Michigan Math. J. 48 (2000)

Analysis of Systematic Scan Metropolis Algorithms Using Iwahori–Hecke Algebra Techniques

PERSI DIACONIS & ARUN RAM

Dedicated to our friend Bill Fulton

1. Introduction

When faced with a complex task, is it better to be systematic or to proceed by making random adjustments? We study aspects of this problem in the context of generating random elements of a finite group. For example, suppose we want to fill n empty spaces with zeros and ones such that the probability of configuration $x = (x_1, \dots, x_n)$ is $\theta^{n-|x|}(1-\theta)^{|x|}$, with $|x|$ the number of ones in x. A systematic scan approach works left to right, filling each successive place with a θ coin toss. A random scan approach picks places at random, and a given site may be hit many times before all sites are hit. The systematic approach takes order n steps and the random approach takes order $\frac{1}{4}n\log n$ steps.

Realistic versions of this toy problem arise in image analysis and Ising-like simulations, where one must generate a random array by a Monte Carlo Markov chain. Systematic updating and random updating are competing algorithms that are discussed in detail in Section 2. There are some successful analyses for random scan algorithms, but the intuitively appealing systematic scan algorithms have resisted analysis.

Our main results show that the binary problem just described is exceptional; for the examples analyzed in this paper, systematic and random scans converge in about the same number of steps.

Let W be a finite Coxeter group generated by simple reflections $s_1, s_2, \dots, s_n$, where $s_i^2 = \text{id}$. For example, W may be the permutation group S_{n+1} with $s_i = (i, i+1)$. The length function $\ell(w)$ is the smallest k such that $w = s_{i_1}s_{i_2}\cdots s_{i_k}$. Fix $0 < \theta \le 1$ and define a probability distribution on W by

$$\pi(w) = \frac{\theta^{-\ell(w)}}{P_W(\theta^{-1})}, \quad \text{where} \quad P_W(\theta^{-1}) = \sum_{w\in W} \theta^{-\ell(w)} \tag{1.1}$$

is the normalizing constant. Thus $\pi(w)$ is smallest when $w = \text{id}$ and, as $\theta \to 1$, π tends to the uniform distribution. These nonuniform distributions arise in statistical work as Mallows models. Background and references are in Section 2e.

Received March 21, 2000. Revision received April 10, 2000.

Research of the first author was supported in part by National Science Foundation grant DMS-9504379. Research of the second author was supported in part by National Science Foundation grant DMS-9622985.

A standard Monte Carlo Markov chain algorithm for sampling from π is the Metropolis algorithm with a systematic scan. This algorithm cycles through the generators in order. If multiplying by the current generator increases length then this multiplication is made. If the length decreases, then the multiplication is made with probability θ and omitted with probability $1-\theta$. One scan uses $s_1, s_2, \ldots, s_{n-1}, s_n, s_n, s_{n-1}, \ldots, s_1$, in order. Define

$$K(w, w') = \text{the chance that a systematic scan started at } w \text{ ends in } w'. \tag{1.2}$$

Repeated scans of the algorithm are defined by

$$K^{\ell}(w, w') = \sum_{w''} K^{\ell-1}(w, w'')K(w'', w'), \quad \ell \geq 2. \tag{1.3}$$

In Section 2c and 4a we show that this Markov chain has π as unique stationary distribution.

The main results of this paper derive sharp results on rates of convergence for these walks. As an example of what our methods give, we show that scans of order n are necessary and suffice to reach stationarity on the symmetric group starting from the identity. More precisely, we prove the following.

THEOREM 1.4. *Let S_n be the permutation group on n letters. Fix θ, $0 < \theta \leq 1$. Let $K_1^{\ell}(w) = K^{\ell}(\mathrm{id}, w)$ be the systematic scan chain on S_n defined by* (1.2) *and* (1.3). *For $\ell = n/2 - (\log n)/(\log \theta) + c$ with $c > 0$,*

$$\|K_1^{\ell} - \pi_\theta\|_{TV}^2 \leq \left(e^{\theta^{2c+1}} - 1\right) + n!\,\theta^{n^2/8 - n(\log n)/(\log\theta) + n(c+1/4)}. \tag{1.5}$$

Conversely, if $\ell \leq n/4$ then, for fixed θ, $\|K_1^{\ell} - \pi\|_{TV}$ tends to 1 *as $n \to \infty$.*

The total variation norm is defined in Section 2a. Note that the upper bound in (1.5) tends to zero for c large, so that about $n/2$ scans suffice to reach stationarity. The lower bound shows that this is of the right order for large n.

Each scan uses $2n$ multiplications. Thus, Theorem 1.4 implies that the systematic scan approach reaches stationarity in n^2 operations up to lower-order terms. We also conjecture that the random scan approach (see Section 2b) for this example takes order n^2 operations. Further, in Section 7, we prove that the scan based on the sequence

$$(s_1, s_2, \ldots, s_n, s_n, \ldots, s_1), (s_1, \ldots, s_{n-1}, s_{n-1}, \ldots, s_1), \ldots, (s_1, s_2, s_2, s_1), (s_1, s_1)$$

converges in one pass. Thus, again, up to lower-order terms, n^2 operations suffice to reach stationarity. These results show that various different scanning strategies take the same number of operations to reach stationarity.

One novel aspect of the present arguments is our use of the Iwahori–Hecke algebra H spanned by the symbols $\{T_w\}_{w \in W}$. This algebra is generated by $T_i = T_{s_i}$ $(1 \leq i \leq n)$ with the relations

$$T_i T_w = \begin{cases} T_{s_i w} & \text{if } \ell(s_i w) > \ell(w), \\ qT_{s_i w} + (q-1)T_w & \text{if } \ell(s_i w) < \ell(w). \end{cases}$$

We have succeeded in giving an algebraic interpretation of the Markov chain $K(w, w')$ as multiplication in the Iwahori–Hecke algebra H. From there, knowledge of the center of H (via a result of Brieskorn–Saito and Deligne) allows us to diagonalize $K(w, w')$ explicitly. Convergence bounds are given in terms of the eigenvalues and the generic degrees of representation theory. Then calculus leads to results like Theorem 1.4.

Section 2 collects together probabilistic background and tools. We explain Markov chains, the Metropolis algorithm, and systematic scans, and we relate the basic Metropolis chain to a natural walk on the chambers of a building. In Section 2e we develop properties of the measures π. Some of these are new even for reflection groups of type A (the symmetric group). These properties will be applied to prove lower bounds for walks as in Theorem 1.4.

Section 3 collects together representation theoretic background and tools and connects the representation theory to Markov chains. Section 4 connects Hecke algebras to the Metropolis algorithm and specializes the results from Section 3. A basic upper bound for convergence is derived by relating two inner products.

Sections 5 and 6 derive results for the hypercube and the dihedral groups. Here we find that both the systematic and random scans converge in about the same number of steps—the differences are only in the lead term constants (which are functions of θ).

Section 7 derives results for two different systematic scanning plans for the symmetric group. Though we do not have the space to treat further examples in this paper, it should be remarked that the methods of Section 7 should also produce analogous results for the Weyl groups of type B_n and the imprimitive complex reflection groups $G(r, 1, n)$. The long and short systematic scans can be defined in a similar way and the representation theory goes through without problems (see [AK; Hf; R]). The remaining necessary ingredient is an analog of Lemma 7.2.

Acknowledgment. We are thankful to Ruth Lawrence for early efforts to help understand deformed random walks.

2. Probabilistic Background

In this section we give background material for Markov chains, the Metropolis algorithm, and systematic scans. In Section 2d we interpret the basic walk as a walk on flags and the chambers of a building, and in Section 2e we derive basic properties of the stationary distributions.

2a. Markov Chains

Background for Markov chains may be found in any standard probability text (see e.g. [F, Chap. XV]). For the quantitative theory developed here, see [S-C] and the references therein.

Let X be a finite set. A *Markov chain* on X is a matrix $K = (K(x, y))_{x,y\in X}$ such that

$$K(x, y) \in [0, 1] \quad \text{and} \quad \sum_{y\in X} K(x, y) = 1.$$

The set X is the *state space,* and $K(x, y)$ gives the probability of moving from x to y in one step. Powers of the matrix K give the probability of moving from x to y in more steps. For example,

$$K^2(x, y) = \sum_{z \in X} K(x, z)K(z, y)$$

indicates that, in order to move from x to y in two steps, the chain must move to z and then from z to y. The chain is *irreducible and aperiodic* if there is an $\ell > 0$ such that $K^\ell(x, y) > 0$ for all $x, y \in X$. The chain K is *reversible* if there is a *stationary distribution* $\pi : X \to [0, 1]$, $\sum_{x \in X} \pi(x) = 1$, such that, for all $x, y \in X$,

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

For irreducible aperiodic K, reversibility implies that, for each $x \in X$, the real numbers $K^\ell(x, y)$ converge to $\pi(y)$ as $\ell \to \infty$.

The quantitative theory of Markov chains studies the speed of convergence. The *total variation distance* of $K^\ell(x, \cdot)$ to π is defined by

$$\|K_x^\ell - \pi\|_{TV} = \max_{A \subseteq X} \left| \sum_{y \in A} K^\ell(x, y) - \pi(y) \right|.$$

Using the set $A = \{y \in X \mid K^\ell(x, y) > \pi(y)\}$, it is easily shown that

$$\|K_x^\ell - \pi\|_{TV} = \frac{1}{2} \sum_{y \in X} |K^\ell(x, y) - \pi(y)|. \tag{2.1}$$

Let $L^2(\pi)$ be the space of functions $f : X \to \mathbb{R}$ with the norm

$$\langle f, g\rangle_2 = \sum_{x \in X} f(x)g(x)\pi(x). \tag{2.2}$$

The following lemma provides a relation between the total variation and the $L^2(\pi)$ norms. This bound is the primary tool for studying rates of convergence of Markov chains.

LEMMA 2.3. *Let $f \in L^2(\pi)$. Then $\|f\|_{TV}^2 \le \frac{1}{4}\|f/\pi\|_2^2$.*

Proof. By the Cauchy–Schwartz inequality,

$$\begin{aligned}\|f\|_{TV}^2 &= \frac{1}{4}\left(\sum_{x \in X} \frac{|f(x)|}{\sqrt{\pi(x)}}\sqrt{\pi(x)}\right)^2 \\ &\le \frac{1}{4}\left(\sum_{x \in X} \frac{f(x)^2}{\pi(x)}\right)\left(\sum_{x \in X} \pi(x)\right) = \frac{1}{4}\left\langle \frac{f}{\pi}, \frac{f}{\pi}\right\rangle_2. \qquad \square\end{aligned}$$

2b. Systematic Scan Algorithms

Let π be a probability distribution on a finite set X, and let $K_1, K_2, \ldots, K_n$ be Markov chains on X each having stationary distribution π. Then any product $K_{i_\ell} K_{i_{\ell-1}} \cdots K_{i_1}$ has stationary distribution π, and a choice of an infinite sequence $\{i_\ell\}_{\ell=1}^\infty$ gives a scanning strategy. A random choice of indices gives a random scanning stategy. If each K_i is reversible for π, then $K_1 K_2 \cdots K_{n-1} K_n K_n K_{n-1} \cdots K_2 K_1$ is an example of a reversible systematic scanning strategy (whereas $K_1 \cdots K_n$ is not necessarily reversible).

In routine applications of the Metropolis algorithm to image analysis and Ising-like models, the state space has coordinates. Randomized strategies choose a coordinate at random and attempt to change it. Systematic strategies cycle through the coordinates in various orders. Fishman [Fi] reviews the literature on scanning strategies and gives some practical comparison. The scheme underlying Theorem 1.4 is Fishman's Plan 3.

There has been some rigorous work on rates of convergence for systematic scans in a related case: Gaussian distribution of coordinates with the stochastic updating done by the heat bath algorithm (also known as Glauber dynamics or the Gibbs sampler). One fascinating study by Goodman and Sokal [GS] relates scanning strategies to standard approaches for solving large linear systems. They show that the systematic scan heat bath algorithm is a stochastic analog of the Gauss–Seidel algorithm. Moreover, they show how previous analyses of Gauss–Seidel give the eigenvalues of its stochastic counterpart. Amit [A1; A2] and Amit and Grenander [AG] have pushed forward and carried out these ideas to give some comparison of systematic and randomized sweeps in the Gaussian case. Their approach uses the fact that the heat bath algorithm is a projection operator. In the Gaussian case, the problem reduces to the computation of angles between subspaces of a Hilbert space; Baronne and Frigessi [BF] and Roberts and Sahu [RS] are related references.

2c. The Metropolis Algorithm

The Metropolis algorithm gives a way of changing the stationary distribution of a given Markov chain into any distribution; it was invented by Metropolis and colleagues [MRRTT]. A clear description is in Hammersley and Handscomb [HH], and a recent survey appears in [DS].

Let X be a finite set. Let $P(x, y) = P(y, x)$ be a symmetric Markov matrix on X, and let π be a fixed probability distribution on X. Form a new chain by the following recipe:

$$M(x, y) = \begin{cases} P(x, y) & \text{if } x \neq y \text{ and } \pi(y) \geq \pi(x), \\ P(x, y)\dfrac{\pi(y)}{\pi(x)} & \text{if } x \neq y \text{ and } \pi(y) < \pi(x), \\ P(x, x) + \displaystyle\sum_{\pi(z)<\pi(x)} P(x, z)\left(1 - \dfrac{\pi(z)}{\pi(x)}\right) & \text{if } x = y. \end{cases} \tag{2.4}$$

In words:

> Form the Metropolis chain from x by choosing y from $P(x, y)$. If $\pi(y) \geq \pi(x)$ then move to x. If $\pi(y) < \pi(x)$, flip a coin with chance of heads $\pi(y)/\pi(x)$. If the coin comes up heads then move to y. In all other cases, stay at x.

