frac{1-f_{\{i,i+1\}}(x)}{3}$ for any $1 \leq i \leq n-1$. Therefore summing over all i, the total number of alternations is*

$$T(x) = \sum_{i=1}^{n-1} 1\{x_i \neq x_{i+1}\} = \frac{n-1}{3} - \frac{1}{3}\sum_{i=1}^{n-1} f_{\{i,i+1\}}(x).$$

Each $f_{\{i,i+1\}}(x)$ is an eigenfunction of K of eigenvalue $\frac{1}{4}$. Thus if $X_0 = x, X_1, \cdots, X_\ell, \cdots$ records the location of the binary Burnside process started at x, we have

$$\begin{aligned}
\mathbb{E}[T(X_\ell)|X_0 = x] &= \mathbb{E}[K^\ell T(x)] \\
&= \frac{n-1}{3} - \frac{1}{3}\sum_{i=1}^{n-1} \mathbb{E}[K^\ell f_{\{i,i+1\}}(x)] \\
&= \frac{n-1}{3} - \frac{1}{3 \cdot 4^\ell}\sum_{i=1}^{n-1} \mathbb{E}[f_{\{i,i+1\}}(x)] \\
&= \frac{n-1}{3} - \frac{1}{4^\ell}\left(\frac{n-1}{3} - T(x)\right);
\end{aligned}$$

in particular, the expected number of alternations is always close to its mean under π after just $\log_4 n + c$ steps from any starting state.

Similar calculations also yield the same type of exponential decay for other "pairwise-coordinate" statistics, such as the covariance of the number of ones between different coordinate sets of x. Indeed, for any sets $S, T \subset [n]$ (remembering that $|x_S|$ denotes the number of ones in x among the set S),

$$\begin{aligned}
\mathrm{Cov}(|x_S|, |x_T|) &= \sum_{i \in S, j \in T} \mathrm{Cov}(x_i, x_j) \\
&= \sum_{i \in S, j \in T} \mathbb{E}\left[\left(x_i - \frac{1}{2}\right)\left(x_j - \frac{1}{2}\right)\right] \\
&= \sum_{i \in S, j \in T} \frac{1}{4}\left(1\{x_i = x_j\} - 1\{x_i \neq x_j\}\right) \\
&= \sum_{i \in S, j \in T} \frac{1}{6}\left(f_{\{i,j\}}(x) + \frac{1}{2}\right),
\end{aligned}$$

so again the deviation from the mean $\mathbb{E}[\mathrm{Cov}(|x_S|, |x_T|)] = \frac{1}{12}|S||T|$ will decay exponentially in the number of steps taken.

Since the number of alternations $T(X_\ell)$ does not concentrate around its mean, it is also informative to compute its variance. For this, we again expand out in terms of eigenvectors:

$$\begin{aligned}
\mathrm{Var}(T(X_\ell)|X_0 = x) &= \mathrm{Var}\left(\frac{1}{3}\sum_{i=1}^{n-1} f_{\{i,i+1\}}(X_\ell) \,\middle|\, X_0 = x\right) \\
&= \frac{1}{9}\mathrm{Cov}\left(\sum_{i=1}^{n-1} f_{\{i,i+1\}}(X_\ell), \sum_{j=1}^{n-1} f_{\{j,j+1\}}(X_\ell) \,\middle|\, X_0 = x\right) \\
&= \frac{1}{9}\sum_{i,j=1}^{n-1}\Bigg(\mathbb{E}\left[f_{\{i,i+1\}}(X_\ell) f_{\{j,j+1\}}(X_\ell) \,\middle|\, X_0 = x\right] \\
&\qquad - \mathbb{E}\left[f_{\{i,i+1\}}(X_\ell) \,\middle|\, X_0 = x\right]\mathbb{E}\left[f_{\{j,j+1\}}(X_\ell) \,\middle|\, X_0 = x\right]\Bigg).
\end{aligned}$$

The latter term decays exponentially because each $f_{\{i,i+1\}}$ is an eigenvector of eigenvalue $\frac{1}{4}$, but for the former term we need to first rewrite $f_{\{i,i+1\}} f_{\{j,j+1\}}$ as a linear combination of eigenvectors, which looks different for each of the cases $i = j$, $|i - j| = 1$, and $|i - j| > 1$. For distinct indices a, b, c, d, we have

$$f_{\{a,b\}} f_{\{c,d\}} = \frac{18}{35} f_{\{a,b,c,d\}} - \frac{2}{7}(f_{\{a,b\}} + f_{\{c,d\}}) + \frac{3}{14}(f_{\{a,c\}} + f_{\{a,d\}} + f_{\{b,c\}} + f_{\{b,d\}}) + \frac{1}{5} f_{\varnothing},$$

$$f_{\{a,b\}} f_{\{a,c\}} = \frac{3}{2} f_{\{b,c\}} - \frac{1}{2}(f_{\{a,b\}} + f_{\{a,c\}}) + \frac{1}{2} f_{\varnothing},$$

$$f_{\{a,b\}}^2 = -f_{\{a,b\}} + 2 f_{\varnothing}.$$

Plugging in all of these formulas and then using that $\mathbb{E}\left[f_{\varnothing}(X_\ell) \,|\, X_0 = x\right] = 1, \mathbb{E}\left[f_{\{a,b\}}(X_\ell) \,\middle|\, X_0 = x\right] = \left(\frac{1}{4}\right)^\ell f_{\{a,b\}}(x)$, and $\mathbb{E}\left[f_{\{a,b,c,d\}}(X_\ell) \,\middle|\, X_0 = x\right] = \left(\frac{9}{64}\right)^\ell f_{\{a,b,c,d\}}(x)$ yields an expression for $\mathrm{Var}(T(X_\ell)|X_0 = x)$ in terms of only exponential factors in ℓ and eigenfunctions evaluated at x. If we only keep the terms corresponding to $f_\varnothing$ (since all other terms are exponentially decaying), we find the variance of $T(x)$ under the stationary distribution π_n, which for all $n \geq 2$ is

$$\mathrm{Var}_{\pi_n}(T(x)) = \frac{1}{9}\left((n-1)\cdot 2 + (2n-4)\cdot\frac{1}{2} + (n^2-5n+6)\cdot\frac{1}{5}\right) = \frac{1}{45}(n^2+10n-14).$$

In particular, $\mathrm{Var}_{\pi_n}\left(\frac{T(x)}{n-1}\right)$ does converge to $\frac{1}{45}$, the variance of $2U(1-U)$ for U uniform on $[0,1]$, as $n \to \infty$.

Remark. *In Theorem 1.2, we claim that f_S is an eigenvector of K for any subset S of even size. In fact, the proof of this result (in Section 4) shows that the formula $f_S(x) = (-1)^{|x_S|}\binom{k}{|x_S|}$ also yields an eigenvector of eigenvalue 0 for any S of odd size k, and that the collection of all such eigenvectors is a basis for the 0-eigenspace. Other applications of our eigenvectors (i.e. expanding other functions as linear combinations involving binomial coefficients) may find this basis useful. For example, letting $|X_\ell|$ be the number of ones after ℓ steps, we find for any starting x and all $\ell \geq 1$ (using $|x| = \frac{n-\sum_{i=1}^n f_{\{i\}}(x)}{2}$) that*

$$\mathbb{E}\left[|X_\ell|\right] = \frac{n}{2}, \quad \mathit{Var}(|X_\ell|) = \frac{n(n+2)}{12}\cdot\left(1-\frac{1}{4^\ell}\right) + \left(|x|-\frac{n}{2}\right)^2\cdot\frac{1}{4^\ell}.$$

That is, the number of ones has the right mean after just one step, but it takes order $\log n$ steps to get the variance right.

2.3 ℓ^1 versus ℓ^2 examples

By now, the literature of carefully worked examples is so large that a serious survey would require a book-length effort; see [SC97] for a start. In general, changing the metric of convergence can result in drastically different results, and Gibbs–Su [GS02] provides a general survey. We content ourselves with examples drawn from our own work.

The first examples of sharp mixing time analyses are found in Aldous [Ald83] and Diaconis–Shahshahani [DS81]. Random transpositions on the symmetric group S_n was found to have a cutoff at the mixing time of $\frac{1}{2}n\log n + cn$, both in ℓ^1 and ℓ^2. Similar results were found for simple random walk on C_2^n, simple random walk on the n-cycle, and the Bernoulli-Laplace urn. All results were proved by "bound ℓ^1 by ℓ^2 and use eigenvalues."

Mixing occurs at the same order, but with different constants for the cutoff in ℓ^1 and ℓ^2, for random walk on certain expanding graphs called Ramanujan graphs; see Lubetzky–Peres [LP15] for precise statements. There are a host of other results where mixing occurs at the same order (up to constants) in ℓ^1 and ℓ^2; most of the examples that use the previously mentioned "comparison theory" fall into this class. For example, random walk on the symmetric group S_n based on choosing either the transposition $(1,2)$ or the n-cycle $(1,2,\cdots,n)$ with probability $\frac{1}{2}$ each (for n odd) mixes in order $n^3\log n$ for both ℓ^1 and ℓ^2.

These classical examples have had considerable development. For example, Olesker-Taylor, Teyssier, and Thévenin [OTTT25] show that any random walk supported on any conjugacy class in S_n (such as random 3-cycles or random n-cycles) have comparable ℓ^1 and ℓ^2 rates of convergence. In a sustained development, Guralnick, Larsen, Liebeck, Shalev, and Tiep (in various combinations) have shown the same fact for walks supported on conjugacy classes for finite groups of Lie type. A convenient reference is [GLT24], detailing a large development.

A different part of the spectrum concerns finite abelian and nilpotent groups. A series of papers by Hermon, Olesker-Taylor, and Huang have close to complete results – see [HOT21, HH24].

The above is a pale summary of a rich literature, but this is not the time or place for further details.

There are also more refined "limit profile" results where "bound ℓ^1 by ℓ^2" does not provide sufficiently refined estimates. For example, the limit shape results for random walk on the hypercube [DGM90], riffle shuffles [BD92], and random transpositions [Tey20] all require more detailed analysis.

In the other camp, here are some results where ℓ^1 and ℓ^2 rates are simply different. The easiest example is lazy simple random walk on the complete graph on n vertices, which has bounded ℓ^1 mixing time but needs order $\log n$ in ℓ^2. For a more striking example, Peres and Revelle [PR04] study simple random walk on the lamplighter group with underlying graph the n-cycle C_n. They show that order n^2 steps are necessary and sufficient for ℓ^1 convergence, while order n^3 steps are necessary and sufficient for ℓ^2 convergence. The present paper offers an even more extreme example where the ℓ^1 and ℓ^2 bounds are exponentially different.

Remark. *As an additional sidenote, continuous-time analogs of discrete-time Markov chains may also have notably different mixing time behavior. While Chen and Saloff-Coste [CSC13] prove that lazy discrete-time Markov chains exhibit total variation cutoff if and only if the associated continuous-time Markov processes do, Hermon and Peres [HP18] show that this is no longer true when using the closely-related metric of separation distance. Turning to ℓ^2 mixing, Saloff-Coste and Zúñiga [SCZ08] show that cutoff times occur at different orders in discrete and continuous time for some conjugacy-class random walks on the symmetric and alternating group, and they remark that this occurs due to the effect of a "very large number of very small eigenvalues." Our calculations in this paper for the binary Burnside process show that high eigenvalue multiplicity can also manifest in differences between ℓ^1 and ℓ^2, even when restricted only to discrete time.*

2.4 The Burnside process

The Burnside process was introduced by Jerrum [Jer93] and studied by Goldberg and Jerrum [GJ98]. Their original motivation was computational complexity, and they produced examples where the Markov chain requires exponentially many steps to converge. More practical applications later appeared: Diaconis–Tung [DT25] and Diaconis–Howes [DH25] use the chain as an extremely effective algorithm for generating uniform partitions of n (of size roughly 10^8) and large contingency tables. Diaconis–Zhong [DZ25] uses the Burnside process to generate random conjugacy classes in the group $U(n, q)$, where for example for $n = 40, q = 2$, the orbit space has order roughly 2^{400}. And Bartholdi–Diaconis [BD24] describes an algorithm for using the chain to generate large uniform unlabeled trees and compare various statistics with the corresponding labeled trees. In all of these examples, extensive empirical testing indicates that the chain converges after just 100 steps. However, *no* explicit rates have been proven for any of these examples.

The binary Burnside process has been carefully studied as a first example towards proving such rates, and the papers of Jerrum, Aldous, and Diaconis–Zhong in the introduction have all contributed to this. The present paper shows that even this simple case has unknown corners. Good results for the Burnside process on conjugacy classes of CA groups are developed by Rahmani [Rah20], and good results for set partitions can be found in the work of Paguyo [Pag23]. Most recently, a sharp analysis of mixing time and limit profile for the Burnside process on certain Sylow double cosets of S_n was obtained by Howes [How25].

3 Properties of the chain

This section develops symmetry and lumping properties of the binary Burnside process. Throughout the rest of the paper, we may write K_n for K for the sake of clarity.

To begin, here is a closed form for the transition matrix:

Proposition 3.1. *Let K denote the transition matrix for the binary Burnside process. Fix $x, y \in C_2^n$, and for $a, b \in \{0, 1\}$, let n_{ab} be the number of coordinates i where $x_i = a$ and $y_i = b$ (so $n_{00} + n_{01} + n_{10} + n_{11} = n$). Then*

$$K(x, y) = \frac{\binom{2n_{00}}{n_{00}}\binom{2n_{01}}{n_{01}}\binom{2n_{10}}{n_{10}}\binom{2n_{11}}{n_{11}}}{4^n\binom{n_{00}+n_{01}}{n_{00}}\binom{n_{10}+n_{11}}{n_{10}}}.$$

Proof. The permutations that fix x can be described by $G_x = S_{n_{00}+n_{01}} \times S_{n_{10}+n_{11}}$, in which we permute the indices where $x_i = 0$ and also the indices where $x_i = 1$. Similarly, $G_x \cap G_y = S_{n_{00}} \times S_{n_{01}} \times S_{n_{10}} \times S_{n_{11}}$ is the set of all permutations whose cycles are each contained entirely within each type of coordinate, and for any such permutation $\sigma = \sigma_{00} \times \sigma_{01} \times \sigma_{10} \times \sigma_{11} \in G_x \cap G_y$ we have

$$\frac{1}{|\mathfrak{X}_g|} = \left(\frac{1}{2}\right)^{C(\sigma_{00})+C(\sigma_{01})+C(\sigma_{10})+C(\sigma_{11})},$$

where $C(\tau)$ denotes the number of cycles in the permutation τ, since a binary n-tuple is fixed by σ if and only if it is constant (either all 0 or all 1) on each cycle. Therefore

$$\begin{aligned} K(x, y) &= \frac{1}{(n_{00}+n_{01})!(n_{10}+n_{11})!}\sum_{\sigma_{00},\sigma_{01},\sigma_{10},\sigma_{11}}\left(\frac{1}{2}\right)^{C(\sigma_{00})+C(\sigma_{01})+C(\sigma_{10})+C(\sigma_{11})} \\ &= \frac{1}{(n_{00}+n_{01})!(n_{10}+n_{11})!}\sum_{\sigma_{00}}\left(\frac{1}{2}\right)^{C(\sigma_{00})}\sum_{\sigma_{01}}\left(\frac{1}{2}\right)^{C(\sigma_{01})}\sum_{\sigma_{10}}\left(\frac{1}{2}\right)^{C(\sigma_{10})}\sum_{\sigma_{11}}\left(\frac{1}{2}\right)^{C(\sigma_{11})}. \end{aligned}$$

Now because the generating function for permutation cycle count is given by

$$C_n(x) = \sum_{\sigma \in S_n} x^{C(\sigma)} = x(x+1)\cdots(x+n-1)$$

(for example by induction, since there are $(n-1)$ ways to insert the number n into an existing cycle and 1 way to add a new cycle with just n), we have that

$$\sum_{\sigma_{00}}\left(\frac{1}{2}\right)^{C(\sigma_{00})} = \frac{1}{2}\left(\frac{1}{2}+1\right)\cdots\left(\frac{1}{2}+n_{00}-1\right) = \frac{(2n_{00}-1)!!}{2^{n_{00}}} = \frac{(2n_{00})!}{4^{n_{00}}n_{00}!}$$

and similar for the other terms. Plugging this in, we thus have

$$\begin{aligned} K(x, y) &= \frac{1}{(n_{00}+n_{01})!(n_{10}+n_{11})!}\frac{(2n_{00})!}{4^{n_{00}}n_{00}!}\frac{(2n_{01})!}{4^{n_{01}}n_{10}!}\frac{(2n_{10})!}{4^{n_{10}}n_{10}!}\frac{(2n_{11})!}{4^{n_{11}}n_{11}!} \\ &= \frac{1}{4^n}\frac{n_{00}!n_{01}!n_{10}!n_{11}!}{(n_{00}+n_{01})!(n_{10}+n_{11})!}\frac{(2n_{00})!}{n_{00}!^2}\frac{(2n_{01})!}{n_{01}!^2}\frac{(2n_{10})!}{n_{10}!^2}\frac{(2n_{11})!}{n_{11}!^2}, \end{aligned}$$

which rearranges to the desired result. ∎

Corollary 3.2. *The transition matrix $K(x, y)$ satisfies*

$$K(x, y) = K(x, \overline{y}) = K(\overline{x}, \overline{y}) = K(\sigma(x), \sigma(y)),$$

where $\overline{x}$ is the binary n-tuple obtained from x by flipping all bits and $\sigma \in S_n$ is any permutation.

Proof. All equalities follow from observing that the quantity in Proposition 3.1 is left invariant under permutation of coordinates or negation of either x or y. □

The first two equalities describe an additional symmetry of the chain which will be helpful in obtaining eigenvectors, while the last equality may be viewed as a consequence of the lumping of the binary Burnside process to its orbits (since $|x| = |\sigma(x)|$ for all $x \in C_2^n$). In particular, K (viewed as an operator) commutes with the S_n-action and thus maps only within isomorphic irreducible subspaces of permutation representations on the vector space of functions $L(C_2^n)$; this is further discussed in Section 6. Similar analysis for other Markov chains with various symmetries, towards comparing the Metropolis algorithm to other random walks on graphs, can be found in [BDPX05].

To state the next (crucial) feature of $K(x, y)$, recall that a function of a Markov chain need not be a Markov chain. For the (general) Burnside process, as in the introduction, [Dia05, Section3] shows that the chain "lumped to orbits" remains a Markov chain with a uniform stationary distribution. Background on lumping, in particular Dynkin's criterion, can be found in [KS83] or [Pan18]. Its application to the binary Burnside process underlies the results for $||K_0^\ell - \pi||_{\text{TV}}$ and $\chi_0^2(\ell)$ stated in the introduction, using that $\underline{0}$ is in its own orbit. This next result shows that a very different set of lumpings also remains valid:

Proposition 3.3. *The restriction of the Burnside process on (C_2^n, S_n) to any $m \leq n$ of its coordinates is also a Markov chain, and its transition probabilities are exactly given by the Burnside process on (C_2^m, S_m).*

Proof. Given any permutation $\sigma \in S_n$, we may write it uniquely in cycle notation by cyclically moving the largest element in each cycle to the beginning and then sorting those largest elements from smallest to largest. For example, the one-line permutation 26375841 becomes (3)(5)(74)(8126). Observe that if we remove the parentheses from this expression, then a new cycle begins at every left-to-right record (that is, every largest number starting from the left), so we can read off the cycle notation simply from the sequence of numbers.

But now if we remove the numbers 1 through t in our cycle notation, this does not change whether each of the numbers $t + 1$ through n is a record or not. Thus regardless of the order in which 1 through t appear, we will get the same permutation on $t + 1$ through n after erasing; in fact, any two elements of $\{t + 1, \cdots, n\}$ end up in the same cycle if and only if they were in the same cycle before the erasure. So in particular, if we sample a uniformly random permutation from S_n and then erase the numbers 1 through t in the cycle notation described above, then the result is a uniformly random permutation from S_{n-t} (on the remaining numbers).

Now consider what happens in one step of the Burnside process. For clarity, first consider the case where we restrict to the last m coordinates. If we are at some state x in C_2^n at the start of a step, then the coordinate set can be partitioned into the locations of 0s and 1s in x; call these sets L_0, L_1. The process specifies that we first pick a uniform group element fixing x, which is a uniform permutation from $S_{L_0} \times S_{L_1}$, write this permutation in cycle notation, and then uniformly assign to each cycle either 0 or 1. But if we erase the numbers $1, \cdots, n - m$ from our cycle notation, by the logic in the previous paragraph, we get a uniform permutation on $S_{L_0 \cap [n-m+1,n]}$ and on $S_{L_1 \cap [n-m+1,n]}$, hence a uniform permutation among all choices of $S_{L_0 \cap [n-m+1,n]} \times S_{L_1 \cap [n-m+1,n]}$. Furthermore, each of the remaining cycles is still uniformly assigned either 0 or 1 because the "being in the same cycle" property is preserved. Therefore, we see that Dynkin's criterion for lumping holds here, since the transition probabilities into each potential orbit (indexed by the values of the tuple on the last m indices) are the same regardless of the values of the first $n-m$ indices, and they are exactly those given by the Burnside process on C_2^m.

This argument also works if we restrict to any other subset of the indices instead of the last m coordinates; the only modification is that instead of sorting the cycles in purely increasing order, we use an ordering of $[n]$ such that all elements of the subset are ordered after all other elements. This completes the proof. □

Our strategy for writing down explicit eigenvectors of K will combine these two lumpings by first lumping down to the binary Burnside process for smaller n and then studying those eigenvectors.

