THEORY OF COMPUTING, Volume 21 (3), 2025, pp. 1–41
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2020

Iterated Decomposition of Biased Permutations Via New Bounds on the Spectral Gap of Markov Chains

Sarah Miracle   Amanda Pascoe Streib   Noah Streib

Received April 2, 2021; Revised January 30, 2025; Published August 22, 2025

Abstract. In this paper, we address a conjecture of Fill (2003) about the spectral gap of a nearest-neighbor transposition Markov chain $\mathcal{M}_{\text{nn}}$ over biased permutations of $[n]$. Suppose we are given an array of input probabilities $\mathcal{P} = \{p_{i,j}\}$ for all $1 \le i, j \le n$ with $p_{i,j} = 1 - p_{j,i}$. The Markov chain $\mathcal{M}_{\text{nn}}$ operates by uniformly choosing a pair of adjacent elements, i and j, and putting i ahead of j with probability $p_{i,j}$ and j ahead of i with probability $p_{j,i}$, independent of their current ordering.

We build on previous work of Miracle and Streib (LATIN'18 and *SIAM J. Discr. Math.*, 2024) that analyzed the spectral gap of $\mathcal{M}_{\text{nn}}$ when the particles in $[n]$ fall into k classes. There, the authors iteratively decomposed $\mathcal{M}_{\text{nn}}$ into simpler chains, but incurred a multiplicative penalty of n^{-2} for each application of the decomposition theorem of Martin and Randall (FOCS '00), leading to an exponentially small lower bound on the gap. We make progress by introducing a new complementary

A preliminary version of this paper appeared in the Proceedings of the 24th International Conference on Randomization and Computation [26].

ACM Classification: Theory of computation, Random walks and Markov chains

AMS Classification: 68Q87, 60J10

Key words and phrases: Markov chains, permutations, decomposition, spectral gap, iterated decomposition

DOI: 10.4086/toc.2025.v021a003

decomposition theorem. We introduce the notion of ϵ-orthogonality, and show that for ϵ-orthogonal chains, the complementary decomposition theorem may be iterated $O(1/\sqrt{\epsilon})$ times while only giving away a constant multiplicative factor on the overall spectral gap. We show that the decomposition given by Miracle and Streib (*op. cit.*) of a related Markov chain $\mathcal{M}_{\text{pp}}$ over k-class particle systems is ϵ-orthogonal for some $\epsilon \le 1/n^2$ as long as the number of particles in each class is at least $C \log n$, where C is a constant not depending on n. We then apply the complementary decomposition theorem iteratively n times to prove nearly optimal bounds on the spectral gap of $\mathcal{M}_{\text{pp}}$ and to further prove the first inverse-polynomial bound on the spectral gap of $\mathcal{M}_{\text{nn}}$ when k is as large as $\Theta(n/\log n)$. The previous best known bound assumed k was bounded.

1 Introduction

For $n \in \mathbb{N}$, the problem of generating permutations of $[n] = \{1, 2 \ldots, n\}$ at random is foundational in the history of computer science [19]. Markov chains for sampling permutations arise in a variety of contexts, including self-organizing lists [17, 32], card shuffling [2, 35], and search engines [5]. The spectral gap of a Markov chain provides a measure of the rate of convergence to stationarity, which is crucial to the efficiency of Markov chain algorithms for sampling.

Suppose we are given an array of input probabilities $\mathcal{P} = \{p_{i,j}\}$ $(1 \le i, j \le n)$ with $p_{i,j} = 1 - p_{j,i}$. A natural Markov chain $\mathcal{M}_{\text{nn}}$ over permutations operates by uniformly choosing a pair of adjacent elements, i and j, and putting i ahead of j with probability $p_{i,j}$ and j ahead of i with probability $p_{j,i}$, regardless of their current ordering. We call $\mathcal{M}_{\text{nn}}$ the *nearest-neighbor* transposition chain.

The Markov chain $\mathcal{M}_{\text{nn}}$ was among the first Markov chains studied in terms of its computational efficiency for sampling [1, 10, 9]. Its spectral gap has been studied extensively, both in the uniform and in the biased settings [3, 4, 8, 10, 35].

A central question is, under what conditions is there a lower bound on the spectral gap that is is an inverse polynomial in n. Such a lower bound implies a polynomial bound on the *mixing time*, or the time until the chain will be "close" to its stationary distribution. We say $\mathcal{P}$ is *positively biased* if $p_{i,j} \ge 1/2$ for all $i < j$. Without the positive bias restriction, it is easy to find examples where the mixing time of $\mathcal{M}_{\text{nn}}$ is not polynomial (see, e. g., [4]); however, even with this restriction the mixing time can still not be polynomial, as demonstrated in [4]. In [4], the authors give an example of positively biased probabilities $\mathcal{P}$ for which they prove the mixing time of $\mathcal{M}_{\text{nn}}$ is exponential in n.

In 2003, Fill [14] introduced the following monotonicity conditions: $p_{i,j+1} \ge p_{i,j}$ for all $1 \le i < j \le n-1$ and $p_{i-1,j} \ge p_{i,j}$ for all $2 \le i < j \le n$. He conjectured that if $\mathcal{P}$ is positively biased and monotone, then the spectral gap of $\mathcal{M}_{\text{nn}}$ is at least an inverse polynomial in n, and moreover that the smallest spectral gap is attained when each $p_{i,j} = 1/2$ and the stationary distribution is uniform.

Despite significant work on this subject, Fill's conjecture remains mostly open after more than 15 years. In the uniform case, Wilson [35] produced a clever path coupling argument showing the mixing time is $\Theta(n^3 \log n)$, with upper and lower bounds within a factor of 2. Fill's conjecture states that this should be the worst case mixing time among all arrays $\mathcal{P}$ which are positively biased and monotone. Various papers [3, 4] have identified different classes of arrays $\mathcal{P}$ for which $\mathcal{M}_{\mathrm{nn}}$ may cleverly be viewed as the direct product of simpler independent Markov chains, and thus may be analyzed easily in terms of those chains. In [3], the authors proved a bound of $O(n^2)$ on the mixing time of $\mathcal{M}_{\mathrm{nn}}$ in the case that $p_{i,j} = p$ for all $i < j$, and in [4] the authors considered the case that $p_{i,j}$ depends only on the smaller of i and j.

Among the positively biased, monotone arrays $\mathcal{P}$ that have proven to be challenging to analyze are those arising in the context of self-organizing lists, where each element i has a frequency w_i of being requested, and then moved ahead by one in the list; in this case, $p_{i,j} = w_i/(w_i + w_j)$. The Markov chain $\mathcal{M}_{\mathrm{nn}}$ was termed a "gladiator chain" in this case by Haddadan and Winkler [16].

$$\text{A partition of [7]:}\quad \begin{aligned} C_1 &= \{1,2,3\} \\ C_2 &= \{4,5\} \\ C_3 &= \{6\} \\ C_4 &= \{7\} \end{aligned} \qquad\qquad 3765241 \longrightarrow 1432121$$

Figure 1: An example of a 4-class permutation and corresponding 4-particle process, with $n = 7$.

It is instructive to consider *k-classes* [25], where $[n]$ is partitioned into k classes and particles from class i and class j interact with a fixed probability $\overline{p}_{i,j}$.[1] When $k = n$, this captures the general setting. In this context, $\mathcal{M}_{\mathrm{nn}}$ can be seen as having dual duties: *whisking*, which uniformly mixes particles of the same type[2], and *sifting*, which mixes particles of different types – that is, which changes relative orders of particles of different types [16]. As the mixing properties of the uniform chain are well-understood, it is sufficient to analyze the sifting operation in isolation [16, 25]. By discarding moves between particles in the same class, we are left with a linear k-particle process that maintains elements within each particle class in fixed relative orders (and therefore we may drop their individual labels and re-index, identifying all elements from class i with the label i, as is done in Figure 1).

In this paper, we use the version of the k-particle process introduced in [25], called $\mathcal{M}_{\mathrm{pp}}$, which is also allowed to make certain non-adjacent transpositions—it may swap i and j if all elements between them are smaller than both i and j. This simplifies our analysis, and as in [25], we compensate by using comparison techniques [9, 30] when evaluating the spectral gap of $\mathcal{M}_{\mathrm{nn}}$.

The *bias* towards having a particle of type i ahead of a particle of type j is $\overline{q}_{i,j} = \overline{p}_{i,j}/\overline{p}_{j,i}$. Let $q = \min_{i,j} \overline{q}_{i,j}$ be the minimum bias. We say that the bias is *bounded* if q is bounded below by a constant greater than 1. After a series of papers [16, 25], it was shown that the mixing time of $\mathcal{M}_{\mathrm{pp}}$ was $O(n^{2k+4})$ (and the spectral gap is at least $\Omega(n^{-2(k-1)})$) which $\mathcal{P}$ is positively biased, monotone, and bounded. These results apply to the gladiator chain (self-organizing lists) with k distinct frequencies. In fact, the result in [25] requires only *weak monotonicity*, and not full

[1] The bar in the notation indicates that we have re-indexed the probabilities by class.

[2] We use the terms "type" and "class" interchangeably.

monotonicity. Weak monotonicity for k-classes is defined as follows.

Definition 1.1 ([4])**.** If $\mathcal{P}$ forms a k-class, then $\mathcal{P}$ is weakly monotonic if properties 1 and either 2 or 3 are satisfied.

1. $\overline{p}_{i,j} \geq 1/2$ for all $1 \leq i < j \leq k$, and
2. $\overline{p}_{i,j+1} \geq \overline{p}_{i,j}$ for all $1 \leq i < j \leq k-1$ or
3. $\overline{p}_{i-1,j} \geq \overline{p}_{i,j}$ for all $2 \leq i < j \leq k$.

The aforementioned results are based on a natural decomposition of $\mathcal{M}_{\text{pp}}$ into simpler chains, but not as a direct product. To get a bound on the overall spectral gap, the authors of [25] used the decomposition theorem of [23], which bounds the spectral gap of a Markov chain in terms of the spectral gaps of the simpler Markov chains. Unfortunately, one incurs an extra factor of n^{-2} each time it is applied in this setting, and in [25] it is applied iteratively $k-2$ times. Thus, this produced a bound of $\Omega(n^{-2(k-1)})$ on the spectral gap, which is an inverse polynomial only for constant k.

To make this iterated decomposition scheme work for larger k requires a stronger decomposition theorem, and that is the main focus of the present paper. We introduce a new decomposition theorem that allows us to achieve nearly optimal bounds of $\Omega(n^{-2})$ on the spectral gap of $\mathcal{M}_{\text{pp}}$ for bounded k-classes, as long as the number of particles of each type is at least $C_q \log n$ (where C_q is a constant depending on the minimum bias q; i. e., not depending on n). We believe this new decomposition theorem is of independent interest and will have other applications.

1.1 The decomposition method

The decomposition method was first introduced by Madras and Randall [21], and has been subsequently used and modified to produce the first polynomial time bounds on the spectral gaps of many interesting Markov chains [6, 7, 11, 13, 15, 16, 18, 22, 23, 24, 27, 28]. Suppose $\mathcal{M}$ is a finite, ergodic Markov chain that is reversible and has stationary distribution π. Let $\Omega = \bigcup_{i=1}^{r} \Omega_i$ be a partition of the state space of $\mathcal{M}$ and γ_i be the spectral gap of $\mathcal{M}$ restricted to Ω_i. The disjoint decomposition theorem of [23] states that the spectral gap γ of $\mathcal{M}$ satisfies $\gamma \geq \frac{1}{2}\gamma_{\min}\bar{\gamma}$, where $\gamma_{\min} = \min_i \gamma_i$ and $\bar{\gamma}$ is the spectral gap of a certain *projection* chain over states $[r]$.

There has been significant effort towards improving the decomposition technique by providing stronger bounds in special cases [7, 13, 15, 18, 22, 23, 27, 28]. While γ may indeed be on the order of $\gamma_{\min}\bar{\gamma}$—one example is the random walk on the path, decomposed into two shorter paths—there are instances in which it may instead satisfy the much larger bound $\gamma \geq c \min\{\gamma_{\min}, \bar{\gamma}\}$, for some constant c. The simplest such example is the direct product of two independent Markov chains [4, 13]; in this case, $c = 1$.

Tight bounds are especially important when applying the decomposition method iteratively, as was done in [25]. At each level of the induction, $\bar{\gamma} = \Theta(n^{-2})$, so the original bound of [23] yields $\gamma = \Omega(n^{-2(n-1)})$ for the final iteration. Even a bound of the form $\gamma \geq c \min\{\gamma_{\min}, \bar{\gamma}\}$, such as the one in [27], would introduce a factor of c for each application, and would thus yield a bound that is an inverse exponential in n if $c < 1$ is constant. The bounds in [18] are iterable in some cases, but $\mathcal{M}_{\mathrm{pp}}$ does not satisfy those conditions.

1.2 Our results

In this paper, we develop a new decomposition framework that yields iterable bounds for a new class of Markov chains. Among our results, we present a complementary decomposition theorem, which achieves a tight bound on γ without appealing to a bound on the gap $\bar{\gamma}$ of the projection chain, but rather the minimum gap $\tilde{\gamma}_{\min}$ of certain *complementary restrictions* $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_{\tilde{r}}$. We first consider the simple setting where the state space Ω can be seen as a product space, i. e., $\Omega = \Omega_1 \times \Omega_2$. In other words, for each $a \in \Omega_1$ and each $b \in \Omega_2$, there is a unique $\sigma = (a, b) \in \Omega$. This setting is similar to the direct product of independent Markov chains, but the transition probabilities are not necessarily independent. We define a restriction chain P_a for each $a \in \Omega_1$ that fixes a and operates only on the second coordinate. Similarly, we define a complementary restriction chain $\tilde{P}_b$, which fixes b and operates only on the first coordinate. Recall π is the stationary distribution of $\mathcal{M}$. We write $\pi(a) = \sum_{b\in\Omega_2} \pi(a, b)$ and $\pi(b) = \sum_{a\in\Omega_1} \pi(a, b)$. Define

$$r(a, b) = \frac{\pi(a, b)}{\pi(a)\pi(b)}\,.$$

The function $r(a, b)$ allows us to capture the degree of dependence between a and b. Let

$$\epsilon = \sum_{(a,b)\in\Omega} \pi(a, b) \left(\sqrt{r(a, b)} - 1/\sqrt{r(a, b)}\right)^2. \tag{1.1}$$

We say a decomposition satisfying the equality above is *ϵ-orthogonal*.

Theorem 1.2. *For any ϵ-orthogonal decomposition of Markov chain $\mathcal{M}$ on product space Ω,*

$$\gamma(\mathcal{M}) \geq \min\{\gamma_{\min}, \tilde{\gamma}_{\min}\} \left(1 - \sqrt{\epsilon}\right)^2.$$

This bound can be iterated t times with only a constant overhead, as long as $\sqrt{\epsilon} \leq 1/t$. We note that parts of the proof of this theorem are similar to a "multi-decomposition" result of Destainville [7], which we discuss in Section 4. There we also present several generalizations of Theorem 1.2, which apply beyond the product space setting.

Favorably, analysis of ϵ-orthogonality requires only a comparison between two stationary distributions and not an analysis of the dynamics of any Markov chain. For example, when $\mathcal{M}$ is a direct product of independent Markov chains, we have that $r(a, b) = 1$ for all pairs $a \in \Omega_1$ and $b \in \Omega_2$ and the decomposition is 0-orthogonal, leading to the bound $\gamma \geq \min\{\gamma_{\min}, \tilde{\gamma}_{\min}\}$, as expected. Notice, however, that we do not actually require a such a strong pointwise bound on $r(a, b)$. The notion of ϵ-orthogonality captures the *average* value of $r(a, b)$, and allows us to achieve tight bounds on γ even when the constituent Markov chains are only *nearly* independent. Indeed, it is possible to prove ϵ-orthogonality for very small ϵ even if $r(a, b)$ is far from 1 for

413112413142

Figure 2: At level 2 of the decomposition, all particles of type 1 are in fixed positions, and the underlined particle 2 may swap with the 3 to its left or the 4 immediately to its right.

1223123112133413143444

Figure 3: A "typical" configuration at level 2 has no 2's appearing after any $j \geq 4$.

pathological pairs a and b, as long as it is close to 1 on average. Importantly, this holds even though the elements in this "bad" space are visited polynomially often.

Armed with our new decomposition theorem, we use the iterated decomposition in [25] to achieve nearly optimal bounds on the spectral gap of the k-particle process $\mathcal{M}_{\text{pp}}$. We prove that at each level of the iteration, the value of ϵ in the decomposition satisfies $\epsilon \leq 1/n^2$. Thus, we may apply Theorem 1.2 iteratively $k \leq n$ times to get a bound of $\Omega(n^{-2})$ on the spectral gap $\gamma(\mathcal{M}_{\text{pp}})$. This bound is optimal up to constants. More formally, let $N^* = \max\left\{\frac{\log(6n^2)+\log((1+q)/(q-1))}{\log((1+q)/2)}, \frac{\log^2(14)}{\log^2(2q/(1+q))}\right\}$, and let c_i denote the number of particles of type i, for $1 \leq i \leq k$. Note that $N^* := C_q \log n$ where C_q depends only on q. We prove the following.