As shown in the references just cited, the Metropolis chain is reversible with stationary distribution π. It is of practical importance that the chain M can be run knowing π only up to a normalizing constant. Irreducibility and aperiodicity of M must be checked on a case-by-case basis.

An example of interest is $X = W$, where W is a finite real reflection group generated by simple reflections $s_1, s_2, \dots, s_n$. Let $P(x, y)$ be the Markov chain given by

$$P(x,y) = \begin{cases} 1/n & \text{if } y = s_i x \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases}$$

Here $P(x, y)$ is the usual random walk based on a generating set. It has uniform stationary distribution. Fix θ, $0 < \theta \leq 1$, and let π be as in (1.1). The Metropolis construction then gives the Markov chain

$$M(x,y) = \begin{cases} 1/n & \text{if } y = s_i x \text{ and } \ell(y) > \ell(x), \\ \theta/n & \text{if } y = s_i x \text{ and } \ell(y) < \ell(x), \\ (1/n)\sum_{\ell(s_i x)<\ell(x)}(1-\theta) & \text{if } y = x, \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

which has stationary distribution π. In Section 4a we demonstrate that this is exactly the chain given by left multiplication by a uniformly chosen generator $\tilde{T}_i$ in the Iwahori–Hecke algebra H with $q = \theta^{-1}$. Similarly, the systematic scan chain of Theorem 1.4 can be interpreted via multiplication in H.

Despite its widespread use there has been very limited success in analyzing the time to stationarity of the Metropolis algorithm. In the present paper we carry this out for the random scan Metropolis algorithm (2.5) on the hypercube (Section 5) and on the dihedral group (Section 6). Though we have not analyzed the random scan Metropolis algorithm on the symmetric group, we conjecture that order n^2 steps are necessary and sufficient to achieve stationarity. A survey of what is rigorously known appears in [DS].

Diaconis and Hanlon [DH] studied the example given by $W = S_n$, the symmetric group (so $c(w) = n - [\#\text{ of cycles in } w]$), with input chain

$$P(x,y) = \begin{cases} 1\big/\binom{n}{2} & \text{if } y = (i,j)x \text{ for some transposition } (i,j), \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

and stationary distribution $\pi(w) = z\theta^{c(w)}$, where z is a normalizing constant and $c(w)$ is the minimum number of transpositions needed to sort w. They showed that all eigenvectors of the resulting Metropolis chain are given by the coefficients of Jack's symmetric functions (expanded in terms of power sum symmetric functions), and they used the corresponding eigenvalues to give a complete analysis of the running time.

Similar analyses were carried out by Belsley [B2] and Silver [Si]. They worked in abelian groups with π proportional to $\theta^{\ell(y)}$, where ℓ is the length function with respect to a natural set of generators. In several cases they found that the eigenfunctions were natural deformations of classical orthogonal polynomials. Ross and Xu [RX] studied the random scan Metropolis algorithm on the hypercube, using its representation as a random walk on a hypergroup. It should be emphasized that, for other choices of π or in nongroup cases, careful analysis of rates of convergence for the Metropolis algorithm is completely open.

2d. Some Other Interpretations of the Walks

We have presented Theorem 1.4 in an algorithmic context. Here we show how the walk (2.5) arises geometrically on the space of flags and as the natural nearest neighbor walk on the chambers of a building. The systematic scan walks have similar interpretations.

Let $\mathbb{F}_q$ be a finite field. A *complete flag* $F = (0 = F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{n-1} \subseteq F_n = V)$ is a nested increasing sequence of subspaces of an n-dimensional vector space V over $\mathbb{F}_q$ with $\dim(F_i) = i$. A natural random walk on complete flags may be performed as follows:

choose i $(1 \le i \le n-1)$ uniformly;
replace F_i by a uniformly chosen subspace $\tilde{F}_i$ with $F_{i-1} \subseteq \tilde{F}_i \subseteq F_{i+1}$.

This walk is symmetric, irreducible, and aperiodic; it therefore has the uniform distribution as its unique stationary measure. It is instructive to think of the $q = 1$ case. Then a flag is a nested increasing chain $\{i_1\} \subseteq \{i_1, i_2\} \subseteq \cdots \subseteq \{i_1, i_2, \ldots, i_n\}$ of elements of an n set or, equivalently, a permutation $(i_1, i_2, \ldots, i_n)$. In this case the walk is multiplication by random pairwise adjacent transpositions.

The space of flags may also be identified as the chambers of a building of type A_{n-1}, and in this formulation the walk is described as follows:

> From a chamber C of the building, choose one of the adjacent chambers uniformly at random and move there.

In an elegant and readable treatment of buildings, Brown [Br] explains that flag space may be represented as G/B, with $G = \mathrm{GL}_n(\mathbb{F}_q)$ and B the subgroup of upper triangular matrices in $\mathrm{GL}_n(\mathbb{F}_q)$. Then two flags g_1B and g_2B differ in the ith step if and only if $g_1P_i = g_2P_i$, where P_i is the parabolic subgroup $P_i = B \cup Bs_iB$ (see [Br, pp. 102–103]). Thus, if flags g_1B and g_2B are i-adjacent then $g_2 = g_1b$ or $g_2 = g_1bs_ib'$ with $b, b' \in B$, so the walk on G/B moves from gB to $gg'B$ with g' uniformly chosen in B or Bs_iB. In this way, choosing an adjacent chamber of the building at random produces a B-invariant walk on G/B. Finally, the walk on flags gives rise to a natural walk on the double coset space $B\backslash G/B$ (described in more detail in Section 3b). The double coset space is identifiable with the symmetric group S_n, and the induced Markov chain is given by (2.5) with $\theta = 1/q$. A similar story holds for the natural walk on any spherical building.

2e. Properties of the Stationary Distribution

Suppose that (X, d) is a finite metric space. A simple way of building probability models on X is to fix $0 < \theta \le 1$ and $x_0 \in X$ and then define

$$\pi(x) = \frac{q^{d(x,x_0)}}{P_X(q)}, \quad \text{where } q = \theta^{-1} \text{ and } P_X(q) = \sum_{x \in X} q^{d(x,x_0)} \tag{2.7}$$

is a normalizing constant. When $q = 1$, the distribution is uniform.

Models of the form (2.7) were introduced by Mallows [M] for the study of permutations. He used the length function as a distance, $\ell(x^{-1}x_0) = d(x, x_0)$, and estimated q and x_0 to match data. Such Mallows models have had application and development for ranked and partially ranked data using a variety of metrics [CR; D; FV; Ma]. They have also been used for phylogenetic trees [BHV], classification trees [SB], and compositions [DHJLP].

One problem in studying Mallows models is that the normalizing constant $P_X(q)$ is uncomputable in general. In such cases, properties of π can be studied by simulation using the Metropolis algorithm given in Section 2c.

For the examples based on reflection groups, the normalizing constants are known; moreover, there is a simple algorithm available for exact generation from π. The properties of π are collected together here. In each case the properties are illustrated for the permutation group; some of our results are new for the original Mallows model. Further, the properties of π (particularly Property 4) are used in proving the lower bounds in Theorem 1.4.

Throughout this section we work with the model whose underlying space $X = W$ is a finite Coxeter group generated by simple reflections, x_0 is the identity element of W, and the length function is the distance on W. Thus the model is

$$\pi(w) = \frac{q^{\ell(w)}}{P_W(q)}, \quad \text{where } q = \theta^{-1} \text{ and } P_W(q) = \sum_{w \in W} q^{\ell(w)} \tag{2.8}$$

is the *Poincaré polynomial* of the group W. It is a classical theorem that the normalizing constant has a simple form:

$$P_W(q) = \prod_{i=1}^{n} \frac{q^{d_i} - 1}{q - 1} \tag{2.9}$$

for known integers d_i, the *degrees* of W (see [Hu, Thm. 3.15]). For the symmetric group S_{n+1}, $d_i = i + 1$ for $1 \le i \le n$. The Poincaré polynomial $P_W(q)$ will be used crucially in what follows.

PROPERTY 1. $\pi(w) = \pi(w^{-1})$, *since* $\ell(w) = \ell(w^{-1})$.

This invariance under inversion was first used by Mallows [M] to characterize Mallows models in a larger class of measures as follows. Suppose n objects are to be ranked by making pairwise comparisons. Suppose that the true ranking is $1 < 2 < 3 < \cdots < n$ and that a subject ranks objects i and j correctly with probability p_{ij}. Let $Q(w)$ be the chance that the comparisons lead to the permutation w, given that they are all consistent. Of course, $Q(w)$ depends on the $\binom{n}{2}$ parameters p_{ij}. Mallows proved that *if* $Q(w) = Q(w^{-1})$ then, for some real numbers q and ϕ, $Q(w) = zq^{\ell(w)}\phi^{r(w)}$ with $r(w) = \sum iw(i)$ and z a normalizing constant. He further showed that the two parameters q and ϕ were practically indistinguishable for large n and suggested setting $\phi = 1$, leading to the distribution $\pi(w)$.

PROPERTY 2. *Let* $J \subseteq \{1, 2, \ldots, n\}$, *and let* W_J *be the subgroup of* W *generated by the generators* s_i *for* $i \in J$. *The group* W_J *is a* parabolic *subgroup of* W. *Each coset of* W_J *in* W *contains a unique coset representative* x_j *of minimal length* [Hu, Prop. 1.10], *and the probability of any such coset is computable via*

$$\pi(x_jW_J) = q^{\ell(x_j)}\frac{P_{W_J}(q)}{P_W(q)}. \tag{2.10}$$

As an example, suppose that W is the symmetric group S_n generated by $s_1, s_2, \ldots, s_{n-1}$, where $s_i = (i,\ i+1)$. If $J = \{1, 2, \ldots, n-2\}$ then W_J is the subgroup of permutations that leave n fixed. The minimal length coset representatives x_j for the cosets of W_J in W have j in position n and the rest of the entries in order. Property 2 says that

$$\pi(\{w \in S_n \mid w(n) = j\}) = q^{n-j}\frac{(1-q)}{(1-q^n)}. \tag{2.11}$$

Similarly, if $J = \{2, 3, \ldots, n-1\}$ then Property 2 yields

$$\pi(\{w \in S_n \mid w(1) = j\}) = q^{j-1}\frac{(1-q)}{(1-q^n)}. \tag{2.12}$$

Similar formulas can be derived for the cases where J consists of the first j or last j elements of $\{1, 2, \ldots, n\}$.

In combination with Property 1, (2.11) also provides a formula for the probability of the set of permutations with j in the nth position and (2.12) gives the probability of the set of permutations with j in the first position. More generally, one can give formulas for the probability of the set of permutations that have $1, 2, \ldots, j$ in any given relative position.

PROPERTY 3. *Let $J_1 \supseteq J_2 \supseteq \cdots \supseteq J_k = \emptyset$ be a sequence of subsets of $\{1, 2, \ldots, n\}$. Then a sequential algorithm for generating w in W from π is to choose, for each $1 \le i \le k-1$, the minimal length coset representative of a coset of $W_{J_{i+1}}$ in W_{J_i} $(1 \le i \le k-1)$ and multiply these together. If x_i is a minimal length coset representative of a coset in $W_{J_i}/W_{J_{i+1}}$, choose x_i with probability $q^{\ell(x_i)}P_{W_{J_{i+1}}}(q)/P_{W_{J_i}}(q)$.*

As an example, suppose W is the symmetric group S_n generated by $s_1, s_2, \ldots, s_{n-1}$. If $J_1 \supseteq J_2 \supseteq \cdots \supseteq J_{n-1}$ is given by $J_i = \{i, i+1, \ldots, n-1\}$, then the algorithm can be realized as the following procedure. Place symbols down sequentially, beginning with 1. If symbols $1, 2, \ldots, i-1$ have been placed in some order, then place i first with probability $q^{i-1}(1-q)/(1-q^i)$, second with probability $q^{i-2}(1-q)/(1-q^i)$, $\ldots$, ith with probability $(1-q)/(1-q^i)$. Continuing until all n elements are placed gives an efficient method of choosing from π.

An application of this is the following clever algorithm suggested by Pak [P] for generating a uniformly chosen element of $\mathrm{GL}_n(\mathbb{F}_q)$. Choose $w \in S_n$ with probability proportional to $q^{\ell(w)}$. Then form b_1wb_2 with b_1 and b_2 uniformly chosen in the lower triangular matrices in $\mathrm{GL}_n(\mathbb{F}_q)$. This yields an efficient algorithm for uniform choice in $\mathrm{GL}_n(\mathbb{F}_q)$. With obvious modifications, this procedure easily adapts to the other finite groups with a BN pair.