4 Eigenvalues and eigenvectors; proof of Theorem 1.2

The determination of the eigenvalues and eigenvectors of the binary Burnside process on C_2^n depends on the explicit diagonalization of the orbit chain K^{lumped} on $\{0, 1, \cdots, n\}$. To be precise, this is the chain which follows the dynamics of the binary Burnside process but then only records the orbit $\mathcal{O}_i$ of the state (which we identify with the integer i). An explicit form for $K^{\text{lumped}}(i, j)$ (which is not needed here) is found in [Dia05, Eqn.(3.1)-(3.3)]. The diagonalization we require is recorded here:

Proposition 4.1 ([DZ21], Theorem 2)**.** *The eigenvectors of the Markov chain K_n^{lumped} are the discrete Chebyshev polynomials T_n^m on $\{0, 1, \cdots, n\}$. The nonzero eigenvalues are $\beta_k = \frac{1}{2^{4k}}\binom{2k}{k}^2$ for $k = 0, 1, \cdots, \lfloor \frac{n}{2} \rfloor$, corresponding to the even-degree Chebyshev polynomials T_n^{2k}, respectively. All remaining eigenvalues are zero, corresponding to the odd-degree Chebyshev polynomials T_n^m for odd $m \leq n$.*

Here, the discrete Chebyshev polynomials are the orthogonal polynomials for the uniform distribution, where T_n^m is the polynomial of degree m (viewed as a vector by evaluating at the points $\{0, 1, \cdots, n\}$). This result is enough to prove ([DZ21], Theorem 1) that for the starting states $\underline{0}$ or $\underline{1}$, a finite number of steps are necessary and sufficient for convergence. However, if we begin the binary Burnside chain at any other starting point, the starting distribution is not uniform within cycles and thus it may take more steps before convergence to stationarity than in the orbit chain. Thus, we would like to find the eigenvectors and eigenvalues of the Markov chain on the full state space.

Our first step for diagonalizing the full chain is to write out the highest-degree Chebyshev polynomial in an explicit form:

Proposition 4.2. *The discrete Chebyshev polynomial T_n^n on $\{0, 1, \cdots, n\}$ of degree n satisfies $T_n^n(i) = (-1)^i\binom{n}{i}$ for all $i \in \{0, 1, \cdots, n\}$.*

Proof. The discrete Chebyshev polynomials satisfy the recurrence relation (plugging in $\alpha = \beta = 0$ into [Ism05, Eq. (6.2.8)])

$$T_n^0(x) = 1, \quad T_n^1(x) = \frac{n-2x}{n}, \quad (j+1)(n-j)T_n^{j+1}(x) = (2j+1)(n-2x)T_n^j(x) - j(j+n+1)T_n^{j-1}(x)$$

for all $1 \leq j \leq n-1$. Since $(j+1)(n-j) = (2j+1)n - j(j+n+1)$, the constant term of T_n^j is 1 for any j. Also, since the T_n^ms are orthogonal with respect to the uniform distribution on $\{0, 1, \cdots, n\}$, we have

$\sum_{i=0}^{n} T_n^n(i) T_n^m(i) = 0$ for all $m < n$ and therefore

$$\sum_{i=0}^{n} T_n^n(i) f(i) = 0 \quad \text{for all polynomials } f \text{ of degree at most } (n-1).$$

But for any such f, the nth finite difference of f started at 0 is exactly $\binom{n}{0} f(0) - \binom{n}{1} f(1) + \binom{n}{2} f(2) - \cdots$, and so setting $T_n^n(i) = (-1)^i \binom{n}{i}$ satisfies orthogonality and also that the constant term is 1. Since T_n^n is uniquely defined by the recurrence relation (and we can perform Lagrange interpolation through these $(n+1)$ points to get a polynomial of degree at most n), this must be the desired polynomial. □

We now use this explicit formula to construct explicit expressions for our eigenvectors.

Proof of Theorem 1.2. Let $m = 2k$ be any even integer less than or equal to n. First, we claim that there is a (right) eigenvector of eigenvalue β_k for the binary Burnside process on C_2^m of the form

$$g_m(x) = T_m^m(|x|) = (-1)^{|x|} \binom{m}{|x|}$$

(last equality by Proposition 4.2). Indeed, for any x, we have (recall that K_m denotes the transition matrix for the process on C_2^m)

$$\begin{aligned} K_m g_m(x) &= \sum_{y \in C_2^m} g_m(y) K_m(x, y) \\ &= \sum_{i=0}^{m} T_m^m(i) \sum_{y : |y| = i} K_m(x, y) \\ &= \sum_{i=0}^{m} T_m^m(i) K_m^{\text{lumped}}(|x|, i), \end{aligned} \tag{4.1}$$

and by the eigenvalue equation on the lumped chain, this last expression is $\beta_k T_m^m(|x|) = \beta_k g_m(x)$, as desired.

Next, choosing any size-m subset $S \in \binom{[n]}{m}$ of the coordinates, we can lift g_m to a function on C_2^n by defining

$$f_S(x) = g_m(x_S) = (-1)^{|x_S|} \binom{m}{|x_S|},$$

where x_S is the restriction of x to the coordinate set S. We claim this is a (right) eigenvector of K_n. Indeed, for any $x \in C_2^n$, we have

$$\begin{aligned} K_n f_S(x) &= \sum_{y \in C_2^n} f_S(y) K_n(x, y) \\ &= \sum_{y_S \in C_2^m} \sum_{y_{\overline{S}} \in C_2^{n-m}} (-1)^{|y_S|} \binom{m}{|y_S|} K_n(x, y) \\ &= \sum_{y_S \in C_2^m} (-1)^{|y_S|} \binom{m}{|y_S|} K_m(x_S, y_S) \end{aligned}$$

since by Proposition 3.3 K_n lumps to K_m when only restricted to the coordinate set S. And now this last expression is exactly $\sum_{y_S} g_m(y_S)K_m(x_S, y_S)$, so by the eigenvalue equation it evaluates to $\beta_k g_m(x_S) = \beta_k f_S(x)$, as desired.

This means that for each of the $\binom{n}{m}$ subsets S, we get an eigenvector f_S for K_n of eigenvalue β_k. Now we prove that $\{f_S : S \in \binom{[n]}{m}\}$ is a linearly independent set for each fixed m via the following steps:

1. *Rephrasing the problem.* The $\binom{n}{m}$ eigenvectors f_S may be written in a $\binom{n}{m}$ by 2^n matrix M, so that the rows are indexed by subsets $S \in \binom{[n]}{m}$ and the columns are indexed by states $x \in C_2^n$. In other words, define
$$M = (M_{S,x})_{S\in\binom{[n]}{m}, x\in C_2^n}, \quad M_{S,x} = (-1)^{|x_S|}\binom{m}{|x_S|},$$
where $|x_S|$ again denotes the number of ones in x among the coordinates S. We then wish to prove that the rows (eigenvectors) of M are linearly independent, or equivalently that the matrix has full column rank.

2. *Using symmetry.* To prove that the matrix has full column rank, it suffices to show that some linear combination of the column vectors is 1 in the entry for the subset $S = \{1, 2, \cdots, m\}$ and 0 in all others. (Then by permuting the role of $1, 2, \cdots, n$ in the coefficients of those vectors, we can get all other subsets S as well.) Fixing notation, let v_x be the column vector of M corresponding to the state $x \in C_2^n$.

3. *Constructing nice linear combinations.* Let $d = \min(m, n-m)$. For all integers $0 \le a, b \le d$, consider the linear combinations of vectors
$$\mathbf{v}_a = \sum_{\substack{\text{states } x \text{ with } a \text{ ones in the first}\\ m \text{ coordinates and no other ones}}} v_x$$
and the sets of subsets
$$\mathcal{S}_b = \left\{ S \in \binom{[n]}{m} : \Big|S \cap \{m+1, \cdots, n\}\Big| = b \right\}.$$
We have $\binom{[n]}{m} = \bigcup_{b=0}^d \mathcal{S}_b$. Now we claim each $\mathbf{v}_a$ is constant on each $\mathcal{S}_b$ and that the value it takes on any $S \in \mathcal{S}_b$ is $n_{ab} = \sum_{r=0}^a (-1)^r \binom{m}{r}\binom{m-b}{r}\binom{b}{a-r}$. Indeed, among all states x with a ones in the first m coordinates (and no other ones), there are $\binom{m-b}{r}\binom{b}{a-r}$ of them with $|x_S| = r$ (because $m-b$ of the numbers $\{1, \cdots, m\}$ are in each $S \in \mathcal{S}_b$), and the value of v_x at S for each of them is $(-1)^r\binom{m}{r}$.
Therefore, each of $\mathbf{v}_0, \cdots, \mathbf{v}_d$ is encoded by its $(d+1)$ values on $\mathcal{S}_0, \cdots, \mathcal{S}_d$, and we would like to show that some linear combination of $\mathbf{v}_0, \cdots, \mathbf{v}_d$ yields 1 on $\mathcal{S}_0 = \{\{1, \cdots, m\}\}$ and 0 on all other $\mathcal{S}_j$s. To do this, it suffices to prove that the $(d+1) \times (d+1)$ matrix
$$N = (n_{ab})_{a,b=0}^d, \quad n_{ab} = \sum_{r=0}^a (-1)^r \binom{m}{r}\binom{m-b}{r}\binom{b}{a-r}$$
is invertible.

4. *Making use of the binomial coefficients.* We now show that row a of the matrix N is a degree a polynomial; more precisely, there is some degree a polynomial f_a such that $n_{ab} = f_a(b)$ for all b. Indeed, for any fixed a, $\binom{m-b}{r}$ is a polynomial of degree r in b, and $\binom{b}{a-r}$ is a polynomial of degree $(a-r)$ in b, so their product is a polynomial of degree a and thus the whole expression is a polynomial. Furthermore, summing over all r, the total coefficient of b^a in n_{ab} is

$$c_a = \sum_{r=0}^{a} (-1)^r \binom{m}{r} \frac{(-1)^r}{r!} \frac{1}{(a-r)!} = \frac{1}{a!} \sum_{r=0}^{a} \binom{m}{r}\binom{a}{a-r} = \frac{1}{a!}\binom{m+a}{a};$$

in particular, it is nonzero, so the polynomial is indeed of degree a. Therefore by row reduction (subtracting off earlier rows from later ones to remove lower-order terms), the determinant of N is equal to the determinant of the matrix $(c_a b^a)_{a,b=0}^d$, which is a nonzero constant times a nonzero Vandermonde determinant, hence nonzero. This proves that we can indeed find a valid linear combination to get the desired column vector, completing the proof of linear independence.

Thus, we have constructed a $\binom{n}{m}$-dimensional eigenspace of eigenvalue β_k for any $0 \le k \le \frac{n}{2}$. Since eigenvectors of different eigenvalues are orthogonal, the span of these eigenspaces has dimension $\sum_{m \text{ even}} \binom{n}{m} = 2^{n-1}$. Furthermore, by construction, notice that any function f in this span satisfies $f(x) = f(\overline{x})$ for all $x \in C_2^n$, where $\overline{x}$ flips each coordinate of x from 0 to 1 or vice versa. This means that all of these eigenvectors are orthogonal to the 2^{n-1}-dimensional space of functions

$$\left\{h \in L(C_2^n) : h(x) = -h(\overline{x}) \text{ for all } x \in C_2^n\right\}.$$

Any such h is an eigenvector of K of eigenvalue zero, because using that $K(x,y) = K(x,\overline{y})$ for any x, y, we have

$$Kh(x) = \sum_{y \in C_2^n} h(y)K(x,y) = \sum_{y \in C_2^n} h(\overline{y})K(x,\overline{y}) = -\sum_{y \in C_2^n} h(y)K(x,y) = -Kh(x) \implies Kh(x) = 0.$$

So the sum of the dimensions of the constructed spaces is $2^{n-1} + 2^{n-1} = 2^n$, meaning we have constructed a basis of C_2^n of eigenvectors with the stated multiplicities, completing the proof. □

As mentioned in the introduction, if we would like to obtain exact expressions for chi-square distance from arbitrary starting points, we would need explicit expressions for an orthonormal basis of eigenvectors. However, the eigenvectors f_S in our proof of Theorem 1.2 are not orthogonal, and in fact we can describe their inner products. First of all, we have $\langle f_S, f_{S'} \rangle = 0$ for any two sets S and S' with $|S| \ne |S'|$, but this does not hold if $|S| = |S'|$. In the latter case, denote $|S| = |S'| = m$ and $|S \cap S'| = \ell$. Under the inner product $\langle f, g \rangle = \sum_x f(x)g(x)\pi(x)$ with $\pi(x) = \frac{1}{n+1}\frac{1}{\binom{n}{|x|}}$, we have

$$\begin{aligned}\langle f_S, f_{S'} \rangle &= \sum_x \pi(x) f_S(x) f'_S(x) \\ &= \sum_{a=0}^{\ell} \sum_{b=0}^{m-\ell} \sum_{c=0}^{m-\ell} \sum_{d=0}^{n-2m+\ell} \binom{\ell}{a}\binom{m-\ell}{b}\binom{m-\ell}{c}\binom{n-2m+\ell}{d} \frac{1}{n+1} \frac{1}{\binom{n}{a+b+c+d}} (-1)^{b+c} \binom{m}{a+b}\binom{m}{a+c},\end{aligned} \tag{4.2}$$

where a, b, c, d respectively represent the number of coordinates among the sets $S \cap S'$, $S \setminus S'$, $S' \setminus S$, and $(S \cup S')^c$ where x has a 1.

Proposition 4.3. *For any even integer m, any $n \geq m$, and any two subsets $S, S' \subseteq \binom{[n]}{m}$ with $|S \cap S'| = \ell$, let $\overline{f_S} = \frac{f_S}{\langle f_S, f_S \rangle^{1/2}}$ and $\overline{f_{S'}} = \frac{f'_S}{\langle f'_S, f'_S \rangle^{1/2}}$ be the $L^2(\pi)$-normalized vectors for S and S'. Then*

$$\langle \overline{f_S}, \overline{f_{S'}} \rangle = \frac{1}{\binom{2m+1-\ell}{m+1}}.$$

Furthermore, for any even integer $m \leq n$ and for any $S \subseteq \binom{[n]}{m}$, we have the normalizing factor

$$\langle f_S, f_S \rangle = \frac{1}{m+1}\binom{2m}{m}.$$

We defer the proof (due to Laurent Bartholdi and Christoph Koutschan, using "creative telescoping") to Appendix A. Given the nice form of these inner products, it is natural to hope for closed-form expressions for linear combinations of the f_Ss which indeed yield an orthonormal basis. Section 6 describes a procedure for doing this, and this procedure actually allows us to obtain eigenvector bases for irreducible subspaces of the permutation representation (as discussed shortly before Proposition 3.3).

The Gram matrix of Proposition 4.3 is so neat that it is natural to try to understand it directly. Richard Stanley points out that matrices indexed by the size-m subsets of $[n]$ with (S, S') entry given by $g(|S \cap S'|)$ (for any function g) are elements of the adjacency algebra of the Johnson scheme. Such matrices can be explicitly diagonalized with dual Hahn polynomial eigenfunctions. A clear, elementary account can be found in [Bur17]; in particular, Theorem 1.1 in that paper shows that the eigenvalues of our Gram matrix are the rational numbers

$$\lambda_t = \sum_{\ell=0}^{m} g(\ell) \sum_{i=0}^{t} (-1)^{t-i} \binom{m-i}{\ell-i}\binom{n-m+i-t}{m-\ell+i-t}\binom{t}{i}, \quad 0 \leq t \leq m,$$

where $g(\ell) = \frac{1}{\binom{2m+1-\ell}{m+1}}$ are our inner products, and where the λ_t-eigenspace has dimension $\binom{n}{t} - \binom{n}{t-1}$. Thus from here, a different approach to orthonormalization is possible: if $\{v_i\}$ are independent column vectors with Gram matrix $G = (\langle v_i, v_j \rangle)_{ij}$, then an orthonormal basis is given by $u_i = \sum_j (G_{ji})^{-1/2} v_j$. (In matrix notation, we have $U = VG^{-1/2}$ where U, V have columns $\{u_i\}$ and $\{v_i\}$.) However, the presence of $G^{-1/2}$ suggests a messy calculation and thus we have not tried to push this further.

5 Proofs of results for chi-square distance

In this section, we compute the average chi-square distance and show that order $\frac{n}{\log n}$ steps are required to be close to stationarity. The argument uses properties of $\chi^2_{\text{avg}}(\ell)$, which we recall here.

As mentioned in the introduction, the average chi-square distance to stationarity after ℓ steps is given by

$$\chi^2_{\text{avg}}(\ell) = \sum_{x \in \mathfrak{X}} \pi(x)\chi^2_x(\ell) = \sum_{x,y \in \mathfrak{X}} \left| \frac{K^\ell(x,y)}{\pi(y)} - 1 \right|^2 \pi(x)\pi(y).$$

We made use of eigenvalues and eigenvectors in Section 2.1 to rewrite chi-square distance in terms of eigenvectors. Again letting $f_0, \cdots, f_{|\mathfrak{X}|-1}$ be an orthonormal basis of eigenvectors for $\ell^2(\pi)$, with f_0 corresponding to the trivial eigenvalue of 1, we have

$$\begin{aligned}\chi^2_{\text{avg}}(\ell) &= \sum_{x\in\mathfrak{X}} \pi(x) \sum_{i=1}^{|\mathfrak{X}|-1} f_i(x)^2 \beta_i^{2\ell} \\ &= \sum_{i=1}^{|\mathfrak{X}|-1} \beta_i^{2\ell} \sum_{x\in\mathfrak{X}} \pi(x) f(x)^2 \\ &= \sum_{i=1}^{|\mathfrak{X}|-1} \beta_i^{2\ell}. \qquad (5.1)\end{aligned}$$

Thus, we may plug in our known eigenvalues and multiplicities directly and obtain estimates for various ℓ. We will use this to first prove all parts of Theorem 1.1 except "slow mixing for the half-zeros, half-ones state" (which requires a different kind of argument).

Proof of Theorem 1.1, except the latter part of statement (2). For (1), we use the eigenvalue multiplicities from Theorem 1.2, along with Eq. (5.1), to get

$$\chi^2_{\text{avg}}(\ell) = \sum_{k=1}^{\lfloor n/2\rfloor} \binom{n}{2k} \beta_k^{2\ell},$$

where $\beta_k = \frac{1}{2^{4k}}\binom{2k}{k}^2$ as in Proposition 4.1. We will lower and upper bound this quantity to show that order $\frac{n}{\log n}$ steps are necessary and sufficient.

To prove (2), we establish a lower bound for $\chi^2_{\text{avg}}(\ell)$ when $\ell \le \frac{0.1n}{\log n}$. We will just consider $n = 4a$ a multiple of 4 for simplicity of notation; the general case can be handled analogously but for instance using one of the middle two binomial coefficients instead of the central binomial coefficient.

We lower bound by the single term $k = \frac{n}{4}$ of the sum, yielding

$$\begin{aligned}\chi^2_{\text{avg}}(\ell) &\ge \binom{4a}{2a}\beta_a^{2\ell} \\ &= \binom{4a}{2a}\left(\frac{1}{2^{4a}}\binom{2a}{a}^2\right)^{2\ell}.\end{aligned}$$

We now use bounds on the central binomial coefficient ([Stă01], Theorem 2.5)

$$\frac{4^a}{\sqrt{\pi a}}\exp\left(-\frac{1}{8a}\right) < \binom{2a}{a} < \frac{4^a}{\sqrt{\pi a}}$$

(which hold for all $a \ge 1$) to get

$$\begin{aligned}\chi^2_{\text{avg}}(\ell) &\ge \binom{4a}{2a}\left(\frac{1}{\pi a}\exp\left(-\frac{1}{4a}\right)\right)^{2\ell} \\ &\ge \frac{2^{4a}}{\sqrt{2\pi a}}\exp\left(-\frac{1}{16a}\right)\left(\frac{1}{\pi a}\exp\left(-\frac{1}{4a}\right)\right)^{2\ell} \\ &= \frac{2^n}{\sqrt{\pi n/2}}\exp\left(-\frac{1}{4n}\right)\left(\frac{4}{\pi n}\exp\left(-\frac{1}{n}\right)\right)^{2\ell}.\end{aligned}$$

In particular, this last quantity is growing exponentially for large n when $\ell = \frac{0.1n}{\log n}$, since $\left(\frac{1}{n}\right)^{2\ell} = e^{-0.2n}$ has less significant exponential decay than the factor of 2^n, and all other terms are of lower order.

Finally, to prove (3), we establish an upper bound on $\chi^2_{\text{avg}}(\ell)$. For this, we include all terms in the sum and use the upper bound on the central binomial coefficient. This time assuming $n = 10b$ for simplicity of notation, we get

$$\begin{aligned}\chi^2_{\text{avg}}(\ell) &\le \sum_{k=1}^{n/2} \binom{n}{2k} (\pi k)^{-\ell} \\ &= \sum_{k=1}^{b} \binom{n}{2k} (\pi k)^{-\ell} + \sum_{k=b+1}^{5b} \binom{n}{2k} (\pi k)^{-\ell} \\ &\le \sum_{k=1}^{b} \binom{n}{2k} (\pi k)^{-\ell} + 2^n (\pi b)^{-\ell}.\end{aligned}$$

Now take $\ell = \frac{10n}{\log n}$. The last term $2^n(\pi b)^{-\ell}$ decays exponentially in n, so we just need to bound the first sum. For this, note that the ratio of consecutive terms (using that $k \le 0.1n$) can be bounded as

$$\begin{aligned}\frac{\binom{n}{2(k+1)}(\pi(k+1))^{-\ell}}{\binom{n}{2k}(\pi k)^{-\ell}} &= \frac{(n-2k)(n-2k-1)}{(2k+1)(2k+2)} \left(1 + \frac{1}{k}\right)^{-\ell} \\ &\le \frac{n^2}{12} \left(1 + \frac{1}{0.1n}\right)^{-10n/\log n} \\ &\le \frac{n^2}{12} \left(\frac{1}{2}\right)^{-n/\log n}\end{aligned}$$

for all $n \ge 10$, and this quantity is bounded uniformly by 0.5. Thus this first sum is upper bounded by an infinite geometric series with first term $\binom{n}{2}(\pi)^{-\ell}$ and common ratio $\frac{1}{2}$, which also decays exponentially in n. Thus the average chi-square distance is decaying for large n when $\ell = \frac{10n}{\log n}$, as desired.

We now wish to upgrade this result to a uniform bound on $\chi^2_x(\ell)$. For this, observe that because both K and π are invariant under the action of S_n, $\chi^2_x(\ell)$ is constant on each orbit $\mathcal{O}_i$, and $\sum_{x \in \mathcal{O}_i} \pi(x) = \frac{1}{n+1}$ for all orbits. Therefore we may write

$$\chi^2_{\text{avg}}(\ell) = \sum_x \pi(x) \chi^2_x(\ell) = \frac{1}{n+1} \sum_{i=0}^{n} \chi^2_{\mathcal{O}_i}(\ell),$$

where $\chi^2_{\mathcal{O}_i}(\ell)$ is the chi-square distance after ℓ steps when started from any state in the orbit $\mathcal{O}_i$. So to prove that $\chi^2_{\mathcal{O}_i}(\ell) \to 0$ for all i, it suffices to prove that $(n+1)\chi^2_{\text{avg}}(\ell) \to 0$. But this is already implied by the calculation above, which in fact showed exponential decay of $\chi^2_{\text{avg}}(\ell)$ in n. This concludes the proof. □

It remains now to show that the "half-zeros, half-ones" state takes a long time to reach stationarity in ℓ^2. For additional context, recall from Example 2.3 that states with all but a constant number of 0s or 1s have ℓ^2 mixing in just $\log n$ steps, and so it is natural to ask whether we can also exhibit explicit states with slow ℓ^2 mixing. The following argument, using an idea of Lucas Teyssier, shows that states with a positive fraction of both 0s and 1s do in fact require $\frac{n}{\log n}$ steps to converge.