Theorem 1.3. *If the probability array $\mathcal{P}$ is weakly monotonic and forms a bounded k-class with $c_i \geq 2N^*$ for all $1 \leq i \leq k$, then the spectral gap γ of the chain $\mathcal{M}_{pp}$ satisfies $\gamma = \Omega(n^{-2})$.*

The iterated decomposition works as follows. Let C_i be the set of particles of type i, $C_{<i}$ the set of particles with type less than i, and $C_{>i}$ the set of particles with type bigger than i. At the ith iteration, all particles in $C_{<i}$ are in fixed positions, and those in $C_i \cup C_{>i}$ are allowed to swap across them – that is, we allow non-adjacent transpositions of, for example, a particle of type $j \geq i$ and a particle of type $j' \geq i$, as long as all particles between them in the permutation are of type less than i. See Figure 2 for an example. This decomposition is designed to exploit the hypothesis that the movement of the particles in C_i is nearly independent of the relative order of the particles in $C_{>i}$. The tool of ϵ-orthogonality allows us to make this precise. We define the complementary restriction chains to contain the moves involving only particles in $C_{>i}$, and we define the restriction chains to contain moves between particles in C_i and particles in $C_{>i}$. We prove that this decomposition is $\epsilon-$orthogonal where $\epsilon \leq 1/n^2$ when $|C_{i+1}| = c_{i+1}$ is large enough. Indeed, the highest probability configuration is the one in which particles are sorted by class, with all particles of smaller type appearing before particles of larger type. Thus, having $|C_{i+1}|$ large ensures that a typical configuration will not have a particle in C_i after a particle in $C_{>(i+1)}$, as this requires many transpositions from the highest probability configuration, and each costs a factor of at least the minimum bias q; see Figure 3. Note that such "bad" configurations are polynomially suppressed (meaning, their total probability is bounded by an inverse polynomial in n), but not exponentially suppressed (bounded by an inverse exponential in n).

We use Theorem 1.3 and comparison techniques to prove the following result for $\mathcal{M}_{\text{nn}}$.

Theorem 1.4. *If the probability array $\mathcal{P}$ is weakly monotonic and forms a bounded k-class with at least $2N^*$ particles in each class, then $\gamma(\mathcal{M}_{nn}) = \Omega(n^{-7})$.*

This is the first inverse-polynomial lower bound on the spectral gap of $\mathcal{M}_{\rm nn}$ for monotone bounded k-classes for k as large as $\Theta(n/\log n)$. This is a significant improvement over the previous result which was inverse polynomial only for constant k [23]. Theorem 1.4 also leads to a bound of $O(n^9)$ on the mixing time of $\mathcal{M}_{\rm nn}$ under the same conditions.

1.3 Outline

The layout of the paper is as follows. In Section 2, we begin with some notation and terminology. In Section 3, we give details on Theorem 1.2 and illustrate its use by applying it to the one-dimensional Ising model. In Section 4, we generalize the notion of ϵ-orthogonality to non-product spaces and present a more general complementary decomposition theorem that applies to all ϵ-orthogonal decompositions. We also present a classical decomposition theorem that generalizes some results of [18] and show how our decomposition theorems relate to previous results. Finally, we give a brief summary of the proof techniques for the decomposition theorems. Section 5 and Section 6 present the proofs of the complementary and classical decomposition theorems, respectively. In Section 7, we apply Theorem 1.2 to the biased permutation problem.

2 Preliminaries

We assume $\mathcal{M}$ is an ergodic Markov chain over a finite state space Ω with transition matrix P. We also assume $\mathcal{M}$ is reversible and has stationary distribution π; that is, it satisfies the *detailed balance* condition: for all $\sigma, \tau \in \Omega$, $\pi(\sigma)P(\sigma,\tau) = \pi(\tau)P(\tau,\sigma)$. Notationally, we write $\pi(S) = \sum_{\sigma\in S}\pi(\sigma)$ for any set $S \subseteq \Omega$.

Let $\Omega = \cup_{i=1}^{r}\Omega_i$ be a partition of the state space into r pieces. For each $i \in [r]$, define $P_i = P(\Omega_i)$ as the restriction of P to Ω_i which rejects moves that leave Ω_i. Formally, P_i is defined as follows: if $\sigma \neq \tau$ and $\sigma, \tau \in \Omega_i$ then $P_i(\sigma,\tau) = P(\sigma,\tau)$; if $\sigma \in \Omega_i$ then $P_i(\sigma,\sigma) = 1 - \sum_{\tau\in\Omega_i,\tau\neq\sigma} P_i(\sigma,\tau)$. The Markov chain $\mathcal{M}_i$ with transition matrix P_i is called a *restriction* Markov chain, and its state space is Ω_i. Let π_i be the normalized restriction of π to Ω_i; i. e., $\pi_i(S) = \pi(S\cap\Omega_i)/\pi(\Omega_i)$ for any $S \subseteq \Omega$. The chain $\mathcal{M}_i$ is ergodic, reversible, and has stationary distribution π_i.

Classical decomposition theorems use the so-called *projection* chain $\bar{\mathcal{M}}$ with transition matrix $\bar{P}$ on the state space $[r]$ defined by $\bar{P}(i,j) = \pi(\Omega_i)^{-1}\sum_{\sigma\in\Omega_i,\tau\in\Omega_j}\pi(\sigma)P(\sigma,\tau)$. The stationary distribution $\bar{\pi}$ of $\bar{\mathcal{M}}$ satisfies $\bar{\pi}(i) := \pi(\Omega_i)$.

In contrast, for our complementary decomposition theorems, we will decompose P in a slightly different way. Define $\hat{P}$ as the block diagonal $|\Omega|\times|\Omega|$ matrix with the P_i matrices along the diagonal; i. e., $\hat{P}$ is obtained from P by rejecting moves between different parts of the partition. Define $\tilde{P}$ to be the transition matrix of the Markov chain defined by rejecting moves from σ to τ if σ and τ are within the same Ω_i. The matrix $\tilde{P}$ defines a *complementary* partition $\Omega = \cup_{j=1}^{\tilde{r}}\tilde{\Omega}_j$, where each $\tilde{\Omega}_j$ is a maximal subset of Ω that is connected by $\tilde{P}$. For each $j \in [\tilde{r}]$, define the *complementary restriction* $\tilde{P}_j = P(\tilde{\Omega}_j)$ as the restriction of P to $\tilde{\Omega}_j$ which rejects moves that leave $\tilde{\Omega}_j$. The complementary restriction $\tilde{P}_j$ is also ergodic, reversible, and its stationary distribution is the normalized restriction of π to $\tilde{\Omega}_j$, which we call $\tilde{\pi}_j$. Notice that the complementary restrictions are defined by the decomposition $P_1, P_2, \ldots, P_r$.

Observe $(\hat{P}+\tilde{P})(\sigma,\tau) = P(\sigma,\tau)$ unless $\sigma = \tau$, and $(\hat{P}+\tilde{P})(\sigma,\sigma) = P(\sigma,\sigma)+1$, since each move of P gets rejected in exactly one of $\hat{P}$ or $\tilde{P}$ (and of course the probability of transitioning from a state is 1). Therefore, we have $P = \hat{P} + \tilde{P} - I_{|\Omega|}$, where we write I_n to mean the $n \times n$ identity matrix.

The efficiency of a Markov chain $\mathcal{M}$ is a function of its spectral gap, denoted $\gamma(\mathcal{M})$, which is defined as the difference of 1 and the second largest eigenvalue of its transition matrix. Letting $\gamma_i = \gamma(\Omega_i)$ and $\tilde{\gamma}_j = \gamma(\tilde{\Omega}_j)$, the complementary decomposition theorem, Theorem 1.2, is proven by analyzing the spectral gaps $\gamma_{\min} = \min_i \gamma_i$ and $\tilde{\gamma}_{\min} = \min_j \tilde{\gamma}_j$. Note that if some restriction or complementary restriction has a single element, its spectral gap is taken to be 1.

3 Introductory examples

In this section, we show how to apply our new complementary decomposition theorem by considering a few simple examples. Recall $r(a,b) = \pi(a,b)/(\pi(a)\pi(b))$ and

$$\epsilon = \sum_{(a,b)\in\Omega} \pi(a,b)\left(\sqrt{r(a,b)} - 1/\sqrt{r(a,b)}\right)^2 .$$

Theorem 1.2 states that the spectral gap γ of $\mathcal{M}$ satisfies $\gamma \geq \min\{\gamma_{\min}, \tilde{\gamma}_{\min}\}\left(1-\sqrt{\epsilon}\right)^2$. A simple application of Theorem 1.2 is to a Markov chain $\mathcal{M}$ that is the direct product of two Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$. It is easy to see that $r(a,b) = 1$ for all a, b, and so this proves $\gamma(\mathcal{M}) = \min\{\gamma(\mathcal{M}_1), \gamma(\mathcal{M}_2)\}$. By iterating on $\tilde{\gamma}_{\min}$, we can immediately prove the following well-known result.

Corollary 3.1. *If $\mathcal{M}$ is the direct product of Markov chains $\{\mathcal{M}_i\}$, then $\gamma(\mathcal{M}) = \min_i \gamma(\mathcal{M}_i)$.*

3.1 One-dimensional Ising model

As a second introductory example prior to our main application, we consider the one-dimensional Ising model. Here each configuration $\sigma \in \Omega$ is an assignment of a "spin" (either +1 or -1) to each of n vertices connected to form a line; see Figure 4. Let $\lambda = \mathrm{e}^{-\beta}$, where $\beta > 0$ represents inverse temperature. We are interested in sampling from the Gibbs distribution given by $\pi(\sigma) = \mathrm{e}^{-\beta H(\sigma)}/Z$, where the Hamiltonian $H(\sigma)$ is the number of edges whose endpoints have different spins and Z is the normalizing constant $\sum_{\sigma\in\Omega} \mathrm{e}^{-\beta H(\sigma)}$, also called the *partition function*. (See [20] for background on the ferromagnetic Ising model.)

Consider the Glauber dynamics Markov chain $\mathcal{M}_{\mathrm{gd}}$.

Glauber Dynamics $\mathcal{M}_{\mathrm{gd}}$

`Starting at any configuration` σ^0`, iterate the following:`

- At time t, choose a vertex $1 \leq i \leq n$ uniformly at random.

- Set the spin of vertex i to $+1$ with probability $p = (\pi(\sigma^{t,i\leftarrow+}))/(\pi(\sigma^{t,i\leftarrow+})+\pi(\sigma^{t,i\leftarrow-}))$ where $\sigma^{t,i\leftarrow+}$ is identical to σ^t with the spin of vertex i set to $+1$ (or -1 for $\sigma^{t,i\leftarrow-}$).

- Otherwise, set the spin of vertex i to -1 with probability $1-p$.

For simplicity, we will assume that n is a power of 2. To apply our theorem, we decompose the state space by breaking configurations in half along the middle edge; again, see Figure 4. Transitions that fix the signs on the left are part of the restriction chains, and transitions that fix signs on the right are part of the complementary restriction chains. Thus, our restrictions and complementary restrictions are both $1 \times n/2$ Ising models for which we can readily apply induction. Let a be the assignment of signs to the left $n/2$ vertices and b be the signs of the right $n/2$ vertices. It is straightforward to see $\sigma = (a, b)$ gives a unique configuration and that the state space is a product space. However, $\mathcal{M}_{\mathrm{gd}}$ is not a direct product of independent Markov chains on a and b because the probability of changing a sign of either of the middle two vertices ($n/2$ or $n/2+1$) depends on the sign of the other middle vertex. In order to apply Theorem 1.2, we first analyze $r(a, b) = \pi(a, b)/(\pi(a)\pi(b))$ and subsequently ϵ. The techniques used here are similar to, but simpler than, those used Section 7.

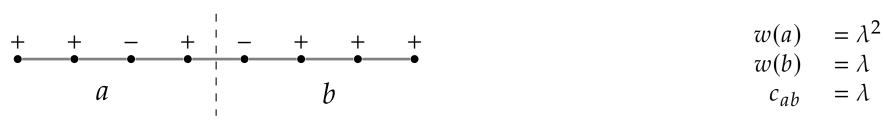


Figure 4: An example configuration of the one-dimensional Ising model.

Define $\lambda = \mathrm{e}^{-\beta}$. Let $w(a) = \lambda^{H(a)}$, where $H(a)$ is the number of edges in a (the left half) with disagreeing signs. Analogously, define $w(b) = \lambda^{H(b)}$. Let $c_{ab} = \lambda$ if the middle signs disagree and $c_{ab} = 1$ otherwise. Thus $\pi(a, b) = w(a)w(b)c_{ab}/Z$. Let $\Omega^* \subset \Omega$ be the configurations where the middle two vertices agree. Define $Z_A = \sum_a w(a)$ and $Z_B = \sum_b w(b)$.

For any fixed a, we have $\sum_{b:(a,b)\in\Omega^*} w(b) = \sum_{b:(a,b)\notin\Omega^*} w(b) = \frac{1}{2}Z_B$, since we can swap spins on all vertices in b to obtain a unique configuration $b' \in \Omega \setminus \Omega^*$ with $w(b') = w(b)$. Thus, we have

$$Z = \sum_{(a,b)\in\Omega^*} w(a)w(b) + \lambda \sum_{(a,b)\notin\Omega^*} w(a)w(b) = (1+\lambda) \sum_{(a,b)\in\Omega^*} w(a)w(b) = \frac{(1+\lambda)Z_A Z_B}{2}.$$

We consider two different cases for $r(a, b)$ depending on whether the sign of the middle two vertices agree. First consider the case where $(a, b) \in \Omega^*$. Here we have

$$\pi(a) = \sum_{b'} \pi(a, b') = \sum_{b':(a,b')\in\Omega^*} \frac{w(a)w(b')}{Z} + \lambda \sum_{b':(a,b')\notin\Omega^*} \frac{w(a)w(b')}{Z} = \frac{w(a)(1+\lambda)Z_B}{2Z},$$

and similarly $\pi(b) = w(b)(1+\lambda)Z_A/(2Z)$. Therefore

$$r(a, b) = \frac{\pi(a, b)}{\pi(a)\pi(b)} = \frac{4Zw(a)w(b)c_{ab}}{w(a)w(b)(1+\lambda)^2 Z_A Z_B} = \frac{2}{1+\lambda}.$$

The next case is almost identical except that $c_{ab} = \lambda$, so we have for $(a, b) \notin \Omega^*$ that $r(a, b) = 2\lambda/(1 + \lambda)$.

Next we use our analysis of $r(a, b)$ to bound ϵ. First notice that since $\sum_{(a,b)\in\Omega^*} \pi(a, b) + \sum_{(a,b)\notin\Omega^*} \pi(a, b) = 1$ and $\sum_{(a,b)\notin\Omega^*} \pi(a, b) = \lambda \sum_{(a,b)\in\Omega^*} \pi(a, b)$, we have that $\sum_{(a,b)\in\Omega^*} \pi(a, b) = 1/(1 + \lambda)$ and $\sum_{(a,b)\notin\Omega^*} \pi(a, b) = \lambda/(1 + \lambda)$. This yields

$$\begin{aligned}\epsilon &\leq \sum_{(a,b)\in\Omega^*} \pi(a, b)(\sqrt{r(a, b)} - 1/\sqrt{r(a, b)})^2 + \sum_{(a,b)\notin\Omega^*} \pi(a, b)(\sqrt{r(a, b)} - 1/\sqrt{r(a, b)})^2 \\ &= \frac{1}{1+\lambda}\left(\sqrt{\frac{2}{1+\lambda}} - \sqrt{\frac{1+\lambda}{2}}\right)^2 + \frac{\lambda}{1+\lambda}\left(\sqrt{\frac{2\lambda}{1+\lambda}} - \sqrt{\frac{1+\lambda}{2\lambda}}\right)^2 = \left(\frac{1-\lambda}{1+\lambda}\right)^2 .\end{aligned}$$

Applying the complementary decomposition theorem (Theorem 1.2) gives the following recurrence: $\gamma_n \geq \gamma_{n/2}\left(\frac{2\lambda}{1+\lambda}\right)^2$. Since the base case has gap $\Omega(n^{-1})$, this solves to $\gamma_n = \Omega(n^{-c})$ for $c = 1 + 2\log_2\left(\frac{1+\lambda}{2\lambda}\right)$. Note that while this does not give a tight bound, the constant c is strictly better than the constant given by [18] and, unlike earlier decomposition approaches, we have not incurred an extra factor of n with each application of the decomposition theorem.