PROPERTY 4. *Consider a finite Coxeter group with probability distribution π as given in* (2.8). *Let Z be the random variable given by $Z(w) = \ell(w)$ for $w \in W$. Then, with d_i as in* (2.9),

$$E_\pi(Z) = \frac{nq}{1-q} - \sum_{i=1}^{n} \frac{d_i q^{d_i}}{1-q^{d_i}},$$
$$\mathrm{Var}_\pi(Z) = \frac{nq}{(1-q)^2} - \sum_{i=1}^{n} \frac{d_i^2 q^{d_i}}{(1-q^{d_i})^2}. \tag{2.13}$$

Proof. The moment generating function of Z is

$$\begin{aligned} M_Z(t) = E_\pi(e^{tZ}) &= \frac{1}{P_W(q)} \sum_{w\in W} (e^t q)^{\ell(w)} \\ &= \frac{P_W(e^t q)}{P_W(q)} = \prod_{i=1}^{n} \frac{(1-(e^t q)^{d_i})}{(1-e^t q)} \frac{(1-q)}{(1-q^{d_i})}. \end{aligned}$$

It follows that Z is the sum of independent random variables $Z_1, \ldots, Z_n$, where

$$M_{Z_i}(t) = \frac{(1-(e^t q)^{d_i})}{(1-e^t q)} \frac{(1-q)}{(1-q^{d_i})}.$$

Then

$$E_\pi(Z_i) = \frac{d}{dt} M_{Z_i}(t)\Big|_{t=0} = \frac{q}{1-q} - \frac{d_i q^{d_i}}{1-q^{d_i}}$$

and

$$E_\pi(Z_i^2) = \frac{d^2}{dt^2} M_{Z_i}(t)\Big|_{t=0} = \frac{q}{1-q} - \frac{2d_i q^{d_i+1}}{(1-q)(1-q^{d_i})} + \frac{2q^2}{(1-q)^2} - \frac{d_i^2 q^{d_i}}{1-q^{d_i}}.$$

It follows that

$$\mathrm{Var}_\pi(Z_i) = \frac{q}{(1-q)^2} - \frac{d_i^2 q^{d_i}}{(1-q^{d_i})^2}. \qquad \square$$

We remark that, for Coxeter groups of type A_n, B_n, D_n under the probability distribution π, $\ell(w)$ has an approximately normal distribution with mean and variance as in (2.13). This follows from its representation as a sum of independent variables in the proof of Property 4. For details, see [D, Chap. 6C, Cor. 1-2].

3. Hecke Algebras

This section introduces Hecke algebras as bi-invariant functions on a group. We develop the needed Fourier analysis and then specialize to the Iwahori–Hecke algebras associated to finite Coxeter groups.

3a. Algebras and Fourier Analysis

Random walks are traditionally analyzed using Fourier analysis [D]. We find this possible in our examples and explain the basic tools here.

An algebra H over $\mathbb{C}$ is (split) *semisimple* if it is isomorphic to a direct sum of matrix algebras. This means that there exist a finite index set $\hat{W}$ and positive integers d_λ ($\lambda \in \hat{W}$) such that

$$H \cong \bigoplus_{\lambda \in \hat{W}} M_{d_\lambda}(\mathbb{C}),$$

where $M_{d_\lambda}(\mathbb{C})$ is the algebra of $d_\lambda \times d_\lambda$ matrices with entries in $\mathbb{C}$. Fix an isomorphism

$$\phi\colon H \to \bigoplus_{\lambda} M_{d_\lambda}(\mathbb{C})$$

and define

$$e^\lambda_{ST} = \phi^{-1}(E^\lambda_{ST}), \quad \lambda \in \hat{W},\ 1 \le S, T \le d_\lambda, \tag{3.1}$$

where E^λ_{ST} is the matrix in $\bigoplus_\lambda M_{d_\lambda}(\mathbb{C})$ that has a 1 in the (S, T) entry of the λth block and 0 everywhere else. The elements $e^\lambda_{ST} \in H$ are a set of *matrix units* for H.

The matrix units $\{e^\lambda_{ST}\}$ form a basis of H, and we write

$$h = \sum_{\lambda \in \hat{W}} \sum_{1 \le S, T \le d_\lambda} \rho^\lambda_{ST}(h) e^\lambda_{ST} \tag{3.2}$$

for $h \in H$. The homomorphisms $\rho^\lambda\colon H \to M_{d_\lambda}(\mathbb{C})$ and the linear functionals $\chi^\lambda_H\colon H \to \mathbb{C}$ given by

$$\rho^\lambda(h) = (\rho^\lambda_{ST}(h))_{1 \le S, T \le d_\lambda} \quad \text{and} \quad \chi^\lambda_H(h) = \mathrm{Tr}(\rho^\lambda(h))$$

are the *irreducible representations* and the *irreducible characters* of H, respectively. The homomorphisms ρ^λ depend on the choice of ϕ, but the irreducible characters χ^λ_H do not.

A *trace* on H is a linear functional $\vec{t}\colon H \to \mathbb{C}$ such that $\vec{t}(h_1h_2) = \vec{t}(h_2h_1)$ for all $h_1, h_2 \in H$. Up to constant multiples, there is a unique trace on $M_{d_\lambda}(\mathbb{C})$; this implies that, for any trace $\vec{t}\colon H \to \mathbb{C}$ on H, there are unique $t_\lambda \in \mathbb{C}$ $(\lambda \in \hat{W})$ such that

$$\vec{t} = \sum_{\lambda \in \hat{W}} t_\lambda \chi^\lambda_H. \tag{3.3}$$

The trace $\vec{t}$ is *nondegenerate* if $t_\lambda \neq 0$ for all $\lambda \in \hat{W}$. Define a symmetric bilinear form $\langle \cdot, \cdot \rangle_H\colon H \times H \to \mathbb{C}$ on H by

$$\langle h_1, h_2 \rangle_H = \vec{t}(h_1h_2) \quad \text{for } h_1, h_2 \in H.$$

The form $\langle \cdot, \cdot \rangle_H$ is nondegenerate if and only if $\vec{t}$ is a nondegenerate trace.

Let $\{T_w\}_{w \in W}$ be a basis of H. The Fourier transform of $h = \sum_{w \in W} h_w T_w$ at the representation ρ is

$$\hat{h}(\rho) = \sum_{w \in W} h_w \rho(T_w). \tag{3.4}$$

The Fourier inversion theorems describe the change of basis matrix between $\{T_w\}$ and $\{e^\lambda_{ST}\}$ and recover h from $\{\hat{h}(\rho^\lambda)\}_{\lambda \in \hat{W}}$.

Theorem 3.5 (Fourier inversion and Plancherel). *Let H be a semisimple algebra over $\mathbb{C}$ with a nondegenerate trace $\vec{t}$. Let $\{T_w\}_{w \in W}$ be a basis for the algebra H. Let $\{\tilde{T}^*_w\}_{w \in W}$ be the dual basis with respect to $\langle \cdot, \cdot \rangle_H$; that is, $\langle \tilde{T}^*_w, T_v \rangle_H = \delta_{wv}$. Then, with notation as in* (3.1)–(3.4),

$$e^\lambda_{ST} = \sum_{w\in W} t_\lambda \rho^\lambda_{TS}(\tilde{T}^*_w)T_w$$

and, for any $h, h_1, h_2 \in H$,

$$h_w = \sum_{\lambda\in\hat{W}} t_\lambda \operatorname{Tr}(\hat{h}(\rho^\lambda)\rho^\lambda(\tilde{T}^*_w)) \quad \text{for } h \in H \tag{3.6}$$

and

$$\langle h_1, h_2\rangle_H = \sum_{\lambda\in\hat{W}} t_\lambda \operatorname{Tr}(\hat{h}_1(\rho^\lambda)\hat{h}_2(\rho^\lambda)) \quad \text{for } h_1, h_2 \in H. \tag{3.7}$$

Proof. Since $\vec{t}$ is nondegenerate, the equation $\vec{t}(e^\lambda_{ST}) = \sum_{\mu\in\hat{W}} t_\mu \chi^\mu_H(e^\lambda_{ST}) = t_\lambda \delta_{ST}$ implies that

$$\left\{\frac{e^\lambda_{TS}}{t_\lambda}\right\} \text{ is the dual basis to } \{e^\lambda_{ST}\} \text{ with respect to } \langle\cdot,\cdot\rangle_H.$$

By (3.2), $\rho^\lambda_{ST}(a) = (1/t_\lambda)\langle a, e^\lambda_{TS}\rangle_H$. Thus

$$e^\lambda_{ST} = \sum_{w\in W} \langle e^\lambda_{ST}, \tilde{T}^*_w\rangle_H T_w = \sum_{w\in W} t_\lambda \rho^\lambda_{TS}(\tilde{T}^*_w)T_w.$$

Then equation (3.6) is

$$h_w = \langle h, \tilde{T}^*_w\rangle_H = \vec{t}(h\tilde{T}^*_w) = \sum_{\lambda\in\hat{W}} t_\lambda \chi^\lambda_H(h\tilde{T}^*_w) = \sum_{\lambda\in\hat{W}} t_\lambda \operatorname{Tr}(\hat{h}(\rho^\lambda)\rho^\lambda(\tilde{T}^*_w)),$$

and (3.7) is

$$\langle h_1, h_2\rangle_H = \vec{t}(h_1h_2) = \sum_\lambda t_\lambda \chi^\lambda_H(h_1h_2) = \sum_\lambda t_\lambda \operatorname{Tr}(\hat{h}_1(\rho^\lambda)\hat{h}_2(\rho^\lambda)). \qquad \square$$

3b. Coset Chains and Hecke Algebras

Let G be a finite group and B a subgroup of G, and let Q be a left B-invariant probability distribution on G. Right multiplication by random picks from Q induces a random walk on G,

$$X_0 = x_0,\ X_1 = x_0g_1,\ X_2 = x_0g_1g_2, \ldots, \tag{3.8}$$

which, in turn, induces a process on B cosets $Y_0, Y_1, Y_2, \ldots$, where Y_i is the coset containing X_i. The chain on G produced by right multiplication by random picks from Q is $\tilde{K}(x, y) = Q(x^{-1}y)$. The chance that this chain winds up in an element of yB is $\tilde{K}(x, yB) = Q(x^{-1}yB)$ and, since Q is left B-invariant, $\tilde{K}(x, yB) = \tilde{K}(xb, yB)$ for any $b \in B$. This invariance is a necessary and sufficient condition for the induced coset process to be a Markov chain for any starting state $x_0B \in G/B$ (see [KeS, Thm. 6.32]). If the support of Q is not a coset of a subgroup of G then the chain in (3.8) is irreducible and aperiodic with uniform stationary distribution. The resulting coset chain is

$$K(xB, yB) = Q(x^{-1}yB) \quad \text{with stationary distribution } \pi(xB) = |B|/|G|.$$

If the probability Q is B bi-invariant, then the right process (3.8) on G induces a process on B double cosets by simply reporting which double coset the element X_i is in. The chance that the G-chain moves from x to an element of ByB in one step is $Q(x^{-1}ByB)$ and, since this depends only on the double coset of x, the induced process is a Markov chain on double cosets for any starting state Bx_0B. Letting W be a set of coset representatives for the double cosets of B in G, the chain is given by

$$K(w, w') = K(w^{-1}Bw'B) \quad \text{with stationary distribution } \pi(w) = |BwB|/|G|,$$

where we view the double coset chain as a Markov chain on the set W.

The *Hecke algebra* of the pair (G, B) is the subalgebra of the group algebra of G consisting of the B bi-invariant functions on G,

$$H = \{f : G \to \mathbb{C} \mid f(g) = f(b_1 g b_2) \text{ for } g \in G \text{ and } b_1, b_2 \in B\}.$$

Background on Hecke algebras may be found in Curtis and Reiner [CR, Sec. 11D]. If H is commutative then (G, B) is called a *Gelfand pair*, and there is a well-developed probabilistic literature surveyed in [L1; L2] and [D, Chap. 3F].

Let W be a set of representatives of the double cosets in $B\backslash G/B$. The functions

$$T_w = \frac{1}{|B|}\delta_{BwB}, \quad w \in W,$$

form a basis of H, where δ_{BwB} is the characteristic function of the double coset BwB. The natural anti-involution on G given by $g \mapsto g^{-1}$ induces an anti-involution $*: H \to H$ given by $T_w \mapsto T_{w^{-1}}$. The trivial representation of G restricts to the *index* representation of H given by

$$\rho^{\mathbf{1}}(T_w) = \text{ind}(w), \quad \text{where } \text{ind}(w) = \frac{|BwB|}{|B|}. \tag{3.9}$$

An example to keep in mind is $G = \text{GL}_n(\mathbb{F}_q)$ with B the subgroup of upper triangular matrices. Then W is the set of permutation matrices and $\text{ind}(w) = q^{\ell(w)}$.

Let $L(G/B) = \{f : G \to \mathbb{C} \mid f(g) = f(gb) \text{ for } g \in G \text{ and } b \in B\}$. The group G acts on the left of $L(G/B)$ and H acts on the right of $L(G/B)$ by convolution. The raison d'être for the Hecke algebra is that $H = \text{End}_G(L(G/B))$ and, as (G, H) bimodules,

$$L(G/B) = \sum_{\lambda \in \hat{W}} G^\lambda \otimes H^\lambda. \tag{3.10}$$

Here λ runs over an index set $\hat{W}$ of all the irreducible representations of H, G^λ is an irreducible G-module, and H^λ is an irreducible H module (see [CR, (11.25)(ii)]). Centralizers of the action of a finite group, in this case G acting on $L(G/B)$, are semisimple (our base field is $\mathbb{C}$) and hence the theory of Section 3a applies to Hecke algebras. The trace of the action of H on $L(G/B)$ is given by

$$\vec{t}(T_w) = \begin{cases} |G|/|B| & \text{if } w = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

The decomposition (3.10) yields

$$\vec{t} = \sum_{\lambda \in \hat{W}} t_\lambda \chi_H^\lambda, \quad \text{where } t_\lambda = \dim(G^\lambda), \tag{3.12}$$

and χ_H^λ are the irreducible characters of H. Define an inner product on H by

$$\langle h_1, h_2 \rangle_H = \vec{t}(h_1 h_2), \quad h_1, h_2 \in H.$$

The basis

$$\left\{ \frac{T_{w^{-1}}}{\mathrm{ind}(w)} \right\}_{w \in W} \quad \text{is the dual basis to } \{T_w\}_{w \in W} \tag{3.13}$$

with respect to $\langle \cdot, \cdot \rangle_H$; that is, $\langle (1/\mathrm{ind}(v)) T_{v^{-1}}, T_w \rangle_H = \delta_{vw}$ for all $v, w \in W$ (see [CR, (11.30)(iii)]).