Proposition 5.1. *Let $x^{(n)} \in C_2^n$ be any sequence of states such that $cn \le |x^{(n)}| \le (1-c)n$ for some $c \in (0,1)$. Then $\chi^2_{x^{(n)}}(\ell) \to \infty$ for $\ell = \Theta_c\left(\frac{n}{\log n}\right)$.*

Proof. We may bound the chi-square distance starting at $x^{(n)}$ using just the term for $x^{(n)}$ itself:

$$\chi^2_{x^{(n)}}(\ell) = \sum_{y \in C_2^n} \left| \frac{K^\ell(x^{(n)}, y)}{\pi(y)} - 1 \right|^2 \pi(y) \ge \left| \frac{K^\ell(x^{(n)}, x^{(n)})}{\pi(x^{(n)})} - 1 \right|^2 \pi(x^{(n)})$$

$$\ge \left| \frac{K\left(x^{(n)}, x^{(n)}\right)^\ell}{\pi(x^{(n)})} - 1 \right|^2 \pi(x^{(n)}).$$

To bound the probability $K(x^{(n)}, x^{(n)})$ from below, recall that in the binary Burnside process, we first pick a uniformly random permutation which fixes $x^{(n)}$ – that is, a permutation among the coordinates in which $x^{(n)}$ is 1, along with an independent permutation among the coordinates in which $x^{(n)}$ is 0. The probability that each of those are just a single cycle is $\frac{1}{|x^{(n)}|} \cdot \frac{1}{n-|x^{(n)}|}$, and the subsequent probability of labeling the cycles with 1 and 0 respectively (hence returning back to $x^{(n)}$ after one step) is $\frac{1}{4}$. Since $cn \le |x^{(n)}| \le (1-c)n$, we therefore have $K(x^{(n)}, x^{(n)}) = \Omega_c(n^{-2})$.

But $\frac{1}{\pi(x^{(n)})} = (n+1)\binom{n}{|x^{(n)}|}$ grows exponentially in n (with constant depending on c), so it takes many steps for $K\left(x^{(n)}, x^{(n)}\right)^\ell$ to get small enough to be of comparable order to $\pi(x^{(n)})$. More precisely, we have $K(x^{(n)}, x^{(n)})^\ell > \pi(x^{(n)})^{1/3}$ for $\ell = \Theta_c\left(\frac{n}{\log n}\right)$, and thus after this many steps $\chi^2_{x^{(n)}}(\ell)$ is still exponentially growing in n, as desired. □

As a corollary, we obtain the final part of our main result by being more careful with the constants in the argument above:

Proof of Theorem 1.1, latter part of (2). Recall that we are considering the state $x^{(n)} \in C_2^n$ with $\lfloor \frac{n}{2} \rfloor$ zeros followed by $\lceil \frac{n}{2} \rceil$ ones. Following the proof of Proposition 5.1 above, we again have

$$\chi^2_{x^{(n)}}(\ell) \ge \left| \frac{K\left(x^{(n)}, x^{(n)}\right)^\ell}{\pi(x^{(n)})} - 1 \right|^2 \pi(x^{(n)})$$

but this time with the explicit bounds $K(x^{(n)}, x^{(n)}) \ge \frac{1}{4} \cdot \frac{1}{\lfloor \frac{n}{2} \rfloor} \cdot \frac{1}{\lceil \frac{n}{2} \rceil} \ge \frac{1}{n^2}$ and $\pi(x^{(n)}) = \frac{1}{(n+1)\binom{n}{\lfloor n/2 \rfloor}} < \frac{n}{2^n}$ for all positive integers n. So if $\ell \le \frac{0.1n}{\log n}$, we have $K\left(x^{(n)}, x^{(n)}\right)^\ell \ge e^{-0.2n} \ge 0.8^n$. Plugging in these bounds (and also using that $\pi(x^{(n)}) \ge \frac{1}{(n+1)2^n}$),

$$\chi^2_{x^{(n)}}(\ell) \ge \left| \frac{2^n \cdot 0.8^n}{n} - 1 \right|^2 \cdot \frac{1}{(n+1)2^n},$$

and this right-hand side indeed goes to infinity as $n \to \infty$. □

The results above can be supplemented with asymptotics. The following shows that the binary Burnside process has a cutoff in ℓ_2 at $\ell = \frac{\log 2}{2} \frac{n}{\log\left(\frac{\pi}{2} n\right)}$:

Theorem 5.2. *For the binary Burnside process on C_2^n, we have the following for all $\varepsilon > 0$ as $n \to \infty$:*

1. *If $\ell \le (1-\varepsilon)\frac{\log 2}{2}\frac{n}{\log n}$, then $\chi^2_{avg}(\ell) \to \infty$.*
2. *If $\ell \ge (1+\varepsilon)\frac{\log 2}{2}\frac{n}{\log n}$, then $\chi^2_{avg}(\ell) \to 0$.*

The proof uses standard asymptotics for the symmetric binomial distribution, as above and in [Fel68, Chapter VII]; we postpone the proof to Appendix B. We also point out that (using the same reasoning as in the end of the proof of Theorem 1.1) this implies that some starting state x has $\chi^2_x(\ell) \to \infty$ even after $(1-\varepsilon)\frac{\log 2}{2}\frac{n}{\log n}$ steps, while all starting states have $\chi^2_x(\ell) \to 0$ after just $(1+\varepsilon)\frac{\log 2}{2}\frac{n}{\log n}$ steps. In particular, this is a better lower bound than what we obtain for the half-zeros, half-ones state in Proposition 5.1, which only works in its current form up until $\frac{\log 2}{4}\frac{n}{\log n}$ steps.

6 An orthonormal eigenvector basis via Schur–Weyl duality

We now describe further the representation theory features mentioned after Corollary 3.2. For relevant references, see [Dia88] (particularly Section 7 for material on the symmetric group) or [Jam78]. The key idea is that the relation $K_n(x,y) = K_n(\sigma(x),\sigma(y))$ implies that K_n, as an operator on the space of functions on C_2^n, maps only within isomorphic irreducible subspaces of the permutation representation, and Schur–Weyl duality allows us to upgrade this to describe an orthonormal basis of eigenvectors.

The key results of this section are as follows. Theorem 6.2 describes a complete set of eigenvectors $\{f_Q^{m,\ell}\}$ indexed by integers m,ℓ (which dictate the eigenvalue) and Young tableaux Q of shape $(n-m,m)$. These $f_Q^{m,\ell}$ are orthogonal – we provide an explicit closed-form formula for their norms in Corollary 6.5 – and they may be thought of as particular linear combinations of the f_S eigenvectors from Theorem 1.2. In particular, we obtain Corollary 6.6, which refines Theorem 1.2 by describing the eigenvalue multiplicities when restricting K_n to copies of each irreducible representation. Finally, Theorem 6.7 provides a probabilistic application of this basis, showing that a bounded number of steps also suffices from the "all-but-one zeros" state by evaluating the $f_Q^{m,\ell}$s at this state (and in fact finding that almost all of them are zero, in stark contrast with our original basis $\{f_S\}$).

Note that in this section only, we make some significant notational changes to make the formulas and proofs easier to parse. In particular, we will view our function space $L(C_2^n)$ as a tensor product $V^{\otimes n}$ over the n coordinates, and we write all functions as linear combinations of the basis vectors v_S.

6.1 Construction of orthonormal eigenvectors

In this subsection, we describe how to construct orthogonal subspaces indexed by Young tableaux and obtain explicit formulas for the vectors that lie in those subspaces. In particular, the orthogonal decomposition here does not depend on our operator K, so it may be useful for other Markov chains on binary n-tuples that are S_n-invariant (see Section 7).

As mentioned above, we first establish new notation. Let $V = \mathbb{C}\text{-span}\{v_0, v_1\}$, so that the tensor space

$$V^{\otimes n} \quad \text{has basis} \quad \{v_{i_1} \otimes \cdots \otimes v_{i_n} \mid i_1, \ldots, i_n \in \{0,1\}\}.$$

For each subset $S \subseteq \{1,\ldots,n\}$, we write

$$v_S = v_{i_1} \otimes \cdots \otimes v_{i_n}, \quad \text{where } i_\ell = \begin{cases} 1 & \text{if } \ell \in S, \\ 0 & \text{if } \ell \notin S. \end{cases}$$

Define the subspaces corresponding to the orbits $\mathcal{O}_\ell = \{x \in C_2^n : |x| = \ell\}$

$$V^{(\ell)} = \mathbb{C}\text{-span}\{v_S \mid |S| = \ell\}, \quad \text{so that} \quad V^{\otimes n} = \bigoplus_{\ell=0}^{n} V^{(\ell)}.$$

Our inner product on $V^{\otimes n}$ can then be written in the form

$$\langle v_S, v_T \rangle_\pi = \frac{1}{n+1} \frac{1}{\binom{n}{|S|}} \delta_{ST}, \quad \text{where in particular } V^{(j)} \perp V^{(\ell)} \text{ if } j \neq \ell. \tag{6.1}$$

In addition to the orthogonal subspaces $V^{(\ell)}$, we may also define a set of different orthogonal subspaces indexed by standard Young tableaux. For this, recall that S_n acts by permutation on the coordinates, and so the group algebra of the symmetric group $\mathbb{C}[S_n]$ acts on $V^{\otimes n}$ by

$$wv_S = v_{wS} \quad \text{for } w \in S_n \text{ and } S \subseteq \{1, \dots, n\}.$$

For $r \in \{2, \dots, n\}$, the *Jucys-Murphy elements* in the group algebra of the symmetric group (see [Ram95, Eq. (3.5)], or the report [DG89], or Murphy's original paper [Mur81]) are the mutually commuting operators

$$M_r = \sum_{i=1}^{r-1} s_{ir}, \quad \text{where } s_{ir} \text{ is the transposition that switches } i \text{ and } r. \tag{6.2}$$

Let $\hat{S}_n^{(n-m,m)}$ denote the set of standard Young tableaux of shape $(n-m, m)$. For $Q \in \hat{S}_n^{(n-m,m)}$, let $Q(r)$ denote the box containing r in Q. Define the *content* of the box via

$$\text{ct}(Q(r)) = y - x \text{ if } Q(r) \text{ is in row } x \text{ and column } y,$$

and also define the subspace of simultaneous eigenvectors

$$V_Q = \{m \in V^{\otimes n} \mid M_r m = \text{ct}(Q(r))m \text{ for all } r \in \{2, \dots, n\}\}.$$

These subspaces are mutually orthogonal for the following reason. For any transposition $w \in S_n$, we have

$$\langle wv_S, wv_T \rangle_\pi = \langle v_S, v_T \rangle_\pi \implies \langle wv_S, v_T \rangle_\pi = \langle v_S, wv_T \rangle_\pi$$

because $w^2 = 1$. Thus any linear combination of transpositions in $\mathbb{C}[S_n]$ is self-adjoint, meaning in particular that all M_j are self-adjoint. Now let $P \neq Q$ be any standard tableaux with n boxes; there must exist some $i \in \{1, \dots, n\}$ such that $\text{ct}(P(i)) \neq \text{ct}(Q(i))$. Then we have for any $p \in V_P$ and $q \in V_Q$ that

$$\text{ct}(P(i))\langle p, q \rangle_\pi = \langle M_i p, q \rangle_\pi = \langle p, M_i q \rangle_\pi = \text{ct}(Q(i))\langle p, q \rangle_\pi,$$

meaning that $\langle p, q \rangle_\pi = 0$. This means that for any Young tableaux P, Q,

$$\text{if } P \neq Q, \text{ then } V_P \perp V_Q. \tag{6.3}$$

We thus obtain a refined decomposition of $V^{\otimes n}$ by defining, for all $\ell \in \{0, 1, \dots, n-m\}$ and $Q \in \hat{S}_n^{(n-m,m)}$, the subspace

$$V_Q^{(\ell)} = V_Q \cap V^{(\ell)}.$$

Combining Eq. (6.1) and Eq. (6.3) gives

$$V_P^{(j)} \perp V_Q^{(\ell)} \text{ unless } P = Q \text{ and } j = \ell. \tag{6.4}$$

By Schur-Weyl duality, as well as the representation theory of $\mathfrak{sl}_2$ and of the symmetric group S_n (see [FH91, Ex. 6.30 and (11.6)] and [Ram95, (3.5), (3.11), and Thm. 3.14]), we have

$$\dim(V_Q^{(\ell)}) = 1 \quad \text{and} \quad V^{\otimes n} = \bigoplus_{m=0}^{\lfloor n/2 \rfloor} \bigoplus_{Q \in \hat{S}_n^{(n-m,m)}} \bigoplus_{\ell=0}^{n-m} V_Q^{(\ell)}. \tag{6.5}$$

Furthermore, it is a consequence of Schur-Weyl duality that

$$\begin{aligned}
&\text{each } V_Q = \bigoplus_{\ell=0}^{n-m} V_Q^{(\ell)} \text{ is an irreducible } \mathfrak{sl}_2\text{-invariant subspace of } V^{\otimes n}, \text{ and} \\
&\text{each } V_{(n-m,m)}^{(\ell)} = \bigoplus_{Q \in \hat{S}_n^{(n-m,m)}} V_Q^{(\ell)} \text{ is an irreducible } S_n\text{-invariant subspace of } V^{\otimes n}.
\end{aligned} \tag{6.6}$$

We will elaborate more on this decomposition in Section 6.2.

We now describe each of the one-dimensional subspaces $V_Q^{(\ell)}$ explicitly. First, we show a useful calculation:

Lemma 6.1. *Viewing all elements of $\mathbb{C}[S_n]$ as operators on $V^{\otimes n}$, define*

$$\tau_j = s_j + \frac{1}{M_j - M_{j+1}} \quad \text{for } j \in \{1, \dots, n-1\}. \tag{6.7}$$

(Here $\frac{1}{M_j - M_{j+1}}$ denotes the inverse of the operator $M_j - M_{j+1}$, which indeed exists because M_j, M_{j+1} are simultaneously diagonalizable with distinct nonzero eigenvalues.) Then we have the relations

$$\tau_j M_j = M_{j+1}\tau_j, \quad M_j\tau_j = \tau_j M_{j+1}, \quad \text{and} \quad \tau_j^2 = \frac{(M_j - M_{j+1} + 1)(M_j - M_{j+1} - 1)}{(M_j - M_{j+1})^2}. \tag{6.8}$$

Proof. We have $M_{j+1} = s_j M_j s_j + s_j$ for all j, so that $M_{j+1}s_j = s_j M_j + 1$. Therefore, using that M_j and M_{j+1} commute,

$$\begin{aligned}
M_{j+1}\tau_j &= M_{j+1}\Big(s_j + \frac{1}{M_j - M_{j+1}}\Big) \\
&= M_{j+1}s_j + \frac{M_{j+1}}{M_j - M_{j+1}} \\
&= s_j M_j + 1 + \frac{M_{j+1}}{M_j - M_{j+1}} \\
&= s_j M_j + \frac{M_j}{M_j - M_{j+1}} \\
&= \left(s_j + \frac{1}{M_j - M_{j+1}}\right) M_j = \tau_j M_j,
\end{aligned}$$

proving the first equality. The second equality follows by an identical argument except with all multiplications in the reverse order. Putting those two facts together yields

$$(M_{j+1} - M_j)\tau_j = \tau_j M_j - M_j\tau_j = \tau_j(M_j - M_{j+1}).$$

Therefore we have $\tau_j(\frac{1}{M_j - M_{j+1}}) = (\frac{1}{M_{j+1} - M_j})\tau_j$, which shows that

$$\begin{aligned}
\tau_j^2 &= \tau_j\Big(s_j + \frac{1}{M_j - M_{j+1}}\Big) \\
&= \tau_j s_j + \frac{1}{M_{j+1} - M_j}\tau_j \\
&= \Big(s_j + \frac{1}{M_j - M_{j+1}}\Big)s_j + \frac{1}{M_{j+1} - M_j}\Big(s_j + \frac{1}{M_j - M_{j+1}}\Big) \\
&= 1 - \frac{1}{(M_j - M_{j+1})^2} \\
&= \frac{(M_j - M_{j+1})^2 - 1}{(M_j - M_{j+1})^2} \\
&= \frac{(M_j - M_{j+1} + 1)(M_j - M_{j+1} - 1)}{(M_j - M_{j+1})^2},
\end{aligned}$$

completing the proof. □

Now if j and $j+1$ are not in the same row or same column of the tableaux Q, then τ_j^2 acts as a nonzero constant because $\mathrm{ct}(Q(j)) - \mathrm{ct}(Q(j+1)) \notin \{-1, 0, 1\}$ (note that 0 is not possible by monotonicity of the rows and columns of Q, meaning that $j+1$ cannot be on the same diagonal as j). Thus Lemma 6.1 implies that

$$\tau_j \colon V_Q^{(\ell)} \to V_{s_jQ}^{(\ell)} \quad \text{is a vector space isomorphism if } j, j+1 \text{ are not in the same row or column of } Q. \tag{6.9}$$

<u>*Definition of eigenvectors*</u>: With this, we are now ready to define the eigenvectors $f_Q^{m,\ell}$ corresponding to the various one-dimensional subspaces. Let $m \in \{0, 1, \ldots, \lfloor n/2 \rfloor\}$. Define the *column reading tableau* of shape $(n-m, m)$ to be the Young tableau

$$T = \begin{array}{|c|c|c|c|c|c|c|c|c|c|}\hline 1 & 3 & 5 & \cdots & 2m-1 & 2m+1 & 2m+2 & \cdots & n-1 & n \\ \hline 2 & 4 & 6 & \cdots & 2m \\ \cline{1-5}\end{array}\;.$$

Let $i, \ell \in \{0, 1, \ldots, n-2m\}$, and let $\mathbb{S}(n-2m)_\ell$ be the set of subsets of $\{1, \ldots, n-2m\}$ with cardinality ℓ. Define the scalars

$$T_{m,n}^{(\ell)}(i) = \sum_{S \in \mathbb{S}(n-2m)_\ell} (-1)^{m+|S\cap\{1,\ldots,i\}|} \binom{2m+\ell}{m + |S \cap \{1, \ldots, i\}|}. \tag{6.10}$$

(These numbers turn out to be the values of certain orthogonal polynomials – see Proposition 6.4 in Section 6.2 – but we will not need that fact for this first proof.) Using these scalars, we first define the following vectors associated to the column reading tableaux T:

$$g_T^{m,i} = (v_{01} - v_{10})^{\otimes m} \otimes \Big(\sum_{S \in \mathbb{S}(n-2m)_i} v_S\Big) \quad \text{and} \quad f_T^{m,\ell} = \sum_{i=0}^{n-2m} T_{m,n}^{(\ell)}(i) g_T^{m,i}. \tag{6.11}$$

Here the subscript of v_{01} stands for the subset $\{2\}$ of $\{1,2\}$, and the tensor product notation stands for concatenation of the subsets (so that, for example, $v_{01} \otimes v_{01} = v_{0101} = v_{\{2,4\}}$). In words, $g_T^{m,i}$ is a linear combination of particular v_Ss with $|S| = m+i$, and $f_T^{m,\ell}$ takes a certain linear combination of these vectors over the various "levels" i with coefficients coming from orthogonal polynomial scalars. These $f_T^{m,\ell}$ are the beginning of our orthogonal eigenvector basis.