3.2 Ising model on bounded-degree trees

As in [18], our proof for the one-dimensional Ising model can be easily generalized to trees with constant maximum degree r. A straightforward induction shows that such a tree T on n vertices has an edge whose deletion cuts T into two components, each with size at least $n/(r + 1)$. We let a represent the spins on one component and b the spins on the other. At each level of the induction, we compute $r(a, b)$ and ϵ using arguments similar to those in Section 3.1 to get $\gamma_n = \Omega(n^{-c})$ for $c = 1 + 2\log_{(r+1)/r}\left(\frac{1+\lambda}{2\lambda}\right)$.

4 Generalizations and comparison with other theorems

In Section 4.1 and Section 4.2, we present several generalizations of Theorem 1.2 and compare these results with related prior work. Our decomposition theorems fall into two categories: complementary decomposition theorems that rely on the notion of ϵ-orthogonality between the restrictions and complementary restrictions, and more classical decomposition theorems based on the projection Markov chain. In Section 4.3, we summarize the proofs, with more details given in Section 5 and Section 6.

4.1 Generalized complementary decomposition theorems

Theorem 1.2 generalizes easily to non-product spaces. Define $r(i, j) = \pi(\Omega_i \cap \tilde{\Omega}_j)/(\pi(\Omega_i)\pi(\tilde{\Omega}_j))$, for any $1 \leq i \leq r, 1 \leq j \leq \tilde{r}$. We say that $\{\Omega_i\}$ and $\{\tilde{\Omega}_j\}$ is an ϵ-orthogonal decomposition of $\mathcal{M}$

if

$$\epsilon = \sum_{(i,j)} \pi(\Omega_i \cap \tilde{\Omega}_j)(\sqrt{r(i,j)} - 1/\sqrt{r(i,j)})^2 .$$

Theorem 4.1. *For any ϵ-orthogonal decomposition of $\mathcal{M}$, $\gamma(\mathcal{M}) \geq \min\{\gamma_{\min}, \tilde{\gamma}_{\min}\}\left(1 - \sqrt{\epsilon}\right)^2$.*

We use Theorem 4.2 to prove Theorem 4.1, which in turn implies Theorem 1.2.

Theorem 4.2. $\gamma(\mathcal{M}) \geq \min_{x \perp \sqrt{\pi}, \|x\|=1} \gamma_{\min} \| \perp\!\!\!\perp x\|^2 + \tilde{\gamma}_{\min} \| \perp\!\!\!\perp \tilde{x}\|^2$.

Here, $\perp\!\!\!\perp x$ and $\perp\!\!\!\perp \tilde{x}$ are orthogonal projections of a vector x onto the complement of the eigenspace of the top eigenvectors of certain matrices (defined in Section 5.3) containing the P_i's and $\tilde{P}_j$'s, respectively. This theorem is similar to a special case of the main result in [7]. Destainville [7] introduced a "multi-decomposition" scheme that uses m different partitions of Ω. In Destainville's result, $\| \perp\!\!\!\perp x\|^2 + \| \perp\!\!\!\perp \tilde{x}\|^2$ is replaced by a function of the norm of a "multi-projection" operator Π. Bounding these norms is essential, as the Markov chain $\mathcal{M}$ can require exponential time to mix even if all of the restrictions and complementary restrictions are polynomially mixing[3].

Unfortunately, bounding these norms can be challenging. Destainville [7] bounds the norm of the projection Π by the spectral gap of a smaller matrix $\bar{\Pi}$. In some cases, this gap can be analyzed directly, or even computationally for particular problem instances. However, for very complex distributions such as the distribution over biased permutations we consider here, it can be challenging to find the spectral gap of $\bar{\Pi}$. We believe one of our main contributions is the definition of ϵ-orthogonality, a concrete combinatorial quantity that may be easier to analyze. This approach is particularly useful when the chain decomposes into pieces that are nearly independent, as in the setting of Theorem 1.2.

4.2 Classical decomposition theorems

The disjoint decomposition theorem of [23] states that the spectral gap γ of $\mathcal{M}$ satisfies $\gamma \geq \frac{1}{2}\gamma_{\min}\bar{\gamma}$, where, as we recall from Section 1, $\gamma_{\min} = \min_i \gamma_i$ and $\bar{\gamma}$ is the spectral gap of a projection chain over states $[r]$. Jerrum, Son, Tetali, and Vigoda [18] considered two quantities related to the spectral gap: the Poincaré and log-Sobolev constants. There, the authors defined a parameter $T = \max_i \max_{\sigma\in\Omega_i} \sum_{\tau\in\Omega\setminus\Omega_i} P(\sigma,\tau)$, which can be seen as the maximum probability of escape from one part of the partition in a single step of P, and used it to produce a bound on the order of the minimum gap when T is on the order of $\bar{\gamma}$. They also provided improved bounds when another parameter η is close to zero; this requires a pointwise regularity condition. More recently, Pillai and Smith [27] introduced other conditions in order to directly bound the mixing time by a constant times the maximum of the mixing times of the projection and the restrictions.

The techniques developed for proving the complementary decomposition theorems introduced in this paper can be further applied to prove the following "classical"-style decomposition theorem.

[3]Indeed, the introduction of the projection chain in [21] was a key insight to the original decomposition theorem.

Theorem 4.3. *Let* $\rho = \sqrt{2T/\bar{\gamma}}$. *Then* $\gamma(\mathcal{M}) \geq \min_{p^2+q^2=1} \gamma_{\min} q^2 + \bar{\gamma}\,(q\rho - p)^2$.

We state a more general version of this theorem, Theorem 6.1, in Section 6. This bound allows us to rederive several known classical decomposition theorems.

Corollary 4.4. *Assume* $\mathcal{M}$ *is lazy. Then* $\gamma \geq \gamma_{\min}\bar{\gamma}/3$.

In fact, one can show that the constant is $1/2$ if $\gamma_{\min}, \bar{\gamma} \leq 1/2$ (which is a common situation) or $\delta_2 \geq 1/2$ (δ_2 is defined in Section 5.2). In Corollary 4.5 we show that Theorem 4.3 can be seen as a generalization of Theorem 1 of [18], except that it instead bounds the spectral gap.

Corollary 4.5. $\gamma \geq \min\left\{\frac{\bar{\gamma}}{3}, \frac{\gamma_{\min}\bar{\gamma}}{3T+\bar{\gamma}}\right\}$.

In particular, if $T/\bar{\gamma}$ is a constant, then we get within a constant of the minimum gap as well. Theorem 4.3 produces slightly improved bounds over Corollary 4.5 when $T \approx \bar{\gamma} \ll \gamma_{\min}$.

4.3 Summary of the proofs of the decomposition theorems

Our proofs are elementary and use only basic facts from linear algebra about eigenvalues and eigenvectors. We have chosen to assume the Markov chains are discrete and finite to keep the proofs as accessible as possible. We utilize the following standard characterization of the second largest eigenvalue λ of a symmetric matrix A with top eigenvector v:

$$\lambda = \max_{x \perp v} \frac{\langle x, xA\rangle}{\|x\|^2} = \max_{x \perp v : \|x\|=1} \langle x, xA\rangle\,. \tag{4.1}$$

For a general reversible Markov chain with transition matrix P, we apply equation (4.1) to a symmetric matrix $A = A(P)$ that has the same eigenvalues as P.

We apply the Vector Decomposition Method from the expander graph literature (see, e. g., [31, 34]), and decompose the vector x into $\perp\!\!\!\perp x + x^{\|}$, where $x^{\|}$ is parallel to the top eigenvector of each restriction matrix. The intuition of this method is that if a particular distribution is far from stationarity, then it will either be far from stationarity on some part of the partition or on the projection, and therefore applying P brings us closer to stationarity. The benefit of this approach is that it allows us to quantify the independence of the restriction chains with the projection or complementary restriction chains. Using equation (4.1), for any $x \perp v$, we need to bound

$$\langle x, xA\rangle = \langle \perp\!\!\!\perp x, \perp\!\!\!\perp xA\rangle + \langle x^{\|}, x^{\|}A\rangle + 2\langle \perp\!\!\!\perp x, x^{\|}A\rangle\,. \tag{4.2}$$

It is easy to bound $\langle \perp\!\!\!\perp x, \perp\!\!\!\perp xA\rangle$ and $\langle x^{\|}, x^{\|}A\rangle$ using ideas from other decomposition results [18, 23]. The term $\langle \perp\!\!\!\perp x, x^{\|}A\rangle$ determines whether the decomposed Markov chain is nearly the direct product of two independent Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$, in which case $\langle \perp\!\!\!\perp x, x^{\|}A\rangle \approx 0$ and $\gamma(\mathcal{M}) \approx \min\{\gamma(\mathcal{M}_1), \gamma(\mathcal{M}_2)\}$, or whether they are far from independent, in which case $\langle \perp\!\!\!\perp x, x^{\|}A\rangle$ is large and $\gamma(\mathcal{M}) = \Theta(\gamma_{\min}\bar{\gamma})$. The key to our decomposition proofs lies in our bounds on $\langle \perp\!\!\!\perp x, x^{\|}A\rangle$, which are different for our complementary decomposition theorems than they are for our classical decomposition theorems. More details are provided in Section 5 and Section 6.

5 Complementary decomposition theorems

In this section, we present the proofs of the complementary decomposition theorems. First, in Section 5.1, we introduce some notation and terminology which will be useful for the proofs. Next, we present key lemmas for both the complementary decomposition theorem and for the classical decomposition theorem in Section 5.2. In Section 5.3, we prove our complementary decomposition theorems, Theorems 1.2, 4.1, and 4.2.

5.1 Extended preliminaries

We first fix some notation and terminology. The symbol $\otimes$ is used for tensor product. We write $(v)_i$ to mean the i^{th} coordinate of a vector v. The second largest eigenvalue of P_i will be denoted λ_i, and $\lambda_{\max} = \max_i \lambda_i$. The "top eigenvector" of a matrix will be the eigenvector corresponding to the eigenvalue of largest absolute value.

In order to prove our decomposition results, we wish to apply equation (4.1) to P. However, since P may not be symmetric, we appeal to the following symmetrization technique that appears in [20, p. 153]. Given P with stationary distribution π, define a matrix $A := A(P)$ by $A(\sigma,\tau) := \pi(\sigma)^{1/2}\pi(\tau)^{-1/2}P(\sigma,\tau)$. A is similar to P (i. e., they have the same eigenvalues), but is symmetric, so we can infer a bound on the second eigenvalue of P by applying equation (4.1) to A. It is easy to check that the top eigenvector of A is $\sqrt{\pi}$, which is the vector with entries $\sqrt{\pi(\sigma)}$ for any $\sigma \in \Omega$.

We apply this same symmetrization technique to other matrices as well. For $i \in [r]$ we let $A_i := A(P_i)$ and for $i \in [\tilde{r}]$ we let $\tilde{A}_i := A(\tilde{P}_i)$. We then write $\hat{A}$ to mean the $|\Omega| \times |\Omega|$ matrix with $\hat{A}(\sigma,\tau) = A_i(\sigma,\tau)$ if $\sigma,\tau \in \Omega_i$ for some $i \in [r]$, and zero otherwise. Analogously, we write $\tilde{A}$ to mean the $|\Omega| \times |\Omega|$ matrix with $\tilde{A}(\sigma,\tau) = \tilde{A}_i(\sigma,\tau)$ if $\sigma,\tau \in \tilde{\Omega}_i$ for some $i \in [\tilde{r}]$, and zero otherwise. It is important to note that $\hat{A} \neq A(\hat{P})$ and $\tilde{A} \neq A(\tilde{P})$.

Proposition 5.1. *The matrix A satisfies $A = \hat{A} + \tilde{A} - I_{|\Omega|}$.*

Let $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{|\Omega|}$ be the eigenvalues of $\tilde{A}$ with corresponding eigenvectors $v_1, v_2, \ldots, v_{|\Omega|}$. As $\tilde{A}$ is symmetric, the real spectral theorem tells us that its eigenvectors form an orthonormal basis of $\mathbb{R}^{|\Omega|}$. We consider the basis representations $\perp\!\!\!\perp x = \sum_i \perp\!\!\!\perp a_i v_i$ and $x^{\|} = \sum_i a_i^{\|} v_i$. More generally, for any $v \in \mathbb{R}^{|\Omega|}$, we write $v = \sum_i a_i v_i$ for some constants $a_1, a_2, \ldots, a_{|\Omega|} \in \mathbb{R}$. Also, $\|v\|^2 = \sum_i a_i^2 \|v_i\|^2$, and $v\tilde{A} = \sum_i a_i \mu_i v_i$.

5.2 Key ideas and lemmas for the proofs

We wish to apply equation (4.1) to A. Recall that $\sqrt{\pi}$ is the top eigenvector of A. Let $x \in \mathbb{R}^{|\Omega|}$ with $x \perp \sqrt{\pi}$ and $\|x\| = 1$. We will decompose x into two vectors $x^{\|}$ and $\perp\!\!\!\perp x$ as follows (note: this is similar to the vector decomposition used for the Zig Zag Product in [31]). For any $i \in [r]$, let $x_i \in \mathbb{R}^{|\Omega_i|}$ be the vector defined by $x_i(\sigma) = x(\sigma)$ for all $\sigma \in \Omega_i$. Then $x = \sum_i e_i \otimes x_i$. We further decompose x_i into $x_i^{\|}$, the part that is parallel to $\sqrt{\pi_i}$, and $\perp\!\!\!\perp x_i$, the part that is perpendicular to

$\sqrt{\pi_i}$; recall that $\sqrt{\pi_i}$ is the top eigenvector of A_i. Finally, define $x^{\parallel}, \perp\!\!\perp x \in \mathbb{R}^{|\Omega|}$ by $x^{\parallel} = \sum_i e_i \otimes x_i^{\parallel}$ and $\perp\!\!\perp x = \sum_i e_i \otimes \perp\!\!\perp x_i$. Hence $x = \sum_i e_i \otimes x_i = x^{\parallel} + \perp\!\!\perp x$. Define $\tilde{x}^{\parallel}$ and $\perp\!\!\perp \tilde{x}$ analogously.

As described in Section 4.3, we will bound $\langle x, xA\rangle$ via equation (4.2):

$$\langle x, xA\rangle = \langle \perp\!\!\perp x, \perp\!\!\perp xA\rangle + \langle x^{\parallel}, x^{\parallel}A\rangle + 2\langle \perp\!\!\perp x, x^{\parallel}A\rangle\,.$$

We need the following simple proposition.

Lemma 5.2. *The following holds:* $x^{\parallel}A = x^{\parallel}\tilde{A}$.

Applying Lemma 5.2, equation (4.2) becomes

$$\langle x, xA\rangle = \langle x^{\parallel}, x^{\parallel}\tilde{A}\rangle + 2\langle \perp\!\!\perp x, x^{\parallel}\tilde{A}\rangle + \langle \perp\!\!\perp x, \perp\!\!\perp x(\hat{A} + \tilde{A} - I_{|\Omega|})\rangle\,.$$

For ease of notation, we define the following quantities:

$$\delta_1 = \frac{\langle \perp\!\!\perp x, \perp\!\!\perp x\hat{A}\rangle}{\|\perp\!\!\perp x\|^2},\ \delta_2 = \frac{\langle \perp\!\!\perp x, \perp\!\!\perp x\tilde{A}\rangle}{\|\perp\!\!\perp x\|^2},\ \delta_3 = \frac{\langle x^{\parallel}, x^{\parallel}\hat{A}\rangle}{\|x^{\parallel}\|^2},\ \delta_4 = \frac{\langle x^{\parallel}, x^{\parallel}\tilde{A}\rangle}{\|x^{\parallel}\|^2}.$$

Plugging these in, we have

$$\langle x, xA\rangle = \delta_4\|x^{\parallel}\|^2 + 2\langle \perp\!\!\perp x, x^{\parallel}\tilde{A}\rangle + (\delta_1 + \delta_2 - 1)\|\perp\!\!\perp x\|^2\,. \tag{5.1}$$

Bounding δ_1 and δ_4 is straightforward, and borrows many of the ideas from classical decomposition results. If $x^{\parallel}\tilde{A}$ were orthogonal to $\perp\!\!\perp x$, then doing so would be sufficient to proving a strong decomposition theorem. However, this is not true in general, so we must also bound $\langle \perp\!\!\perp x, x^{\parallel}\tilde{A}\rangle$. Our two types of theorems do so in different ways, which are presented in Section 5.3 and Section 6.

The next lemma makes concrete the intuition that if a particular distribution is far from stationarity, then it will either be far from stationarity on some restriction—in which case $\hat{A}$ will bring it closer to stationarity (as in part 1)—or on the projection—in which case $\tilde{A}$ will bring it closer to stationarity (as in part 2). The proof is straightforward from the definitions.

Lemma 5.3. *With the above notation,*

1. $\delta_1 \le \lambda_{\max}$.

2. $\delta_4 \le \bar{\lambda}$.

Note that λ is formally defined in equation (4.1) and $\lambda_{\max}$ is defined at the beginning of Section 5.1.