3c. Iwahori–Hecke Algebras

The Hecke algebras associated to finite Chevalley groups G and their Borel subgroups B have a remarkable structure theory for their double cosets—they are indexed by the elements of a finite Coxeter group W. For example, in the case of the group $G = \mathrm{GL}_n(\mathbb{F}_q)$ and its Borel subgroup B of upper triangular matrices, the group W is the symmetric group. There are many wonderful references for this material; see [Bo, Chap. IV, Sec. 2, Ex. 22–27; Br; C, Sec. 10.8–10.11; CR, Sec. 67–68]. We develop what we need in this section and give the relation to probability theory.

Let W be a finite Coxeter group generated by *simple reflections* $s_1, \dots, s_n$. These define a length function with $\ell(\mathrm{id}) = 0$, $\ell(s_i) = 1$, and $\ell(s_i w) = \ell(w) \pm 1$ for each $w \in W$, $1 \le i \le n$. The *Iwahori–Hecke algebra H* corresponding to W is the vector space with basis $\{T_w \mid w \in W\}$ and multiplication given by

$$T_i T_w = \begin{cases} T_{s_i w} & \text{if } \ell(s_i w) = \ell(w) + 1, \\ (q-1) T_w + q T_{s_i w} & \text{if } \ell(s_i w) = \ell(w) - 1, \end{cases} \tag{3.14}$$

where $T_i = T_{s_i}$ for $1 \le i \le n$. When $w = s_i$ we have $T_i^2 = (q-1)T_i + q$ or, equivalently, $(T_i - q)(T_i + 1) = 0$.

Let $\hat{W}$ be an index set for the irreducible representations of W. For each $\lambda \in \hat{W}$, let χ_W^λ be the corresponding irreducible character of W and let $d_\lambda = \chi_W^\lambda(1)$ be the dimension of this representation. The irreducible representations of the Iwahori–Hecke algebra H are in one-to-one correspondence with the irreducible representations of W in such a way that, if χ_H^λ is the character of the irreducible representation of H indexed by $\lambda \in \hat{W}$, then

$$\chi_H^\lambda(T_w)\big|_{q=1} = \chi_W^\lambda(w)$$

for all $w \in W$; see [CR, (68.21)]. In particular, the dimension of the irreducible representation of H indexed by λ is d_λ.

Define a trace $\vec{t}\colon H \to \mathbb{C}$ on H by

$$\vec{t}(T_w) = \begin{cases} P_W(q) & \text{if } w = 1, \\ 0 & \text{otherwise}, \end{cases} \quad \text{where} \quad P_W(q) = \sum_{w \in W} q^{\ell(w)}$$

is the Poincaré polynomial of W. Then $\vec{t}$ is the trace on H given by (3.11) and the *generic degrees* are the constants t_λ defined by

$$\vec{t} = \sum_{\lambda \in \hat{W}} t_\lambda \chi_H^\lambda, \tag{3.15}$$

where χ_H^λ $(\lambda \in \hat{W})$ are the irreducible characters of H (see (3.2)). Now, if $\langle \cdot, \cdot \rangle_H\colon H \times H \to \mathbb{C}$ is the inner product on H given by $\langle h_1, h_2 \rangle_H = \vec{t}(h_1 h_2)$ for all $h_1, h_2 \in H$, then

$$\langle T_x, T_{y^{-1}} \rangle_H = \delta_{xy} q^{\ell(y)} P_W(q) \quad \text{for all } \ x, y \in W; \tag{3.16}$$

see [CR, (68.29)].

The "trivial" representation $\rho^{\mathbf{1}}$ of the Iwahori–Hecke algebra H is the 1-dimensional representation corresponding to the trivial representation of W. For $w \in W$,

$$\rho^{\mathbf{1}}(T_w) = \chi_H^{\mathbf{1}}(T_w) = q^{\ell(w)}, \quad \text{and} \quad \pi = \frac{1}{P_W(q)} \sum_{w \in W} T_w$$

is the corresponding minimal central idempotent of H (cf. (3.9)). Since $t_{\mathbf{1}} = 1$,

$$\vec{t}(h\pi) = t_{\mathbf{1}} \chi^{\mathbf{1}}(h) \quad \text{and} \quad T_w \pi = q^{\ell(w)} \pi \tag{3.17}$$

for all $h \in H$ and $w \in W$; see [CR, (68.23) and (68.28)].

Let tr be the trace of the regular representation of H—that is, $\operatorname{tr}(h)$ is the trace of the linear transformation obtained from the action of h on H by left multiplication. Then

$$\operatorname{tr} = \sum_{\lambda \in \hat{W}} d_\lambda \chi_H^\lambda, \tag{3.18}$$

where d_λ are the dimensions of the irreducible representations of H (see [CR, (3.37)(iii)]). Both traces tr and $\vec{t}$ are important in our analysis of Metropolis walks (see e.g. the proof of Proposition 4.8).

4. Metropolis Walks and Systematic Scans

This section brings together previous results in the form needed to prove our main theorems. We show that the various systematic scans are precisely represented as multiplication in the Iwahori–Hecke algebra. Then representation theory yields tractable expressions for the norms involved.

4a. Metropolis Walks on W

Let W be a finite Coxeter group generated by simple reflections $s_1, \dots, s_n$ and, for each $1 \le i \le n$, let $P_i(x, y)$ be the Markov chain on W given by

$$P_i(x, y) = \begin{cases} 1 & \text{if } \ y = s_i x, \\ 0 & \text{otherwise.} \end{cases}$$

Fix θ $(0 < \theta \le 1)$ and let π be as in (1.1). Then the Metropolis construction produces the Markov chain

$$K_i(x, y) = \begin{cases} 1 & \text{if } y = s_i x \text{ and } \ell(y) > \ell(x), \\ \theta & \text{if } y = s_i x \text{ and } \ell(y) < \ell(x), \\ 1-\theta & \text{if } y = x. \end{cases} \tag{4.1}$$

The chain K_i can be interpreted as follows.

> From w, try to multiply by s_i. If this increases the length, carry out the multiplication. If it decreases the length then flip a θ coin. If the coin comes out heads, carry out the multiplication; if it comes up tails then the chain stays at w.

Of course, the chain based on a fixed value of i is not irreducible. However, any convex linear combination and any symmetric product of reversible Markov chains with a fixed stationary distribution is reversible with the same stationary distribution. If W is the symmetric group then the following chains are reversible for π:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} K_i &\quad \text{(random scan Metropolis)}, \\ K_1K_2\cdots K_nK_n\cdots K_2K_1 &\quad \text{(short systematic scan)}, \\ (K_1K_2\cdots K_nK_n\cdots K_2K_1)\cdots(K_1K_2K_2K_1)(K_1K_1) &\quad \text{(long systematic scan)}. \end{aligned} \tag{4.2}$$

Note that $K_1K_2\cdots K_n$ is an irreducible Markov chain with π stationary. However, it is not reversible in general.

The following theorem (a direct consequence of our setup) shows that many Markov chains on W can be obtained by left multiplication by elements of H on the basis $\{\tilde{T}_w\}$. The remaining results in this subsection provide the necessary tools for studying the convergence of these chains by using the representation theory of the Iwahori–Hecke algebra H. Though we have chosen to focus here on the Iwahori–Hecke algebras related to finite reflection groups W, the results of this section hold in a general Hecke algebra context.

THEOREM 4.3. *Let W be a finite Coxeter group, and let H be the Iwahori–Hecke algebra with basis $\{T_w\}_{w\in W}$ as defined in* (3.14). *Let*

$$q = \theta^{-1}, \quad \tilde{T}_i = T_i/q, \quad \text{and} \quad \tilde{T}_w = q^{-\ell(w)}T_w \ \text{ for } w \in W.$$

The Metropolis chain K_i in (4.1) *is the same as the matrix of left multiplication by $\tilde{T}_i$ with respect to the basis $\{\tilde{T}_w\}_{w\in W}$ of H:*

$$\tilde{T}_i\tilde{T}_w = \begin{cases} \tilde{T}_{s_iw} & \text{if } \ell(s_iw) > \ell(w), \\ (1-\theta)\tilde{T}_w + \theta\tilde{T}_{s_iw} & \text{if } \ell(s_iw) < \ell(w). \end{cases} \tag{4.4}$$

Identify functions $f\colon X \to \mathbb{R}$ in $L^2(\pi)$ with elements of the Iwahori–Hecke algebra H via

$$f = \sum_{x\in W} f(x)\tilde{T}_x. \tag{4.5}$$

The following proposition shows that we can use the inner product $\langle\cdot,\cdot\rangle_H$ on H (defined in Section 3c) to compute norms in $L^2(\pi)$. Coupled with Lemma 2.3, it gives bounds on rates of convergence in total variation distance.

PROPOSITION 4.6. *Let W be a finite Coxeter group, and let π be as in* (1.1). *With the identification of $L^2(\pi)$ and the Iwahori–Hecke algebra H given by* (4.5),

$$\langle f/\pi, g/\pi\rangle_2 = \langle f, g^*\rangle_H \quad \textit{for all } f, g \in L^2(\pi),$$

where $*\colon H \to H$ is the involutive anti-automorphism of H defined by $T_w^* = T_{w^{-1}}$.

Proof. Use the notation

$$f = \sum_{x\in W} f(x)\tilde{T}_x = \sum_{x\in W} f(x)q^{-\ell(x)}T_x = \sum_{x\in W} f_x T_x.$$

Then, since $\theta = q^{-1}$, (2.2) and (1.1) give

$$\begin{aligned}\left\langle \frac{f}{\pi}, \frac{g}{\pi}\right\rangle_2 &= \sum_{x\in W} \frac{f(x)g(x)}{\pi(x)}\\ &= \sum_{x\in W} \frac{f_x q^{\ell(x)} g_x q^{\ell(x)}}{\theta^{-\ell(x)}} P_W(\theta^{-1}) = \sum_{x\in W} f_x g_x q^{\ell(x)} P_W(q).\end{aligned}$$

Thus, by (3.16),

$$\left\langle \frac{f}{\pi}, \frac{g}{\pi}\right\rangle_2 = \sum_{x\in W} f_x g_x \langle T_x, T_{y^{-1}}\rangle_H = \sum_{x,y\in W} f_x g_y \langle T_x, T_{y^{-1}}\rangle_H = \langle f, g^*\rangle_H. \qquad \square$$

The following lemma shows that the inner product in $L^2(\pi)$, reversibility, and the involution $*\colon H \to H$ are simply related.

LEMMA 4.7. *Let H be the Iwahori–Hecke algebra corresponding to a finite real reflection group W, and let π be as in* (1.1). *Let K be a reversible Markov chain on W determined by left multiplication by an element of H (also denoted by K). The chain K operates on $L^2(\pi)$ by $Kf(x) = \sum_{y\in W} K(x,y)f(y)$. Then the following are equivalent*:

(a) *K is reversible with respect to π*;
(b) *K is self-adjoint with respect to $\langle\cdot,\cdot\rangle_2$; and*
(c) *$K = K^*$ in the Iwahori–Hecke algebra H.*

Here $\langle\cdot,\cdot\rangle_2$ is the norm on $L^2(\pi)$ defined in (2.2) *and $*\colon H \to H$ is the involutive anti-automorphism of H defined by $T_w^* = T_{w^{-1}}$.*

Proof. If K is reversible, then

$$\begin{aligned}\langle Kf, g\rangle_2 &= \sum_{x,y\in X} K(x,y)f(y)g(x)\pi(x)\\ &= \sum_{x,y\in X} f(y)K(y,x)g(x)\pi(y) = \langle f, Kg\rangle_2;\end{aligned}$$

conversely, if K is self-adjoint then

$$\pi(x)K(x,y) = \langle \delta_x, K\delta_y\rangle = \langle K\delta_x, \delta_y\rangle = \pi(y)K(y,x),$$

where δ_z denotes the delta function at z given by $\delta_z(x) = \delta_{zx}$ (Kronecker delta). Hence K is reversible if and only if K is self-adjoint.

If K is self-adjoint with respect to $\langle\cdot,\cdot\rangle_2$ then, by Proposition 4.6,

$$\langle Kf, g^*\rangle_H = \langle Kf/\pi, g/\pi\rangle_2 = \langle f/\pi, Kg/\pi\rangle_2 = \langle f, (Kg)^*\rangle_H = \langle f, g^*K^*\rangle_H.$$

Thus, for all $w \in W$,

$$\langle K, T_w\rangle_H = \langle 1, T_wK^*\rangle_H = \langle T_w, K^*\rangle_H = \langle K^*, T_w\rangle_H.$$

Therefore, $K = K^*$. □

The following proposition is a primary tool for studying rates of convergence of Markov chains on Iwahori–Hecke algebras, and it bounds the $L^2(\pi)$ norm of Lemma 2.3 in terms of characters of that algebra. In contrast with the way that random walks are often analyzed (see e.g. [DS]), the following proposition also shows that the Markov chain given by K can be analyzed without knowing the eigenvalues of K—it is necessary only to compute traces.