We now explain how to define the vectors associated to any other tableau Q of shape $(n-m,m)$. If Q is the standard Young tableau of shape $(n-m,m)$ with $\boxed{a_1}\boxed{a_2}\boxed{\quad\cdots\quad}\boxed{a_m}$ in the second row, then $a_r \geq 2r$ for all $1 \leq r \leq m$. Therefore, we can apply a sequence of adjacent transpositions of the boxes of T to get to Q:

$$Q = c_Q^{(1)} \cdots c_Q^{(m)} T, \quad \text{where} \quad c_Q^{(r)} = \begin{cases} s_{a_r-1} \cdots s_{2r+1} s_{2r} & \text{if } a_r > 2r, \\ 1 & \text{if } a_r = 2r. \end{cases}$$

Thanks to Eq. (6.9), the corresponding map $\tau_Q \colon V_T^{(m+i)} \to V_Q^{(m+i)}$ defined by

$$\tau_Q = \tau_Q^{(1)} \cdots \tau_Q^{(m)}, \quad \text{where} \quad \tau_Q^{(r)} = \begin{cases} \tau_{a_r-1} \cdots \tau_{2r+1} \tau_{2r} & \text{if } a_r > 2r, \\ 1 & \text{if } a_r = 2r, \end{cases} \tag{6.12}$$

is a vector space isomorphism. In particular, if $a_r > 2r$, then $\tau_Q^{(r)} = \left(s_{a_r-1} - \frac{1}{a_r-2r+1}\right) \cdots (s_{2r+1} - \frac{1}{3})(s_{2r} - \frac{1}{2})$ is an expression for $\tau_Q^{(r)}$ in terms of the simple reflections in S_n. We then define the following vectors associated to the tableau Q (of shape $(n-m,m)$), for any $m \in \{0,1,\ldots,\lfloor n/2 \rfloor\}$ and $\ell, i \in \{0,1,\ldots,n-2m\}$:

$$g_Q^{m,i} = \tau_Q g_T^{m,i} \quad \text{and} \quad f_Q^{m,\ell} = \tau_Q f_T^{m,\ell}. \tag{6.13}$$

These vectors $f_Q^{m,\ell}$ will form the (unnormalized) orthogonal basis of eigenvectors in Theorem 6.2. Before we state that result, we define some constants that appear crucially in the norms. For the Young tableau Q notated above, define the scalar

$$\gamma_Q = \gamma_Q^{(1)} \cdots \gamma_Q^{(m)}, \quad \text{where} \quad \gamma_Q^{(r)} = \begin{cases} \frac{((a_r-2r+1)^2-1)}{(a_r-2r+1)^2} \cdots \frac{(3^2-1)}{3^2} \cdot \frac{(2^2-1)}{2^2} & \text{if } a_r > 2r, \\ 1 & \text{if } a_r = 2r. \end{cases} \tag{6.14}$$

Theorem 6.2. *Using notation as in Eq. (6.13) and Eq. (6.14), with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$, the set*

$$\left\{ f_Q^{m,\ell} : m \in \{0,1,\ldots,\lfloor n/2 \rfloor\}, \quad Q \in \hat{S}_n^{(n-m,m)}, \quad \ell \in \{0,1,\ldots,n-2m\} \right\}$$

is an orthogonal basis of $V^{\otimes n}$. With the notation for eigenvalues β_k as in Theorem 1.2,

$$f_Q^{m,\ell} \text{ is an eigenvector of } K_n \text{ of eigenvalue } \begin{cases} \beta_{(m+\ell)/2} & \text{if } m+\ell \text{ is even,} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\begin{aligned} \langle f_Q^{m,\ell}, f_Q^{m,\ell} \rangle_\pi &= \gamma_Q \langle f_T^{m,\ell}, f_T^{m,\ell} \rangle_\pi \\ &= \gamma_Q \frac{2^m}{n+1} \sum_{i=0}^{n-2m} \left(T_{m,n}^{(\ell)}(i)\right)^2 \frac{\binom{n-2m}{i}}{\binom{n}{m+i}}. \end{aligned} \tag{6.15}$$

We also express this sum is closed form in Corollary 6.5.

m	ℓ	Q	000	001	010	011	100	101	110	111
0	0	$\begin{matrix}1&2&3\end{matrix}$	1	0	0	0	0	0	0	0
0	1	$\begin{matrix}1&2&3\end{matrix}$	0	1	1	0	1	0	0	0
1	0	$\begin{matrix}1&2\\3&\end{matrix}$	0	1	-0.5	0	-0.5	0	0	0
1	0	$\begin{matrix}1&3\\2&\end{matrix}$	0	0	1	0	-1	0	0	0
0	2	$\begin{matrix}1&2&3\end{matrix}$	0	0	0	1	0	1	1	0
1	1	$\begin{matrix}1&2\\3&\end{matrix}$	0	0	0	0.5	0	0.5	-1	0
1	1	$\begin{matrix}1&3\\2&\end{matrix}$	0	0	0	1	0	-1	0	0
0	3	$\begin{matrix}1&2&3\end{matrix}$	0	0	0	0	0	0	0	1

Figure 2: The eight vectors $g_Q^{m,\ell}$ for $n = 3$. Observe that we have $g_Q^{m,\ell} \in V^{(m+\ell)}$ in all cases; that is, the vectors are only supported on the states S where $|S| = m + \ell$.

To illustrate a concrete example, we write out the list of g and f vectors in the case $n = 3$. In Fig. 2, the first three columns indicate the values of m, ℓ, and the tableau Q, and the last eight columns are the entries of the vector $g_Q^{m,\ell}$ evaluated at each state. (For example, the column for 011 corresponds to $v_{\{2,3\}} = v_0 \otimes v_1 \otimes v_1$.) The analogous table for the $f_Q^{m,\ell}$ vectors appears in Fig. 3, along with an extra column for displaying the normalizing factors $\langle f_Q^{m,\ell}, f_Q^{m,\ell}\rangle$.

In words, it turns out (as will be shown in the proof) that the relation between the f and g vectors is that f are linear combinations of the f_S vectors from Theorem 1.2, with coefficients given by g. For instance, $g^{0,2}_{\boxed{1}\boxed{2}\boxed{3}}$ has a 1 in each of the entries $011, 101$, and 110, so $f^{0,2}_{\boxed{1}\boxed{2}\boxed{3}} = 1f_{\{1,2\}} + 1f_{\{1,3\}} + 1f_{\{2,3\}}$.

The remainder of this subsection is dedicated to proving Theorem 6.2. We first prove a crucial lemma, which is the first step of the observation in the above paragraph. To adapt notation, note that the formula for the eigenvectors f_S from Theorem 1.2 can be written

$$f_S = \sum_{T \subseteq \{1,\ldots,n\}} (-1)^{|S\cap T|} \binom{|S|}{|S \cap T|} v_T. \tag{6.16}$$

Lemma 6.3. *Let T be the column reading tableau of shape $(n-m, m)$ and let $f_T^{m,\ell}$ and $g_T^{m,\ell}$ be as defined in Eq. (6.11). Define the vector space isomorphism $\Phi : V^{\otimes n} \to V^{\otimes n}$ by setting $\Phi(v_S) = f_S$ and extending by linearity. Then*

$$f_T^{m,\ell} = \Phi(g_T^{m,\ell}).$$

Proof. We begin by constructing careful linear combinations of our original eigenvectors f_S. For any

m	ℓ	Q	000	001	010	011	100	101	110	111	$\langle f_Q^{m,\ell}, f_Q^{m,\ell}\rangle$
0	0	1 2 3	1	1	1	1	1	1	1	1	1
0	1	1 2 3	3	1	1	-1	1	-1	-1	-3	5
1	0	1 2 / 3	0	-2	1	-1	1	-1	2	0	1
1	0	1 3 / 2	0	0	-2	-2	2	2	0	0	$\frac{4}{3}$
0	2	1 2 3	3	-3	-3	-3	-3	-3	-3	3	9
1	1	1 2 / 3	0	-3	1.5	1.5	1.5	1.5	-3	0	$\frac{9}{4}$
1	1	1 3 / 2	0	0	-3	3	3	-3	0	0	3
0	3	1 2 3	1	-3	-3	3	-3	3	3	-1	5

Figure 3: The eight vectors $f_Q^{m,\ell}$ for $n = 3$. Observe that the vector with $m + \ell = 0$ is an eigenvector of K_3 with eigenvalue $\beta_0 = 1$, the three vectors with $m + \ell = 2$ are eigenvectors with eigenvalue $\beta_1 = \frac{1}{4}$, and all other vectors are eigenvectors with eigenvalue 0.

subset $S \subseteq \{3, 4, \ldots, n\}$, we have (here the notation f_{01S} is shorthand for $f_{\{2\}\cup S}$)

$$\begin{aligned}
f_{01S} = \sum_{T\subseteq\{3,\ldots,n\}} & (-1)^{|S\cap T|}\binom{1+|S|}{|S\cap T|}v_{00T} + (-1)^{|S\cap T|}\binom{1+|S|}{|S\cap T|}v_{10T} \\
& + (-1)^{1+|S\cap T|}\binom{1+|S|}{1+|S\cap T|}v_{01T} + (-1)^{1+|S\cap T|}\binom{1+|S|}{1+|S\cap T|}v_{11T}, \\
f_{10S} = \sum_{T\subseteq\{3,\ldots,n\}} & (-1)^{|S\cap T|}\binom{1+|S|}{|S\cap T|}v_{00T} + (-1)^{1+|S\cap T|}\binom{1+|S|}{1+|S\cap T|}v_{10T} \\
& + (-1)^{|S\cap T|}\binom{1+|S|}{|S\cap T|}v_{01T} + (-1)^{1+|S\cap T|}\binom{1+|S|}{1+|S\cap T|}v_{11T},
\end{aligned}$$

so subtracting these equations yields

$$\begin{aligned}
f_{(01-10)S} = f_{01S} - f_{10S} = \sum_{T\subseteq\{3,\ldots,n\}} & (-1)^{|S\cap T|}\left(\binom{1+|S|}{|S\cap T|} + \binom{1+|S|}{1+|S\cap T|}\right)v_{10T} \\
& - (-1)^{|S\cap T|}\left(\binom{1+|S|}{1+|S\cap T|} + \binom{1+|S|}{|S\cap T|}\right)v_{01T} \\
= (v_{01} - v_{10}) \otimes & \left(\sum_{T\subseteq\{3,\ldots,n\}} (-1)^{1+|S\cap T|}\binom{2+|S|}{1+|S\cap T|}v_T\right).
\end{aligned}$$

Iterating this process on subsequent pairs of coordinates, we see that for any subset $S \subseteq \{2k+1, \ldots, n\}$, we have

$$
\begin{aligned}
f_{(01-10)^m S} &= (v_{01}-v_{10})^{\otimes m} \otimes \left(\sum_{T \subseteq \{2m+1,\ldots,n\}} (-1)^{m+|S\cap T|} \binom{2m+|S|}{m+|S\cap T|} v_T \right) \\
&= (v_{01}-v_{10})^{\otimes m} \otimes \left(\sum_{T \subseteq \{1,\ldots,n-2m\}} (-1)^{m+|S\cap T|} \binom{2m+|S|}{m+|S\cap T|} v_T \right), \qquad (6.17)
\end{aligned}
$$

where this last line is only a change in notation (notating a tensor product by starting both sets of coordinates from 1). As before, let $\mathbb{S}(n-2m)$ denote the set of all subsets of $\{1,\ldots,n-2m\}$ and $\mathbb{S}(n-2k)_i$ the set of such subsets of cardinality i. Letting T denote the column reading tableau of shape $(n-m,m)$, and defining

$$
g^{(i)}_{1\cdots(n-2m)} = \sum_{S\in\mathbb{S}(n-2m)_i} v_S
$$

(note that this is *not* a vector of the form $g_T^{m,\ell}$), we have that

$$
\begin{aligned}
f_T^{m,\ell} &= \sum_{i=0}^{n-2m} T^{(\ell)}_{m,n}(i) g_T^{m,i} \\
&= \sum_{i=0}^{n-2m} T^{(\ell)}_{m,n}(i) \left((v_{01}-v_{10})^{\otimes m} \otimes g^{(i)}_{1\cdots(n-2m)} \right) \\
&= (v_{01}-v_{10})^{\otimes m} \otimes \left(\sum_{i=0}^{n-2m} g^{(i)}_{1\cdots(n-2m)} T^{(\ell)}_{m,n}(i) \right) \\
&= (v_{01}-v_{10})^{\otimes m} \otimes \left(\sum_{i=0}^{n-2m} g^{(i)}_{1\cdots(n-2m)} \sum_{S\in\mathbb{S}(n-2m)_\ell} (-1)^{m+|S\cap\{1,\ldots,i\}|} \binom{2m+\ell}{m+|S\cap\{1,\ldots,i\}|} \right),
\end{aligned}
$$

where in the last line we plugged in the definition of our scalars $T^{(\ell)}_{m,n}(i)$. Therefore expanding out and then swapping the order of summation yields

$$
\begin{aligned}
f_T^{m,\ell} &= (v_{01}-v_{10})^{\otimes m} \otimes \left(\sum_{i=0}^{n-2m} \sum_{T\in\mathbb{S}(n-2m)_i} v_T \sum_{S\in\mathbb{S}(n-2m)_\ell} (-1)^{m+|S\cap\{1,\ldots,i\}|} \binom{2m+\ell}{m+|S\cap\{1,\ldots,i\}|} \right) \\
&= \sum_{\substack{S\subseteq\{2m+1,\cdots,n\} \\ |S|=\ell}} \left((v_{01}-v_{10})^{\otimes m} \otimes \sum_{T\in\mathbb{S}(n-2m)} (-1)^{m+|S\cap T|} \binom{2m+|S|}{m+|S\cap T|} v_T \right) \\
&= \sum_{S\in\mathbb{S}(n-2m)_\ell} f_{(10-01)^m S},
\end{aligned}
$$

where before swapping summation in the middle step we use that the innermost sum over S is always preserved if we replace $\{1,\cdots,i\}$ with the i-element subset T, and where the last equality follows from

Eq. (6.17). Now because $f_{(10-01)^m S}$ is just shorthand for a particular linear combination of the f_Ss, we have $\Phi(v_{(10-01)^m S}) = f_{(10-01)^m S}$. Thus

$$f_T^{m,\ell} = \sum_{S \in \mathbb{S}(n-2m)_\ell} f_{(10-01)^m S} = \sum_{S \in \mathbb{S}(n-2m)_\ell} \Phi(v_{(10-01)^m S}) = \Phi(g_T^{m,\ell}),$$

completing the proof. □

We are now ready to prove the main result of this subsection. Rephrasing the previous lemma, we have just shown that $f_T^{m,\ell}$ is a linear combination of eigenvectors f_S each with $|S| = m + \ell$, and so our proof will show that many of the nice properties relating f_T and g_T are still preserved when we apply transpositions to the boxes of the Young tableaux.

Proof of Theorem 6.2. Fix $m \in \{0, 1, \dots, \lfloor n/2 \rfloor\}$, and let T be the column reading tableau of shape $(n - m, m)$. Also fix $i \in \{0, 1, \dots, n - 2m\}$. We first claim that for all $2 \le r \le n$, we have (recall that s_{tr} denotes the transposition switching t and r)

$$\begin{aligned} M_r g_T^{m,i} &= \sum_{t=1}^{r-1} s_{tr} \, (v_{01} - v_{10})^{\otimes m} \otimes \left(\sum_{S \in \mathbb{S}(n-2m)_i} v_S \right) \\ &= \mathrm{ct}(T(r)) g_T^{m,i}. \end{aligned}$$

Indeed for $r \le 2m$, the only nonzero contributions to this sum are if $t < r$ are of the same parity (yielding $g_T^{m,i}$) or if r is even and $t = r - 1$ (yielding $-g_T^{m,i}$). And for $r > 2m$, the nonzero contributions are from $2m + 1 \le t < r$ (yielding $g_T^{m,i}$) and also from of the pairs $t \in \{1, 2\}, \{3, 4\}, \cdots, \{2m - 1, 2m\}$ (each of which yield $g_T^{m,i}$ when added together). In all cases, the total coefficient of $g_T^{m,i}$ is indeed the column number of the box r minus the row number. Therefore $g_T^{m,i} \in V_T^{(m+i)}$ by definition.

Now suppose Q is a standard tableau of shape $(n - m, m)$ with $\boxed{a_1}\boxed{a_2}\boxed{\quad\cdots\quad}\boxed{a_m}$ in the second row. Let $\tau_Q \colon V_T^{(m+i)} \to V_Q^{(m+i)}$ be the vector space isomorphism defined in Eq. (6.12). Since $g_Q^{m,i} = \tau_Q g_T^{m,i}$, this implies that

$$g_Q^{m,i} \in V_Q^{(m+i)} \quad \text{and} \quad g_Q^{m,i} \neq 0.$$

Thus, by Eq. (6.4) and Eq. (6.5), we have that

$$\left\{ g_Q^{m,i} \;\middle|\; m \in \{0, \dots, \lfloor n/2 \rfloor\}, \ Q \in \hat{S}_n^{(n-m,m)}, \ i \in \{0, \dots n - 2m\} \right\}$$

is an orthogonal basis of $V^{\otimes n}$ with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$.

We will now use this orthogonality to also prove orthogonality of the $f_Q^{m,\ell}$s. Recall that Φ is the vector space isomorphism mapping v_S to f_S for all S. For any $w \in S_n$, we have

$$w\Phi(v_S) = w f_S = f_{wS} = \Phi(v_{wS}) = \Phi(w v_S),$$

so Φ is an S_n-module isomorphism. Therefore Lemma 6.3 tells us that

$$f_Q^{m,\ell} = \tau_Q f_T^{m,\ell} = \tau_Q \Phi(g_T^{m,\ell}) = \Phi(\tau_Q g_T^{m,\ell}) = \Phi(g_Q^{m,\ell})$$

for all tableaux Q. Since $\left\{ g_Q^{m,\ell} \,\middle|\, m \in \{0, \dots, \lfloor n/2 \rfloor\},\ Q \in \hat{S}_n^{(n-m,m)},\ \ell \in \{0, \dots, n-2m\} \right\}$ is a basis of $V^{\otimes n}$, we must have that

$$\left\{ f_Q^{m,\ell} \,\middle|\, m \in \{0, \dots, \lfloor n/2 \rfloor\},\ Q \in \hat{S}_n^{(n-m,m)},\ \ell \in \{0, \dots, n-2m\} \right\} \quad \text{is a basis of } V^{\otimes n}.$$

Recall that by Theorem 1.2, $\{f_S \mid S \in \mathbb{S}(n)\}$ is a basis of $V^{\otimes n}$ where

$$f_S \text{ is an eigenvector of } K_n \text{ of eigenvalue } \begin{cases} \beta_{|S|/2} & \text{if } |S| \text{ is even,} \\ 0 & \text{otherwise.} \end{cases} \tag{6.18}$$

So since $g_T^{m,\ell}$ is a linear combination of v_Ss with $|S| = m + \ell$ and $wv_S = v_{wS}$ for all $w \in S_n$, $g_Q^{m,\ell}$ must also be a linear combination of v_S with $|S| = m + \ell$. Thus $f_Q^{m,\ell} = \Phi(g_Q^{m,\ell})$ is a linear combination of f_Ss all with $|S| = k + \ell$, meaning that Eq. (6.18) implies

$$f_Q^{m,\ell} \text{ is an eigenvector of } K_n \text{ with eigenvalue } \begin{cases} \beta_{(m+\ell)/2} & \text{if } m+\ell \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}$$

We now prove that these eigenvectors are all orthogonal. Because Φ is an S_n-module isomorphism, we have for any r that

$$M_r f_Q^{m,\ell} = M_r \Phi(g_Q^{m,\ell}) = \Phi(M_r g_Q^{m,\ell}) = \Phi(\mathrm{ct}(Q(r)) g_Q^{m,\ell}) = \mathrm{ct}(Q(r)) \Phi(g_Q^{m,\ell}) = \mathrm{ct}(Q(r)) f_Q^{m,\ell}.$$

So in fact $f_Q^{m,\ell} \in V_Q$, so it follows from Eq. (6.3) that

$$P \neq Q \implies \langle f_P^{m_1,\ell_1}, f_Q^{m_2,\ell_2} \rangle_\pi = 0.$$

Therefore, it just remains to show orthogonality among eigenvectors of the same tableau Q. Remembering that m is determined by the shape of Q, it thus remains to show that $\langle f_Q^{m,\ell_1}, f_Q^{m,\ell_2} \rangle_\pi = 0$ for all ℓ_1, ℓ_2. Because K_n is self-adjoint with respect to $\langle \cdot, \cdot \rangle_\pi$, eigenvectors of K_n of different eigenvalues are automatically orthogonal, but this does not account for the case where both vectors are zero eigenfunctions.

For this, we claim first that for the column reading tableau T of shape $(n-m, m)$, we have

$$\langle f_T^{m,\ell_1}, f_T^{m,\ell_2} \rangle_\pi = 0$$

for any $\ell_1 \neq \ell_2$. We defer the proof of this fact (which reduces to a "WZ-pair" binomial coefficient calculation done with computer assistance) to Lemma A.2. Then recalling that all transpositions are self-adjoint, each $\tau_j = s_j + \frac{1}{M_j - M_{j+1}}$ is also self adjoint. Therefore, for any tableau Q and any transposition s_i, we have

$$\begin{aligned} \langle f_{s_iQ}^{m,\ell_1}, f_{s_iQ}^{m,\ell_2} \rangle_\pi &= \langle \tau_i f_Q^{m,\ell_1}, \tau_i f_Q^{m,\ell_2} \rangle_\pi \\ &= \langle f_Q^{m,\ell_1}, \tau_i^2 f_Q^{m,\ell_2} \rangle_\pi \\ &= \frac{(\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1)) + 1)(\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1)) - 1)}{(\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1)))^2} \langle f_Q^{m,\ell_1}, f_Q^{m,\ell_2} \rangle_\pi, \end{aligned}$$

where the last equality follows from Lemma 6.1 and the fact that $f_Q^{m,\ell} \in V_Q$. Since we obtain any tableau Q from T via a sequence of such transpositions, we find inductively that $\langle f_Q^{m,\ell_1}, f_Q^{m,\ell_2}\rangle_\pi = 0$ for all Q of shape $(n-m,m)$, completing the verification of orthogonality.

Finally, we compute the norms of our eigenvectors $f_Q^{m,\ell}$. Again, let T be the column reading tableau of shape $(n-m,m)$ and let $i \in \{0,1,\ldots,n-2m\}$. We can compute explicitly the norm of g_T, since

$$\begin{aligned}
\langle g_T^{m,i}, g_T^{m,i}\rangle_\pi &= \left\langle (v_{01}-v_{10})^{\otimes m} \otimes \Big(\sum_{S\in\mathbb{S}(n-2m)_i} v_S\Big), (v_{01}-v_{10})^{\otimes m} \otimes \Big(\sum_{S\in\mathbb{S}(n-2m)_i} v_S\Big)\right\rangle_\pi \\
&= \frac{1}{n+1}\sum_{S\in\mathbb{S}(n-2m)_i} \frac{1}{\binom{n}{m+i}} 2^m \\
&= \frac{2^m}{n+1}\frac{\binom{n-2m}{i}}{\binom{n}{m+i}},
\end{aligned} \tag{6.19}$$

where the middle step uses that $g_T^{m,i} \in V^{(m+i)}$ and so the inner product weight $\pi(x)$ is in fact constant. But then because the $g_T^{m,i}$s are orthogonal, our definition of $f_T^{m,\ell}$ as a linear combination yields

$$f_T^{m,\ell} = \sum_{i=0}^{n-2m} T_{m,n}^{(\ell)}(i) g_T^{m,i} \implies \langle f_T^{m,\ell}, f_T^{m,\ell}\rangle_\pi = \frac{2^m}{n+1}\sum_{i=0}^{n-2m}\big(T_{m,n}^{(\ell)}(i)\big)^2 \frac{\binom{n-2m}{i}}{\binom{n}{m+i}}.$$

And now we again apply the logic for transferring inner products from T to Q: for any tableau Q and any transposition s_i, we again have

$$\begin{aligned}
\langle f_{s_iQ}^{m,\ell}, f_{s_iQ}^{m,\ell}\rangle_\pi &= \langle \tau_i f_Q^{m,\ell}, \tau_i f_Q^{m,\ell}\rangle_\pi \\
&= \langle f_Q^{m,\ell}, \tau_i^2 f_Q^{m,\ell}\rangle_\pi \\
&= \frac{(\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1)) + 1)(\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1)) - 1)}{(\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1)))^2}\langle f_Q^{m,\ell}, f_Q^{m,\ell}\rangle_\pi,
\end{aligned}$$

and this time we extract the constants from Eq. (6.12). For each $\tau_Q^{(r)} = \tau_{a_r-1}\cdots\tau_{2r}$ performed from right-to-left, note that $\mathrm{ct}(Q(i)) - \mathrm{ct}(Q(i+1))$ will successively take on the values $2,3,4,\cdots,a_r-2r+1$. Indeed, before we perform these transpositions, box $2r+1$ will always be one row above and one column to the right of box $2r$, and then boxes $2r+2, 2r+3, \cdots$ will be immediately to the right of box $2r+1$. Therefore the product of all factors is exactly $\gamma_Q^{(r)}$, and multiplying all contributions together indeed yields

$$\langle f_Q^{m,\ell}, f_Q^{m,\ell}\rangle_\pi = \gamma_Q \langle f_T^{m,\ell}, f_T^{m,\ell}\rangle_\pi,$$

completing the proof. □

6.2 More on orthogonal polynomials and eigenvalue multiplicities

This subsection outlines some useful consequences and observations related to Theorem 6.2. We first describe a connection between the scalars $T_{m,n}^{\ell}(i)$ defined in Eq. (6.10) and the discrete Chebyshev polynomials used in Section 4. Recall that the former were the coefficients used in Eq. (6.11) to obtain our eigenvectors $f_T^{m,\ell}$, and the latter were used (via Proposition 4.2) to get explicit binomial-coefficient formulas for the eigenvectors f_S. Now that we have proven that the eigenvectors $f_T^{m,\ell}$ are orthogonal, we can obtain a vast generalization of Proposition 4.2 and write out formulas for all degrees and a wider class of underlying weights.