5.3 Proofs of complementary decomposition theorems

Next we use the technology developed in Section 5.2 to prove Theorem 4.2. Recall that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{|\Omega|}$ are the eigenvalues of $\tilde{A}$ with corresponding eigenvectors $v_1, v_2, \ldots, v_{|\Omega|}$, and that for any $v \in \mathbb{R}^{|\Omega|}$, we write $v = \sum_i a_i v_i$ for some constants $a_1, a_2, \ldots, a_{|\Omega|} \in \mathbb{R}$. Also, $\|v\|^2 = \sum_i a_i^2 \|v_i\|^2$, and $v\tilde{A} = \sum_i a_i \mu_i v_i$. Define the set $S = \{i : \mu_i = 1\}$. Let $\tilde{\delta}_1 = \frac{\langle \perp\!\!\!\perp \tilde{x}, \perp\!\!\!\perp \tilde{x}\tilde{A}\rangle}{\|\perp\!\!\!\perp \tilde{x}\|^2}$. Now we can make an explicit statement about the gap of $\mathcal{M}$; notice the equality in equation (5.2).

Theorem 4.2.

$$\gamma(\mathcal{M}) = \min_{x \perp \sqrt{\pi}, \|x\|=1} (1-\delta_1)\| \perp\!\!\!\perp x\|^2 + (1-\tilde{\delta}_1)\| \perp\!\!\!\perp \tilde{x}\|^2 . \tag{5.2}$$

In particular,

$$\gamma(\mathcal{M}) \geq \min_{x \perp \sqrt{\pi}, \|x\|=1} \gamma_{\min}\| \perp\!\!\!\perp x\|^2 + \tilde{\gamma}_{\min}\| \perp\!\!\!\perp \tilde{x}\|^2 . \tag{5.3}$$

Proof. Notice $\tilde{\delta}_1 \| \perp\!\!\!\perp \tilde{x}\|^2 = \sum_{i \in S} \mu_i (\perp\!\!\!\perp a_i + a_i^{\|})^2$. Thus,

$$(1-\tilde{\delta}_1)\| \perp\!\!\!\perp \tilde{x}\|^2 = \sum_i (1-\mu_i)(\perp\!\!\!\perp a_i + a_i^{\|})^2 = (1-\delta_2)\| \perp\!\!\!\perp x\|^2 + (1-\delta_4)\|x^{\|}\|^2 - 2\langle \perp\!\!\!\perp x, x^{\|}\tilde{A}\rangle .$$

On the other hand, from equation (6.1), we have

$$1 - \langle x, xA\rangle = (1-\delta_1)\| \perp\!\!\!\perp x\|^2 + (1-\delta_2)\| \perp\!\!\!\perp x\|^2 + (1-\delta_4)\|x^{\|}\|^2 - 2\langle \perp\!\!\!\perp x, x^{\|}\tilde{A}\rangle .$$

Thus, for all $x \perp \sqrt{\pi}$ with norm 1, we have

$$1 - \langle x, xA\rangle = (1-\delta_1)\| \perp\!\!\!\perp x\|^2 + (1-\tilde{\delta}_1)\| \perp\!\!\!\perp \tilde{x}\|^2 .$$

Applying equation (4.1), we get equation (5.2). To get equation (5.3), we apply Lemma 5.3, which yields $1 - \delta_1 \geq 1 - \lambda_{\max} = \gamma_{\min}$. An analogous statement to Lemma 5.3 holds for $\tilde{\delta}_1$, and shows $1 - \tilde{\delta}_1 \geq \tilde{\gamma}_{\min}$. ◻

It remains to prove Theorem 4.1. By Theorem 4.2, if $\gamma_{\min}$ and $\tilde{\gamma}_{\min}$ are not too small, it suffices to show that $\| \perp\!\!\!\perp x\|^2$ and $\| \perp\!\!\!\perp \tilde{x}\|^2$ cannot both be small. To this end, we further decompose $\perp\!\!\!\perp x$ and $\tilde{x}^{\|}$ based on the eigenvectors of $\tilde{A}$. Define $S = \{i : \mu_i = 1\}$ and vectors $x_{11} = \sum_{i \in S} a_i^{\|} v_i$ and $x_{12} = \sum_{i \notin S} a_i^{\|} v_i$. Similarly, let $x_{21} = \sum_{i \in S} \perp\!\!\!\perp a_i v_i$ and $x_{22} = \sum_{i \notin S} \perp\!\!\!\perp a_i v_i$. Notice $\tilde{x}^{\|} = x_{11} + x_{21}$ and $\perp\!\!\!\perp \tilde{x} = x_{12} + x_{22}$, so that the vectors in each row (respectively, column) of the following table sum to the vector in its row (respectively, column) label.

	$\tilde{x}^{\|}$	$\perp\!\!\!\perp \tilde{x}$
$x^{\|}$	x_{11}	x_{12}
$\perp\!\!\!\perp x$	x_{21}	x_{22}

The vectors within each row are orthogonal, as they are in the span of eigenvectors with distinct eigenvalues. However, the vectors within each column are not necessarily orthogonal.

The idea of the proof of Theorem 4.1 is that if $\|x_{11}\|$ is small, then $\| \perp\!\!\!\perp x\|^2 + \| \perp\!\!\!\perp \tilde{x}\|^2$ is large. The following lemma states that ϵ-orthogonality is sufficient to guarantee $\|x_{11}\|$ is small.

Lemma 5.4. *Let ϵ be as defined in equation (1.1). Then $\|x_{11}\|^2 \leq \epsilon$.*

Proof. Recall x_{11} is the projection of $x^{\|}$ onto the top eigenvectors of $\tilde{A}$. The top eigenvectors of $\tilde{A}$ are precisely the set of all $\sqrt{\tilde{\pi}_j}$ for $j \in [\tilde{r}]$. Therefore,

$$x_{11} = \sum_j \frac{\langle x^{\|}, \sqrt{\tilde{\pi}_j}\rangle}{\|\sqrt{\tilde{\pi}_j}\|^2}\sqrt{\tilde{\pi}_j}\,.$$

As the eigenvectors of $\tilde{A}$ are an orthonormal basis, we have

$$\|x_{11}\|^2 = \sum_j \langle x^{\|}, \sqrt{\tilde{\pi}_j}\rangle^2\,.$$

For any $j \neq j' \in [\tilde{r}]$ and any $\sigma \in \tilde{\Omega}_{j'}$, $\sqrt{\tilde{\pi}_j(\sigma)} = 0$ and for $i \in [r]$, $\tilde{\pi}_j(\Omega_i \cap \tilde{\Omega}_j) = \pi(\Omega_i \cap \tilde{\Omega}_j)/\pi(\tilde{\Omega}_j)$. Therefore,

$$\langle x^{\|}, \sqrt{\tilde{\pi}_j}\rangle = \sum_i \sum_{\sigma\in\Omega_i} x^{\|}(\sigma)\sqrt{\tilde{\pi}_j(\sigma)} = \sum_i \alpha_i \sum_{\sigma\in\Omega_i\cap\tilde{\Omega}_j} \sqrt{\pi_i(\sigma)\tilde{\pi}_j(\sigma)} = \sum_{i\in[r]} \alpha_i \frac{\pi(\Omega_i \cap \tilde{\Omega}_j)}{\sqrt{\pi(\Omega_i)\pi(\tilde{\Omega}_j)}}\,. \tag{5.4}$$

Since $x \perp \sqrt{\pi}$ and $\amalg\ x \perp \sqrt{\pi}$ by definition, it follows that $x^{\|} \perp \sqrt{\pi}$ as well. This implies that $\alpha \perp \sqrt{\overline{\pi}}$, as

$$0 = \langle x^{\|}, \sqrt{\pi}\rangle = \sum_i \alpha_i \sum_{\sigma\in\Omega_i} \sqrt{\pi_i(\sigma)\pi(\sigma)} = \sum_i \alpha_i \sum_{\sigma\in\Omega_i} \frac{\sqrt{\pi(\sigma)}}{\sqrt{\pi(\Omega_i)}}\sqrt{\pi(\sigma)} = \sum_i \alpha_i\sqrt{\pi(\Omega_i)}\,, \tag{5.5}$$

and this final term is equal to $\sum_i \alpha_i\sqrt{\overline{\pi}_i} = \langle \alpha, \sqrt{\overline{\pi}}\rangle$. Multiplying equation (5.5) by $\pi(\tilde{\Omega}_j)$ and subtracting it from equation (5.4), we have

$$\langle x^{\|}, \sqrt{\tilde{\pi}_j}\rangle = \sum_{i\in[r]} \alpha_i \left(\frac{\pi(\Omega_i \cap \tilde{\Omega}_j)}{\sqrt{\pi(\Omega_i)\pi(\tilde{\Omega}_j)}} - \sqrt{\pi(\Omega_i)\pi(\tilde{\Omega}_j)}\right) = \langle \alpha, V_j\rangle,$$

where

$$V_j(i) := \left(\frac{\pi(\Omega_i \cap \tilde{\Omega}_j)}{\sqrt{\pi(\Omega_i)\pi(\tilde{\Omega}_j)}} - \sqrt{\pi(\Omega_i)\pi(\tilde{\Omega}_j)}\right) = \sqrt{\pi(\Omega_i \cap \tilde{\Omega}_j)}\left(\sqrt{r(i,j)} - 1/\sqrt{r(i,j)}\right).$$

By the Cauchy–Schwarz inequality and the fact that $\|\alpha\| = \|x^{\|}\| \leq \|x\| = 1$, we have $\langle \alpha, V_j\rangle \leq \|\alpha\|\|V_j\| = \|V_j\|$. Therefore we get,

$$\|x_{11}\|^2 = \sum_j \langle x^{\|}, \sqrt{\tilde{\pi}_j}\rangle^2 \leq \sum_j \|V_j\|^2 = \sum_{i,j} \pi(\Omega_i \cap \tilde{\Omega}_j)\left(\sqrt{r(i,j)} - 1/\sqrt{r(i,j)}\right)^2\,. \qquad \square$$

To prove Theorem 4.1 from Theorem 4.2, we must show that if $\|x_{11}\|^2 \leq \epsilon$, then $\| \perp\!\!\!\perp x\|^2 + \| \perp\!\!\!\perp \tilde{x}\|^2 \geq (1-\sqrt{\epsilon})^2$. As the sum of the squared norms of the vectors in the above table is 1, it is reasonable to expect that if $\|x_{11}\|^2$ is small, then $\| \perp\!\!\!\perp x\|^2 + \| \perp\!\!\!\perp \tilde{x}\|^2$ is large. However, this is not as straightforward as one might expect, as the vectors within each column are not necessarily orthogonal, so we may have $\| \perp\!\!\!\perp \tilde{x}\|^2 < \|x_{12}\|^2 + \|x_{22}\|^2$.

Proof of Theorem 4.1. By equation (5.3) from Theorem 4.2, it suffices to show that $\| \perp\!\!\!\perp x\|^2 + \| \perp\!\!\!\perp \tilde{x}\|^2 \geq (1-\sqrt{\epsilon})^2$. If $\| \perp\!\!\!\perp x\|^2 \geq (1-\sqrt{\epsilon})^2$ we are done, so we may assume not. As the vectors within each row of the table are orthogonal, we have $(1-\sqrt{\epsilon})^2 > \| \perp\!\!\!\perp x\|^2 = \|x_{22}\|^2 + \|x_{21}\|^2$. Furthermore, since

$$0 = \langle \perp\!\!\!\perp x, x^{\|} \rangle = \sum_i \perp\!\!\!\perp a_i a_i^{\|} \|v_i\|^2$$

we have

$$\langle x_{22}, x_{12} \rangle = \sum_{i \notin S} \perp\!\!\!\perp a_i a_i^{\|} \|v_i\|^2 = -\sum_{i \in S} \perp\!\!\!\perp a_i a_i^{\|} \|v_i\|^2 = -\langle x_{21}, x_{11} \rangle \,.$$

Thus,

$$\begin{aligned} \| \perp\!\!\!\perp \tilde{x}\|^2 &= \|x_{22} + x_{12}\|^2 \\ &= \|x_{22}\|^2 + \|x_{12}\|^2 + 2\langle x_{22}, x_{12} \rangle \\ &= \|x_{22}\|^2 + \|x_{12}\|^2 - 2\langle x_{21}, x_{11} \rangle \\ &\geq \|x_{22}\|^2 + \|x_{12}\|^2 - 2\|x_{21}\|\|x_{11}\| \,. \end{aligned} \tag{5.6}$$

We used Cauchy–Schwarz for the final inequality. Putting everything together,

$$\begin{aligned} \| \perp\!\!\!\perp x\|^2 + \| \perp\!\!\!\perp \tilde{x}\|^2 &\geq \|x_{22}\|^2 + \|x_{21}\|^2 + \|x_{22}\|^2 + \|x_{12}\|^2 - 2\|x_{21}\|\|x_{11}\| \\ &= \|x_{22}\|^2 + \|x_{21}\|^2 + (1 - \|x_{21}\|^2 - \|x_{11}\|^2) - 2\|x_{21}\|\|x_{11}\| \\ &= 1 + \|x_{22}\|^2 - \|x_{11}\|^2 - 2\|x_{21}\|\|x_{11}\| \\ &\geq 1 + \|x_{22}\|^2 - \epsilon - 2\sqrt{\epsilon}\sqrt{(1-\sqrt{\epsilon})^2 - \|x_{22}\|^2} \\ &\geq 1 - \epsilon - 2\sqrt{\epsilon}(1-\sqrt{\epsilon}) \\ &= (1-\sqrt{\epsilon})^2. \end{aligned}$$

□

6 Classical decomposition theorems

In this section, we will prove our classical decomposition theorem, Theorem 4.3. It will be a corollary of the following theorem.

Theorem 6.1. *Let $\rho = \sqrt{(1-\delta_2)/\bar{\gamma}}$. Then*

$$\gamma(\mathcal{M}) \geq \min_{p^2+q^2=1} \gamma_{\min} q^2 + \left(q\sqrt{1-\delta_2} - p\sqrt{\bar{\gamma}}\right)^2 .$$

With the technology developed in Section 5.2, there is one critical piece remaining to prove Theorem 6.1, which is to bound the cross terms generated by applying the matrix $\tilde{A}$ to $x^{\parallel}$.

Lemma 6.2. *With the above notation,* $|\langle \perp\!\!\!\perp x, x^{\parallel}\tilde{A}\rangle| \leq \sqrt{(1-\delta_4)(1-\delta_2)}\|x^{\parallel}\|\|\perp\!\!\!\perp x\|$.

Proof. Recall that $\{\mu_i\}$ are the eigenvalues of $\tilde{A}$ with corresponding eigenvectors $\{v_i\}$, and $\perp\!\!\!\perp x = \sum_i \perp\!\!\!\perp a_i v_i$ and $x^{\parallel} = \sum_i a_i^{\parallel} v_i$ are the basis representations of $\perp\!\!\!\perp x$ and $x^{\parallel}$.

Since $\perp\!\!\!\perp x$ is perpendicular to $x^{\parallel}$, we have $0 = \langle \perp\!\!\!\perp x, x^{\parallel}\rangle = \sum_i \perp\!\!\!\perp a_i a_i^{\parallel}\|v_i\|^2$. Notice

$$\begin{aligned}\langle \perp\!\!\!\perp x, x^{\parallel}\tilde{A}\rangle &= \sum_i \mu_i \perp\!\!\!\perp a_i a_i^{\parallel}\|v_i\|^2 \\ &= \sum_i \perp\!\!\!\perp a_i a_i^{\parallel}\|v_i\|^2 - \sum_i (1-\mu_i)\perp\!\!\!\perp a_i a_i^{\parallel}\|v_i\|^2 \\ &= 0 - \sum_i (1-\mu_i)\perp\!\!\!\perp a_i a_i^{\parallel}\|v_i\|^2\,.\end{aligned}$$

Define vectors $z^{\parallel} := \sum_i \sqrt{1-\mu_i}a_i^{\parallel}v_i$ and $\perp\!\!\!\perp z := \sum_i \sqrt{1-\mu_i} \perp\!\!\!\perp a_i v_i$. Then because the v_i are mutually orthogonal, we have

$$\langle \perp\!\!\!\perp z, z^{\parallel}\rangle = \sum_i (1-\mu_i)\perp\!\!\!\perp a_i a_i^{\parallel}\|v_i\|^2 = -\langle \perp\!\!\!\perp x, x^{\parallel}A\rangle\,.$$

By Cauchy–Schwarz, $|\langle \perp\!\!\!\perp z, z^{\parallel}\rangle| \leq \|\perp\!\!\!\perp z\|\|z^{\parallel}\|$. Moreover,

$$\|z^{\parallel}\|^2 = \sum_i (1-\mu_i)(a_i^{\parallel})^2\|v_i\|^2 = \|x^{\parallel}\|^2 - \sum_i \mu_i (a_i^{\parallel})^2\|v_i\|^2 = \|x^{\parallel}\|^2 - \langle x^{\parallel}, x^{\parallel}\tilde{A}\rangle = (1-\delta_4)\|x^{\parallel}\|^2.$$

Similarly, $\|\perp\!\!\!\perp z\|^2 = (1-\delta_2)\|\perp\!\!\!\perp x\|^2$. Taken together, this proves the lemma. □

Finally, we are ready to prove Theorem 6.1.