PROPOSITION 4.8. *Let H be the Iwahori–Hecke algebra corresponding to a finite real reflection group W. Let K be a reversible Markov chain on W with stationary distribution π determined by left multiplication by an element of H (also denoted by K). Let K_x^ℓ denote the Markov chain started at x after ℓ steps. Then*

(a) $\|K_x^\ell/\pi - 1\|_2^2 = q^{-2\ell(x)}\sum_{\lambda\neq\mathbf{1}} t_\lambda\chi_H^\lambda(T_{x^{-1}}K^{2\ell}T_x)$ *and*

(b) $\sum_{x\in W}\pi(x)\|K_x^\ell/\pi - 1\|_2^2 = \sum_{\lambda\neq\mathbf{1}} d_\lambda\chi_H^\lambda(K^{2\ell})$,

where χ_H^λ are the irreducible characters, t_λ the generic degrees (3.15), *and d_λ the dimensions of the irreducible representations of H.*

Proof. Equation (3.17) says that $\vec{t}(h\pi) = t_{\mathbf{1}}\chi_H^{\mathbf{1}}(h)$ and $\tilde{T}_w\pi = \pi$ for all $h \in H$ and $w \in W$. Thus, since K_x^ℓ is a probability, Proposition 4.6 gives

$$\begin{aligned}1 = \langle K_x^{2\ell}/\pi, 1\rangle_2 &= \langle K^{2\ell}\tilde{T}_x, \pi\rangle_H = \langle K^{2\ell}\tilde{T}_x, \tilde{T}_{x^{-1}}\pi\rangle_H \\ &= \vec{t}(K^{2\ell}\tilde{T}_x\tilde{T}_{x^{-1}}\pi) = t_{\mathbf{1}}\chi_H^{\mathbf{1}}(K^{2\ell}\tilde{T}_x\tilde{T}_{x^{-1}}) = t_{\mathbf{1}}\chi_H^{\mathbf{1}}(\tilde{T}_{x^{-1}}K^{2\ell}\tilde{T}_x).\end{aligned}$$

Then, by Proposition 4.6,

$$\langle K_x^\ell/\pi, K_x^\ell/\pi\rangle_2 = \langle K^\ell\tilde{T}_x, (K^\ell\tilde{T}_x)^*\rangle_H = \langle K^\ell\tilde{T}_x, \tilde{T}_{x^{-1}}(K^\ell)^*\rangle_H.$$

Thus, by (3.12) and Lemma 4.7(c),

$$\begin{aligned}\langle K_x^\ell/\pi, K_x^\ell/\pi\rangle_2 = \langle K^\ell\tilde{T}_x, \tilde{T}_{x^{-1}}K^\ell\rangle_H &= \vec{t}(\tilde{T}_{x^{-1}}K^{2\ell}\tilde{T}_x) \\ &= q^{-2\ell(x)}\sum_{\lambda\in\hat{W}} t_\lambda\chi_H^\lambda(T_{x^{-1}}K^{2\ell}T_x).\end{aligned}$$

Now (a) follows because $\langle K_x^\ell/\pi - 1, K_x^\ell/\pi - 1\rangle_2 = \langle K_x^\ell/\pi, K_x^\ell/\pi\rangle_2 - 1$. Part (b) follows similarly from the following calculation. Using the definition (2.2) of the norm on $L^2(\pi)$,

$$\begin{aligned}\sum_{x\in W}\pi(x)\left\langle\frac{K_x^\ell}{\pi},\frac{K_x^\ell}{\pi}\right\rangle_2 &= \sum_{x,y\in W}\pi(x)\frac{K^\ell(x,y)K^\ell(x,y)}{\pi(y)}\\ &= \sum_{x,y\in W}\pi(y)\frac{K^\ell(y,x)K^\ell(x,y)}{\pi(y)}\\ &= \sum_{y\in W}K^{2\ell}(y,y)=\operatorname{tr}(K^{2\ell})=\sum_{\lambda\in\hat{W}}d_\lambda\chi_H^\lambda(K^{2\ell}),\end{aligned}$$

where tr is the trace of the regular representation of H given in (3.18). □

4b. Systematic Scans

One case of Proposition 4.8 that can be analyzed for all finite Coxeter groups W is the case when the Markov chain K is a (generalized) *systematic scan*. This is when K is given by left multiplication by the element $\tilde{T}_{w_0}^2$ in the Iwahori–Hecke algebra. In terms of the geometry of the Coxeter group, this chain is the Metropolis walk on the chambers that tries to move a chamber to its opposite chamber and back again by successive reflections in the walls of chambers. Since each step is a Metropolis step, the chance that the chamber returns to its original position after one pass is not 1 but instead depends on the parameter θ. When W is the symmetric group, this chain is the long systematic scan defined in (4.2).

Let z be the sum of all the reflections in W. Then z is a central element (since it is a conjugacy class sum) in the group algebra of W and thus, by Schur's lemma, z acts by a constant c_λ on the irreducible representation of W labeled by $\lambda\in\hat{W}$. The following well-known result shows that the element $\tilde{T}_{w_0}^2$, where w_0 is the longest element of W, is an Iwahori–Hecke algebra analog of the element z. From the point of view provided by Theorem 4.3, the following proposition determines the eigenvalues (with their multiplicities) of the systematic scan Metropolis chain K on W.

Proposition 4.9. *Let z be the sum of all the reflections in W, and let w_0 be the longest element of W. Then the element $\tilde{T}_{w_0}^2$ is in the center of the Iwahori–Hecke algebra H and*

$$\rho^\lambda(\tilde{T}_{w_0}^2)=q^{c_\lambda-\ell(w_0)}\,\mathrm{Id},\quad \textit{where}\ \ c_\lambda=\frac{\chi_W^\lambda(z)}{d_\lambda},$$

ρ^λ is the irreducible representation of H indexed by λ, χ_W^λ is the irreducible character of W labeled by λ, and $d_\lambda=\chi_W^\lambda(1)$ is the dimension of this representation.

Proof. This result is standard (see [R, (2.4) and (2.5)], so we only sketch the proof here. A result of Brieskorn–Saito [BS] and Deligne [De] states that $T_{w_0}^2$ is in the center of the corresponding braid group. Since the Iwahori–Hecke algebra H is a quotient of the group algebra of the braid group, it follows that $T_{w_0}^2$ is in the center of H. The constant by which it acts on the irreducible representation labeled by λ can be checked as follows. The element $T_{w_0}^2-q^{\ell(w_0)}$ is divisible by $(q-1)$ and

$$z = \left.\frac{T_{w_0}^2 - q^{\ell(w_0)}}{q-1}\right|_{q=1}.$$

Since

$$\left.\frac{q^{\ell(w_0)+c_\lambda} - q^{\ell(w_0)}}{q-1}\right|_{q=1} = c_\lambda$$

and z acts by the constant c_λ, the element $T_{w_0}^2$ must act by the constant $q^{\ell(w_0)+c_\lambda}$. The result of the proposition now follows, since $\tilde{T}_{w_0}^2 = q^{-2\ell(w_0)} T_{w_0}^2$. An alternative way to obtain the constant $q^{\ell(w_0)+c_\lambda}$ by which $T_{w_0}^2$ acts is to note that

$$\det(\rho^\lambda(T_i)) = (-1)^{(d_\lambda - \chi_W^\lambda(s_i))/2} q^{(d_\lambda + \chi_W^\lambda(s_i))/2}$$

for all $1 \le i \le n$; this and [Bo, Chap. VI, Sec. 1, Cor. 2] imply that

$$\det(T_{w_0}^2) = q^{2\ell(w_0)d_\lambda + 2\chi_W^\lambda(z))/2} = q^{d_\lambda(\ell(w_0)+c_\lambda)}. \qquad \square$$

Combining Propositions 4.9 and 4.8 gives the following bounds on the convergence of the systematic scan Metropolis walk on a finite Coxeter group W. Explicit analyses of these bounds in examples are given in Sections 5, 6, and 7.

THEOREM 4.10. *Let H be the Iwahori–Hecke algebra corresponding to a finite real reflection group W. Let K be the systematic scan Metropolis chain on W—that is, the reversible Markov chain on W with stationary distribution π determined by left multiplication by the element $\tilde{T}_{w_0}^2$ of H, where w_0 is the longest element of W. Then*

(a) $\|K_1^\ell/\pi - 1\|_2^2 = \sum_{\lambda \ne \mathbf{1}} t_\lambda d_\lambda \theta^{2\ell(\ell(w_0)-c_\lambda)}$ *and*

(b) $\sum_{x\in W} \pi(x) \|K_x^\ell/\pi - 1\|_2^2 = \sum_{\lambda\ne\mathbf{1}} d_\lambda^2 \theta^{2\ell(\ell(w_0)-c_\lambda)}$

where $\ell(w_0)$ is the length of w_0, t_λ are the generic degrees (see (3.15)), d_λ are the dimensions of the irreducible representations of H, and the constants c_λ are as given in Proposition 4.9.

5. The Hypercube

We begin with a simple but instructive example where all details can be carried out. We are able to analyze and compare both randomized and systematic scans. We show that both kinds of scans take order $n \log n$ operations to converge to stationarity. For small values of θ, the systematic scan converges faster; for θ close to 1, the random scan converges faster.

5a. Preliminaries

The Coxeter group $W = (\mathbb{Z}/2\mathbb{Z})^n$ has generators $s_1, s_2, \dots, s_n$ and relations

$$s_i^2 = 1 \quad \text{and} \quad s_i s_j = s_j s_i \quad \text{for all } 1 \le i, j \le n.$$

The set $X = W = (\mathbb{Z}/2\mathbb{Z})^n$ is the space of binary n-tuples, s_i is the vector with 1 in the ith coordinate and 0 elsewhere, and the length function is given by

$\ell(x) = |x| = (\#\text{ of ones in } x)$. The longest element of W is $w_0 = s_1 s_2 \cdots s_n$, and $\ell(w_0) = n$.

The irreducible representations ρ^λ of the Iwahori–Hecke algebra of $(\mathbb{Z}/2\mathbb{Z})^n$ are all 1-dimensional and are indexed by n-tuples $\lambda = (\lambda_1, \dots, \lambda_n)$, $\lambda_i \in \{0, 1\}$. Let $|\lambda| = \lambda_1 + \cdots + \lambda_n$. Then

$$\rho^\lambda(T_i) = \begin{cases} q & \text{if } \lambda_i = 0, \\ -1 & \text{if } \lambda_i = 1, \end{cases} \qquad c_\lambda = n - 2|\lambda|, \qquad t_\lambda = q^{|\lambda|}, \tag{5.1}$$

where c_λ and t_λ are the constants defined in Proposition 4.9 and (3.15), respectively.

Fix $0 < \theta \le 1$ and let

$$\pi(x) = \frac{q^{\ell(x)}}{P_W(q)}, \quad \text{where } q = \theta^{-1} \text{ and } P_W(q) = (1+q)^n \tag{5.2}$$

is a normalizing constant. Then π is a product measure on $(\mathbb{Z}/2\mathbb{Z})^n$, since

$$\pi(x) = \left(\frac{q}{1+q}\right)^{\ell(x)} \left(\frac{1}{1+q}\right)^{n-\ell(x)}.$$

5b. Random Scan Metropolis

The random scan Metropolis algorithm proceeds by choosing a coordinate at random and attempting to change to its opposite mod 2. If this results in a 1, the change is made. If the change results in a 0, flip a coin with parameter θ. If the flip comes up heads then change the chosen coordinate to 0; if it comes up tails then the coordinate stays at 1. The resulting chain is

$$K(x, y) = \begin{cases} (1/n) & \text{if } \ell(y) = \ell(x) + 1, \\ (1/n)\theta & \text{if } \ell(y) = \ell(x) - 1, \\ (\ell(x)/n)(1-\theta) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

The following theorem shows that order $n \log n$ steps are necessary and sufficient to reach stationarity.

THEOREM 5.4. *Let the random scan Metropolis algorithm on $(\mathbb{Z}/2\mathbb{Z})^n$ be defined by* (5.3) *with* $0 < \theta \le 1$. *Then, for any starting state x and any ℓ,*

$$\left\| \frac{K_x^\ell}{\pi} - 1 \right\|_2^2 = \sum_{\lambda \ne 0} \theta^{2\lambda \cdot x - |\lambda|} \left(1 - \frac{|\lambda|}{n}(1+\theta)\right)^{2\ell}. \tag{5.5}$$

For $0 < \theta < 1$ *and* $\ell = n(\log n - \log\theta + c)/2(1+\theta)$ *with* $c > 0$,

$$\|K_x^\ell - \pi\|_{TV}^2 \le \left(e^{e^{-c}} - 1\right) + e^{-c/2}. \tag{5.6}$$

The bound in (5.6) *is sharp in the sense that if* $\ell = n(\log n - \log\theta + c)/2(1+\theta)$ *then, for all* $\varepsilon > 0$, *there exists a* $c < 0$ *such that* $\|K_0^\ell - \pi\|_{TV} > 1 - \varepsilon$ *for all sufficiently large* n.