For more on orthogonal polynomial theory, see [Chi78] for a very readable exposition or [KS96] for a detailed compendium (we will follow the notational conventions of the latter). Define the rising factorial

$$(a)_0 = 1, \quad (a)_j = a(a+1)\cdots(a+j-1).$$

The (α,β)-**Hahn polynomials** (see [KS96, Section 1.5] for a definition in terms of the hypergeometric ${}_3F_2$ function)

$$Q^{\ell}_{n;\alpha,\beta}(x) = \sum_{k=0}^{n} \frac{1}{k!} \frac{(-\ell)_k(\ell+\alpha+\beta+1)_k(-x)_k}{(\alpha+1)_k(-n)_k}$$

are the orthogonal polynomials on $\{0,1,\cdots,n\}$ with respect to the beta-binomial distribution $m(i) = \binom{n}{i}\frac{(\alpha+1)_i(\beta+1)_{n-i}}{(\alpha+\beta+2)_n}$, normalized so that $Q^{\ell}_{n;\alpha,\beta}(0) = 1$. (Note that this is a different convention from the one used in [DZ21, Section 2.3].) In particular, the $(0,0)$-Hahn polynomials are exactly the discrete Chebyshev polynomials, since the corresponding beta-binomial distribution is uniform on $\{0,\cdots,n\}$.

Proposition 6.4. *The scalars $T^{(\ell)}_{m,n}(i)$ are related to the (m,m)-Hahn polynomials on $\{0,\cdots,n-2m\}$ as follows:*

$$Q^{\ell}_{n-2m;m,m}(i) = \frac{(-1)^m}{\binom{n-2m}{\ell}\binom{2m+\ell}{m}} T^{(\ell)}_{m,n}(i).$$

In particular, recall that $T^{\ell}_n(x)$ denotes the discrete Chebyshev polynomial on $\{0,1,\cdots,n\}$ of degree ℓ. Then the $m=0$ case says that for all $0 \le j \le n$,

$$T^{\ell}_n(j) = \frac{1}{\binom{n}{\ell}} T^{(\ell)}_{0,n}(j) = \frac{1}{\binom{n}{\ell}} \sum_{S\in\mathbb{S}(n)_\ell} (-1)^{|S\cap\{1,\cdots,j\}|}\binom{\ell}{|S\cap\{1,\cdots,j\}|}.$$

Proof. First, write out Eq. (6.10) by casework on the value of $|S\cap\{1,\ldots i\}|$ to find

$$\begin{aligned} T^{(\ell)}_{m,n}(i) &= \sum_{S\in\mathbb{S}(n-2m)_\ell} (-1)^{m+|S\cap\{1,\ldots,i\}|}\binom{2m+\ell}{m+|S\cap\{1,\ldots,i\}|} \\ &= \sum_{j=0}^{i} (-1)^{m+j}\binom{2m+\ell}{m+j}\cdot\binom{i}{j}\binom{n-2m-i}{\ell-j}. \end{aligned} \tag{6.20}$$

Now fix n and m. For each $\ell \in \{0,\cdots,n-2m\}$, observe that this last expression is a polynomial in i of degree ℓ, since each term in the sum is of degree ℓ and as the same sign as $(-1)^{m+j}(-1)^{\ell-j} = (-1)^{m+\ell}$, which is constant for all terms. Moreover, because of the orthogonality guaranteed by Theorem 6.2, we have that for any $\ell_1 \neq \ell_2$ that

$$\begin{aligned} 0 &= \langle f_T^{m,\ell_1}, f_T^{m,\ell_2}\rangle \\ &= \sum_{i,j=0}^{n-2m} T^{(\ell_1)}_{m,n}(i) T^{(\ell_2)}_{m,n}(j)\langle g_T^{m,i}, g_T^{m,j}\rangle \\ &= \sum_{i=0}^{n-2m} T^{(\ell_1)}_{m,n}(i) T^{(\ell_2)}_{m,n}(i) \frac{2^m}{n+1}\frac{\binom{n-2m}{i}}{\binom{n}{m+i}} \end{aligned}$$

where we have used Eq. (6.11) and Eq. (6.19) (along with the fact that the $g_T^{m,i}$s are orthogonal over i). And now

$$\begin{aligned}\frac{\binom{n-2m}{i}}{\binom{n}{m+i}} &= \frac{(n-2m)!(m+i)!(n-m-i)!}{i!(n-2m-i)!n!}\\ &= \frac{m!^2}{n!}\binom{n-2m}{i}\frac{(m+i)!}{m!}\frac{(n-m-i)!}{m!}\\ &= \frac{m!^2}{n!}\binom{n-2m}{i}(m+1)_i(m+1)_{n-2m-i}\end{aligned}$$

is proportional in i to the beta-binomial distribution on $\{0,\cdots,n-2m\}$ with parameters $(\alpha,\beta)=(m,m)$, so in fact our calculation above shows that $T_{m,n}^{(\ell_1)}$ and $T_{m,n}^{(\ell_2)}(i)$ are orthogonal with respect to this distribution. Thus up to a constant, $T_{m,n}^{(\ell)}$ is exactly the degree-ℓ (m,m)-Hahn polynomial, and we can evaluate the constant by evaluating at $i=0$. We know that $Q_{n-2m;m,m}^{(\ell)}(0)=1$, and plugging in $i=0$ into Eq. (6.20) leaves only the $j=0$ term, which is $(-1)^m\binom{2m+\ell}{m}\binom{n-2m}{\ell}$. Thus dividing by this constant yields the result. □

This connection to orthogonal polynomials lets us write the sum appearing in Eq. (6.15) in closed form, since orthogonality relations for such polynomials are readily available. Indeed, in our notation, the orthogonality equation [KS96, (1.5.2)] reads, for any $\alpha,\beta>-1$,

$$\sum_{i=0}^{n}\binom{\alpha+i}{i}\binom{n+\beta-i}{n-i}Q^{\ell}_{n;\alpha,\beta}(i)Q^{\ell'}_{n;\alpha,\beta}(i)=\delta_{\ell\ell'}\frac{(-1)^\ell\ell!(\beta+1)_\ell(\ell+\alpha+\beta+1)_{n+1}}{n!(2\ell+\alpha+\beta+1)(-n)_\ell(\alpha+1)_\ell}. \tag{6.21}$$

Corollary 6.5. *Using the notation in Theorem 6.2, we have that*

$$\langle f_Q^{m,\ell},f_Q^{m,\ell}\rangle_\pi=\gamma_Q\frac{2^m}{n+1}\cdot\frac{(2m+\ell)!}{(2m+2\ell+1)(m+\ell)!^2\ell!}\cdot\frac{(n-2m)!}{n!}\cdot\frac{(n+\ell+1)!}{(n-2m-\ell)!}.$$

It may be easier to parse this formula by treating m and ℓ as constants and considering the dependence on n. In particular, for the column reading tableau T (so that $\gamma_T=1$), setting $m=0$ yields

$$\begin{aligned}\langle f_T^{0,\ell},f_T^{0,\ell}\rangle_\pi &= \frac{1}{n+1}\cdot\frac{\ell!}{(2\ell+1)\ell!^3}\frac{(n+\ell+1)!}{(n-\ell)!}\\ &= \frac{1}{(2\ell+1)\ell!^2}\cdot\frac{1}{n+1}\prod_{i=-\ell+1}^{\ell+1}(n+i)\end{aligned} \tag{6.22}$$

and similarly setting $m=1$ yields

$$\begin{aligned}\langle f_T^{1,\ell},f_T^{1,\ell}\rangle_\pi &= \frac{2}{n+1}\cdot\frac{(\ell+2)!}{(2\ell+3)(\ell+1)!^2\ell!}\cdot\frac{1}{n(n-1)}\cdot\frac{(n+\ell+1)!}{(n-2-\ell)!}\\ &= \frac{2(\ell+2)}{(\ell+1)(2\ell+3)\ell!^2}\cdot\frac{1}{(n-1)n(n+1)}\prod_{i=-\ell-1}^{\ell+1}(n+i).\end{aligned} \tag{6.23}$$

The key interpretation is that these expressions are always rational functions of n which nicely factor into terms of the form $(n+i)$. (We will use these later on to bound the complicated expression for chi-square distance in Eq. (6.31).)

Proof. Plugging in $\alpha = \beta = m, \ell = \ell'$, and replacing n with $n - 2m$ in Eq. (6.21) yields

$$\sum_{i=0}^{n-2m} \binom{m+i}{i}\binom{n-m-i}{n-2m-i} Q^{\ell}_{n-2m;m,m}(i)^2 = \frac{(-1)^\ell \ell!(\ell+2m+1)_{n-2m+1}}{(n-2m)!(2\ell+2m+1)(-(n-2m))_\ell}$$
$$= \frac{\ell+2m+1}{2\ell+2m+1}\frac{\binom{\ell+n+1}{n-2m}}{\binom{n-2m}{\ell}}.$$

Thus the sum in Eq. (6.15) can be written as

$$\begin{aligned}
\sum_{i=0}^{n-2m} (T^{(\ell)}_{m,n}(i))^2 \frac{\binom{n-2m}{i}}{\binom{n}{m+i}} &= \binom{n-2m}{\ell}^2\binom{2m+\ell}{m} \cdot \sum_{i=0}^{n-2m} (Q^{(\ell)}_{n-m;m,m}(i))^2 \frac{\binom{n-2m}{i}}{\binom{n}{m+i}} \\
&= \binom{n-2m}{\ell}^2\binom{2m+\ell}{m}^2 \cdot \sum_{i=0}^{n-2m} (Q^{(\ell)}_{n-m;m,m}(i))^2 \frac{(n-2m)!(m+i)!(n-m-i)!}{i!(n-2m-i)!n!} \\
&= \binom{n-2m}{\ell}^2\binom{2m+\ell}{m}^2 \cdot \frac{(n-2m)!m!^2}{n!}\sum_{i=0}^{n-2m} (Q^{(\ell)}_{n-m;m,m}(i))^2 \binom{m+i}{i}\binom{n-m-i}{n-2m-i} \\
&= \binom{n-2m}{\ell}^2\binom{2m+\ell}{m}^2 \frac{(n-2m)!m!^2}{n!}\frac{\ell+2m+1}{2m+2\ell+1}\frac{\binom{\ell+n+1}{n-2m}}{\binom{n-2m}{\ell}} \\
&= \frac{(n-2m)!m!^2}{n!}\frac{\ell+2m+1}{2m+2\ell+1}\binom{2m+\ell}{m}^2\binom{n-2m}{\ell}\binom{\ell+n+1}{n-2m} \\
&= \frac{(2m+\ell)!}{(2m+2\ell+1)(m+\ell)!^2\ell!} \cdot \frac{(n-2m)!}{n!} \cdot \frac{(n+\ell+1)!}{(n-2m-\ell)!},
\end{aligned}$$

and plugging this back into the remaining expression of Eq. (6.15) yields the result. □

We'll now elaborate on how this eigenvector decomposition sheds light on the structure of the eigenspaces for the binary Burnside process. We claimed in Eq. (6.6) that the subspaces that our $f_Q^{m,\ell}$s span are exactly invariant subspaces under certain actions, and we elaborate now. On each "orbit level" subspace $V^{(i)}$ for $i \le \frac{n}{2}$, the action of S_n permutes the locations of the v_1s (equivalently, the coordinates of the ones in the n-tuple), so that $V^{(i)}$ is isomorphic to the permutation representation $M^{(n-i,i)} = \mathrm{Ind}_{S_i \times S_{n-i}}^{S_n}(1)$ on size-i subsets. (Similarly for $i > \frac{n}{2}$, the action permutes the locations of the zeros and thus is isomorphic to $M^{(i,n-i)}$. These permutation representations decompose into irreducible representations as

$$M^{(n-i,i)} = S^{(n)} \oplus S^{(n-1,1)} \oplus \cdots \oplus S^{(n-i,i)}, \tag{6.24}$$

where for any partition λ of n, S^λ is the irreducible *Specht module* associated to that partition. (The above equation is a special case of Young's rule, or more generally the Littlewood–Richardson rule [Jam78, Sections 14-17].) In particular, this means $V^{\otimes n}$ has $(n+1)$ copies of $S^{(n)}$, $(n-1)$ copies of $S^{(n-1,1)}$, $(n-3)$ copies of $S^{(n-2,2)}$, and so on. The fact that the binary Burnside transition matrix K_n commutes with the action of S_n implies (by Schur's lemma) that as a map from any S^λ to any S^μ, K_n must act as a constant multiple of the identity, and that constant is only nonzero if $\lambda = \mu$. This alone already implies that eigenvalues will appear with high multiplicity, since restricting K_n to the copies of $S^{(n-i,i)}$ yields an operator whose eigenvalues repeat with multiplicity $\dim(S^{(n-i,i)}) = \binom{n}{i} - \binom{n}{i-1}$.

However, the eigenvalues further degenerate beyond those irreducible subspaces in the binary Burnside process, and this is how Eq. (6.6) makes its appearance. The Lie algebra $\mathfrak{sl}_2$ acts on the two-dimensional vector space spanned by $\{v_0, v_1\}$ via matrix multiplication

$$e = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad h = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and so therefore $\mathfrak{sl}_2$ also acts on $V^{\otimes n}$ by

$$g(v_1 \otimes \cdots \otimes v_n) = \sum_{i=1}^{n} v_1 \otimes \cdots \otimes v_{i-1} \otimes g v_i \otimes v_{i+1} \otimes \cdots \otimes v_n.$$

The statement of Schur–Weyl duality is that this action and the S_n action commute, and each is the full centralizer of the other, leading to the decomposition

$$L(C_2^n) = \bigoplus_{\lambda} S^\lambda \otimes L^\lambda \tag{6.25}$$

where the L^λs are irreducible representations of $\mathfrak{sl}_2$. Since K_n commutes with S_n, this implies that in fact K_n is in the universal enveloping algebra of $\mathfrak{sl}_2$; slightly imprecisely, it may be written as a polynomial in the es, fs, and hs which acts on the L^λ component of the decomposition. This is discussed further in Conjecture 6.10 below.

Since Theorem 6.2 consists of eigenvalues that lie in irreducible $\mathfrak{sl}_2$-invariant and S_n-invariant subspaces, it thus implies the following description (and thus how the eigenvalue multiplicities manifest across the different irreducible subspaces):

Corollary 6.6. *Let K_n^λ be the binary Burnside operator K_n restricted to the copies of S^λ. By Schur's lemma, we know that K_n^λ acts as $\tilde{K}_n^\lambda \otimes I_{\dim(S^\lambda)}$ for some operator $\tilde{K}_n^\lambda$ (which is typically not a Markov chain). Let $\beta_k = \frac{1}{2^{4k}}\binom{2k}{k}^2$ be as in Theorem 1.2. Then for any $k \le \frac{n}{2}$, β_k is an eigenvalue of $\tilde{K}_n^\lambda$ of multiplicity 1 (meaning it is an eigenvalue of K_n^λ of multiplicity $\dim(S^\lambda)$) for $\lambda = (n-m, m)$ when $m \le \min(2k, n-2k)$, and there are no other nonzero eigenvalues.*

The decompositions for $n = 4$ and $n = 5$ are shown below for illustration. Note that $\dim(S^{(n-m,m)})$ is exactly the number of standard Young tableaux of shape $(n-m, m)$, so that each value in the array corresponds to a particular eigenvector. The rows with an even value of $m + \ell$ correspond to eigenspaces for the nonzero eigenvalues.

$m+\ell$	1 2 / 3 4	1 3 / 2 4	1 2 3 / 4	1 2 4 / 3	1 3 4 / 2	1 2 3 4
0						1
1			0	0	0	0
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
3			0	0	0	0
4						$\frac{9}{64}$

$m+\ell$	1 2 3 / 4 5	1 2 4 / 3 5	1 3 4 / 2 5	1 3 5 / 2 4	1 2 5 / 3 4	1 2 3 4 / 5	1 2 3 5 / 4	1 2 4 5 / 3	1 3 4 5 / 2	1 2 3 4 5
0										1
1						0	0	0	0	0
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
3	0	0	0	0	0	0	0	0	0	0
4						$\frac{9}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{9}{64}$
5										0

Proof. Theorem 6.2 exhibits a full basis of eigenvectors $\{f_Q^{m,\ell}\}$, such that restricted to the copies of $S^{(n-m,m)}$, the eigenvectors are indexed by the $\dim(S^{(n-m,m)})$ Young tableaux Q of shape $(n-m,m)$ and by the parameter ℓ, and the eigenvalues depend only on $m+\ell$. Since $\ell \in \{0,1,\cdots,n-2m\}$, eigenvectors with $m+\ell \in \{m, m+1, \cdots, n-m\}$ appear, and in particular this means that β_k appears as an eigenvector in $S^{n-m,m}$ if and only if $m \le 2k \le n-m$, or equivalently $m \le \min(2k, n-2k)$, as desired. $\square$

Remark. *For any representation ρ of a group G on a vector space V and any irreducible character χ of G of degree d, the projection of ρ onto the copies of χ that appear is given by*

$$P = \frac{d}{|G|} \sum_{g \in G} \chi^*(g)\rho(g).$$

(This formula is the "canonical decomposition" described in [Ser77, Section 2.6].) In particular, if χ corresponds to the trivial representation, P is exactly averaging over the entire orbit, so K_n restricted to the $(n+1)$ copies of the trivial representation is exactly K^{lumped}; this is consistent with the fact that all eigenvalues $\beta_0, \beta_1, \cdots, \beta_{\lfloor n/2 \rfloor}$ each appear exactly once in the lumped chain. The fact that Hahn polynomials appear in the eigenvectors corresponding to the other irreducible representations indicates that there may be nice interpretations for the other projections as well, even if they are not Markov operators.

6.3 Mixing time analysis from the one-ones state

In this subsection, we will take the orthogonal basis of eigenfunctions $\{f_Q^{m,\ell}\}$ and apply it to the identity $\chi_x^2(\ell) = \sum_{i=1}^{|\mathfrak{X}|-1} f_i^2(x)\beta_i^{2\ell}$ from Eq. (2.1). Rewriting this in our new notation, we have that the chi-square distance to stationarity started from x after s steps is

$$\chi_x^2(s) = \sum_{\substack{m \in \{0,\cdots,\lfloor n/2 \rfloor\} \\ Q \in \hat{S}_n^{(n-m,m)} \\ \ell \in \{0,\cdots,n-2m\} \\ m+\ell \text{ even} \\ (m,\ell) \neq (0,0)}} \left(\overline{f_Q^{m,\ell}}(x)\right)^2 \left(\beta_{(m+\ell)/2}\right)^{2s}, \tag{6.26}$$

where $\overline{f_Q^{m,\ell}}(x)$ is the $L^2(\pi)$-normalized multiple of $f_Q^{m,\ell}$ (whose squared norm is given by Eq. (6.15)). Of course, this can be a very complicated sum in general, but we will now demonstrate that the specific form of our orthonormal basis $\{f_Q^{m,\ell}\}$ can be very convenient for computations. To do this, we will compute the chi-square distance to stationarity started from the "one-ones" state $e_n = (0, \cdots, 0, 1)$; in particular, we will need to compute the value of each $f_Q^{m,\ell}$ at this state.

Remark. *By symmetry of the original Markov chain under permutation of coordinates, we know that the chi-square distance to stationarity is identical for any of the "one-ones" states. However, the individual values of $f_Q^{m,\ell}(x)$ are different for those different states x, which will become clear in our computation.*

The result we will prove is the following:

Theorem 6.7. *For the binary Burnside process started from the state $e_n = (0, 0, \cdots, 1)$ (or any other state with a single 1), a constant number of steps is necessary and sufficient in ℓ^2 (and therefore also in ℓ^1). More precisely, for all $n, s \geq 3$, the chi-square distance to stationarity after s steps satisfies*

$$5\left(\frac{1}{4}\right)^{2s} \leq \chi^2_{e_n}(s) \leq 270\left(\frac{1}{4}\right)^{2s}.$$

Proof. First, let $S = \{n\}$, so that $v_S = v_0^{\otimes(n-1)} \otimes v_1$. We note that by definition, for any $v \in V^{\otimes n}$ we have that

$$\begin{aligned} v(e_n) &= \text{the coefficient of } v_S \text{ in } v \\ &= (n+1)n\langle v, v_S\rangle_\pi, \end{aligned}$$

since our stationary distribution assigns mass $\frac{1}{(n+1)\binom{n}{1}}$ to the state e_n. Thus, we may alternatively think of evaluating our vectors at e_n as computing the inner products $n(n+1)\langle f_Q^{m,\ell}, v_S\rangle_\pi$, though we will not take this perspective here.