Proof of Theorem 6.1. Let $x \in \mathbb{R}^{|\Omega|}$ with $x \perp \sqrt{\pi}$ and $\|x\| = 1$. By equation (4.1), $\gamma(\mathcal{M}) \geq 1 - \langle x, xA\rangle$. Applying Lemma 5.2 and the definitions of δ_1, δ_4, and δ_2, equation (4.2) becomes

$$\begin{aligned}\langle x, xA\rangle &= \langle x^{\parallel}, x^{\parallel}\tilde{A}\rangle + 2\langle \perp\!\!\!\perp x, x^{\parallel}\tilde{A}\rangle + \langle \perp\!\!\!\perp x, \perp\!\!\!\perp x(\hat{A} + \tilde{A} - I_{|\Omega|})\rangle \\ &= \delta_4\|x^{\parallel}\|^2 + 2\langle \perp\!\!\!\perp x, x^{\parallel}\tilde{A}\rangle + (\delta_1+\delta_2-1)\|\perp\!\!\!\perp x\|^2\,. \end{aligned} \tag{6.1}$$

Applying Lemma 6.2, we have

$$\gamma(\mathcal{M}) \geq 1 - \left(\delta_4\|x^{\parallel}\|^2 + 2\sqrt{(1-\delta_4)(1-\delta_2)}\|x^{\parallel}\|\|\perp\!\!\!\perp x\| + (\delta_1+\delta_2-1)\|\perp\!\!\!\perp x\|^2\right).$$

Rearranging terms and using the fact that $1 = \|x\|^2 = \|\perp\!\!\!\perp x\|^2 + \|x^{\parallel}\|^2$, we have

$$\gamma(\mathcal{M}) \geq \min_{x\perp\sqrt{\pi}:\|x\|=1} (1-\delta_1)\|\perp\!\!\!\perp x\|^2 + \left(\sqrt{1-\delta_2}\|\perp\!\!\!\perp x\| - \sqrt{1-\delta_4}\|x^{\parallel}\|\right)^2\,. \tag{6.2}$$

By setting $q = \|\perp\!\!\!\perp x\|$ and $p = \|x^{\|}\|$, we immediately get

$$\gamma(\mathcal{M}) \geq \min_{p^2+q^2=1} (1-\delta_1)q^2 + \left(\sqrt{1-\delta_2}q - \sqrt{1-\delta_4}p\right)^2 .$$

By Lemma 5.3 and the definition of the spectral gap, $(1-\delta_1) \geq \gamma_{\min}$ and $(1-\delta_4) \geq \bar{\gamma}$. Thus we have, for any values of q and p,

$$(1-\delta_1)q^2 + \left(\sqrt{1-\delta_2}q - \sqrt{1-\delta_4}p\right)^2 \geq \gamma_{\min}q^2 + \left(\sqrt{1-\delta_2}q - \sqrt{1-\delta_4}p\right)^2 .$$

Define $f(q,p) = aq^2 + (bq-cp)^2$, where $a,b,c \geq 0$. Assume $c \geq c_* \geq 0$, and define $f' = aq^2 + (bq - c_*p)^2$. We want to show that

$$\min_{q^2+p^2=1} f \geq \min_{q^2+p^2=1} f' . \tag{6.3}$$

If $c_* = 0$ then f' is minimized at the point $(0,1)$, and $f'(0,1) = 0 \leq \min_{q^2+p^2=1} f$. Thus, we may assume $c \geq c_* > 0$. If $a = 0$ then $\min_{q^2+p^2=1} f = \min_{q^2+p^2=1}(bq-cp)^2 = 0$, since we can choose $p/q = b/c$. Similarly, in this case, $\min_{q^2+p^2=1} f' = 0$. Hence we may assume $a > 0$.

The minimum of the function f with respect to q and $p = \sqrt{1-q^2}$ occurs either at the endpoints $q = 0$ or $q = 1$, or at the critical point (q_*, p_*) satisfying $\frac{\partial f}{\partial q}(q_*, p_*) = 0$. Note that $f(0,1) = c^2 \geq c_*^2 = f'(0,1) \geq \min_{q^2+p^2=1} f'$, and $f(1,0) = a + b^2 = f'(1,0) \geq \min_{q^2+p^2=1} f'$. Now we may assume $q, p > 0$.

Now,

$$\frac{\partial f}{\partial q} = 2aq + 2(bq - cp)\left(b + \frac{cq}{p}\right) .$$

Since $c, q, p > 0$ and $b \geq 0$, we have $b + \frac{cq}{p} > 0$. Thus, the point (q_*, p_*) satisfies

$$bq_* - cp_* = \frac{-aq_*}{b + \frac{cq_*}{p_*}} < 0 ,$$

since $a, q_* > 0$.

We will now take the derivative of f with respect to c:

$$\frac{\partial f}{\partial c} = 2(cp - bq)p .$$

Thus, $\frac{\partial f}{\partial c}(q_*, p_*) = 2(cp_* - bq_*)p_* > 0$. This implies f is increasing with c in the neighborhood around the point (q_*, p_*), and so by decreasing c by ϵ we will lower the value of $\min_{q^2+p^2=1} f$; this is true for every $c > 0$, so we have $\min_{q^2+p^2=1} f \geq \min_{q^2+p^2=1} f'$. This proves Equation 6.3.

Therefore, we have

$$\gamma(\mathcal{M}) \geq \min_{p^2+q^2=1} \gamma_{\min}q^2 + \left(q\sqrt{1-\delta_2} - p\sqrt{\bar{\gamma}}\right)^2 . \qquad \square$$

The statement of Theorem 6.1 is admittedly technical. However, from it we may derive several corollaries, as listed in Section 4.

To see that Theorem 4.3 follows from Theorem 6.1, we prove that $\delta_2 \geq 1 - 2T$.

Proof of Theorem 4.3. Recall $T = \max_i \max_{\sigma \in \Omega_i} \sum_{\tau \in \Omega \setminus \Omega_i} P(\sigma, \tau)$. This is the parameter γ from [18]. We will now show that $\delta_2 \geq 1 - 2T$. Notice the probability of a move in $\hat{A}$ is at least $1 - T$, so every element has a self-loop probability of at least $1 - T$ in $\tilde{A}$. Thus, $\tilde{A}$ can be written as $\tilde{A} = T\tilde{A}' + I(1 - T)$ for some transition matrix $\tilde{A}'$ with minimum eigenvalue -1. This implies that the minimum eigenvalue of $\tilde{A}$ satisfies $\mu_{\min} \geq 1 - 2T$. On the other hand, $\delta_2 \geq \mu_{\min}$. □

Next we prove Corollary 4.4, which states that $\gamma \geq \gamma_{\min}\bar{\gamma}/3$. In fact, one can show that the constant is $1/2$ if $\gamma_{\min}, \bar{\gamma} \leq 1/2$ or $1 - \delta_2 \leq 1/2$.

Proof of Corollary 4.4. Since the bound in Theorem 4.3 is minimized when δ_2 is minimized, we may assume $\delta_2 = 0$. Thus, $\rho = 1/\sqrt{\bar{\gamma}}$. We will show that for all p, q satisfying $p^2 + q^2 = 1$, we have

$$\gamma_{\min} q^2 + \bar{\gamma}\left(p - \frac{q}{\sqrt{\bar{\gamma}}}\right)^2 \geq \frac{\gamma_{\min}\bar{\gamma}}{3}\,.$$

Clearly if $q^2 \geq \bar{\gamma}/3$, then we are done. So we may assume $q^2/\bar{\gamma} < 1/3$. Notice that since $\bar{\gamma} \leq 1$,

$$\left(p - \frac{q}{\sqrt{\bar{\gamma}}}\right)^2 \;=\; \left(\sqrt{1 - q^2} - \frac{q}{\sqrt{\bar{\gamma}}}\right)^2 \;\geq\; \left(\sqrt{1 - \frac{q^2}{\bar{\gamma}}} - \frac{q}{\sqrt{\bar{\gamma}}}\right)^2 .$$

As $q^2/\bar{\gamma} < 1/3$, we have $\left(\sqrt{1 - \frac{q^2}{\bar{\gamma}}} - \frac{q}{\sqrt{\bar{\gamma}}}\right)^2 \geq \frac{1}{3} - \frac{q^2}{\bar{\gamma}}$. Therefore,

$$\begin{aligned}
\gamma_{\min} q^2 + \bar{\gamma}\left(p - \frac{q}{\sqrt{\bar{\gamma}}}\right)^2 &\geq \gamma_{\min} q^2 + \frac{\bar{\gamma}}{3} - q^2 \\
&= \frac{\bar{\gamma}}{3} - q^2(1 - \gamma_{\min}) \\
&> \frac{\bar{\gamma}}{3}(1 - (1 - \gamma_{\min})) \\
&= \gamma_{\min}\bar{\gamma}/3\,.
\end{aligned}$$
□

Next, we will show that a variant of Theorem 1 of [18] follows from Theorem 4.3 from this paper, which is the content of Corollary 4.5. Theorem 4.3 produces slightly improved bounds over Corollary 4.5 when $T \approx \bar{\gamma} \ll \gamma_{\min}$.

Proof of Corollary 4.5. We wish to show $\gamma \geq \min\left\{\frac{\bar{\gamma}}{3}, \frac{\gamma_{\min}\bar{\gamma}}{3T + \bar{\gamma}}\right\}$. As $\rho^2 \leq 2T/\bar{\gamma}$, it suffices to show that for all $p^2 + q^2 = 1$,

$$\gamma_{\min} q^2 + \bar{\gamma}(p - \rho q)^2 \geq \min\left\{\frac{\bar{\gamma}}{3}, \frac{\gamma_{\min}}{1 + 1.5\rho^2}\right\}.$$

If $q^2 \geq \frac{1}{1+1.5\rho^2}$ then we are done, so we may assume $\frac{1}{1+1.5\rho^2} - q^2 \geq 0$. Define

$$f_1 = \frac{(p-\rho q)^2}{1/(1+1.5\rho^2) - q^2} \quad \text{and} \quad f_2 = \frac{1/3 - (p-\rho q)^2}{q^2} \,.$$

Notice that $\gamma_{\min} q^2 + \bar{\gamma}(p-\rho q)^2 \geq \frac{\gamma_{\min}}{1+1.5\rho^2}$ if and only if $\gamma_{\min}/\bar{\gamma} \leq f_1$. On the other hand, $\gamma_{\min} q^2 + \bar{\gamma}(p-\rho q)^2 \geq \bar{\gamma}/3$ if and only if $\gamma_{\min}/\bar{\gamma} \geq f_2$. Thus, it suffices to show $f_1 \geq f_2$ for all parameter choices. First, notice that since $\frac{1}{1+1.5\rho^2} - q^2 \geq 0$, we have $f_1 \geq f_2$ whenever

$$1 \geq \left(\frac{1}{3(p-\rho q)^2} - 1\right)\left(\frac{1}{(1+1.5\rho^2)q^2} - 1\right) .$$

This is satisfied whenever $3(p-\rho q)^2 \geq 1 - (1+1.5\rho^2)q^2$. Expanding and bringing all to the left hand side shows this is true because $(2p - 3\rho q)^2 \geq 0$. □

While we do not currently have a comparison between our Theorem 4.3 and Corollary 2 of [18], which requires a pointwise bound on π_i^j, we believe that Corollary 2 of [18] is insufficient for our application to permutations. We expound upon this in Remark 7.5.

7 Application to permutations

In this section, we apply the complementary decomposition theorem, Theorem 1.2, to the problem of sampling biased permutations. We give the proofs of Theorem 1.3 and Theorem 1.4 which bound the spectral gap of the nearest-neighbor Markov chain $\mathcal{M}_{\text{nn}}$ and the k-particle process Markov chain $\mathcal{M}_{\text{pp}}$, respectively. The section is laid out as follows. In Section 7.1, we describe the state spaces and stationary distributions of $\mathcal{M}_{\text{nn}}$ and $\mathcal{M}_{\text{pp}}$, and formally give the definitions of the weakly monotonic property for each chain. Next, in Section 7.2 we give the proof of Theorem 1.3, and then in Section 7.3 we complete the proof by proving that our decomposition is ϵ-orthogonal, where $\epsilon \leq 1/n^2$ (Lemma 7.6). Finally, in Section 7.4 we use comparison techniques [9, 30] to bound the spectral gap of $\mathcal{M}_{\text{nn}}$ (Theorem 1.4) using the bound on the spectral gap of $\mathcal{M}_{\text{pp}}$ (Theorem 1.3).

7.1 Biased permutations and k-particle processes

We begin by giving the formal definition for the nearest neighbor Markov chain $\mathcal{M}_{\text{nn}}$ over biased permutations. Here we are interested in the problem of sampling from the symmetric group S_n, the permutations of $[n]$, where we interpret the elements of S_n as strings of length n with each symbol in $[n]$ appearing exactly once. We are given as input a set of probabilities $\mathcal{P} = \{p_{i,j}\}$ for all $1 \leq i, j \leq n$ with $p_{i,j} = 1 - p_{j,i}$. Let $\sigma(i)$ denote the element in position i of $\sigma \in S_n$.

The Nearest Neighbor Markov chain $\mathcal{M}_{\text{nn}}$

`Starting at any permutation` σ^0`, iterate the following:`

- At time $t \geq 0$, choose a position $1 < i \leq n$ uniformly at random in permutation σ^t.
- With probability $p_{\sigma^t(i),\sigma^t(i-1)}/2$, exchange the elements $\sigma^t(i)$ and $\sigma^t(i-1)$ to obtain σ^{t+1}.
- Otherwise, do nothing so that $\sigma^{t+1} = \sigma^t$.

The chain $\mathcal{M}_{\text{nn}}$ connects the state space S_n and has the following stationary distribution (see, e. g., [4]):

$$\pi_{\text{nn}}(\sigma) = \prod_{i<j:\sigma(i)>\sigma(j)} \frac{p_{\sigma(i),\sigma(j)}}{p_{\sigma(j),\sigma(i)}} Z_{\text{nn}}^{-1} = \prod_{i<j:\sigma(i)>\sigma(j)} q_{\sigma(i),\sigma(j)} Z_{\text{nn}}^{-1}$$

where Z_{nn} is a normalizing constant and $q_{\sigma(i),\sigma(j)} = \frac{p_{\sigma(i),\sigma(j)}}{p_{\sigma(j),\sigma(i)}}$.

Our central question is under what conditions does $\mathcal{M}_{\text{nn}}$ mix in polynomial time. Fill [14] introduced monotonicity conditions under which he conjectured $\mathcal{M}_{\text{nn}}$ would be rapidly mixing. In our analysis of $\mathcal{M}_{\text{nn}}$ we use the weakly monotonic condition that appears in [4, 25]:

Definition 7.1 ([4])**.** The set $\mathcal{P}$ is weakly monotonic if properties 1 and either 2 or 3 are satisfied.

1. $p_{i,j} \geq 1/2$ for all $1 \leq i < j \leq n$, and
2. $p_{i,j+1} \geq p_{i,j}$ for all $1 \leq i < j \leq n-1$ or
3. $p_{i-1,j} \geq p_{i,j}$ for all $2 \leq i < j \leq n$.

We consider the special case of k-classes where $[n]$ is partitioned into k classes $C_1, C_2, \ldots, C_k$, and assume elements in class C_i interact with elements in class C_j with the same probability. That is, if $i_1, i_2 \in C_i$ and $j_1, j_2 \in C_j$ then $p_{i_1,j_1} = p_{i_2,j_2}$. In this case we define $\overline{p}_{i,j}$ to be this shared probability for classes C_i and C_j (the bar indicates that we have reindexed the set of probabilities by the classes) and we say that $\mathcal{P}$ forms a *k-class*. Note that $\overline{p}_{i,i}$ is assumed to be $1/2$, so that $\mathcal{M}_{\text{nn}}$ swaps elements within the same class with probability $1/2$. Define $\overline{\mathcal{P}} = \{\overline{p}_{i,j}\}$ as the set of probabilities over pairs of classes C_i and C_j where $i, j \leq k$. In this case, we say $\mathcal{P}$ forms a *k-class*. When $k = n$, the k-class assumption does not lose any generality, but this structure allows us to simplify the problem by considering $k < n$, as was done in [25, 16].

Define $\overline{q}_{i,j} = \overline{p}_{i,j}/\overline{p}_{j,i}$ to be the *bias* towards having a particle of type i ahead of a particle of type j. We say that $\mathcal{P}$ is *bounded* if there exists a constant $q > 1$ such that $\overline{q}_{i,j} \geq q$ for all $i < j \leq k$. The constant q is called the *minimum bias*.