Proof. From the definitions of the irreducible representations of H,

$$\begin{aligned}
&\chi_H^\lambda\big(\tilde{T}_{x^{-1}}\big(\big(\tfrac{1}{n}\big)(\tilde{T}_1+\cdots+\tilde{T}_n)\big)^{2\ell}\tilde{T}_x\big)\\
&\quad= n^{-2\ell}q^{-2\ell(x)-2\ell}\chi_H^\lambda(T_1+\cdots+T_n)^{2\ell}\chi_H^\lambda(T_x)^2\\
&\quad= n^{-2\ell}q^{-2\ell}((n-|\lambda|)q-|\lambda|)^{2\ell}q^{2(|x|-\lambda\cdot x)}q^{-2|x|}\\
&\quad= (1-(|\lambda|/n)(1+\theta))^{2\ell}\theta^{2\lambda\cdot x}.
\end{aligned}$$

The first statement then follows from Theorem 4.10(a) with the value for t_λ given in (5.1). For the second statement, we need to bound the sum on the right-hand side of (5.5). Since $\theta\le 1$, it follows that $\theta^{\lambda\cdot x}\le\theta^0$ and

$$\left\|\frac{K_x^\ell}{\pi}-1\right\|_2^2\le\sum_{j=1}^n\binom{n}{j}\theta^{-j}\left(1-\frac{j}{n}(1+\theta)\right)^{2\ell}.$$

Break the sum at $n/2$. For the first half, use $\binom{n}{j}\le n^j/j!$ and $1-x\le e^{-x}$ to give an upper bound

$$\sum_{j=1}^{n/2}\frac{1}{j!}\left(\frac{n}{\theta}\right)^j e^{-j(1+\theta)2\ell/n}=\sum_{j=1}^{n/2}\frac{e^{-jc}}{j!}\le e^{e^{-c}}-1.$$

For the second half, change j to $n-k$ and use the same inequalities to get an upper bound

$$\sum_{k=0}^{n/2}\frac{n^k}{k!}\theta^{k-n}e^{-(n-k)(1+\theta)2\ell/n}=\sum_{k=0}^{n/2}\frac{1}{k!}e^{k(\log n+\log\theta+2\ell(1+\theta)/n)-n\log\theta-2\ell n(1+\theta)/n}.$$

Using $\sum_{k=0}^m(A^k/k!)\le A^m$ for $A\ge 2$, the bound for the second half becomes

$$e^{(n/2)(\log n+\log\theta+2\ell(1+\theta)/n)-n\log\theta-2\ell n(1+\theta)/n}=e^{(n/2)(\log n-\log\theta-2\ell(1+\theta)/n)}=e^{-nc/2}.$$

To show that this upper bound is sharp, we use the second moment method. With respect to the action of K on $L^2(\pi)$ defined in Lemma 4.7, the matrix $K(x,y)$ of (5.3) has an orthonormal basis of eigenfunctions

$$f_\lambda(y)=\theta^{-|\lambda|/2}(-\theta)^{\lambda\cdot y}\quad\text{with eigenvalues }1-\frac{|\lambda|}{n}(1+\theta),\quad\lambda\in(\mathbb{Z}/2\mathbb{Z})^n.\tag{5.7}$$

Let $e_i\in(\mathbb{Z}/2\mathbb{Z})^n$ be the vector with 1 in the ith entry and 0 elsewhere. We shall use the test function

$$\begin{aligned}
T(y)&=\sum_i f_{e_i}(y)=\theta^{-1/2}\sum_{i=1}^n(-\theta)^{y_i}\\
&=\theta^{-1/2}(n-|y|(1+\theta))=\frac{n}{\sqrt{\theta}}\left(1-\frac{|y|(1+\theta)}{n}\right).
\end{aligned}\tag{5.8}$$

The expectation and the variance of T with respect to the distribution π are

$$E_\pi(T)=0\quad\text{and}\quad\operatorname{Var}_\pi(T)=E_\pi(T^2)=n.\tag{5.9}$$

For $i\ne j$,

$$f_{e_i} f_{e_j} = f_{e_i + e_j} \quad \text{and} \quad f_{e_i}^2 = f_0 + \frac{1-\theta}{\sqrt{\theta}} f_{e_i},$$

where the second identity is verified by checking that both sides agree when evaluated at each of the two cases: y such that $y_i = 1$ and y such that $y_i = 0$. We can compute the expectation and variance of T under the distribution K_0^ℓ as follows:

$$E_{\ell,0}(T) = \sum_y K^\ell(0, y) T(y) = \sum_{i=1}^n (K^\ell f_{e_i})(0) = \frac{n}{\sqrt{\theta}} \left(1 - \frac{1+\theta}{n}\right)^\ell \tag{5.10}$$

and

$$\begin{aligned}
\mathrm{Var}_{\ell,0}(T) &= E_{\ell,0}(T^2) - E_{\ell,0}(T)^2 \\
&= E_{\ell,0}\left(\sum_{i=1}^n f_{e_i}^2 + \sum_{i \neq j} f_{e_i} f_{e_j}\right) - \frac{n^2}{\theta}\left(1 - \frac{1+\theta}{n}\right)^{2\ell} \\
&= n + \frac{n(1-\theta)}{\theta}\left(1 - \frac{1+\theta}{n}\right)^\ell + \frac{n(n-1)}{\theta}\left(1 - \frac{2(1+\theta)}{n}\right)^\ell \\
&\quad - \frac{n^2}{\theta}\left(1 - \frac{1+\theta}{n}\right)^{2\ell}.
\end{aligned}$$

We want to use these formulas to show that $\ell = (n/2(1+\theta))(\log n - \log\theta + c)$ steps are sharp. Fixing k and using $\log(1-x) = -x - x^2/2 + O(x^3)$ and $e^{-x^2/2} = 1 - x^2/2 + O(x^4)$, it follows that

$$\left(1 - \frac{k(1+\theta)}{n}\right)^\ell \sim e^{\ell(-k(1+\theta)/n - \ell k^2(1+\theta)^2/2n^2)} \sim \left(\frac{\theta e^{-c}}{n}\right)^{k/2}\left(1 - \frac{\ell k^2(1+\theta)^2}{2n^2}\right)$$

when n is large. Thus, for $\ell = (n/2(1+\theta))(\log n - \log\theta + c)$ and n large,

$$E_{\ell,0}(T) \sim \sqrt{n} e^{-c/2}$$

and

$$\begin{aligned}
\mathrm{Var}_{\ell,0}(T) &\sim n + \sqrt{\frac{n}{\theta}}(1-\theta)e^{-c/2} + (n-1)e^{-c}\left(1 - \frac{\ell 4(1+\theta)^2}{2n^2}\right) \\
&\quad - \frac{n^2}{\theta}\frac{\theta}{n}e^{-c}\left(1 - \frac{2\ell(1+\theta)^2}{2n^2}\right) \\
&\sim n + O_{c,\theta}(\sqrt{n}) + ne^{-c}\left(-\frac{\ell 2(1+\theta)^2}{2n^2}\right) - e^{-c} \\
&= n + O_{c,\theta}(\sqrt{n}) + O_{c,\theta}(\log n),
\end{aligned}$$

with the error terms depending on c and θ. By first choosing c to be a fixed (large) negative number and then letting $n \to \infty$, we see that, if b is large, the set $A_b = \{x \mid |T(x)| \le b\sqrt{n}\}$ has probability $1 - 1/b^2$ under π and probability $O(1/b^2)$ under K_0^ℓ. This completes the proof of the last statement. □

5c. Systematic Scan Metropolis

We turn next to the systematic scan version. Order $n \log n$ steps are required here, too. Lest the reader think this contradicts the example that began this paper, we note that the opening example (which actually corresponds to the heat bath updating setup) replaces each coordinate with a freshly chosen pick. Thus a chosen zero coordinate can remain zero with probability θ. For the Metropolis version analyzed here, a chosen 0 must change to a 1.

With notation as in Section 5a, let N be the chain on $\mathbb{Z}/2\mathbb{Z}$ with matrix $\begin{pmatrix} 0 & 1 \\ \theta & 1-\theta \end{pmatrix}$. On $(\mathbb{Z}/2\mathbb{Z})^n$ define K_i acting as N on the ith coordinate. Let

$$K = K_1 K_2 \cdots K_n K_n \cdots K_1. \tag{5.11}$$

This is the systematic scan Metropolis algorithm with stationary distribution π. The following theorem gives bounds on the distance to stationarity. The proof is similar to the proof of Theorem 5.4; for further details see [DR].

THEOREM 5.12. *Let the systematic scan Metropolis algorithm on* $(\mathbb{Z}/2\mathbb{Z})^n$ *be defined by* (5.11) *with* $0 < \theta \le 1$. *Then, for any starting state* x *and any* ℓ,

$$\left\| \frac{K_x^\ell}{\pi} - 1 \right\|_2^2 = \sum_{\lambda \neq 0} \theta^{(4\ell-1)|\lambda| + 2\lambda\cdot x}. \tag{5.13}$$

For $0 < \theta < 1$ *and*

$$\ell = \frac{1}{4}\left(\frac{\log n + c}{\log(1/\theta)} + 1 \right)$$

with $c > 0$,

$$\|K_x^\ell - \pi\|_{TV}^2 \le \tfrac{1}{4}\left(e^{e^{-c}} - 1\right). \tag{5.14}$$

The bound in (5.14) *is sharp in the sense that if*

$$\ell = \frac{1}{4}\left(\frac{\log n + c}{\log(1/\theta)} + 1 \right)$$

then, for all $\varepsilon > 0$, *there exists a* $c < 0$ *such that* $\|K_0^\ell - \pi\|_{TV} > 1 - \varepsilon$ *for all sufficiently large* n.

After ℓ passes, the systematic scan algorithm makes $2\ell n$ basic steps. Thus, the results of Theorems 5.4 and 5.12 show that both scanning strategies converge in order $n \log n$ basic steps. The following table compares the lead term constants for the two scanning strategies at various values of θ.

θ	random	systematic
general	$\dfrac{n\log(n/\theta)}{2(1+\theta)}$	$\dfrac{n\log n}{2\log(1/\theta)}$
$\dfrac{1}{1+\varepsilon}$	$\dfrac{n\log n}{4}$	$\dfrac{n\log n}{2\log(1+\varepsilon)}$
$\dfrac{1}{2}$	$\dfrac{n\log 2n}{3}$	$\dfrac{n\log n}{2\log 2}$
ε	$\dfrac{n\log(n/\varepsilon)}{2}$	$\dfrac{n\log n}{2\log(1/\varepsilon)}$

We see that the lead term constants make the random scan faster as $\theta \to 1$ whereas the systematic scan is faster as $\theta \to 0$.

6. The Dihedral Group

The hypercube of Section 5 is commutative. This section treats the simplest non-commutative example—the dihedral group D_{2n}. We completely analyze the convergence of both the randomized and systematic scans. We find that both scanning strategies take order n operations to converge to stationarity.

The dihedral group of order $2n$ is the group W given by generators s_1, s_2 and relations

$$s_1^2 = 1, \quad s_2^2 = 1, \quad \text{and} \quad \underbrace{s_1 s_2 s_1 \cdots}_{n \text{ factors}} = \underbrace{s_2 s_1 s_2 \cdots}_{n \text{ factors}}.$$

This is the group of symmetries of a regular $2n$-gon, where s_1 and s_2 act by reflection in axes through the center of the $2n$-gon that form an angle of $2\pi/2n$ (see Figure 1).

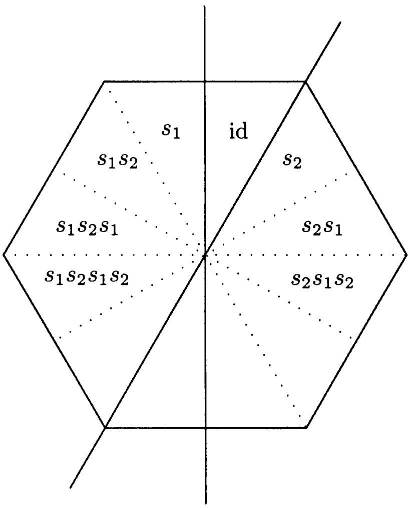


Figure 1

Fix $0 < \theta \leq 1$ and consider the distribution on W given by

$$\pi(w) = \frac{q^{\ell(w)}}{P_W(q)}, \quad \text{where} \ \ P_W(q) = \frac{(q^2-1)}{(q-1)} \frac{(q^n-1)}{(q-1)} \ \ \text{and} \ \ q = \theta^{-1}.$$

This measure is largest at the longest element of W, $w_0 = s_1 s_2 s_1 \cdots$ (n factors), and $\ell(w_0) = n$. The walks to be analyzed will all start at the identity.

One may picture the walks described in this section on the $2n$ chambers of an n-gon. Pick one chamber (labeled with identity) and identify the two internal sides with s_1, s_2. Reflecting the fundamental chamber around gives each edge and each chamber a label. The distance $\ell(w)$ is the smallest number of chambers required to walk from the chamber labeled by w to the identity. For example, in D_{12} pictured in Figure 1, $\ell(s_1 s_2 s_1 s_2) = 4$.

The random scan Metropolis walk proceeds from w by choosing one of s_1, s_2 with probability $1/2$. If $\ell(s_i w) > \ell(w)$, the move is accepted. If $\ell(s_i w) < \ell(w)$ then the move is accepted with probability θ and rejected with probability $1 - \theta$.

One pass of the systematic scan Metropolis algorithm chooses n generators in the order $s_1, s_2, s_1, s_2, \ldots$. Geometrically, starting from the identity, this amounts to marching around the n-gon. If no rejections are made then one complete scan ends in w_0.

Our bounds result in explicit expressions for the convergence of the two walks. One of these has been carefully analyzed by Belsey [B1, Chap. VI, Thm. 2-10]. He showed the following.

PROPOSITION 6.1. *For the random scan Metropolis algorithm starting from the identity,*

$$\left\| \frac{K_1^\ell}{\pi} - 1 \right\|_2^2 \le \theta^{-n} \sqrt{\frac{1+\theta}{1-\theta}} \left(1 - \frac{1}{2}(1-\sqrt{\theta})^2\right)^{2\ell}. \tag{6.2}$$

For $0 < \theta < 1$, *the right-hand side of* (6.2) *is small for* k *of order*

$$\frac{n \log \theta}{2 \log(1 - (1/2)(1-\sqrt{\theta})^2)}.$$

Belsley [B1, Chap. VI, Thm. 4-12] further showed that the random scan Metropolis algorithm has a total variation cutoff at

$$\ell = \frac{2n}{1-\theta} + c\sqrt{n}.$$

For the systematic scan algorithm, Theorem 4.10 and a computation of the relevant constants from that theorem for the dihedral group shows that

$$\begin{aligned} \left\| \frac{K_1^\ell}{\pi} - 1 \right\|_2^2 &= \theta^{(4\ell-1)n} + \theta^{(2\ell-1)n}\left(\frac{\theta^2-1}{\theta-1} \cdot \frac{\theta^n - 1}{\theta - 1} - 1\right) - \theta^{2\ell n}, \\ \sum_{x \in W} \pi(x) \left\| \frac{K_x^\ell}{\pi} - 1 \right\|_2^2 &= \theta^{4\ell n} + (2n-2)\theta^{2\ell n}. \end{aligned} \tag{6.3}$$

(For details of these calculations see [DR] or http://math.wisc.edu/~ram/pub/persi3.21.00.ps.) Thus, for large n and fixed $0 < \theta \le 1$, a single scan suffices to achieve stationarity for typical starting states under the systematic scan. For typical starting states, the random scan converges in order $\log n$ steps.