Since $f_Q^{m,\ell}$ is a linear combination of $g_Q^{m,i}$ terms, it will be simplest to compute using those latter vectors first. As usual, we begin with T, the column reading tableau of shape $(n-m, m)$. By inspection of the definition of $g_T^{m,i}$, we have that

$$g_T^{m,i}(e_j) = \begin{cases} 1 & \text{if } (m,i) = (0,1), \\ -1 & \text{if } (m,i) = (1,0) \text{ and } j = 1, \\ 1 & \text{if } (m,i) = (1,0) \text{ and } j = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{6.27}$$

The key property used here is that for $m \geq 2$, the $(v_{01} - v_{10})^{\otimes m}$ term means that all nonzero terms have at least two coordinates with v_1s and so the coefficient of v_S (as well as all other singleton sets) must be zero. Since $f_T^{m,\ell}$ is a linear combination of $g_T^{m,i}$s, this then also implies that $f_T^{m,\ell}(e_j) = 0$. Further recalling Eq. (6.13), since each $g_Q^{m,i}$ (resp. $f_Q^{m,\ell}$) is a linear combination of permuted $g_T^{m,i}$s (resp. $f_T^{m,\ell}$s), this immediately implies that for all $1 \leq j \leq n$ (and in particular $j = n$),

$$f_Q^{m,\ell}(e_j) = 0 \text{ for all } m \geq 2 \text{ (and all } \ell \in \{0, \cdots, n-2m\} \text{ and } Q \in \hat{S}_n^{(n-m,m)}). \tag{6.28}$$

Thus we need only compute $f_Q^{m,\ell}(e_n)$ for $m = 0, 1$. The case $m = 0$ is simpler, since the only tableau of this shape is the standard tableau. We thus have, for any $0 \le \ell \le n$ and $1 \le j \le n$, that

$$\begin{aligned} f_T^{0,\ell}(e_j) &= \sum_{i=0}^{n} T_{0,n}^{\ell}(i) g_T^{0,i}(e_1) \\ &= T_{0,n}^{\ell}(1) g_T^{0,1}(e_1) \\ &= \sum_{S \in \mathbb{S}(n)_\ell} (-1)^{|S| \cap \{1\}} \binom{\ell}{|S \cap \{1\}|} \\ &= \binom{n-1}{\ell} - \ell \binom{n-1}{\ell-1}. \end{aligned} \tag{6.29}$$

Now for the case $m = 1$, we begin with a similar calculation for the column reading tableau and say that

$$\begin{aligned} f_T^{1,\ell}(e_1) &= \sum_{i=0}^{n-2} T_{1,n}^{\ell}(i) g_T^{1,i}(e_1) \\ &= T_{1,n}^{\ell}(0) g_T^{1,0}(e_1) \\ &= - \sum_{S \in \mathbb{S}(n-2)_\ell} (-1)^{1+0} \binom{2+\ell}{1+0} \\ &= (2+\ell) \binom{n-2}{\ell}, \end{aligned}$$

and by the exact same calculation

$$f_T^{1,\ell}(e_2) = -(2+\ell)\binom{n-2}{\ell}, \quad f_T^{1,\ell}(e_j) = 0 \text{ for } j \ge 3.$$

Now a general tableau Q of shape $(n-1, 1)$ has a single entry $a_1 \ge 2$ in the second row, and so we have by the definition in Eq. (6.12) that

$$\begin{aligned} f_Q^{1,\ell} &= \tau_{a_1 - 1} \cdots \tau_2 f_T^{1,\ell} \\ &= \left(s_{a_1 - 1} - \frac{1}{a_1 - 1} \right) \cdots \left(s_2 - \frac{1}{2} \right) f_T^{1,\ell}. \end{aligned}$$

We can now see why picking a particular state e_j can simplify calculations; we will demonstrate this by evaluating our vectors at e_1 and at e_n. First of all, the transpositions $s_2, \cdots, s_{a_1 - 1}$ do not alter the value of the vectors at e_1, and so in fact we have

$$\begin{aligned} f_Q^{1,\ell}(e_1) &= \left(1 - \frac{1}{a_1 - 1} \right) \cdots \left(1 - \frac{1}{2} \right) f_T^{1,\ell}(e_1) \\ &= \frac{1}{a_1 - 1} f_T^{1,\ell}(e_1) \\ &= \frac{1}{a_1 - 1} (2+\ell) \binom{n-2}{\ell}. \end{aligned}$$

Thus any tableau Q of this shape $(n-1,1)$ contributes to the final sum. On the other hand, take any $n \geq 3$. If we instead choose to evaluate $f_Q^{1,\ell}$ at e_n, then the only nonzero contribution comes from applying the sequence of transpositions $s_2, s_3, \cdots, s_{n-1}$ (so that the nonzero entry beginning in e_2 is now in e_n). In other words, we have

$$f_Q^{1,\ell}(e_n) = \begin{cases} -(2+\ell)\binom{n-2}{\ell} & \text{if } a_1 = n, \\ 0 & \text{otherwise.} \end{cases} \tag{6.30}$$

Thus evaluating our eigenvectors at e_n results in the smallest number of terms required, and we can now plug into Eq. (6.26) to get, for any $n \geq 3$,

$$\begin{aligned} \chi^2_{e_n}(s) &= \sum_{\substack{m \in \{0,\cdots,\lfloor n/2 \rfloor\} \\ Q \in \hat{S}_n^{(n-m,m)} \\ \ell \in \{0,\cdots,n-2m\} \\ m+\ell \text{ even} \\ (m,\ell) \neq (0,0)}} \left(\overline{f_Q^{m,\ell}}(e_n)\right)^2 \left(\beta_{(m+\ell)/2}\right)^{2s} \\ &= \sum_{\ell \in \{1,\cdots,n\} \text{ even}} \frac{f_{Q_{(0)}}^{0,\ell}(e_n)^2}{\langle f_{Q_{(0)}}^{0,\ell}, f_{Q_{(0)}}^{0,\ell} \rangle_\pi} (\beta_{\ell/2})^{2s} + \sum_{\ell \in \{0,\cdots,n\} \text{ odd}} \frac{f_{Q_{(1)}}^{1,\ell}(e_n)^2}{\langle f_{Q_{(1)}}^{1,\ell}, f_{Q_{(1)}}^{1,\ell} \rangle_\pi} (\beta_{(1+\ell)/2})^{2s}, \end{aligned} \tag{6.31}$$

where the two sums are the contributions from $m = 0, 1$ respectively, $Q_{(0)}$ is the only tableau of shape (n), and $Q_{(1)}$ is the tableau of shape $(n-1,1)$ with n in the second row. We have just computed all of the values in the numerators of the fractions, and we can use Corollary 6.5 to compute the denominators. By Eq. (6.22) we have

$$\langle f_{Q_{(0)}}^{0,\ell}, f_{Q_{(0)}}^{0,\ell} \rangle_\pi = \langle f_T^{0,\ell}, f_T^{0,\ell} \rangle_\pi = \frac{1}{(2\ell+1)\ell!^2} \cdot \frac{1}{n+1} \prod_{i=-\ell+1}^{\ell+1} (n+i). \tag{6.32}$$

Similarly by Eq. (6.23), and using that $\gamma_{Q_{(1)}} = \left(1 - \frac{1}{(n-1)^2}\right) \cdots \left(1 - \frac{1}{2^2}\right) = \frac{n}{2(n-1)}$ by a telescoping sum,

$$\langle f_{Q_{(1)}}^{1,\ell}, f_{Q_{(1)}}^{1,\ell} \rangle_\pi = \frac{n}{2(n-1)} \langle f_T^{1,\ell}, f_T^{1,\ell} \rangle_\pi = \frac{(\ell+2)}{(\ell+1)(2\ell+3)\ell!^2} \cdot \frac{1}{(n-1)^2(n+1)} \prod_{i=-\ell-1}^{\ell+1} (n+i). \tag{6.33}$$

We will now compute a lower bound by taking only the term $\ell = 2$ from the first sum and $\ell = 1$ from the second sum in Eq. (6.31). (Indeed, $(m,\ell) = (0,2)$ and $(1,1)$ are the only terms corresponding to the largest nontrivial eigenvalue β_1.) The numerators here are the squares of $f_{Q_{(0)}}^{0,2}(e_n) = \frac{(n-1)(n-6)}{2}$ and $f_{Q_{(1)}}^{1,1}(e_n) = -3(n-2)$, and the denominators simplify to

$$\langle f_{Q_{(0)}}^{0,2}, f_{Q_{(0)}}^{0,2} \rangle_\pi = \frac{1}{20} \cdot \frac{1}{n+1} \prod_{i=-1}^{3} (n+i) = \frac{1}{20}(n-1)n(n+2)(n+3)$$

and

$$\langle f_{Q_{(1)}}^{1,1}, f_{Q_{(1)}}^{1,1} \rangle_\pi = \frac{3}{10} \cdot \frac{1}{(n-1)^2(n+1)} \prod_{i=-2}^{2} (n+i) = \frac{3}{10} \frac{(n-2)n(n+2)}{n-1}.$$

The most important feature to notice is that the numerator and denominator of these fractions are polynomials of the same order. Thus, the lower bound

$$\chi^2_{e_n}(s) \geq \left(\frac{\left(\frac{(n-1)(n-6)}{2}\right)^2}{\frac{1}{20}(n-1)n(n+2)(n+3)} + \frac{(-3(n-2))^2}{\frac{3}{10}\frac{(n-2)n(n+2)}{n-1}} \right) \left(\frac{1}{4}\right)^{2s}$$

is asymptotically $35 \cdot \left(\frac{1}{4}\right)^{2s}$, and in particular it is at least $5 \cdot \left(\frac{1}{4}\right)^{2s}$ for all $n \geq 3$.

Finally, for the upper bound, we include all terms and upper bound each one independently of n. Plugging in Eq. (6.29) and Eq. (6.32), we have

$$\begin{aligned}
\frac{f^{0,\ell}_{Q_{(0)}}(e_n)^2}{\langle f^{0,\ell}_{Q_{(0)}}, f^{0,\ell}_{Q_{(0)}} \rangle_\pi} &= (n+1)(2\ell+1)\ell!^2 \frac{\left(\binom{n-1}{\ell} - \ell\binom{n-1}{\ell-1}\right)^2}{\prod_{i=-\ell+1}^{\ell+1}(n+i)} \\
&\leq (2\ell+1)\frac{(n+1)\prod_{i=1}^{\ell}(n-i)^2}{\prod_{i=-\ell+1}^{\ell+1}(n+i)} + (2\ell+1)\ell^4 \frac{(n+1)\prod_{i=1}^{\ell-1}(n-i)^2}{\prod_{i=-\ell+1}^{\ell+1}(n+i)}
\end{aligned}$$

using that $(a-b)^2 \leq a^2+b^2$ for $a, b > 0$. But now we can pair up each linear factor in n in the numerators with a larger factor in the denominator, meaning that this is simply upper bounded by $(2\ell+1)+\frac{(2\ell+1)\ell^4}{n^2} \leq (2\ell+1)+(2\ell+1)\ell^2 = 2\ell^3+\ell^2+2\ell+1$. Similarly, plugging in Eq. (6.30) and Eq. (6.33),

$$\begin{aligned}
\frac{f^{1,\ell}_{Q_{(1)}}(e_n)^2}{\langle f^{1,\ell}_{Q_{(1)}}, f^{1,\ell}_{Q_{(1)}} \rangle_\pi} &= \frac{(\ell+1)(2\ell+3)\ell!^2}{(\ell+2)} \cdot \frac{(n-1)^2(n+1)(2+\ell)^2\binom{n-2}{\ell}^2}{\prod_{i=-\ell-1}^{\ell+1}(n+i)} \\
&= (\ell+1)(\ell+2)(2\ell+3)\frac{(n-1)^2(n+1)\prod_{i=1}^{\ell}(n-1-i)^2}{\prod_{i=-\ell-1}^{\ell+1}(n+i)} \\
&\leq (\ell+1)(\ell+2)(2\ell+3) \\
&= 2\ell^3+9\ell^2+13\ell+6.
\end{aligned}$$

Our upper bound of Eq. (6.31) therefore reads

$$\begin{aligned}
\chi^2_{e_n}(s) &\leq \sum_{\ell\in\{1,\cdots,n\}\text{ even}} (2\ell^3+\ell^2+2\ell+1)(\beta_{\ell/2})^{2s} + \sum_{\ell\in\{0,\cdots,n\}\text{ odd}} (2\ell^3+9\ell^2+13\ell+6)(\beta_{(1+\ell)/2})^{2s} \\
&\leq \sum_{\ell\in 2\mathbb{N}} (2\ell^3+\ell^2+2\ell+1)(\beta_{\ell/2})^{2s} + \sum_{\ell\in 2\mathbb{N}-1} (2\ell^3+9\ell^2+13\ell+6)(\beta_{(1+\ell)/2})^{2s} \\
&\leq \sum_{k\in\mathbb{N}} (32k^3+16k^2+6k+1)(\beta_k)^{2s}.
\end{aligned}$$

But remembering that as in the proof of Theorem 1.1, we have $\beta_k < \frac{1}{\pi k}$ for all k, we can use the crude bound

$$\beta_k^{2s} \leq \left(\frac{1}{4}\right)^{2(s-3)} \beta_k^6 < \frac{4^6}{\pi^6}\left(\frac{1}{4}\right)^{2s} \cdot \frac{1}{k^6}$$

for all $s \geq 3$ so that the series converges. This yields

$$\begin{aligned}\chi_{e_n}^2(s) &\leq \frac{4^6}{\pi^6} \sum_{k\in\mathbb{N}} \frac{32k^3+16k^2+6k+1}{k^6} \left(\frac{1}{4}\right)^{2s} \\ &\leq 270 \left(\frac{1}{4}\right)^{2s},\end{aligned}$$

completing the proof. □

A similar analysis may be carried out from other starting states as well, though there will be even more nonzero terms in the expression for $\chi_x^2(s)$ for general states x. We leave these calculations as potential future work, though we believe that a similar strategy as what we have described here may be fruitful and that the asymptotics will be relatively well-behaved due to the nice form of the norms $\langle f_Q^{m,\ell}, f_Q^{m,\ell}\rangle_\pi$.

6.4 Miscellaneous remarks

In this final subsection, we collect some additional facts about our matrix K_n in this more algebraic framework. Recall the closed-form expression $K_n(x,y)$ from Proposition 3.1. We first write out a version of Proposition 3.3, showing that "lumping over the final coordinate still yields the Burnside process on the remaining coordinates:"

Proposition 6.8. *Let I denote the identity 2×2 matrix. We have*

$$K_n(I^{\otimes(n-1)} \otimes K_1) = K_{n-1} \otimes K_1.$$

Proof. Recall the closed-form expression

$$K_n(x,y) = \frac{\binom{2n_{00}}{n_{00}}\binom{2n_{01}}{n_{01}}\binom{2n_{10}}{n_{10}}\binom{2n_{11}}{n_{11}}}{4^n\binom{n_{00}+n_{01}}{n_{00}}\binom{n_{10}+n_{11}}{n_{10}}}$$

from Proposition 3.1. We have

$$\frac{\binom{2(n_{00}+1)}{n_{00}+1}}{\binom{n_{00}+n_{01}+1}{n_{00}+1}} = \frac{\frac{(2n_{00}+1)(2n_{00}+2)}{(n_{00}+1)(n_{00}+1)}}{\frac{(n_{00}+n_{01}+1)}{(n_{00}+1)}} \frac{\binom{2n_{00}}{n_{00}}}{\binom{n_{00}+n_{01}}{n_{00}}} = \frac{2(2n_{00}+1)}{(n_{00}+n_{01}+1)} \frac{\binom{2n_{00}}{n_{00}}}{\binom{n_{00}+n_{01}}{n_{00}}},$$

as well as the same identity with n_{00} and n_{01} switched. Thus for any $x, y \in C_2^{n-1}$ (and defining $n_{00}, n_{01}, n_{10}, n_{11}$ relative to x, y),

$$\begin{aligned}K_n(x0,y0) + K_n(x0,y1) &= \left(\frac{2(2n_{00}+1)}{n_{00}+n_{01}+1} + \frac{2(2n_{01}+1)}{n_{00}+n_{00}+1}\right) \cdot \frac{\binom{2n_{00}}{n_{00}}\binom{2n_{01}}{n_{01}}\binom{2n_{10}}{n_{10}}\binom{2n_{11}}{n_{11}}}{4^n\binom{n_{00}+n_{01}}{n_{00}}\binom{n_{10}+n_{11}}{n_{10}}} \\ &= \frac{\binom{2n_{00}}{n_{00}}\binom{2n_{01}}{n_{01}}\binom{2n_{10}}{n_{10}}\binom{2n_{11}}{n_{11}}}{4^{n-1}\binom{n_{00}+n_{01}}{n_{00}}\binom{n_{10}+n_{11}}{n_{10}}} \\ &= K_{n-1}(x,y),\end{aligned}$$

where we have used that n_{00} increments from (x, y) to $(x0, y0)$ and n_{01} increments from (x, y) to $(x0, y1)$, but all other values stay the same. Using the same strategy, we also have that $K_n(x1, y0) + K_n(x1, y1) = K_{n-1}(x, y)$ for all x, y. Putting this together and using that $K_1 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$, we arrive at the desired

$$K_n(I^{\otimes(n-1)} \otimes K_1) = K_{n-1} \otimes K_1$$

(in words, this says that we can either average over the last coordinate before or after applying the binary Burnside matrix). □

Next, we show (analogously to the proof of Theorem 1.2) an algebraic proof that the "lifted vectors" $f_S = \sum_{T \subseteq \{1,\dots,n\}} (-1)^{|S \cap T|} \binom{|S|}{|S \cap T|} v_T$ are indeed eigenvectors of our Burnside matrix:

Proposition 6.9. *Let $S \subseteq \{1, \dots, n\}$ and let $\ell = |S|$. Then f_S is an eigenvector of K_n of eigenvalue $\beta_{\ell/2}$ if ℓ is even and 0 otherwise.*

Proof. First, assume that $S = \{1, \cdots, \ell\}$ (meaning $v_S = v_1^{\otimes \ell} \otimes v_0^{\otimes(n-\ell)}$). Write

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad K_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Applying K_1 to each of the last $(n - \ell)$ coordinates yields

$$K_1^{\otimes(n-\ell)} v_{0^{(n-\ell)}} = \frac{1}{2^{n-\ell}} (v_0 + v_1)^{\otimes(n-\ell)} = \frac{1}{2^{n-\ell}} \sum_{T \subseteq \{1,\dots,n-\ell\}} v_T.$$

Therefore, we can write

$$f_S = f_{\{1,\cdots,\ell\}} = 2^{n-\ell} (I^{\otimes \ell} \otimes K_1^{\otimes(n-\ell)})(f_{\{1,\cdots,\ell\}} \otimes v_0^{\otimes(n-\ell)}).$$

Hence letting β denote the corresponding eigenvalue (either $\beta_{\ell/2}$ if ℓ is even or 0 otherwise), we have

$$\begin{aligned} K_n f_{\{1,\cdots,\ell\}} &= 2^{n-\ell} K_n (I^{\otimes \ell} \otimes K_1^{\otimes(n-\ell)})(f_{\{1,\cdots,\ell\}} \otimes v_0^{\otimes(n-\ell)}) \\ &= 2^{n-\ell} (K_\ell \otimes K_1^{\otimes(n-\ell)})(f_{\{1,\cdots,\ell\}} \otimes v_0^{\otimes(n-\ell)}) \\ &= 2^{n-\ell} \beta (I^{\otimes \ell} \otimes K_1^{\otimes(n-\ell)})(f_{\{1,\cdots,\ell\}} \otimes v_0^{\otimes(n-\ell)}) \\ &= \beta f_{\{1,\cdots,\ell\}}, \end{aligned}$$

where we used Proposition 6.8 in the second line and that $f_{\{1,\cdots,\ell\}}$ is an eigenfunction of K_ℓ in the third line. This proves that $f_{\{1,\cdots,\ell\}}$ is indeed an eigenvector of the correct eigenvalue.

Finally, for the general case, for any $S \subseteq \{1, \dots, n\}$ of size ℓ, then there exists $\sigma \in S_n$ such that $f_S = \sigma f_{\{1,\cdots,\ell\}}$. Therefore

$$K_n f_S = K_n \sigma f_{\{1,\cdots,\ell\}} = \sigma K_n f_{\{1,\cdots,\ell\}} = \beta \sigma f_{\{1,\cdots,\ell\}} = \beta f_S,$$

as desired. □

Lastly, we describe one more curious property of the matrices K_n. As previously discussed, K_n commuting with the action of S_n implies that it may be viewed as an element of the universal enveloping algebra $U(\mathfrak{sl}_2)$. Thus, K_n may be rewritten as some polynomial expression in the basis elements e, f, h. In particular, the fact that the nonzero eigenvalues obtained in Theorem 1.2 do not depend on n suggests that there may be a single element of $U(\mathfrak{sl}_2)$ which agrees with K_n as an operator on $V^{\otimes n}$, or alternatively that there is some recursive relation among the K_ns which explains Proposition 3.3 algebraically. The following conjecture (which has been checked up to $n = 10$) is one direction in which this idea could be further explored (though we do not do so here):

Conjecture 6.10. *Define the* 2×2 *matrices*

$$p^+ = \frac{1}{2}(1+e+f) = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad p^- = \frac{1}{2}(1-e-f) = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}, \quad p^+h = \frac{1}{2}(1+e+f)h = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & -1/2 \end{bmatrix}.$$

Let $f(x, y, z)$ *be the sum over all ways (orders) of taking the matrix tensor product of* x *copies of* p^+, y *copies of* p^-, *and* z *copies of* p^+h. *Then*

$$K_n = \sum_{x+y+z=n} c_{y,z} f(x, y, z), \quad c_{y,z} = \begin{cases} \left(\frac{y!z!}{(\frac{y}{2})!(\frac{z}{2})!(\frac{y+z}{2})!2^{y+z}} \right)^2 & \text{if } y, z \text{ are nonnegative even integeres,} \\ 0 & \text{otherwise.} \end{cases}$$

For example, the expressions for $n = 4$ and $n = 6$ read

$$K_4 = f(4,0,0) + \frac{1}{4}(f(2,2,0) + f(2,0,2)) + \frac{9}{64}(f(0,4,0) + f(0,0,4)) + \frac{1}{64}f(0,2,2),$$

$$K_6 = f(6,0,0) + \frac{1}{4}(f(4,2,0) + f(4,0,2)) + \frac{9}{64}(f(2,4,0) + f(2,0,4)) + \frac{1}{64}f(2,2,2) + \frac{25}{256}(f(0,6,0) + f(0,0,6)) + \frac{1}{256}(f(0,2,4) + f(0,4,2)).$$

Also note that the constants $c_{k,k}$ are exactly the eigenvalues β_k of our Markov chain. It may be interesting to write out more explicit expressions for the various terms $f(x, y, z)$, or to find probabilistic interpretations for the off-diagonal constants $c_{y,z}$; note in particular that $f(n, 0, 0)$ is always $\frac{1}{2^n}$ times the all-ones matrix, while all other terms have all row sums equal to zero.