The chain $\mathcal{M}_{\text{nn}}$ samples over S_n using these probabilities, and in particular the order of elements within each class approaches the uniform distribution. The spectral gap of this uniform sampling is well-understood and may be analyzed separately (see Section 7.4). In order to isolate the biased moves, we define a new Markov chain $\mathcal{M}_{\text{pp}}$ that eliminates swaps within each class. As $\mathcal{M}_{\text{pp}}$ maintains a fixed order on particles within each class, it makes sense to relabel each element of $[n]$ by the index of the class it is in. That is, we let $c_i = |C_i|$ and we consider a linear array of length n with c_i particles labeled i for each $1 \leq i \leq k$. We call this a k-particle system for the given set $\{c_i\}$, and the Markov chain $\mathcal{M}_{\text{pp}}$ is called a k-particle process. We view the new state space as the set Ω of k-particle systems for $\{c_i\}$.

The Markov chain $\mathcal{M}_{\text{pp}}$ over k-particle systems also allows certain non-adjacent transpositions. In particular, we let a particle of type i and a particle of type j swap across particles of type less than i and j. More formally, the chain $\mathcal{M}_{\text{pp}}$ is defined as follows.

The particle process Markov chain $\mathcal{M}_{\text{pp}}$

`Starting at any` k`-particle system` σ^0`, iterate the following:`

- At time t, choose a position $1 \leq i \leq n$ and direction $d \in \{L, R\}$ uniformly at random.
- If $d = L$, find the largest j less than i with $\sigma^t(j) \geq \sigma^t(i)$ (if one exists). If $\sigma^t(j) > \sigma^t(i)$, then with probability 1/2, exchange $\sigma^t(i)$ and $\sigma^t(j)$ to obtain σ^{t+1}.
- If $d = R$, find the smallest j with $j > i$ and $\sigma^t(j) \geq \sigma^t(i)$ (if one exists). If $\sigma^t(j) > \sigma^t(i)$, then exchange $\sigma^t(i)$ and $\sigma^t(j)$ to obtain σ^{t+1} with probability

$$\frac{1}{2}\,\overline{q}_{\sigma_t(j),\sigma_t(i)} \prod_{i<l<j} \overline{q}_{\sigma_t(j),\sigma_t(l)} \overline{q}_{\sigma_t(l),\sigma_t(i)} \, .$$

- With all remaining probability, $\sigma^{t+1} = \sigma^t$.

The chain $\mathcal{M}_{\text{pp}}$ connects the space Ω and has the stationary distribution (see, e. g., [4])

$$\pi(\sigma) = \prod_{i<j:\sigma(i)>\sigma(j)} \frac{\overline{p}_{\sigma(i),\sigma(j)}}{\overline{p}_{\sigma(j),\sigma(i)}} Z^{-1} = \prod_{i<j:\sigma(i)>\sigma(j)} \overline{q}_{\sigma(i),\sigma(j)} Z^{-1}$$

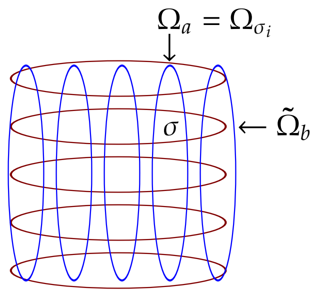
where Z is a normalizing constant and $\overline{q}_{\sigma(i),\sigma(j)} = \frac{\overline{p}_{\sigma(i),\sigma(j)}}{\overline{p}_{\sigma(j),\sigma(i)}}$.

We adapt the definition of weakly monotonic to the setting of k-particle systems as follows:

Definition 7.2 ([4])**.** The set $\overline{\mathcal{P}}$ is weakly monotonic if properties 1 and either 2 or 3 are satisfied.

1. $\overline{p}_{i,j} \geq 1/2$ for all $1 \leq i < j \leq k$, and
2. $\overline{p}_{i,j+1} \geq \overline{p}_{i,j}$ for all $1 \leq i < j \leq k-1$ or
3. $\overline{p}_{i-1,j} \geq \overline{p}_{i,j}$ for all $2 \leq i < j \leq k$.

As in [25], we will assume that property (2) holds. If instead property (3) holds, then as described in [25] we would modify $\mathcal{M}_{\text{pp}}$ (and $\mathcal{M}_{\text{T}}$ defined below) to allow swaps between elements of different particle types across elements whose particle types are larger (instead of smaller) and modify the induction so that at each step σ_i restricts the location of particles larger than i (instead of smaller).



$$\begin{aligned} \sigma_2 &= 12\_1\_2\_1\_\,\_ \\ \sigma_3 &= 1231\_2\_13\_ \\ a &= 3\_\,\_3\_ \\ b &= 454 \\ \sigma &= 1231425134 \end{aligned}$$

Figure 5: The state space $\Omega_{\sigma_{i-1}}$ decomposed, with $i = 3$.

7.2 Details of the decomposition and proof of Theorem 1.3

In this section, we present the proof of Theorem 1.3, which bounds the mixing time of the particle process chain $\mathcal{M}_{\text{pp}}$.

Theorem 1.3. If the probabilities $\mathcal{P}$ are weakly monotonic and form a bounded k-class with $c_i \geq 2N^*$ for all $1 \leq i \leq k$, then the spectral gap γ of the chain $\mathcal{M}_{\text{pp}}$ satisfies $\gamma = \Omega(n^{-2})$.

Our proof uses the same inductive technique as [25], where at each level of the induction we fix the locations of particles in one less particle class. Recalling notation from Section 1, let $C_{<=i} = C \cup C_{<i}$ be the set of particles with type less than or equal to i (i. e., particles of type $1, 2, \ldots, i$), and similarly, $C_{>=i} = C \cup C_{>i}$. For $i \geq 0$, let σ_i represent a fixed location of the particles in $C_{<=i}$ (σ_0 represents no restriction); for example, in Figure 5, we set $\sigma_2 = 12\_1\_2\_1\_\_$, where "_" represents locations that can be filled with particles in $C_{>=3}$. We will consider the chain $\mathcal{M}_{\sigma_i}$ whose state space Ω_{σ_i} is the set of all k-particle systems σ where the location of particles in $C_{<=i}$ are consistent with σ_i. The moves of $\mathcal{M}_{\sigma_i}$ are those moves from $\mathcal{M}_{\text{pp}}$ that do not involve particles in $C_{<=i}$. We prove by induction that $\mathcal{M}_{\sigma_i}$ has spectral gap $\Omega(n^{-2}(1-1/n)^{2(k-2-i)})$ for all choices of σ_i. To be clear, we assume that the spectral gap of $\mathcal{M}_{\sigma_i}$ are bounded for all σ_i by induction, and then prove our bound on the spectral gap of $\mathcal{M}_{\sigma_{i-1}}$.

To start, we show that $\Omega_{\sigma_{i-1}}$ is a product space, which is required to apply Theorem 1.2. Let A consist of all 2-particle systems with $c_i = |C_i|$ particles of type i and $\sum_{j=i+1}^{k} c_j$ particles of type "_". Let B consist of all $k-i$ particle systems with $c_j = |C_j|$ particles of type j for all $i+1 \leq j \leq k$. For example, using our running example where $\sigma_2 = 12\_1\_2\_1\_\_$, an example $a \in A$ is _3_3_ and $b \in B$ is 544. See Figure 5 for an additional example. Our goal is to show that the set of permutations σ consistent with σ_{i-1} on particles in $C_{<i}$ is in bijection with $A \times B$. To this end, we can write $\sigma = (a, b)$, where $a \in A$ is the 2-particle system obtained from σ_i by removing particles in $C_{<i}$ (see Figure 5), as those particles are in a fixed position for all of $\Omega_{\sigma_{i-1}}$. Next, define $b \in B$ to be the restriction of σ to particles in $C_{>i}$. For the other direction, given any (a, b) pair, it is clear that there is a unique $\sigma \in \Omega_{\sigma_{i-1}}$ corresponding to that pair. Returning to our example, given $\sigma_2 = 12\_1\_2\_1\_\_$, $a = \_3\_3\_$, and $b = 544$, then the unique $\sigma \in \Omega_{\sigma_2}$ is 1251324134.

We next describe the decomposition. Note that the moves of $\mathcal{M}_{\sigma_i}$ fix an $a \in A$ and perform (j_1, j_2) transpositions, where $j_1, j_2 > i$; i. e., they operate exclusively on B. Thus, the Markov chain $\mathcal{M}_{\sigma_i}$ is a restriction of $\mathcal{M}_{\sigma_{i-1}}$ with state space Ω_a. On the other hand, the remaining moves

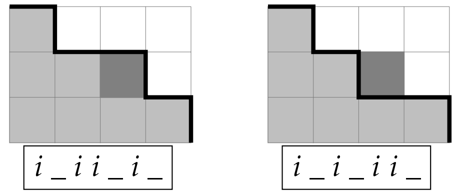


Figure 6: An exclusion process on staircase walks operates by adding or removing a square.

of $\mathcal{M}_{\sigma_{i-1}}$ are (i, j) transpositions for $j > i$. These are the complementary restrictions; these moves fix a $b \in B$ and operate on A, so we label the state space of this Markov chain $\tilde{\Omega}_b$. As these moves fix the relative order of all particles in $C_{>i}$, the complementary restriction chains can be seen as a 2-particle process on particles in C_i with particles in $C_{>i}$. This process is easily seen as in bijection with staircase walks by mapping each particle in C_i to a step right and each particle in $C_{>i}$ to a step down, as in Figure 6. In [25], they define a *bounded generalized biased exclusion process* as a processes that operate on staircase walks as in Figure 6, where every square has a different bias but they are all bounded by some q. More formally, an exclusion process is a process on n_1 1's and n_0 0's which occupy linear positions: $1, \ldots, n_0 + n_1$.. The set of all distinct linear orderings of these particles are called *2-particle systems*. In the generalized setting, for a configuration σ, the probability $p_{\sigma,i}$ of swapping two particles at position i and $i + 1$ can depend on both the current ordering and the particles being exchanged. The process is *bounded* if there exists a constant $\gamma > 1$ such that for all configurations σ if $\sigma(i) = 1$ and $\sigma(i + 1) = 0$ and τ is obtained from σ by swapping particles $\sigma(i)$ and $\sigma(i + 1)$, then $p_{\sigma,i}/p_{\tau_i} \geq q$. Bounded generalized biased exclusion processes were analyzed in [25].

Theorem 7.3 ([25]). *Let $\mathcal{M}_{ex}$ be a bounded generalized exclusion process on n_1 1's and n_0 0's. The spectral gap of $\mathcal{M}_{ex}$ is $\Omega((n_0 n_1)^{-1})$.*

Next we prove the following lemma.

Lemma 7.4. *The complementary restrictions at each level of the induction are bounded generalized biased exclusion processes with spectral gap $\Omega(n^{-2})$.*

Proof. Since b is fixed, each particle in b corresponds to a given row of squares in Figure 6. Each column in the figure corresponds to a particular element of type i in a. The bias of swapping an element i with a particular "_" depends only on the position of those two elements in the array σ, and not on the relative order of all other elements in $C_{>i-1}$. In this way, that bias corresponds to a particular square in Figure 6. Elements in $C_{<i}$ change the bias of all squares along a given diagonal in Figure 6. For example, let $\sigma = 3243\underline{3}1\underline{5}324$ and suppose $i = 3$. Then $a = 3\_33\_3\_$ and $b = 454$. The highlighted square in Figure 6 corresponds to swapping the 3 and 5 in positions 5 and 7, respectively. The probability of making that move is $\overline{q}_{5,3}\overline{q}_{5,1}/2\overline{q}_{3,1}$. Thus, each complementary restriction chain can be viewed as a generalized exclusion process acting on $a \in A$. The minimum bias is $q := \min_{i<j} \overline{q}_{i,j}$, which we have assumed is a constant bigger than 1. We use the following result from [25].

Since the probabilities $\overline{\mathcal{P}}$ are weakly monotonic (specifically, they satisfy condition (1) of weak monotonicity) and bounded, the exclusion process involving the particles in C_i and the particles in $C_{<i}$ is a bounded generalized exclusion process and we can apply Theorem 7.3. There are c_i particles of type i, $\sum_{j=i+1}^{k} c_j < n$ particles of type "_", and the moves are selected with probability $c_i/4n$. Applying Theorem 7.3 shows that the spectral gap of each complementary restriction chain is $\Omega(\frac{c_i}{n}\frac{1}{nc_i}) = \Omega(n^{-2})$. □

The chain $\mathcal{M}_{\sigma_{i-1}}$ is not the direct product of the chains on A and B because, e. g., for $(a, b) \in A \times B$, $P((a, b), (a', b))$ depends on b. However, in Section 7.3 we prove Lemma 7.6 which states that the above decomposition is ϵ-orthogonal where $\epsilon \leq 1/n^2$.Specifically, Lemma 7.6 says that if the probabilities $\mathcal{P}$ are weakly monotonic and bounded with $c_i \geq 2N^*$ for all $1 \leq i \leq k$, then at each step of the induction, $\epsilon \leq 1/n^2$. Here, we use Lemma 7.6 to complete the proof of Theorem 1.3 which says that if the probabilities $\overline{\mathcal{P}}$ are weakly monotonic and bounded, and $c_i \geq 2N^*$ for $1 \leq i \leq k$ then the spectral gap γ of the chain $\mathcal{M}_{\text{pp}}$ satisfies $\gamma = \Omega(n^{-2})$. Note that $\mathcal{P}$ being weakly monotonic implies that $\overline{\mathcal{P}}$ is as well.

Proof of Theorem 1.3. At each step of the induction, we apply the complementary decomposition theorem (Theorem 1.2). The restrictions of each decomposition will be rapidly mixing by induction and the complementary restriction chains are bounded generalized exclusion processes, by Lemma 7.4. The base case is $i = k - 2$ and the final decomposition is $i = 0$.

We begin with our base case, $i = k - 2$. Let σ_{k-2} be any fixed location of the particles in $C_{<=k-2}$. The Markov chain $\mathcal{M}_{\sigma_{k-2}}$ rejects all moves of $\mathcal{M}_{\text{pp}}$ unless they exchange a particle in C_{k-1} with a particle in C_k, making it a generalized exclusion process. Thus, $\mathcal{M}_{\sigma_{k-2}}$ is a bounded generalized exclusion process slowed down by a factor of $c_{k-1}/4n$, as this is the probability that these transitions are chosen. By Theorem 7.3 for any such σ_{k-2}, $\mathcal{M}_{\sigma_{k-2}}$ has spectral gap $\Omega(\frac{c_{k-1}}{n}\frac{1}{nc_{k-1}}) = \Omega(n^{-2})$.

We assume by induction the mixing time bound holds for all $\mathcal{M}_{\sigma_i}$ for some $i \leq k - 2$, and we will use this result to prove that our mixing time bound holds for all $\mathcal{M}_{\sigma_{i-1}}$. Let σ_{i-1} represent any fixed choice of locations for all particles in $C_{<i}$. In order to bound the spectral gap $\gamma(\mathcal{M}_{\sigma_{i-1}})$ of the chain $\mathcal{M}_{\sigma_{i-1}}$ we will apply Theorem 1.2. Given any σ_i that is consistent with σ_{i-1} (i. e., they agree on the locations of all particles in $C_{<=i}$), the Markov chain $\mathcal{M}_{\sigma_i}$ will be a restriction Markov chain of $\mathcal{M}_{\sigma_{i-1}}$. By induction, we have that the spectral gap $\gamma(\mathcal{M}_{\sigma_i}) = \Omega\left(n^{-2}\left(1-\frac{1}{n}\right)^{2(k-2-i)}\right)$. Lemma 7.4 shows that the complementary restrictions have minimum gap $\Omega(n^{-2})$. By Lemma 7.6, we know that the decomposition is ϵ-orthogonal, with $\epsilon \leq 1/n^2$. Combining these with Theorem 1.2 implies $\gamma(\mathcal{M}_{\sigma_{i-1}}) = \Omega\left(n^{-2}\left(1-\frac{1}{n}\right)^{2(k-2-i)+2}\right)$. Substituting $i = 0$ gives the desired theorem

$$\gamma(\mathcal{M}_{\text{pp}}) = \Omega\left(n^{-2}\left(1-\frac{1}{n}\right)^{2k-2}\right) = \Omega\left(n^{-2}\left(1-\frac{1}{n}\right)^{n}\right) = \Omega\left(n^{-2}\right). \qquad \square$$

Remark 7.5. It is worth pointing out that this decomposition of $\mathcal{M}_{\sigma_{i-1}}$ does not satisfy the regularity conditions of [18] needed to obtain a better bound. For any $v \in \Omega_j$, define

$$\pi_j^{j'}(v) = \pi_j(v)\frac{\sum_{v' \in \Omega_{j'}} P(v, v')}{\bar{P}(j, j')} .$$

We need to bound $\pi_j^{j'}(v)/\pi_j(v)$ for any j, j', and $v \in \Omega_j$. For example, let $\sigma_2 = 12\_11111\_2\_1\_\ \_$ and $\sigma_3 = 12311111\_231\_\ \_$. Notice that the two permutations $v_1 = 12311111423156$ and $v_2 = 12311111523146$ are in the same restriction Ω_j (i. e., they are both consistent with σ_3). They each have a single move to $\Omega_{j'}$: the move of swapping the first 3 with the 4 (in the case of v_1) or 5 (in the case of v_2). However, the probability of these moves differ by a factor of $(\overline{q}_{4,3}/\overline{q}_{5,3})(\overline{q}_{4,1}/\overline{q}_{5,1})^5$, as there are five 1's between them. In principle, there could be order n smaller numbers between the two numbers we are swapping. Thus, $\pi_j^{j'}(v)/\pi_j(v)$ cannot be uniformly bounded to within $1 \pm \eta$ unless η is exponentially large.