A comparison of the results in (6.2) and (6.3) shows a mild advantage for systematic scans. The effect is most pronounced as θ approaches 1.

7. The Symmetric Group

This section proves Theorem 1.4 and a similar result for a different scanning strategy. The results show that both scanning strategies require n^2 operations up to lead term constants.

7a. Preliminaries

The symmetric group S_n is generated by the simple transpositions $s_i = (i, i+1)$, $1 \le i \le n-1$, and the longest element of S_n is the reversal permutation

$$w_0 = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ n & n-1 & \cdots & 2 & 1 \end{pmatrix} \quad \text{with } \ell(w_0) = \binom{n}{2}.$$

The book of Fulton [Fu] provides a review of the representation theory of S_n, and we will adopt the conventions for tableaux used there. The irreducible representations of the Iwahori–Hecke algebra H are indexed by partitions λ with n boxes (see Figure 2).

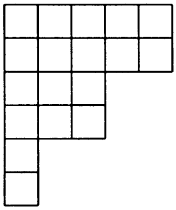
$\lambda =$ 

Figure 2

Number the rows and columns of λ as for matrices. If λ_i and λ'_j denote the length of the ith row and jth column (respectively) of λ, then the *content* and the *hook length* of a box b in position (i, j) of λ are

$$c(b) = j - i \quad \text{and} \quad h_b = \lambda_i - i + \lambda'_j - j + 1,$$

respectively. Set

$$n(\lambda) = \sum_{i=1}^{\ell(\lambda)} (i-1)\lambda_i,$$

and let

$$[k]_q = \frac{q^k - 1}{q - 1} \quad \text{and} \quad [k]_q! = [k]_q[k-1]_q \cdots [2]_q[1]_q$$

for each positive integer k. Using this notation, the dimensions d_λ of the irreducible representations of H, the generic degrees t_λ defined in (3.15), and the constants c_λ defined in Proposition 4.9 are given by

$$d_\lambda = \frac{n!}{\prod_{b\in\lambda} h_b}, \quad t_\lambda = q^{n(\lambda)} \frac{[n]_q!}{\prod_{b\in\lambda} [h_b]_q}, \quad \text{and} \quad c_\lambda = \sum_{b\in\lambda} c(b) \tag{7.1}$$

(see [Fu, Sec. 7.2, Prop. 2; Hf, 3.4.14; Mac, I, Sec. 7, Ex. 7; Mac, I, Sec. 1, Ex. 3; Mac, IV, (6.7)]). The dimension d_λ is also equal to the number of standard tableaux of shape λ—that is, fillings of the boxes of λ with $1, 2, \ldots, n$ such that the rows are increasing left to right and the columns are increasing top to bottom (see [Fu, p. 53]).

The next lemma provides bounds on the constants in (7.1) that will be useful for proving bounds on the convergence times of the systematic scan Metropolis walks that we analyze here. The bounds on c_λ given in part (c) are essentially those given by Diaconis and Shahshahani (see [D, 3D, Lemma 2].

LEMMA 7.2. *For each partition λ, let t_λ, c_λ, and d_λ be as defined in* (7.1).

(a) *When $\theta = 1/q$ and $0 < \theta \le 1$, $t_\lambda \le \theta^{\binom{\lambda_1}{2}-\binom{n}{2}} d_\lambda$.*

(b) $\displaystyle\sum_{\lambda\vdash n} d_\lambda^2 = n!$ *and* $\displaystyle\sum_{\substack{\lambda\\ \lambda_1=n-j}} d_\lambda^2 \le \frac{n^{2j}}{j!}$ *for each $1 \le j \le n$.*

(c) $c_\lambda \le \begin{cases} \binom{\lambda_1}{2} + \frac{1}{2}(n-\lambda_1)(n-\lambda_1-3) & \text{if } \lambda_1 \ge n/2, \\ n^2/4 - n & \text{if } \lambda_1 \le n/2. \end{cases}$

Proof. Set $\theta = 1/q$ and use [Mac, I, Sec. 1, Ex. 2] and [Mac, III, Sec. 6, Ex. 2] to obtain

$$t_\lambda = \theta^{-n(\lambda)-\binom{n}{2}-n+(\sum h_b)} \frac{[n]_\theta!}{\prod_{b\in\lambda}[h_b]_\theta} = \theta^{n(\lambda')-\binom{n}{2}} \frac{[n]_\theta!}{\prod_{b\in\lambda}[h_b]_\theta} = \theta^{-\binom{n}{2}} \sum_Q \theta^{r(Q)},$$

where the sum is over all standard tableaux Q of shape λ and where $r(Q)$ is the sum of i such that $i+1$ is to the right of i in Q. Thus t_λ is a sum of d_λ monomials, where the lowest-degree term has degree $n(\lambda') - \binom{n}{2} \ge \binom{\lambda_1}{2} - \binom{n}{2}$. Part (a) follows.

For (b), we can bound the number of standard tableaux Q of shape λ with $\lambda_1 = n-j$ by noting that there are $\binom{n}{j}$ ways of picking the elements not in the first row of Q and at most $\sqrt{j!}$ ways of arranging these to complete a standard tableau. Thus

$$\sum_{\substack{\lambda\\ \lambda_1=n-j}} d_\lambda^2 \le \Bigg(\sum_{\lambda_1=n-j} d_\lambda\Bigg)^2 \le \left(\binom{n}{j}\sqrt{j!}\right)^2 \le \frac{n^{2j}}{(j!)^2} j! = \frac{n^{2j}}{j!}.$$

The inequalities in (c) are direct consequences of

$$c_\lambda \le \begin{cases} c_{(\lambda_1, n-\lambda_1)} & \text{if } \lambda \ge n/2, \\ c_\lambda \le c_{(n/2,n/2)} & \text{if } \lambda_1 \le n/2. \end{cases} \qquad \square$$

7b. *Long Systematic Scan*

As in Section 4a, we fix $0 < \theta \le 1$ and consider the Markov chain

$$K_i(x,y) = \begin{cases} 1 & \text{if } y = s_i x \text{ and } \ell(y) > \ell(x), \\ \theta & \text{if } y = s_i x \text{ and } \ell(y) < \ell(x), \\ 1-\theta & \text{if } y = x, \end{cases} \tag{7.3}$$

which is produced by applying the Metropolis construction to the base chain

$$P_i(x,y) = \begin{cases} 1 & \text{if } y = s_i x, \\ 0 & \text{otherwise,} \end{cases}$$

with the distribution π as given in (1.1). Recall that the chain K_i can be interpreted as follows.

> From w, try to multiply by s_i. If this increases the number of inversions of w, carry out the multiplication. If it decreases the the number of inversions then flip a θ coin and carry out the multiplication if the coin comes up heads; otherwise stay at w.

The long systematic scan Metropolis chain is the chain given by

$$K = (K_1K_2\cdots K_nK_n\cdots K_2K_1)\cdots(K_1K_2K_2K_1)(K_1K_1).$$

The following theorem bounds the rate of convergence of this Markov chain. It shows that a single scan suffices to be close to stationarity.

PROPOSITION 7.4. *Let K be the long systematic scan Metropolis walk on the symmetric group S_n defined by* (4.2). *Let d_λ, t_λ, and c_λ be the constants given in* (7.1), *and let* $0 < \theta \le 1$. *Then the following statements hold.*

(a) $\|K_1^\ell/\pi - 1\|_2^2 = \sum_{\lambda\neq(n)} t_\lambda d_\lambda \theta^{2\ell(\binom{n}{2}-c_\lambda)}$ *and, with* $\ell = 1$,

$$\|K_1^1 - \pi\|_{TV}^2 \le \left(e^{n^2\theta^{n/2}} - 1\right) + n!\,\theta^{n^2/8+5n/4},$$

which, when $\theta < 1$, *approaches* 0 *as* $n \to \infty$.

(b) $\sum_{x\in W} \pi(x)\|K_x^\ell/\pi - 1\|_2^2 = \sum_{\lambda\neq(n)} d_\lambda^2 \theta^{2\ell(\binom{n}{2}-c_\lambda)}$ *and, with* $\ell = 1$,

$$\sum_{x\in W}\pi(x)\left\|\frac{K_x^\ell}{\pi} - 1\right\|_{TV}^2 \le \left(e^{n^2\theta^n} - 1\right) + n!\,\theta^{n^2/2+n},$$

which, when $\theta < 1$, *approaches* 0 *as* $n \to \infty$.

Proof. By Theorem 4.3, this walk is equivalent to the walk on the Iwahori–Hecke algebra H defined by multiplication by $\tilde{T}_{w_0}^2$ with respect to the basis $\{\tilde{T}_w \mid w \in S_n\}$. Thus the equalities in (a) and (b) are consequences of Theorem 4.10.

Fix $\ell = 1$. If $\lambda_1 = n-j$ and $j \le n/2$, then the bound on c_λ from Lemma 7.2(c) gives

$$\theta^{\binom{\lambda_1}{2}-\binom{n}{2}+2\binom{n}{2}-2c_\lambda} \le \theta^{j(n-j/2-1/2)-j(j-3)} = \theta^{j(n-3j/2+5/2)} \le \theta^{j(n/4+5/2)};$$

by using the bounds in Lemma 7.2(a) and (b), it follows that

$$\sum_{j=1}^{n/2}\sum_{\substack{\lambda\neq(n)\\ \lambda_1=n-j}} t_\lambda d_\lambda \theta^{2\ell(\binom{n}{2}-c_\lambda)} \le \sum_{j=1}^{n/2}\frac{n^{2j}}{j!}\theta^{j(n/4+5/2)} \le e^{n^2\theta^{n/4}} - 1.$$

When $\lambda_1 \le n/2$, the bound in Lemma 7.2(c) gives

$$\sum_{\substack{\lambda\\ \lambda_1\le n/2}} d_\lambda^2\theta^{\binom{\lambda_1}{2}-\binom{n}{2}+2\binom{n}{2}-2c_\lambda} \le n!\,\theta^{n^2/8+5n/4}.$$

The upper bound on $\|K_1^1 - \pi\|_{TV}$ follows by combining these expressions. The upper bound in (b) is proved similarly. $\square$

7c. Short Systematic Scan

We now analyze the convergence of the short systematic scan and prove Theorem 1.4 of the introduction. The short systematic scan Metropolis chain on the symmetric group is given by

$$K = K_1 K_2 \cdots K_n K_n \cdots K_2 K_1,$$

where K_i is as in (7.3). The theorem shows that order n short systematic scans are necessary and suffice to reach stationarity when starting from the identity. In part (b′) it is shown that, for typical starting values, this chain converges in order $\log n$ scans.

THEOREM 7.5. *Let K be the short systematic scan Metropolis algorithm on the symmetric group defined by* (4.2). *Let d_λ, t_λ, and c_λ be the constants given in* (7.1).

(a) $\|K_1^\ell/\pi - 1\|_2^2 = \sum_{\lambda\neq(n)} t_\lambda \sum_S \theta^{2\ell(n-1-c(S(n)))}$, *where the sum is over standard tableaux of shape λ and where $S(n)$ denotes the box of S containing n.*

(a′) *For* $\ell = n/2 - (\log n/\log\theta) + c$ *with* $c > 0$,

$$\|K_1^\ell - \pi\|_{TV}^2 \le \left(e^{\theta^{2c+1}} - 1\right) + n!\,\theta^{n^2/8 - n(\log n)/(\log\theta) + n(c+1/4)}.$$

Conversely, if $\ell < n/4$ then, for fixed $0 < \theta < 1$, $\|K_1^\ell - \pi\|_{TV}$ tends to 1 as $n \to \infty$.

(b) $\sum_{x\in W} \pi(x)\|K_1^\ell/\pi - 1\|_2^2 = \sum_{\lambda\neq(n)} d_\lambda \sum_S \theta^{2\ell(n-1-c(S(n)))}$, *where the sum is over standard tableaux of shape λ and where $S(n)$ denotes the box of S containing n.*

(b′) *For $0 < \theta < 1$ and $\ell = -(\log n)/(\log\theta) + c$ with $c > 0$,*

$$\sum_{x\in S_n} \pi(x)\left\|\frac{K_x^\ell}{\pi} - 1\right\|_2^2 \le \left(e^{\theta^{2c}} - 1\right) + \left(\frac{\theta^c}{e}\right)^n e^{1/12}\sqrt{2\pi n}.$$

Proof. By Theorem 4.3, the Markov chain K is the random walk on $\{\tilde{T}_w \mid w \in S_n\}$ defined by multiplication by $\tilde{T}_{n-1}\cdots\tilde{T}_2\tilde{T}_1^2\tilde{T}_2\cdots\tilde{T}_{n-1}$ in the Iwahori–Hecke algebra H corresponding to the symmetric group S_n. Let H' be the Iwahori–Hecke algebra corresponding to the symmetric group S_{n-1}, and let H be the Iwahori–Hecke algebra corresponding to S_n. Let w_0' be the longest element of S_{n-1} and let w_0 be the longest element of S_n. The inclusion $S_{n-1} \subseteq S_n$ induces an inclusion $H' \subseteq H$ of the corresponding Iwahori–Hecke algebras. In H, the generators T_i are invertible with $T_i^{-1} = q^{-1}T_i + (1 - q^{-1})$. Then

$$\tilde{T}_{w_0}^2\tilde{T}_{w_0'}^{-2} = \tilde{T}_{n-1}\cdots\tilde{T}_2\tilde{T}_1^2\tilde{T}_2\cdots\tilde{T}_{n-1}$$

and so it follows from Proposition 4.9 that, in the representation ρ^λ of H indexed by the partition λ, the element

$$K = \tilde{T}_{n-1}\cdots\tilde{T}_2\tilde{T}_1^2\tilde{T}_2\cdots\tilde{T}_{n-1} \quad \text{has eigenvalues } \theta^{n-1-c(S(n))},$$

where S runs over standard tableaux of shape λ and $S(n)$ denotes the box of S containing n. This determines the eigenvalues of $K^{2\ell}$ in the representation ρ^λ and so

$$\chi_H^\lambda(K^{2\ell}) = \sum_S \theta^{2\ell(n-1-c(S(n)))},$$

where χ_H^λ is the irreducible character of the Iwahori–Hecke algebra corresponding to the partition λ. Parts (a) and (b) now follow from Proposition 4.8.