7 Related chains

7.1 Other Markov chains with similar properties

The Burnside process in this paper is a Markov chain on the hypercube which lumps to the orbits $\mathcal{O}_i$. Here, we mention some other chains that share this property and highlight some differences in their rates of convergence.

First of all, consider the *nearest-neighbor random walk on the hypercube*. The lumped chain in that setting is the *Ehrenfest urn*, in which there are two urns and a uniform ball is moved from one urn to the other at each step. While both chains can be lumped to the same orbits $\mathcal{O}_i$ and also both satisfy

$K(x,y) = K(\sigma(x),\sigma(y))$, this lumped chain behaves quite differently from the lumped binary Burnside process. Specifically (see [Dia96] for more details and references to proofs, as well as [KZ09] for some generalizations), $\frac{1}{4}n\log n + cn$ steps are necessary and sufficient for convergence in both ℓ^1 and ℓ^2 when started from 0, while only cn steps are necessary and sufficient when started from $\frac{n}{2}$. In contrast, explicit computations using the discrete Chebyshev polynomials show that for the lumped binary Burnside chain, a constant number of steps are necessary and sufficient both when started from 0 and from $\frac{n}{2}$, with no cutoff occurring in either case.

Continuing this comparison, we may also compare behavior of the two unlumped chains on C_2^n. Started from any vertex, nearest-neighbor random walk with holding converges to stationarity in $\frac{1}{4}n\log n + cn$ steps (the exact profile is computed in [DGM90]), and the choice of starting state does not matter since we have a random walk on a group. In contrast, our main results show that the starting state drastically affects rates of convergence: Theorem 1.1 shows that some states take order $\frac{n}{\log n}$ steps to converge (and in fact Proposition 5.1 shows that most of them do), while Eq. (1.2) and Theorem 6.7 prove that the all-zeros state $\underline{0}$, as well as any one-ones state, take just a constant number of steps.

For a second example, consider the *uniform-prior beta-binomial chain*, first studied in [DKSC08] as an example of a two-component Gibbs sampler. Briefly, this chain may be described as follows. Consider (j,θ) sampled from the joint distribution $f(j,\theta)d\theta = \binom{n}{j}\theta^j(1-\theta)^{n-j}d\theta$, where $j \in \{0,1,\cdots,n\}$ and $d\theta$ is Lebesgue measure on $[0,1]$. (This is indeed a probability measure, since summing over j yields 1 for all θ and then integrating over θ yields 1 overall.) We may form a Markov chain on the j-state space $\{0,1,\cdots,n\}$ as follows:

- From j, sample θ from the distribution conditioned on j (which is Beta with parameters $(j+1, n-j+1)$).
- From θ, sample j' from the distribution conditioned on θ (which is binomial with parameters (n,θ)).

Much like one step of the binary Burnside chain consists of performing the steps $x \mapsto s \mapsto y$, one step of this chain consists of performing the steps $j \mapsto \theta \mapsto j'$. The resulting chain has a uniform stationary distribution, and [DKSC08, Proposition 1.1] shows that it also has the discrete Chebyshev polynomials as eigenvectors (just like our lumped chain, as shown in Proposition 4.1). However, the eigenvalues in the beta-binomial chain do have an explicit dependence on n (unlike in our problem), and the chain requires order n steps to converge in chi-square distance when started from either 0 or n.

We may unlump this chain to get a Markov chain on C_2^n in a straightforward way:

- From $x \in C_2^n$, sample θ from the Beta distribution with parameters $(|x|+1, n-|x|+1)$.
- From θ, sample x' as a sequence of n Bernoulli(θ) random variables, viewed as a binary n-tuple.

This chain thus also has the same constant-on-orbits stationary distribution $\pi(x) = \frac{1}{(n+1)\binom{n}{|x|}}$ as our binary Burnside process. But in this case, "unlumping" the chain does not lead to higher eigenvalue multiplicities or longer mixing times. Indeed, since θ depends only on the orbit of x, we have (letting $\tilde{K}^{\text{lumped}}$ and $\tilde{K}^{\text{unlumped}}$ denote the transition matrices of the lumped and unlumped chains, respectively)

$$\tilde{K}_n^{\text{unlumped}}(x,x') = \frac{1}{\binom{n}{|x'|}}\tilde{K}_n^{\text{lumped}}(|x|,|x'|),$$

which implies that the nonzero eigenvalues and multiplicities of the unlumped chain are identical to those of the lumped chain – all additional eigenvalues are zero. Additionally, the symmetry of the binary Burnside chain described in Proposition 3.3 does not hold for the unlumped beta-binomial chain. Together, these facts mean that the same set of eigenvectors $\{f_Q^{m,\ell}\}$ as in Theorem 6.2 are in fact also an orthogonal basis of eigenvectors for the unlumped beta-binomial chain, but the only nonzero eigenvalues come from the eigenvectors with $m = 0$. In summary, despite the identical stationary distributions and eigenvectors coming from orthogonal polynomials, these two Markov chains behave quite differently.

7.2 Generalizing beyond the binary case

This paper discusses the binary Burnside process, which is a Markov chain on the hypercube C_2^n. An analogous definition can also be made for a Burnside process on (C_k^n, S_n) for $k \geq 2$, and we discuss how some symmetries of the binary case still persist and propose some ideas for extending our results.

In one step of this more general Burnside process, we begin with an n-tuple $x \in C_k^n$, uniformly pick a permutation permuting the coordinates within each value, write it as a product of disjoint cycles, and label each cycle uniformly with one of the k values in the alphabet. Proposition 3.3 generalizes directly in this setting, with the only modification to the proof being that we partition the coordinate set into k sets of locations rather than just the locations of zeros and ones:

Proposition 7.1. *The restriction of the Burnside process on (C_k^n, S_n) to any $m \leq n$ of its coordinates is also a Markov chain, and its transition probabilities are exactly given by the Burnside process on (C_k^m, S_m).*

In place of the decomposition in Eq. (6.24), we must now consider permutation representations M^λ for partitions λ of n of up to k parts, and we now have

$$M^\lambda = \bigoplus_{\mu} K_{\mu\lambda} S^\mu$$

where $K_{\mu\lambda}$ are the *Kostka numbers* (which are positive if and only if μ dominates λ). Towards understanding how these (many) copies of S^μ are arranged in the function space, the Schur–Weyl decomposition of Eq. (6.25) now reads

$$L(C_k^n) = \bigoplus_{\substack{\lambda \text{ partition of } n \\ \text{with at most } k \text{ parts}}} S^\lambda \otimes L^\lambda,$$

where the S^λs are again Specht modules and the L^λs are now irreducible representations of $\mathfrak{sl}_k$. One question is whether there is an explicit expression for the transition matrix as an element of the universal enveloping algebra $U(\mathfrak{sl}_k)$ in the same way as Conjecture 6.10, and whether that expression can be written in a way that demonstrates how K_ns of different n relate to each other.

This representation theory connection may be of particular interest because the methods used to prove Theorem 1.2 in the $k = 2$ case run into significant challenges for $k > 2$. Instead of considering a lumped chain on $\{0, 1, \cdots, n\}$, we must now consider the Bose-Einstein orbit chain of [Dia05] mentioned in Section 3. This Markov chain generally has irrational eigenvalues (even for small values like $k = 3, n = 6$), making explicit descriptions of the eigenvectors (as we had with the discrete Chebyshev polynomials) difficult. Additionally, the unlumped Burnside chain now exhibits eigenvalues not present in the lumped chain, meaning we cannot just "lift" lumped eigenvectors to unlumped ones and expect a full eigendecomposition.

Conjecture 7.2. *Fix k, and let λ be any nonzero eigenvalue of the Burnside chain on (C_k^n, S_n) for any n. Then λ occurs with multiplicity $a_\lambda\binom{n}{b_\lambda}$ for some integers a_λ, b_λ.*

For the $k = 2$ case, Theorem 1.2 shows that for the eigenvalue $\lambda = \beta_i$, we have $a_\lambda = 1$ and $b_\lambda = 2i$, and no other eigenvalues appear. In contrast, consider the eigenvalue $\frac{1}{18}$ for $k = 3$. This eigenvalue does not appear in the orbit Bose-Einstein chain for any $n \leq 8$, but it occurs with multiplicity $2, 10, 30, 70$ for $n = 4, 5, 6, 7$, suggesting that $a_\lambda = 2$ and $b_\lambda = 4$. It is possible that this conjecture may be resolved by proving some appropriate generalization of Corollary 6.6; however, the expressions for $\dim(S^\lambda)$ (and thus the contributions to the total eigenvalue multiplicities) are in general the number of standard Young tableaux of shape λ, which may be more cumbersome to work with than the simpler expressions $\dim(S^{(n-i,i)}) = \binom{n}{i} - \binom{n}{i-1}$.

Even without eigenvalues and eigenvectors, some information about mixing time can still be proved. Aldous showed (as a generalization of Eq. (1.1)) that

$$||K_x^\ell - \pi||_{\text{TV}} \leq n\left(1 - \frac{1}{k}\right)^\ell,$$

meaning that $k(\log n + c)$ steps are sufficient for ℓ^1 mixing from any starting state. The argument ℓ^2 mixing Proposition 5.1 also generalizes to prove an analogous bound:

Proposition 7.3. *Let $x^{(n)} \in C_k^n$ be any sequence of states such that $x^{(n)}$ contains at least cn coordinates of each of the k values. Then for the Burnside process on (C_k^n, S_n), we have $\chi^2_{x^{(n)}}(\ell) \to \infty$ for $\ell = \Theta_{c,k}(\frac{n}{\log n})$.*

Proof. As in the proof of Proposition 5.1, we again bound $\chi^2_{x^{(n)}}(\ell)$ using only the term $y = x^{(n)}$. We now have $K(x^{(n)}, x^{(n)}) = \Omega_c\left(\frac{1}{(nk)^k}\right)$ (because there is a probability $\Theta_c\left(\frac{1}{n^k}\right)$ for the permutation fixing $x^{(n)}$ to just be k full cycles, and then a probability $\frac{1}{k^k}$ of each cycle to be labeled with its original value), while $\frac{1}{\pi(x^{(n)})}$ still grows exponentially in n (with constant depending on c and k). Thus $\Theta_{c,k}\left(\frac{n}{\log n}\right)$ steps are required until $K\left(x^{(n)}, x^{(n)}\right)^\ell < \pi(x^{(n)})^{1/3}$, and so $\chi^2_{x^{(n)}}(\ell)$ is still exponentially growing as $n \to \infty$ for this value of ℓ. □

So even in this more general case, $\frac{n}{\log n}$ steps are necessary for ℓ^2 mixing from most starting states. However, without eigenvalues and eigenvectors, neither the eigenvalue bound of Theorem 1.1 nor the "ℓ^2 by ℓ^1 upper bound" of Corollary 2.2 is admissible for proving that this is also sufficient. Thus, it would be interesting to prove Conjecture 7.2 (through the Schur–Weyl duality connection or otherwise) and provide matching upper bounds for mixing time. Along those lines, we conclude with a final unified conjecture for the Burnside process:

Conjecture 7.4. *For any fixed $k \geq 2$, let K_n denote the Burnside process on (C_k^n, S_n). Then K_n has cutoff in both ℓ^1 and ℓ^2 when started from states with a positive limiting proportion of at least two different values in C_k.*

A Proofs of binomial coefficient identities

The formulas for $\langle\overline{f_S}, \overline{f_{S'}}\rangle$ and $\langle f_S, f_S\rangle$, as well as the more complicated $\langle f_T^{m,\ell_1}, f_T^{m,\ell_2}\rangle$, can all be rewritten as binomial coefficient summation identities. Thus the "creative telescoping" method can be used to verify many of these formulas with computer assistance. The key ideas of this algorithm come from Wilf and Zeilberger's WZ method – an overview can be found in [NPWZ97] – and subsequent work has been done to speed up the algorithm with various heuristics and a careful ansatz [Kou10]. The Mathematica package `HolonomicFunctions` that we used, along with further literature references, may be found at the following link.

We first show how to compute the simplest case, which is the normalizing factor $\langle f_S, f_S \rangle$, without needing this machinery. (This illustrates the concept of "showing that a certain quantity is independent of one of its parameters.") To do this, we first note down a useful binomial coefficient computation:

Lemma A.1. *For any $c_1, c_2, c_3 \in \mathbb{Z}_{\geq 0}$ with $c_1 \geq c_2$, we have the identity*

$$\frac{1}{c_3 + c_1 + 1} \sum_{i=0}^{c_3} \frac{\binom{c_3}{i}}{\binom{c_3+c_1}{i+c_2}} = \frac{1}{c_1 + 1} \frac{1}{\binom{c_1}{c_2}}.$$

Proof. We have

$$\begin{aligned}
\sum_{i=0}^{c_3} \frac{\binom{c_3}{i}\binom{c_1}{c_2}}{\binom{c_3+c_1}{i+c_2}} &= \sum_{i=0}^{c_3} \frac{c_3! c_1!}{(c_3+c_1)!} \cdot \frac{(i+c_2)!}{i! c_2!} \cdot \frac{(c_3 - i + c_1 - c_2)!}{(c_3 - i)!(c_1 - c_2)!} \\
&= \frac{1}{\binom{c_1+c_3}{c_1}} \sum_{i=0}^{c_3} \binom{c_2 + i}{c_2} \binom{c_3 + c_1 - (c_2 + i)}{c_1 - c_2} \\
&= \frac{1}{\binom{c_1+c_3}{c_1}} \cdot \binom{c_3 + c_1 + 1}{c_1 + 1} \\
&= \frac{c_3 + c_1 + 1}{c_1 + 1},
\end{aligned}$$

where between the second and third lines we use the identity $\sum_{i=0}^{a} \binom{a-i}{b}\binom{c+i}{d} = \binom{a+c+1}{b+d+1}$. Now rearranging the equality between the first and last expression yields the result. □

Proof of formula for $\langle f_S, f_S \rangle$ in Proposition 4.3. We are considering the case $\ell = m$ of Eq. (4.2), so that our computation simplifies to

$$\begin{aligned}
\langle f_S, f_S \rangle &= \sum_{a=0}^{m} \sum_{d=0}^{n-m} \binom{m}{a} \binom{n-m}{d} \frac{1}{n+1} \frac{1}{\binom{n}{a+d}} \binom{m}{a}^2 \\
&= \sum_{a=0}^{m} \binom{m}{a}^3 \cdot \frac{1}{n+1} \sum_{d=0}^{n-m} \frac{\binom{n-m}{d}}{\binom{n}{a+d}}.
\end{aligned}$$

Plugging in $c_3 = n - m$, $c_1 = m$, and $c_2 = a$ into Lemma A.1, we thus have

$$\begin{aligned}
\langle f_S, f_S \rangle &= \sum_{a=0}^{m} \binom{m}{a}^3 \frac{1}{m+1} \cdot \frac{1}{\binom{m}{a}} \\
&= \frac{1}{m+1} \sum_{a=0}^{m} \binom{m}{a}^2 \\
&= \frac{1}{m+1} \binom{2m}{m},
\end{aligned}$$

as desired. □

We now turn to the formula for $\langle \overline{f_S}, \overline{f_{S'}} \rangle$. While plugging in Lemma A.1 into Eq. (4.2) does prove that $\langle f_S, f_{S'} \rangle$ is independent of n (since it allows us to perform the sum over the index d), computing the remaining sum by hand is a much harder task. Instead, the following argument due to Laurent Bartholdi (using Christoph Koutschan's `HolonomicFunctions` package) may be employed:

Proof of formula for $\langle \overline{f_S}, \overline{f_{S'}} \rangle$ in Proposition 4.3. Since the quadruple sum in Eq. (4.2) is independent of n by Lemma A.1, we select $n = 2m - \ell$ and rewrite $\ell = m - s$. Rearranging, we thus have that

$$\binom{2m+1-\ell}{m+1} \langle \overline{f_S}, \overline{f'_S} \rangle = \sum_{a=0}^{m-s} \sum_{b=0}^{s} \sum_{c=0}^{s} \binom{m-s}{a} \binom{s}{b} \binom{s}{c} (-1)^{b+c} \binom{m}{a+b} \binom{m}{a+c} \frac{(a+b+c)!(m+s-a-b-c)!m!}{s!(2m)!},$$

and we wish to prove that this quantity is identically 1 for all m and s. Write the summand as $H(m, s, a, b, c)$; our goal is to prove that $\overline{H}(m, s) = \sum_{a,b,c} H(m, s, a, b, c) = 1$.

When $m = s = 0$, we can compute directly that $\overline{H}(0, 0) = 1$. With the assistance of the Mathematica program `HolonomicFunctions` by Christoph Koutschan (which may be found at the following link), we can produce rational expressions $F_a, F_b, F_c, G_a, G_b, G_c$ so that

$$H(m+1) - H(m) = F_a(a+1) - F_a(a) + F_b(b+1) - F_b(b) + F_c(c+1) - F_c(c),$$

$$H(s+1) - H(s) = G_a(a+1) - G_a(a) + G_b(b+1) - G_b(b) + G_c(c+1) - G_c(c),$$

where in both of these expressions all functions implicitly depend on the remaining parameters. But this means that the sum $\sum_{a,b,c} H(m+1) - H(m)$ telescopes (and in fact evaluates to zero, since $F_a(0) = \lim_{x\to\infty} F_a(x) = 0$ and similar for F_b, F_c), meaning that in fact $\overline{H}(m, s)$ is independent of m. Similarly the sum $\sum_{a,b,c} H(s+1) - H(s)$ telescopes and thus $\overline{H}(m, s)$ is independent of s, meaning that $\overline{H}(m, s) = 1$ for all m, s as desired. $\square$

We list the "certificates" $F_a, F_b, F_c, G_a, G_b, G_c$ below. In all cases, $H = H(m, s, a, b, c)$ is the summand defined above.

$$\frac{F_a}{H} = \frac{2abc - 2a^2bc + 4abcm - 2a^2bcm + 2abcm^2 + a^2s - a^3s - 2abs + a^2bs - 2acs + a^2cs - 2abcs + 2a^2ms - a^3ms - 4abms + a^2bms - 4acms + a^2cms - 2abcms + a^2m^2s - 2abm^2s - 2acm^2s + 3as^2 - 2a^2s^2 + abs^2 + acs^2 + 6ams^2 - 2a^2ms^2 + abms^2 + acms^2 + 3am^2s^2 - as^3 - ams^3}{2(-1+a+b-m)(-1+a+c-m)(1+2m)s(-1+a-m+s)},$$

$$\frac{F_b}{H} = \frac{-2ab + 2a^2b + 2ab^2 + 2abc + 2bm - 8abm + 4a^2bm - 2b^2m + 4ab^2m - 2bcm + 4abcm + 2b^2cm + 4bm^2 - 6abm^2 - 2b^2m^2 - 2bcm^2 + 2bm^3 - bs - abs + b^2s + bcs - 2bms - abms + b^2ms + bcms - bm^2s - bs^2 - bms^2}{2(-1+a+b-m)(-1+a+c-m)(1+2m)s},$$

$$\frac{F_c}{H} = \frac{2abc - 2bcm + 4abcm + 2bc^2m - 2bcm^2 + cs - 3acs + bcs - c^2s + 4cms - 5acms + bcms - 3c^2ms + 3cm^2s - cs^2 - cms^2}{2(1+a+b)(-1+a+c-m)(1+2m)s},$$

$$\frac{G_a}{H} = \frac{a}{m-s},$$

$$\frac{G_b}{H} = \frac{ab - a^2b - ab^2 - abc - bm + 2abm + b^2m + bcm - bm^2 + bs - b^2s - bcs + bs^2}{(-1+b-s)(1-c+s)(-m+s)},$$

$$\frac{G_c}{H} = \frac{a^2c + abc - 2acm - bcm + cm^2 + acs + bcs - cms}{(1+a+b)(m-s)(1-c+s)}.$$

Finally, we verify one more formula which is used in proving the orthogonality of certain eigenvectors in Theorem 6.2.