7.3 Bounding the orthogonality of the decomposition.

In order to complete the proof of Theorem 1.3, it remains to prove Lemma 7.6, which states that at each step of the induction, the value of ϵ is bounded by $1/n^2$.

Lemma 7.6. *If the probabilities $\mathcal{P}$ are weakly monotonic and bounded with $c_i \geq 2N^*$ for all $1 \leq i \leq k$, then at each step of the induction the quantity ϵ defined in equation (1.1) satisfies $\epsilon \leq 1/n^2$.*

Before proceeding with the proof, recall that the chain $\mathcal{M}_{\sigma_{i-1}}$ is not the direct product of the chains on A and B because, e. g., for $(a, b) \in A \times B$, $P((a, b),(a', b))$ depends on b. However, we show that the decomposition given in Section 7.2 is $1/n^2$-orthogonal by bounding $r(a, b) = \pi(a, b)/(\pi(a)\pi(b))$. Recall that at each step of the induction, a is a 2-particle system obtained from σ_i by removing all particles in $C_{<i}$, $c_i = |C_i|$, and there are $\sum_{j=i+1}^{k} c_j$ particles of type "_". Also recall that b is the $(k - i)$-particle system consisting of the particles in $C_{>i}$, and the location of the particles in $C_{<i-1}$ are fixed throughout this step of the induction (see Figure 5).

We define a to be "good" when it has fewer than $N^* = C_q \log n$ inversions; otherwise a is "bad." Similarly, we define b to be "good" when it has fewer than N^* inversions involving particles in C_{i+1}; otherwise it is "bad." Thus, as $|C_{i+1}| \geq 2N^*$, (a, b) has no inversions between i and j for $j > i + 1$ when a and b are both good. For such pairs, $r(a, b)$ is very close to 1. We will show that almost all weight contributing to π comes from pairs (a, b) where a and b are good. For all other pairs we show $r(a, b)\pi(\Omega_a \cap \tilde{\Omega}_b)$ is small.

By viewing b as a staircase walk on C_{i+1} and $C_{>i+1}$, we see that for either a or b, the probability it is bad is smaller than the weighted sum of all biased exclusion processes with more than N^* inversions (equivalently, area N^* under the curve). We start with the following lemma.

Lemma 7.7. *For any biased exclusion process with minimum bias constant $q > 1$, the total weight of staircase walks with area larger than N^* satisfies*

$$\sum_{\sigma : N(\sigma) \geq N^*} q^{-N(\sigma)} \leq \frac{1}{6n^2} .$$

Proof. The number of staircase walks with area N under the curve is precisely the partition number $p(N)$. By a theorem of Erdős [12], $p(N) < \mathrm{e}^{\pi\sqrt{2N/3}}$, and as $\mathrm{e}^{\pi\sqrt{2/3}} < 14$, we have $p(N) < 14^{\sqrt{N}}$. We can use this bound on $p(N)$ to bound the weight of a set of staircase walks, where each inversion contributes a factor of $1/q$ to the weight of the staircase walk (recall $q := \min_{i<j} \overline{q}_{i,j}$). Define

$$N^* = \max\left\{\frac{\log(6n^2) + \log((1+q)/(q-1))}{\log((1+q)/2)}, \frac{\log^2(14)}{\log^2(2q/(1+q))}\right\} = \Theta(\log n)\,.$$

Let $N(\sigma)$ be the number of inversions in a particular staircase walk σ. For all $N \geq \frac{\log^2(14)}{\log(2q/(1+q))}$, we have $14^{\sqrt{N}}q^{-N} \leq (2/(1+q))^N$. Therefore

$$\sum_{\sigma:N(\sigma)\geq N^*} q^{-N(\sigma)} \leq \sum_{N\geq N^*} p(N)q^{-N} \leq \sum_{N\geq N^*} 14^{\sqrt{N}}q^{-N} \leq \sum_{N\geq N^*}\left(\frac{2}{1+q}\right)^N = \left(\frac{2}{1+q}\right)^{N^*}\frac{1}{1-\frac{2}{1+q}} \leq \frac{1}{6n^2}$$

by the choice of N^*. □

7.3.1 High-level summary

We are now ready to use Lemma 7.7 to prove that $\epsilon \leq 1/n^2$. For improved readability, we illustrate the main ideas of the proof here in the simplified $k = 3$ case before proving Lemma 7.6. In this case, there is no recursion but instead just a single application of the decomposition theorem. The restriction chains of $\mathcal{M} = \mathcal{M}_{\sigma_0}$ are the set $\{\mathcal{M}_{\sigma_1}\}$, which fix all elements in class C_1. The stationary distribution of $\mathcal{M}$ is

$$\pi(\sigma) = Z^{-1}\prod_{\substack{i<j:\\ \sigma(i)>\sigma(j)}} \overline{q}_{\sigma(i),\sigma(j)}, \tag{7.1}$$

where Z is a normalizing constant.

Let $w(a)$ and $w(b)$ be the parts of this product that depend only on a and only on b, respectively, and let $w(a,b)$ be a correction factor that depends on both a and b. Let $t_{1,3}$ (respectively, $t_{1,2}$ and $t_{2,3}$) denote the number of inversions in σ between a 1 and a 3 (respectively a 1 and a 2 and a 2 and a 3). For example, let $\sigma = 111221312323$, which has stationary probability $Z^{-1}(\overline{q}_{1,2})^4(\overline{q}_{2,3})^3(\overline{q}_{1,3})$. Then $a = 111\_\ \_1\_1\_\ \_\ \_\ \_$ $b = 2232323$. From b we find that $t_{2,3} = 3$ and $w(b) = (\overline{q}_{2,3})^3$; more generally, define $w(b)$ to be the product $(\overline{q}_{2,3})^{t_{2,3}}$. From a, we can see that there are five inversions involving 1, but the number of those that are inversions with a 3 versus a 2 depends on b as well. Ignoring this for a moment, we define $w(a)$ to be the product $(\overline{q}_{1,2})^{t_{1,2}+t_{1,3}}$. In our example, $w(a) = (\overline{q}_{1,2})^5$. Since we have made the false assumption that there were no inversions between a 1 and a 3 in σ, we need a correction factor $w(a,b) = (\overline{q}_{1,3}/\overline{q}_{1,2})^{t_{1,3}}$. With these definitions, it is clear that $\pi(\sigma) = Z^{-1}w(a)w(b)w(a,b)$.

A key idea in the proof of Lemma 7.6 is that if a and b are both good, then $t_{1,3} = 0$—indeed, the total number of inversions is less than $2N^*$ and the number of 2's is at least $2N^*$—and thus the correction factor satisfies $w(a, b) = 1$, implying $\pi(a, b) \approx \pi(a)\pi(b)$. Moreover, the probability that a or b is bad is very small, so these pairs (a, b) do not contribute much to the sum in equation (1.1).

7.3.2 Proof of Lemma 7.6

We may now prove Lemma 7.6 in its full generality.

Proof of Lemma 7.6. Assume that at this step in the induction, all particles in $C_{<i}$ are fixed in the same position in all k-particle systems according to some σ_{i-1}. The stationary distribution of the chain $\mathcal{M}_{\sigma_{i-1}}$ is

$$\pi_{\sigma_{i-1}}(\sigma) = \prod_{i<j:\sigma(i)>\sigma(j)} \overline{q}_{\sigma(i),\sigma(j)} Z^{-1}_{\sigma_{i-1}}$$

where $Z_{\sigma_{i-1}}$ is the normalizing constant $\sum_{\sigma\in\Omega_{\sigma_{i-1}}} \pi(\sigma)$, the set $\Omega_{\sigma_{i-1}}$ contains the k-particle systems consistent with σ_{i-1}, and $\overline{q}_{\sigma(i),\sigma(j)} = \frac{\overline{p}_{\sigma(j),\sigma(i)}}{\overline{p}_{\sigma(j),\sigma(i)}}$. For ease of notation, throughout the remainder of this section we will let $\pi = \pi_{\sigma_{i-1}}$ and $Z = Z_{\sigma_{i-1}}$.

Let a^* and b^* be the highest weight elements in A and B, respectively. Using these definitions, the k-particle system (a^*, b^*) has the particles in $C_{<i}$ fixed according to σ_{i-1} and all other higher particles in sorted order. In our example, $b^* = 445$, $(a, b^*) = 1231423145$, $a^* = 33\_\ \_\ \_$, $(a^*, b) = 1231324154$, and $(a^*, b^*) = 1231324145$.

Next, we will decompose the product over inversions in equation (7.1) into several quantities. Notice that $\pi(a^*, b^*)$ is the product over inversions that are in every $\sigma \in \Omega$, normalized by $Z_{\sigma_{i-1}}$. Define $w(a) = \pi(a, b^*)/\pi(a^*, b^*)$. This is the product over inversions that are in every $\sigma \in \Omega_a$ that are not in every $\sigma \in \Omega$. Similarly, $w(b) = \pi(a^*, b)/\pi(a^*, b^*)$ is the product over inversions in every $\sigma \in \tilde{\Omega}_b$, and $w(a, b) = \pi(a, b)\pi(a^*, b^*)/(\pi(a, b^*)\pi(a^*, b))$ is the product over inversions in $\sigma = (a, b)$ beyond those that are required by being in Ω_a and $\tilde{\Omega}_b$. From these definitions, it is clear that $\pi(a, b) = w(a)w(b)w(a, b)\pi(a^*, b^*)$. We will prove in the following lemma that if both a and b are good then $w(a, b) = 1$; i. e., that the weight of (a, b) is determined entirely by being in Ω_a and $\tilde{\Omega}_b$.

Lemma 7.8. *If a and b are both good then $w(a, b) = 1$.*

Proof. From the definitions, it is sufficient to show that $\pi(a, b) = \pi(a, b^*)\pi(a^*, b)/\pi(a^*, b^*)$. First consider the inversions (j, l) for $j, l < i$. These are present in every term—$\pi(a, b)$, $\pi(a, b^*)$, $\pi(a^*, b)$, and $\pi(a^*, b^*)$—and thus cancel. The inversions $(i + 1, i)$ are exactly the same in $\pi(a, b)$ and $\pi(a, b^*)$, and there are none in $\pi(a^*, b)$, and $\pi(a^*, b^*)$; thus these also cancel. Similarly, inversions (j, l) for $j \geq i + 1, l \geq i$ are exactly the same in $\pi(a, b)$ and $\pi(a^*, b)$, and there are none in $\pi(a, b^*)$, and $\pi(a^*, b^*)$; thus these cancel. Next consider the inversions (j, l) for $j > i + 1, l < i$. Since there are no (j, i) inversions, these are the same in $\pi(a, b)$ and $\pi(a^*, b)$ and the same in

$\pi(a^*, b^*)$ and $\pi(a, b^*)$. Similarly, the (i, j) for $j < i$ inversions are the same in $\pi(a^*, b)$ and $\pi(a^*, b^*)$ and the same in $\pi(a, b)$ and $\pi(a, b^*)$.

Finally, the $(i+1, j)$ for $j < i$ inversions are the most complicated. Assume the particles in C_{i+1} are numbered and in order in all 4 k-particle systems. Consider any particular particle in C_{i+1} and look at its position in $\pi(a, b)$. If it is to the left of any particle in C_i (i. e., it is involved in an $(i+1, i)$ inversion) then it must come before all particles in $C_{>i+1}$ and it will be in the same position in $\pi(a, b)$ and $\pi(a, b^*)$ and the same positions (which may be different positions) in $\pi(a^*, b)$ and $\pi(a^*, b^*)$; thus any $(i+1, j)$ inversions will cancel. If not, then it comes after all particles in C_i and it will be in the same position in $\pi(a, b)$ and $\pi(a^*, b)$ and again the same positions in $\pi(a, b^*)$ and $\pi(a^*, b^*)$.

As there are no (j, i) inversions for $j > i+1$, we conclude that $w(a, b) = 1$ when a and b are good. □

Next, define $Z_A = \sum_a w(a)$, $Z_B = \sum_b w(b)$, and $\epsilon_1 = 1/6n^2$. We show that $\sum_{a \text{ bad}} w(a) \leq \epsilon_1$ and $\sum_{b \text{ bad}} w(b) \leq \epsilon_1 / Z_B$. Thus, we find $Z_A \approx \sum_{a \text{ good}} w(a)$ and $Z_B \approx \sum_{b \text{ good}} w(b)$, and moreover $Z = \sum_{a,b} w(a)w(b)w(a,b) \approx Z_A Z_B$. We show that when a and b are both good, $\pi(a) \approx w(a) Z_B / Z$ and $\pi(b) \approx w(b) Z_A / Z$. Thus, when a and b are both good, $r(a, b) = \frac{\pi(a,b)}{\pi(a)\pi(b)} \approx \frac{Z}{Z_A Z_B} \approx 1$.

Let $Z_A^{good} = \sum_{a \text{ good}} w(a)$ be the sum Z_A restricted to only good configurations a and $Z_B^{good} = \sum_{b \text{ good}} w(b)$ be the sum Z_B restricted to only good configurations b. By Lemma 7.7, this is a bound on the total weight of staircase walks with area more than N^* under the curve. We will use this to bound the weight of the bad a's contributing to Z_A and the weight of the bad b's contributing to Z_B. Specifically, we will prove the following lemma.

Lemma 7.9.

1. *$Z_A - Z_A^{good} = \sum_{a \text{ bad}} w(a) \leq \epsilon_1$ and $Z_B - Z_B^{good} = \sum_{b \text{ bad}} w(b) \leq \epsilon_1 Z_B$.*
2. *For all a, b we have $\pi(a) \leq w(a)\pi(a^*, b^*) Z_B$ and $\pi(b) \leq w(b)\pi(a^*, b^*) Z_A$.*
3. *For* good *a and* good *b we have $\pi(a) \geq w(a)\pi(a^*, b^*) Z_B^{good}$ and $\pi(b) \geq w(b)\pi(a^*, b^*) Z_A^{good}$.*
4. *For all b we have $\pi(b) \geq w(b)\pi(a^*, b^*) w(a^*, b)$.*

Proof. **Part (1.)** Recall that $w(a) = \pi(a, b^*)/\pi(a^*, b^*)$. Both k-particle systems (a, b^*) and (a^*, b^*) have the same inversions (j, l) for $j, l < i$, so when we consider the ratio $w(a)$, this contains the (i, j) for $j > i$ inversions and the (j, l) for $j \geq i, l < i$ inversions. Since there are no (i, j) for $j > i$ inversions in (a^*, b^*), and the weight of the (j, l) for $j \geq i, l < i$ inversions is less in (a, b^*) than it is in (a^*, b^*), we have

$$Z_A - Z_A^{good} = \sum_{a \text{ bad}} w(a) \leq \sum_{a \text{ bad}} \prod_{\substack{j<l:a(j)>i \\ a(l)=i}} q_{a(j),i} \leq \sum_{a \text{ bad}} \prod_{\substack{j<l:a(j)>i \\ a(l)=i}} q_{i+1,i} \leq \epsilon_1 \,.$$

The last two steps follow from $\mathcal{P}$ being weakly monotonic and from Lemma 7.7, respectively.