Using $\theta^{2\ell(n-1-c(S(n)))} \le \theta^{2\ell(n-\lambda_1)}$ and the bound for t_λ in Lemma 7.2, we have

$$\left\| \frac{K_1^\ell}{\pi} - 1 \right\|_2^2 \le \sum_{\lambda \ne (n)} \theta^{\binom{\lambda_1}{2} - \binom{n}{2}} d_\lambda \sum_S \theta^{2\ell(n-\lambda_1)} \le \sum_{\lambda \ne (n)} \theta^{(1/2)(n-\lambda_1)(n-\lambda_1+4\ell+1-2n)} d_\lambda^2,$$

since d_λ is the number of standard tableaux of shape λ. Fix $\ell = n/2 - (\log n)/(\log \theta) + c$. Then, using the bound on the sum of d_λ^2 from Lemma 7.2,

$$\begin{aligned}\sum_{j=1}^{n/2} \sum_{\substack{\lambda \ne (n) \\ \lambda_1 = n-j}} \theta^{(1/2)j(j+4\ell+1-2n)} d_\lambda^2 &\le \sum_{j=1}^{n/2} \frac{\theta^{(1/2)j(1+4\ell+1-2n+4(\log n/\log\theta))}}{j!} \\ &= \sum_{j=1}^{\infty} \frac{\theta^{(2c+1)j}}{j!} = e^{\theta^{2c+1}} - 1.\end{aligned}$$

The function $(n-\lambda_1)(n-\lambda_1+4\ell+1-2n)$ has a minimum at $n-\lambda_1 = (-1/2)(4\ell+1-2n)$. At this minimum, $\lambda_1 = n - 2(\log n)/(\log\theta) + 2c + 1/2 \ge n/2$ and so

$$\begin{aligned}\sum_{\substack{\lambda \\ \lambda_1 \le n/2}} \theta^{(1/2)(n-\lambda_1)(n-\lambda_1+4\ell+1-2n)} d_\lambda^2 &\le \theta^{(1/2)(n/2)(n/2+4\ell+1-2n)} n! \\ &= n!\, \theta^{n^2/8-(n\log n)/(\log\theta)+n(c+1/4)}.\end{aligned}$$

Combining these sums establishes the upper bound in (a′).

To prove the lower bound in (a′) let

$$\begin{aligned}A &= \{w \in S_n \mid \ell(w) > 2\ell(n-1)\} \\ &= \left\{ w \in S_n \;\middle|\; \left| \ell(w) - \binom{n}{2} \right| < 2(n-1)(n/4-\ell) \right\}.\end{aligned}$$

Since each pass of the systematic scan can change the length of a permutation by at most $2(n-1)$,

$$K_1^\ell(A) = 0. \tag{7.6}$$

From equation (2.13),

$$\begin{aligned}E_\pi(\ell(w)) &= -\frac{(n-1)}{1-\theta} + \sum_{j=2}^{n} \frac{j}{1-\theta^j} \\ &= \sum_{j=2}^{n} j - \frac{(n-1)}{1-\theta} + \sum_{j=2}^{n} \frac{j\theta^j}{1-\theta^j} = \binom{n}{2} + O(n)\end{aligned}$$

and

$$\mathrm{Var}_\pi(\ell(w)) = \frac{(n-1)\theta}{(1-\theta)^2} - \sum_{j=2}^{n} \frac{j^2\theta^j}{(1-\theta^j)^2} = \frac{(n-1)\theta}{(1-\theta)^2} + O(1).$$

Hence Chebychev's inequality implies that, when n is large,

$$\begin{aligned} \pi(A) &\sim \left(1 - \frac{\mathrm{Var}_\pi(\ell(w))}{(2(n-1)(n/4-\ell))^2}\right) \\ &\sim \left(1 - \frac{\theta}{4(n-1)(1-\theta)^2(n/4-\ell)^2}\right). \end{aligned} \tag{7.7}$$

If $\ell < n/4$ then the right-hand side approaches 1 as $n \to \infty$. Thus (7.5) and (7.6) imply that $\|K_1^\ell - \pi\|_{TV} \to 1$ as $n \to \infty$, and this proves the second statement of (a$'$).

For (b$'$), use the bound $\theta^{2\ell(n-1-c(S(n)))} \le \theta^{2\ell(n-\lambda_1)}$ to obtain

$$\sum_{x\in S_n} \pi(x) \left\| \frac{K_1^\ell}{\pi} - 1 \right\|_2^2 \le \sum_{\lambda \ne (n)} \theta^{2\ell(n-\lambda_1)} d_\lambda^2.$$

Fix $\ell = -(\log n)/(\log\theta) + c$. Using the bound in Lemma 7.2(a) gives

$$\sum_{j=1}^{n/2} \sum_{\substack{\lambda\ne(n) \\ \lambda_1 = n-j}} \theta^{2\ell j} d_\lambda^2 \le \sum_{j=1}^{n/2} \theta^{2\ell j} \frac{n^{2j}}{j!} = \sum_{j=1}^{n/2} \frac{\theta^{2j(\ell + (\log n/\log\theta))}}{j!} = e^{\theta^{2c}} - 1$$

and, by using the bound in Lemma 7.2(b) and the bound on $n!$ given in [F, (9.15)],

$$\sum_{\substack{\lambda \\ \lambda_1 \le n/2}} \theta^{2\ell(n-\lambda_1)} d_\lambda^2 \le \theta^{\ell n} n! = n!\, n^{-n}\theta^{cn} \le \left(\frac{\theta^c}{e}\right)^n e^{1/12}\sqrt{2\pi n}.$$

The result follows by combining the bounds for these two sums. □

References

[A1] Y. Amit, *Convergence properties of the Gibbs sampler for perturbations of Gaussians,* Ann. Statist. 24 (1996), 122–140.

[A2] ———, *On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions,* J. Multivariate Anal. 38 (1991), 82–99.

[AG] Y. Amit and U. Grenander, *Comparing sweep strategies for stochastic relaxation,* J. Multivariate Anal. 37 (1991), 197–222.

[AK] S. Ariki and K. Koike, *A Hecke algebra of $(Z/rZ) \wr S_n$ and construction of its irreducible representations,* Adv. Math. 106 (1994), 216–243.

[BF] P. Barone and A. Frigessi, *Improving stochastic relaxation for Gaussian random fields,* Probab. Engrg. Inform. Sci. 4 (1990), 369–389.

[B1] E. Belsley, *Rates of convergence of random walk on distance regular graphs,* Probab. Theory Related Fields 112 (1998), 493–533.

[B2] ———, *Rates of convergence of Markov chains related to association schemes,* Ph.D. dissertation, Harvard Univ., Cambridge, MA, 1993.

[BHV] L. Billera, S. Holmes, and K. Vogtman, *Geometry of the space of phylogenetic trees,* technical report, Stanford Univ., Stanford, CA, 1999.

[Bo] N. Bourbaki, *Elements de mathematique. Groupes et algèbres de Lie,* Masson, Paris, 1981.

[BS] E. Brieskorn and K. Saito, *Artin-Gruppen und Coxeter-Gruppen,* Invent. Math. 17 (1972), 245–271.

[Br] K. S. Brown, *Buildings,* Springer-Verlag, New York, 1989.

[C] R. Carter, *Finite groups of Lie type,* Pure Appl. Math., Wiley, New York, 1985.

[Cr] D. E. Critchlow, *Metric methods for analyzing partially ranked data,* Lecture Notes in Statist., 34, Springer-Verlag, Berlin, 1985.

[CR] C. W. Curtis and I. Reiner, *Methods of representation theory,* vols. I and II, Wiley, New York, 1981 and 1987.

[De] P. Deligne, *Les immeubles des groupes de tresses généralisés,* Invent. Math. 17 (1972), 273–302.

[D] P. Diaconis, *Group representations in probability and statistics,* Institute of Mathematical Statistics, Hayward, CA, 1988.

[DH] P. Diaconis and P. Hanlon, *Eigen-analysis for some examples of the Metropolis algorithm,* Contemp. Math., 138, pp. 99–117, Amer. Math. Soc., Providence, RI, 1992.

[DHJLP] P. Diaconis, S. Holmes, S. Janson, S. P. Lalley, and R. Pemantle, *Metrics on compositions and coincidences among renewal sequences,* IMA Vol. Math. Appl., 76, Springer, New York, 1996.

[DR] P. Diaconis and A. Ram, *Analysis of systematic scan Metropolis algorithms using Iwahori–Hecke algebra techniques,* technical report, Stanford Univ., Stanford, CA, 2000.

[DS] P. Diaconis and L. Saloff-Coste, *What do we know about the Metropolis algorithm?,* J. Comput. System Sci. 57 (1998), 20–36.

[F] W. Feller, *An introduction to probability theory and its applications,* vol. I, Wiley, New York, 1968.

[FV] M. A. Figner and J. S. Verducci, *Probability models and statistical analyses for ranking data,* Springer-Verlag, New York, 1992.

[Fi] G. S. Fishman, *Coordinate selection rules for Gibbs sampling,* Ann. Appl. Probab. 6 (1996), 444–465.

[Fu] W. Fulton, *Young tableaux,* London Math. Soc. Stud. Texts, 35, Cambridge Univ. Press, Cambridge, U.K., 1997.

[GS] J. Goodman and A. Sokal, *Multigrid Monte-Carlo method,* Phys. Rev. D 40 (1989), 2035–2071.

[HH] J. Hammersley and D. Handscomb, *Monte Carlo methods,* Methune, London, 1965.

[Hf] P. N. Hoefsmit, *Representations of Hecke algebras of finite groups with BN-pairs of classical type,* Ph.D. dissertation, Univ. of British Columbia, Vancouver, BC, 1974.

[Hu] J. E. Humphreys, *Reflection groups and Coxeter groups,* Cambridge Univ. Press, Cambridge, U.K., 1990.

[KeS] J. G. Kemeny and J. L. Snell, *Finite Markov chains,* Van Nostrand, New York, 1960.

[KiS] R. Kilmoyer and L. Solomon, *On the theorem of Feit–Higman,* J. Combin. Theory Ser. A 15 (1973), 310–322.

[L1] G. Letac, *Problèmes classiques de probabilité sur un couple de Gelfand,* Lecture Notes in Math., 861, Springer-Verlag, New York, 1981.

[L2] ———, *Les fonctions sphériques d'un couple de Gelfand symétrique et les chaînes de Markov,* Adv. in Appl. Probab. 14 (1982), 272–294.

[Mac] I. G. Macdonald, *Symmetric functions and Hall polynomials,* Oxford Math. Monogr., Oxford Univ. Press, New York, 1995.

[M] C. L. Mallows, *Non-null ranking models. I,* Biometrika 44 (1957), 114–130.

[Ma] J. I. Marden, *Analyzing and modeling rank data,* Monogr. Statist. Appl. Probab., 64, Chapman & Hall, London, 1995.

[MRRTT] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *Equations of state calculations by fast computing machines,* J. Chem. Phys. 21 (1953), 1087–1092.

[P] I. M. Pak, *When and how n choose k,* Ser. Discrete Math. Theoret. Comput. Sci., 43, pp. 191–238, Amer. Math. Soc., Providence, RI, 1999.

[R] A. Ram, *Seminormal representations of Weyl groups and Iwahori–Hecke algebras,* Proc. London Math. Soc. (3) 75 (1997), 99–133

[RS] G. Roberts and S. Sahu, *Updating schemes, correlation schemes, correlation structure, blocking and parameterization for the Gibbs sampler,* J. Roy. Statist. Soc. B 59 (1997), 291–318.

[RX] K. Ross and D. Xu, *Hypergroup deformations and Markov chains,* J. Theoret. Probab. 7 (1994), 813–830.

[S-C] L. Saloff-Coste, *Lectures on probability theory and statistics,* Lecture Notes in Math., 1665, pp. 301–413, Springer-Verlag, Berlin, 1997.

[S] C. Schoolfield, *Random walks on wreath products of groups and Markov chains on related homogeneous spaces,* Ph.D. dissertation, Johns Hopkins Univ., Baltimore, MD, 1998.

[SB] W. Shannon and D. Banks, *Combining classification trees using MLE,* Statist. in Medicine 18 (1999), 727–740.

[Si] J. Silver, *Weighted Poincaré and exhaustive approximation techniques for scaled Metropolis–Hastings and spectral total variation convergence bounds in infinite commutable Markov chain theory,* Ph.D. dissertation, Harvard Univ., Cambridge, MA, 1996.

P. Diaconis
Department of Mathematics
Stanford University
Stanford, CA 94305

A. Ram
Department of Mathematics
University of Wisconsin, Madison
Madison, WI 53706

ram@math.wisc.edu