Lemma A.2. *Using the notation in Section 6, in particular Eq. (6.1), Eq. (6.10) and Eq. (6.11), we have*

$$\langle f_T^{m,\ell_1}, f_T^{m,\ell_2} \rangle = 0$$

for any integer $m \in \{0, 1, \cdots, \lfloor n/2 \rfloor\}$ *and* $\ell_1 \neq \ell_2 \in \{0, 1, \cdots, n-2m\}$*, where* T *is the column reading tableau of shape* $(n-m, m)$*.*

Proof. First, we unpack the definitions to write the left-hand side as an explicit sum. We have

$$\begin{aligned} \langle f_T^{m,\ell_1}, f_T^{m,\ell_2} \rangle &= \sum_{i,j=0}^{n-2m} T_{m,n}^{(\ell_1)}(i) T_{m,n}^{(\ell_2)}(j) \langle g_T^{m,i}, g_T^{m,j} \rangle \\ &= \frac{2^m}{n+1} \sum_{i=0}^{n-2m} T_{m,n}^{(\ell_1)}(i) T_{m,n}^{(\ell_2)}(i) \frac{\binom{n-2m}{i}}{\binom{n}{m+i}} \end{aligned}$$

by the orthogonality of the $g_T^{m,i}$s and the computation Eq. (6.19). Writing out the $T_{m,n}^{(\ell)}(i)$s using Eq. (6.20), we wish to prove whether we have the triple sum

$$\sum_{i=0}^{n-2m} \sum_{j_1=0}^{i} \sum_{j_2=0}^{i} (-1)^{j_1+j_2} \binom{2m+\ell_1}{m+j_1} \binom{2m+\ell_2}{m+j_2} \binom{i}{j_1} \binom{i}{j_2} \binom{n-2m-i}{\ell_1-j_1} \binom{n-2m-i}{\ell_2-j_2} \frac{\binom{n-2m}{i}}{\binom{n}{m+i}} \stackrel{?}{=} 0$$

whenever $\ell_1 \neq \ell_2$. Multiplying by a factor of $(\ell_2 - \ell_1)$ and also noting that $\frac{\binom{n-2m}{i}}{\binom{n}{m+i}} = \frac{1}{\binom{n}{n-2m,m,m}} \binom{m+i}{i} \binom{n-m-i}{m}$, it suffices to prove that for all nonnegative integers n, m, ℓ_1, ℓ_2 with $n \geq 2m$ and $\ell_1, \ell_2 \in \{0, 1, \cdots, n-2m\}$, we have

$$\sum_{i=0}^{n-2m} \sum_{j_1=0}^{i} \sum_{j_2=0}^{i} (-1)^{j_1+j_2} \binom{2m+\ell_1}{m+j_1} \binom{2m+\ell_2}{m+j_2} \binom{i}{j_1} \binom{i}{j_2} \binom{n-2m-i}{\ell_1-j_1} \binom{n-2m-i}{\ell_2-j_2} \binom{m+i}{i} \binom{n-m-i}{m} (\ell_1-\ell_2) \stackrel{?}{=} 0.$$

However, we can produce rational expressions Q_i, Q_{j_1}, Q_{j_2} such that (letting the summand be $P = P(n, m, \ell_1, \ell_2, i, j_1, j_2)$

$$P(n+1) - P(n) = Q_i(i+1) - Q_i(i) + Q_{j_1}(j_1+1) - Q_{j_1}(j_1) + Q_{j_2}(j_2+1) - Q_{j_2}(j_2).$$

All of these rational expressions are well-defined within our range of valid n, m, ℓ_1, ℓ_2 (that is, the denominators are nonzero), and a similar telescoping argument shows that $\sum_{i,j_1,j_2} P(n+1) - P(n)$ is therefore zero. (There is one additional detail here, which is that $Q_i(0)$ is not identically zero but is antisymmetric in j_1, j_2, so the total sum is again zero.) Thus the sum is independent of n whenever $\ell_1 \neq \ell_2$. So in particular we can plug in $n = 2m$, so that the only allowed term in the summation is $i = 0, j_1 = 0, j_2 = 0$. Then the $\binom{n-2m-i}{\ell_1-j_1}$ and $\binom{n-2m-i}{\ell_2-j_2}$ factors show that the only nonzero contribution can come if $\ell_1 = \ell_2 = 0$ (and in fact that term is also zero because of the $(\ell_1 - \ell_2)$ factor in the summand). Thus we've proven our desired identity, concluding the proof. □

We conclude by listing out the certificates Q_i, Q_{j_1}, Q_{j_2}, where P is the full summand described above.

$$\frac{Q_i}{P} = \frac{-j_1 + 2ij_1 - i^2 j_1 + j_2 - 2ij_2 + i^2 j_2 + 3j_1 m - 3ij_1 m - 3j_2 m + 3ij_2 m - 2j_1 m^2 + 2j_2 m^2 - 2j_1 n + 2ij_1 n + 2j_2 n - 2ij_2 n + 3j_1 mn - 3j_2 mn - j_1 n^2 + j_2 n^2}{(\ell_1 - \ell_2)(-1 + i - j_1 + \ell_1 + 2m - n)(-1 + i - j_2 + \ell_2 + 2m - n)},$$

$$\frac{Q_{j_1}}{P} = \frac{-j_1^2 - j_1 m}{(1 + i - j_1)(\ell_1 - \ell_2)},$$

$$\frac{Q_{j_2}}{P} = \frac{j_2^2 + j_2 m}{(1 + i - j_2)(\ell_1 - \ell_2)}.$$

B Proof of ℓ_2 cutoff of the binary Burnside process

In this last section, we show the calculations for ℓ^2 cutoff of the average mixing time for the binary Burnside process. This involves some careful bounds with binomial coefficients, similarly to those used in the proof of Theorem 1.1 but with more refined estimates.

Proof of Theorem 5.2. From Theorem 1.1, we have the expression

$$\chi^2_{\text{avg}}(\ell) = \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{2k} \left(\frac{\binom{2k}{k}^2}{2^{4k}} \right)^{2\ell}.$$

As in the previous proof, we have the bounds on the central binomial coefficient

$$\frac{1}{\pi k} \exp\left(-\frac{1}{4k}\right) < \frac{\binom{2k}{k}^2}{2^{4k}} < \frac{1}{\pi k}.$$

In particular, since $\exp\left(-\frac{1}{4k}\right)^{2\ell}$ is bounded by a constant for the values $\ell \sim \frac{n}{\log n}$ that we are considering, it suffices to prove the result with $\chi^2_{\text{avg}}(\ell)$ replaced by

$$\overline{\chi}^2_{\text{avg}}(\ell) = \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{2k} \frac{1}{(\pi k)^{2\ell}}.$$

Like before, for simplicity of notation, take n to be a multiple of 4 and use analogous bounds with slightly offset indices otherwise. First we do the simpler proof of (1); that is, for $\ell = (1-\varepsilon)\frac{\log 2}{2}\frac{n}{\log n}$, we have $\overline{\chi}^2_{\text{avg}}(\ell) \to \infty$. Indeed, the middle term ($k = \frac{n}{4}$) of the summation has asymptotics

$$\begin{aligned}\binom{n}{n/2} \frac{1}{(\frac{\pi}{2}n)^{2\ell}} &\sim \frac{2^n}{\sqrt{\pi n/2}} \exp\left(-(1-\varepsilon)\log 2 \frac{n}{\log n} \log\left(\frac{\pi}{2}n\right)\right) \\ &= \frac{2^{\varepsilon n} 2^{(\varepsilon-1)\log\frac{\pi}{2}\frac{n}{\log n}}}{\sqrt{\pi n/2}},\end{aligned}$$

and this expression indeed diverges to $+\infty$ as $n \to \infty$.

Now for the proof of (2), we will choose $2\ell = \left(1 + \frac{c}{\log n}\right)\frac{n \log 2}{\log(\frac{\pi}{4}n)}$, where $c = c_n$ is of order 1 and is determined in Eq. (B.2). (The result then follows if we show that the sum tends to zero for this choice of ℓ.) If we parametrize $k = \frac{n}{4} + j$, meaning that $2k = \frac{n}{2} + 2j$ (and $-\frac{n}{4} < j < \frac{n}{4}$), then we have

$$\begin{aligned}\frac{1}{(\pi k)^{2\ell}} &= e^{-2\ell \log(\pi k)} \\ &= \exp\left(-\left(1+\frac{c}{\log n}\right)\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)} \log\left(\pi\left(\frac{n}{4}+j\right)\right)\right) \\ &= \exp\left(-\left(1+\frac{c}{\log n}\right)\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\left(\log\left(\frac{\pi}{4}n\right) + \log\left(1+\frac{4j}{n}\right)\right)\right) \\ &= \exp\left(-\left[n\log 2 + \frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\log\left(1+\frac{4j}{n}\right) + \frac{c}{\log n}n\log 2 + \frac{c}{\log n}\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\log\left(1+\frac{4j}{n}\right)\right]\right).\end{aligned} \tag{B.1}$$

This last expression in Eq. (B.1) must be multiplied by $\binom{n}{\frac{n}{2}+2j}$ and then summed over $-\frac{n}{4} < j < \frac{n}{4}$, and we must show that the sum tends to zero. This is shown in zones:

- *Zone 1:* $0 \leq j < \frac{n}{4}$. For these cases, we can bound the binomial coefficient crudely by 2^n (which cancels out the first term in the exponent of Eq. (B.1)). Using that $\frac{x}{1+x} < \log(1+x) < x$ for all $x > 0$ and plugging in $x = \frac{4j}{n}$ yields that $-\log\left(1+\frac{4j}{n}\right) < -\frac{4j/n}{1+4j/n} < -\frac{2j}{n}$. Thus, the second and fourth terms in the exponent of Eq. (B.1) may be bounded via

$$-\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\log\left(1+\frac{4j}{n}\right) < -\frac{(2\log 2)j}{\log\left(\frac{\pi}{4}n\right)},$$

$$-\frac{c}{\log n}\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\log\left(1+\frac{4j}{n}\right) < -\frac{c}{\log n}\frac{(2\log 2)j}{\log\left(\frac{\pi}{4}n\right)}.$$

  Further observe that for any $A > 0$, we have

$$\sum_{j=0}^{\infty} e^{-Aj/\log n} = \frac{1}{1-e^{-A/\log n}} \sim \frac{\log n}{A},$$

  so that combining bounds together, the total contribution to $\overline{\chi}^2_{\text{avg}}(\ell)$ over $0 \leq j < \frac{n}{4}$ is asymptotically bounded from above by

$$\frac{\log n}{2\log 2}\exp\left(-\frac{cn\log 2}{\log n}\right),$$

  and (with the choice of c below in Eq. (B.2)) this indeed goes to zero as $n \to \infty$.

- *Zone 2:* $-\frac{n}{6} < j < 0$. Write $j' = -j$ for clarity. Because $-\log(1-y) < \frac{y}{1-y}$ for $0 < y < 1$, we have

$$-\log\left(1-\frac{4j'}{n}\right) < \frac{4j'/n}{1-4j'/n} < \frac{12j'}{n}$$

  for all $j' < \frac{n}{6}$. If we write $\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\frac{12j'}{n} = \alpha j'$ (so $\alpha = \frac{12\log 2}{\log\left(\frac{\pi}{4}n\right)}$), then

$$\sum_{j'=0}^{\frac{n}{6}-1} e^{\alpha j'} = \frac{e^{\alpha n/6}-1}{e^{\alpha}-1} \sim \frac{1}{\alpha}e^{\alpha n/6} = \frac{\log\left(\frac{\pi}{4}n\right)}{12\log 2}\exp\left(\frac{2n\log 2}{\log\left(\frac{\pi}{4}n\right)}\right).$$

  In this zone, we still bound all binomial coefficients crudely by $\binom{n}{2k} < 2^n$. Since the expression Eq. (B.1) has third term $-\frac{cn\log 2}{\log n}$ in the exponent, it follows that if

$$c = 2 + \frac{\log n}{n\log 2}\left(\log\log\left(\frac{\pi}{4}n\right)+\theta\right), \tag{B.2}$$

  then the total contribution to $\overline{\chi}^2_{\text{avg}}(\ell)$ over $-\frac{n}{6} < j < 0$ is bounded above by a constant times $e^{-\theta}$, which tends to zero by choosing $\theta = \theta_n$ increasing (but increasing slowly enough so that c_n stays bounded).

- *Zone 3: $-\frac{n}{4}+n^{0.9}<j\le-\frac{n}{6}$.* For this zone, it is important to bound the binomial coefficient. We have

$$\binom{n}{a}=\frac{n(n-1)\cdots(n-a+1)}{a!}=\frac{n^a}{a!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{a-1}{n}\right)$$
$$\le n^a\exp\left(-\frac{1}{n}\sum_{j=1}^{a-1}j\right)$$
$$\le\frac{n^a e^{-\binom{a}{2}/n}}{a!}.$$

Thus for $a=\theta n$ for $0<\theta<\frac{1}{2}$, we have by the above bound and Stirling's formula that

$$\binom{n}{a}\le\exp\left(a\log n-\frac{1}{2n}a^2-a\log a+a-\frac{1}{2}\log a+O(1)\right)$$
$$=\exp\left(-\theta n\log\theta-\frac{\theta^2}{2}n+\theta n-\frac{1}{2}\log(\theta n)+O(1)\right)$$
$$=\exp\left(n\left(\theta-\frac{\theta^2}{2}-\theta\log\theta\right)-\frac{1}{2}\log n+O(1)\right).$$

In our expression for $\overline{\chi}^2_{\text{avg}}(\ell)$, this factor is being multiplied by $\frac{1}{\left(\frac{\pi a}{2}\right)^{2\ell}}=\exp\left(-2\ell\log\left(\frac{\pi}{2}\theta n\right)\right)$, so that if $2\ell=(1+\varepsilon)\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}$, this factor is of the form

$$\exp\left(-(1+\varepsilon)\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\log\left(\frac{\pi}{2}\theta n\right)\right)=\exp\left(-(1+\varepsilon)n\log 2\left(1+\frac{\log(2\theta)}{\log\left(\frac{\pi}{4}n\right)}\right)\right).$$

Multiplying these exponentials together, the contribution to $\overline{\chi}^2_{\text{avg}}(\ell)$ has lead term in the exponent $n\left(\theta-\frac{\theta^2}{2}-\theta\log\theta-(1+\varepsilon)\log 2\right)$, and for $\theta<\frac{1}{6}$ this is bounded from above by $-0.2n$. Furthermore, the remaining positive factor in the exponential $\exp\left(-(1+\varepsilon)n\log 2\frac{\log(2\theta)}{\log\left(\frac{\pi}{4}n\right)}\right)$ is asymptotically bounded from above by $e^{0.15n}$ as long as $\theta\gtrsim n^{-0.1}$ and $\varepsilon<1$. Thus in this zone each term is exponentially small and so the sum is also exponentially small.

- *Zone 4: $-\frac{n}{4}<j\le-\frac{n}{4}+n^{0.9}$* (that is, $1\le 2k\le 2n^{0.9}$). Here we can just bound $\frac{1}{\pi k}\le\frac{1}{\pi}$, so that in this zone we have

$$\binom{n}{2k}\frac{1}{(\pi k)^{2\ell}}\le n^{2k}\frac{1}{\pi^{2\ell}}$$
$$=\exp\left(2k\log n-2\ell\log\pi\right)$$
$$\le\exp\left(2n^{0.9}\log n-\frac{n\log 2}{\log\left(\frac{\pi}{4}n\right)}\log\pi\right).$$

Since $n^{0.9}$ times this quantity goes to zero as $n\to\infty$, the contribution to $\overline{\chi}^2_{\text{avg}}(\ell)$ in this zone also goes to zero.

Combining the bounds across the different zones yields the result of (2), completing the proof. □

Remark. *The delicate choice of c for zone 2 was developed as part of an argument to prove a limiting "shape theorem" for $\chi^2_{avg}(\ell)$. Directly using $2\ell = (1+\varepsilon)\log 2\frac{n}{\log n}$, perhaps the calculations for zones may be merged. We leave the argument in its current form in case a reader wants to work more and prove a shape theorem.*

References

[AF02] D. Aldous and J. A. Fill. Reversible Markov Chains and Random Walks on Graphs, 2002. Unfinished monograph, recompiled 2014, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[Ald83] D. J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de probabilités de Strasbourg*, 17:243–297, 1983.

[BD92] D. Bayer and P. Diaconis. Trailing the Dovetail Shuffle to its Lair. *The Annals of Applied Probability*, 2(2):294 – 313, 1992.

[BD24] L. Bartholdi and P. Diaconis. An algorithm for uniform generation of unlabeled trees (Pólya trees), with an extension of Cayley's formula, 2024, 2411.17613.

[BDPX05] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao. Symmetry Analysis of Reversible Markov Chains. *Internet Mathematics*, 2(1):31 – 71, 2005.

[Bur17] A. Burcroff. Johnson Schemes and Certain Matrices with Integral Eigenvalues, 2017.

[Chi78] T. Chihara. *An Introduction to Orthogonal Polynomials*. Ellis Horwood series in mathematics and its applications. Gordon and Breach, 1978.

[CSC13] G.-Y. Chen and L. Saloff-Coste. Comparison of cutoffs between lazy walks and Markovian semigroups. *Journal of Applied Probability*, 50(4):943 – 959, 2013.

[DG89] P. Diaconis and C. Greene. Applications of Murphy's elements. Technical report, Stanford University, 1989.

[DGM90] P. Diaconis, R. L. Graham, and J. A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures & Algorithms*, 1(1):51–72, 1990, https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.3240010105.

[DH25] P. Diaconis and M. Howes. Random sampling of partitions and contingency tables: Two practical examples of the Burnside process, 2025, 2503.02818.

[Dia88] P. Diaconis. Group Representations in Probability and Statistics. *Lecture Notes-Monograph Series*, 11:i–192, 1988.

[Dia96] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

[Dia05] P. Diaconis. Analysis of a Bose–Einstein Markov chain. *Annales de l'I.H.P. Probabilités et statistiques*, 41(3):409–418, 2005.

[DKSC08] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs Sampling, Exponential Families and Orthogonal Polynomials. *Statistical Science*, 23(2):151 – 178, 2008.

[DS81] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57:159–179, 1981.

[DSC93] P. Diaconis and L. Saloff-Coste. Comparison Theorems for Reversible Markov Chains. *The Annals of Applied Probability*, 3(3):696 – 730, 1993.

[DSC96a] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695 – 750, 1996.

[DSC96b] P. Diaconis and L. Saloff-Coste. Nash Inequalities for Finite Markov Chains. *Journal of Theoretical Probability*, 9(2):459–510, 1996.

[DT25] P. Diaconis and N. Tung. Poisson approximation for large permutation groups. *Adv. Appl. Math.*, 167(C), May 2025.

[DZ21] P. Diaconis and C. Zhong. Hahn polynomials and the Burnside process. *The Ramanujan Journal*, 61(2):567–595, September 2021.

[DZ25] P. Diaconis and C. Zhong. Counting the number of group orbits by marrying the Burnside process with importance sampling, 2025, 2501.11731.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume 1*. An Introduction to Probability Theory and Its Applications. Wiley, 1968.

[FH91] W. Fulton and J. Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York, 1991.

[GJ98] L. A. Goldberg and M. Jerrum. The "Burnside Process" Converges Slowly. In M. Luby, J. D. P. Rolim, and M. Serna, editors, *Randomization and Approximation Techniques in Computer Science*, pages 331–345, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg.

[GLT24] R. M. Guralnick, M. Larsen, and P. H. Tiep. Character levels and character bounds for finite classical groups. *Inventiones Mathematicae*, 235(1):151–210, January 2024.

[GS02] A. L. Gibbs and F. E. Su. On Choosing and Bounding Probability Metrics. *International Statistical Review*, 70(3):419–435, 2002, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-5823.2002.tb00178.x.

[GVT85] T. Gilovich, R. Vallone, and A. Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3):295–314, 1985.

[HH24] J. Hermon and X. Huang. Cutoff for random Cayley graphs of nilpotent groups, 2024, 2403.12355.

[HOT21] J. Hermon and S. Olesker-Taylor. Cutoff for Almost All Random Walks on Abelian Groups, 2021, 2102.02809.

[How25] M. Howes. Limit profiles for the Burnside process on Sylow double cosets, 2025. In preparation.

[HP18] J. Hermon and Y. Peres. On sensitivity of mixing times and cutoff. *Electronic Journal of Probability*, 23(none):1 – 34, 2018.

[Ism05] M. E. H. Ismail. *Classical and Quantum Orthogonal Polynomials in One Variable*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2005.

[Jam78] G. James. *The Representation Theory of the Symmetric Groups*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 1978.

[Jer93] M. Jerrum. Uniform sampling modulo a group of symmetries using Markov chain simulation. In *Expanding Graphs*, volume 10 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 1993.

[Kou10] C. Koutschan. A Fast Approach to Creative Telescoping. *Mathematics in Computer Science*, 4:259–266, 2010.

[KS83] J. Kemeny and J. Snell. *Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix"*. Undergraduate Texts in Mathematics. Springer New York, 1983.

[KS96] R. Koekoek and R. F. Swarttouw. The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue, 1996, math/9602214.

[KZ09] K. Khare and H. Zhou. Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *The Annals of Applied Probability*, 19(2):737 – 777, 2009.

[LP15] E. Lubetzky and Y. Peres. Cutoff on all Ramanujan graphs. *Geometric and Functional Analysis*, 26:1190–1216, 2015.

[LP17] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, second edition, 2017. With contributions by Elizabeth L. Wilmer, with a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.

[MSC23] M. Murugan and L. Saloff-Coste. Harnack inequalities and Gaussian estimates for random walks on metric measure spaces. *Electronic Journal of Probability*, 28(none):1 – 81, 2023.

[Mur81] G. Murphy. A new construction of Young’s seminormal representation of the symmetric groups. *Journal of Algebra*, 69(2):287–297, 1981.

[Nes24] E. Nestoridi. Comparing limit profiles of reversible Markov chains. *Electronic Journal of Probability*, 29(none):1 – 14, 2024.

[NPWZ97] I. Nemes, M. Petkovšek, H. S. Wilf, and D. Zeilberger. How to do Monthly Problems With Your Computer. *The American Mathematical Monthly*, 104(6):505–519, 1997, https://doi.org/10.1080/00029890.1997.11990673.

[OTTT25] S. Olesker-Taylor, L. Teyssier, and P. Thévenin. Sharp character bounds and cutoff for symmetric groups, 2025, 2503.12735.

[Pag23] J. E. Paguyo. Mixing times of a Burnside process Markov chain on set partitions, 2023, 2207.14269.

[Pan18] C. Y. A. Pang. Lumpings of Algebraic Markov Chains Arise from Subquotients. *Journal of Theoretical Probability*, 32(4):1804–1844, June 2018.

[PR04] Y. Peres and D. Revelle. Mixing Times for Random Walks on Finite Lamplighter Groups. *Electronic Journal of Probability*, 9(none):825 – 845, 2004.

[Rah20] J. Rahmani. Mixing Times for the Commuting Chain on CA Groups. *Journal of Theoretical Probability*, 35:457–483, 2020.

[Ram95] A. Ram. Seminormal Representations of Weyl Groups and Iwahori-Hecke Algebras. *Proceedings of the London Mathematical Society*, 75, 1995.

[RT96] D. A. Redelmeier and A. Tversky. On the belief that arthritis pain is related to the weather. *Proceedings of the National Academy of Sciences*, 93(7):2895–2896, 1996, https://www.pnas.org/doi/pdf/10.1073/pnas.93.7.2895.

[SC97] L. Saloff-Coste. *Lectures on finite Markov chains*, pages 301–413. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

[SCZ08] L. Saloff-Coste and J. Zúñiga. Refined estimates for some basic random walks on the symmetric and alternating groups. *ALEA Latin American Journal of Probability and Mathematical Statistics*, 4:359–392, 2008.

[Ser77] J.-P. Serre. *Linear Representations of Finite Groups*. Graduate Texts in Mathematics. Springer New York, 1977.

[Stă01] P. Stănică. Good lower and upper bounds on binomial coefficients. *Journal of Inequalities in Pure & Applied Mathematics*, 2(3), 2001.

[SV23] D. Stefankovic and E. Vigoda. Lecture Notes on Spectral Independence and Bases of a Matroid: Local-to-Global and Trickle-Down from a Markov Chain Perspective, 2023, 2307.13826.

[Tey20] L. Teyssier. Limit profile for random transpositions. *The Annals of Probability*, 48(5):2323 – 2343, 2020.

[TK74] A. Tversky and D. Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 1974.