Next, we consider $Z_B - Z_B^{good}$ and recall that $w(b) = \pi(a^*, b)/\pi(a^*, b^*)$. The two k-particle systems (a^*, b) and (a^*, b^*) have the same inversions (j, l) for $j, l \le i$. When considering the ratio $w(b)$, there are several types of inversions remaining. There are the (j, l) for $j, l > i+1$ inversions, which arise due to the order of the particles in $C_{>=i+2}$. We will represent these τ_{i+2}. Additionally, there are the inversions $(j, i+1)$ for $j > i+1$, of which there are at least N^* if b is *bad*, and the (j, l) for $j \ge i+1, l \le i$ inversions, which are maximized in τ^*_{i+2}, which we will define as the highest weight configurations consistent with σ_{i-1} and τ_{i+2}. In other words, τ^*_{i+2} is the configuration that has the particles in $C_{<i}$ ordered according to σ_{i-1}, the particles in C_i as far forward as possible, then the particles in C_{i+1} again as far forward as possible, and finally the particles in $C_{>i+1}$ in the remaining positions ordered according to the $(k-i-1)$-particle system τ_{i+2}. Given these definitions, we have the following:

$$
\begin{aligned}
Z_B - Z_B^{good} &= \sum_{b \text{ bad}} w(b) = \sum_{\tau_{i+2}} w(\tau^*_{i+2}) \sum_{EX(i+1, j:j>i+1, \text{ bad})} w(b)/w(\tau^*_{i+2}) \\
&\le \sum_{\tau_{i+2}} w(\tau^*_{i+2}) \sum_{EX(i+1, j:j>i+1, \text{ bad})} \prod_{\substack{j<l: a(j)>i+1 \\ a(l)=i+1}} q_{a(j), i+1} \\
&\le \sum_{\tau_{i+2}} w(\tau^*_{i+2}) \sum_{EX(i+1, j:j>i+1, \text{ bad})} \prod_{\substack{j<l: a(j)>i+1 \\ a(l)=i+1}} q_{i+2, i+1} \\
&\le \sum_{\tau_{i+2}} w(\tau^*_{i+2}) \epsilon_1 \le \epsilon_1 Z_B,
\end{aligned}
$$

where $EX(i+1, j : j > i+1, \text{ bad})$ is an exclusion process (2-particle system) with more than N^* inversions; one particle in this process is C_{i+1} and the other is the elements of $C_{>i+1}$ with more than N^* inversions. In other words, we are dividing the $(k-i)$-particle system b based on the location of the particles in C_{i+1}.

Part (2.) Both of these bounds are straightforward from the definitions. For all a we have the following:

$$\pi(a) = \sum_{b'} \pi(a, b') = \sum_{b'} w(a) w(b') w(a, b') \pi(a^*, b^*) \le w(a) \pi(a^*, b^*) \sum_{b'} w(b') = w(a) \pi(a^*, b^*) Z_B \, .$$

Similarly, for all b we have the following:

$$\pi(b) = \sum_{a'} \pi(b, a') = \sum_{a'} w(a') w(b) w(a', b) \pi(a^*, b^*) \le w(b) \pi(a^*, b^*) \sum_{a'} w(a') = w(b) \pi(a^*, b^*) Z_A \, .$$

Part (3.) Next, we give a lower bound on $\pi(a)$ and $\pi(b)$ for good a and good b. Using the property that if a and b are both good then $w(a,b) = 1$, we have the following bound for good a:

$$\begin{aligned}\pi(a) &= \sum_{b'} \pi(a,b') = \sum_{b'} w(a)w(b')w(a,b') \\ &\geq w(a)\pi(a^*,b^*) \sum_{b' \text{ good}} w(b')w(a,b') \\ &= w(a)\pi(a^*,b^*) \sum_{b' \text{ good}} w(b') = w(a)\pi(a^*,b^*)Z_B^{good}.\end{aligned}$$

For good b we have a similar lower bound:

$$\begin{aligned}\pi(b) &= \sum_{a'} \pi(a',b) = \sum_{a'} w(a')w(b)w(a',b)\pi(a^*,b^*) \\ &\geq w(b)\pi(a^*,b^*) \sum_{a' \text{ good}} w(a')w(a',b) \\ &= w(b)\pi(a^*,b^*) \sum_{a' \text{ good}} w(a') = w(b)\pi(a^*,b^*)Z_A^{good}.\end{aligned}$$

Part (4.) Finally, we prove a weaker lower bound on b that holds for all b:

$$\begin{aligned}\pi(b) &= \sum_{a'} \pi(a',b) = \sum_{a'} w(a')w(b)w(a',b)\pi(a^*,b^*) \\ &= w(b)\pi(a^*,b^*) \sum_{a} w(a)w(a,b) \\ &\geq w(b)\pi(a^*,b^*)w(a^*)w(a^*,b) = w(b)\pi(a^*,b^*)w(a^*,b).\end{aligned}$$ □

Returning to the proof of Lemma 7.6, we start by assuming both a and b are good. We will handle the case where either a or b is bad next. Recall from Lemma 7.8 that if a and b are both good then $w(a,b) = 1$. Then, by parts (3) and (1) of Lemma 7.9, we have

$$\begin{aligned}r(a,b) &= \frac{\pi(a,b)}{\pi(a)\pi(b)} = \frac{w(a)w(b)\pi(a^*,b^*)}{\pi(a)\pi(b)} \\ &\leq \frac{w(a)w(b)\pi(a^*,b^*)}{(w(a)\pi(a^*,b^*)Z_B^{good})(w(b)\pi(a^*b^*)Z_A^{good})} \\ &= \frac{1}{\pi(a^*,b^*)Z_B^{good}Z_A^{good}} \leq \left(\frac{1}{(1-\epsilon_1)^2}\right)\frac{1}{Z_BZ_A\pi(a^*,b^*)} \leq \frac{1}{(1-\epsilon_1)^2}.\end{aligned}$$

The last step uses the following lower bound on $Z_BZ_A\pi(a^*,b^*)$.

$$Z_BZ_A\pi(a^*,b^*) = \sum_{(a,b)} w(a)w(b)\pi(a^*,b^*) > \sum_{(a,b)} w(a)w(b)w(a,b)\pi(a^*,b^*) = \sum_{(a,b)} \pi(a,b) = 1\,.$$

From part (2) of Lemma 7.9 we have the following lower bound when a and b are both good:

$$\begin{aligned} r(a,b) = \frac{\pi(a,b)}{\pi(a)\pi(b)} &= \frac{w(a)w(b)\pi(a^*,b^*)}{\pi(a)\pi(b)} \\ &\geq \frac{w(a)w(b)\pi(a^*,b^*)}{(w(a)\pi(a^*,b^*)Z_B)\,(w(b)\pi(a^*,b^*)Z_A)} \\ &= \frac{1}{Z_B Z_A \pi(a^*,b^*)} \geq (1-\epsilon_1)^2. \end{aligned}$$

In the last step we upper bound $Z_B Z_A \pi(a^*,b^*)$ as follows using part (1) of Lemma 7.9:

$$Z_B Z_A \pi(a^*,b^*) \leq \frac{Z_B^{good} Z_A^{good} \pi(a^*,b^*)}{(1-\epsilon_1)^2} = \sum_{\substack{a \text{ good}, \\ b \text{ good}}} \frac{w(a)w(b)\pi(a^*,b^*)}{(1-\epsilon_1)^2} < \sum_{(a,b)} \frac{\pi(a,b)}{(1-\epsilon_1)^2} = \frac{1}{(1-\epsilon_1)^2}.$$

In order to apply Lemma 5.4, we need to bound the quantity $\sum_{(a,b)} \pi(a,b) \left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2$. There are two cases depending on whether $r(a,b) \leq 1$, but either way we have

$$\left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \leq \left(\frac{1}{1-\epsilon_1} - (1-\epsilon_1)\right)^2 \leq 5\epsilon_1^2,$$

as long as $\epsilon_1 \leq .191$ (this is true since $\epsilon_1 \leq 1/(6n^2)$ and $n \geq 2$). Therefore

$$\sum_{\substack{a \text{ good}, \\ b \text{ good}}} \pi(a,b) \left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \leq 5\epsilon_1^2.$$

Next, we address the case where at least one of a or b is bad. In this case, we will show that the weight of these configurations is so small that it overcomes the fact that $w(a,b)$ and $r(a,b)$ may be exponentially small. If $r(a,b) \leq 1$ then $\pi(a,b) \left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \leq \pi(a)\pi(b)$. Otherwise, $\pi(a,b) \left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \leq \frac{\pi(a,b)^2}{\pi(a)\pi(b)}$. Either way,

$$\pi(a,b) \left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \leq \pi(a)\pi(b) + \frac{\pi(a,b)^2}{\pi(a)\pi(b)}.$$

In order to upper bound $\frac{\pi(a,b)^2}{\pi(a)\pi(b)}$, we will bound the conditional probabilities $\pi(a,b)/\pi(a)$ for *good* a and $\pi(a,b)/\pi(b)$ for *all* b. Using part (3) of Lemma 7.9 we have for a good:

$$\frac{\pi(a,b)}{\pi(a)} \leq \frac{w(a)w(b)w(a,b)\pi(a^*,b^*)}{w(a)\pi(a^*,b^*)Z_B^{good}} \leq \frac{w(b)}{Z_B^{good}}.$$

Using part (4) of Lemma 7.9 we have for all b:

$$\frac{\pi(a,b)}{\pi(b)} \leq \frac{w(a)w(b)w(a,b)\pi(a^*,b^*)}{w(b)w(a^*,b)\pi(a^*,b^*)} \leq w(a)\,.$$

Now we may use these facts to bound $\sum_{a,b} \frac{\pi(a,b)^2}{\pi(a)\pi(b)}$ for *bad* a and *all* b using part (1) of Lemma 7.9:

$$\begin{aligned}
\sum_{a \text{ bad},b} \frac{\pi(a,b)^2}{\pi(a)\pi(b)} &\leq \sum_{a \text{ bad},b} \frac{\pi(a,b)}{\pi(a)} w(a) \\
&= \sum_{a \text{ bad}} w(a) \frac{\sum_b \pi(a,b)}{\pi(a)} \\
&= \sum_{a \text{ bad}} w(a) \quad \leq \quad \epsilon_1.
\end{aligned}$$

Similarly, we bound $\sum_{a,b} \frac{\pi(a,b)^2}{\pi(a)\pi(b)}$ for good a and bad b using part (1) of Lemma 7.9.

$$\begin{aligned}
\sum_{a \text{ good},b \text{ bad}} \frac{\pi(a,b)^2}{\pi(a)\pi(b)} &\leq \sum_{a \text{ good},b \text{ bad}} \frac{\pi(a,b)}{\pi(b)} \frac{w(b)}{Z_B^{good}} \\
&\leq \sum_{b \text{ bad}} \frac{w(b)}{Z_B^{good}} \\
&= \frac{Z_B - Z_B^{good}}{Z_B^{good}} \\
&\leq \frac{\epsilon_1}{1-\epsilon_1}.
\end{aligned}$$

We continue our analysis of the case where at least one of a or b is bad and bound $\sum_{a,b} \pi(a)\pi(b)$. Notice that the summations over a and b are separable, so we have the following:

$$\sum_{a \text{ bad},b} \pi(a)\pi(b) = \Pr(a \text{ bad})$$

and

$$\sum_{b \text{ bad},a} \pi(a)\pi(b) = \Pr(b \text{ bad}).$$

We may bound these probabilities via the following inequalities.

$$Z_B\pi(a^*,b^*) \leq \frac{Z_B^{good} Z_A^{good} \pi(a^*,b^*)}{(1-\epsilon_1)} = \frac{1}{(1-\epsilon_1)} \sum_{\substack{a \text{ good},\\ b \text{ good}}} w(a)w(b)\pi(a^*,b^*) < \frac{1}{1-\epsilon_1}\,.$$

Similarly, we have $Z_A\pi(a^*, b^*) \le 1/(1-\epsilon_1)$. Using part (2) of Lemma 7.9, this implies that

$$\Pr(a \text{ bad}) = \sum_{a \text{ bad}} \pi(a) \le \sum_{a \text{ bad}} w(a)\pi(a^*, b^*) Z_B \le \sum_{a \text{ bad}} \frac{w(a)}{1-\epsilon_1} \le \frac{\epsilon_1}{1-\epsilon_1}\,.$$

We also have $\Pr(b \text{ bad}) \le \epsilon_1/(1-\epsilon_1)$. Thus when at least one of a and b is bad we have

$$\sum_{a,b} \pi(a,b)\left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \le \sum_{a,b}\left(\pi(a)\pi(b) + \frac{\pi(a,b)^2}{\pi(a)\pi(b)}\right) \le \epsilon_1 + \frac{3\epsilon_1}{1-\epsilon_1}\,.$$

Putting both cases together, we have

$$\sum_{a,b} \pi(a,b)\left(\sqrt{r(a,b)} - \frac{1}{\sqrt{r(a,b)}}\right)^2 \le 5\epsilon_1^2 + \epsilon_1 + \frac{3\epsilon_1}{1-\epsilon_1} \le 6\epsilon_1 = 1/n^2\,,$$

as long as $\epsilon_1 \le .225$, which is true since $n \ge 2$. Thus, this decomposition is ϵ-orthogonal, where $\epsilon \le 1/n^2$ This completes the proof of Lemma 7.6. □

7.4 The Markov chain $\mathcal{M}_{\mathbf{nn}}$

Here we give the details to show how we can use Theorem 1.3, which gives a lower bound on the spectral gap of the particle process Markov chain $\mathcal{M}_{\text{pp}}$, to obtain an upper bound on the mixing time of the nearest-neighbor Markov chain $\mathcal{M}_{\text{nn}}$.

First, we begin with some preliminaries on Markov chains and mixing times. The time a Markov chain takes to converge to its stationary distribution, or *mixing time*, is measured in terms of the distance between the distribution at time t and the stationary distribution π. The *total variation distance* at time t is $\|P^t, \pi\|_{tv} = \max_{x\in\Omega} \frac{1}{2}\sum_{y\in\Omega} |P^t(x,y) - \pi(y)|$, where $P^t(x,y)$ is the t-step transition probability, and Ω is the state space of the Markov chain. For all $\epsilon > 0$, the *mixing time* $\tau(\epsilon)$ of $\mathcal{M}$ is defined as

$$\tau(\epsilon) = \min\{t : \|P^{t'}, \pi\|_{tv} \le \epsilon, \forall t' \ge t\}\,.$$

In order to use the lower bound on the spectral gap to obtain an upper bound on the mixing time we will use the following well-known result. Again, let $\gamma(P) = 1 - |\lambda_1|$ denote the spectral gap, where $\lambda_0, \lambda_1, \ldots, \lambda_{|\Omega|-1}$ are the eigenvalues of the transition matrix P and $1 = \lambda_0 > |\lambda_1| \ge |\lambda_i|$ for all $i \ge 2$. The following result relates the spectral gap with the mixing time (see, e. g., [33],[29]):

Theorem 7.10 ([29]). *Let $\pi_* = \min_{x\in\Omega} \pi(x)$. For all $\epsilon > 0$ we have*

$$(a) \qquad \tau(\epsilon) \le \frac{1}{\gamma(P)} \log\left(\frac{1}{\pi_*\epsilon}\right).$$

$$.(b) \qquad \tau(\epsilon) \ge \frac{|\lambda_1|}{2\gamma(P)} \log\left(\frac{1}{2\epsilon}\right).$$

We also require the following comparison theorem from [9]. Let P' and P be two reversible Markov chains on the same state space Ω with the same stationary distribution π and let $E(P) = \{(x, y) : P(x, y) > 0\}$ and $E(P') = \{(x, y) : P'(x, y) > 0\}$ denote the sets of edges of the two chains, viewed as directed graphs. For each x, y with $P'(x, y) > 0$, define a path γ_{xy} using a sequence of states $x = x_0, x_1, \cdots, x_k = y$ with $P(x_i, x_{i+1}) > 0$, and let $|\gamma_{xy}|$ denote the length of the path. Let $\Gamma(z, w) = \{(x, y) \in E(P') : (z, w) \in \gamma_{xy}\}$ be the set of paths that use the transition (z, w) of P. Finally, define

$$\zeta = \max_{(z,w)\in E(P)} \left\{ \frac{1}{\pi(z)P(z,w)} \sum_{\Gamma(z,w)} |\gamma_{xy}|\pi(x)P'(x,y) \right\}. \tag{7.2}$$

Theorem 7.11 ([9]). *With the above notation,* $\gamma(P) \geq \frac{1}{\zeta}\gamma(P')$.

Now we may prove Theorem 1.4.

Proof of Theorem 1.4. Given our improved bound on the spectral gap of $\mathcal{M}_{\text{pp}}$ from Theorem 1.3, the remainder of this proof uses exactly the same approach as [25], except that we use Corollary 3.1 to eliminate one factor of n. We include a summary here for completeness. The complete details can be found in [25]. Instead of analyzing $\mathcal{M}_{\text{nn}}$ directly we will analyze an auxiliary chain $\mathcal{M}_{\text{T}}$ that allows a larger set of transpositions (including those allowed by $\mathcal{M}_{\text{pp}}$) and then use the comparison theorem, Theorem 7.11, to obtain a bound for $\mathcal{M}_{\text{nn}}$. In what follows, we write $C(x)$ to mean the index of the class containing element $x \in [n]$, so if $x \in C_i$ then $C(x) = i$.

The Transposition Markov chain $\mathcal{M}_{\text{T}}$

`Starting at any permutation` σ^0, `iterate the following:`

- At time $t \geq 0$, choose a position $1 \leq i \leq n$ in permutation σ^t and a direction $d \in \{L, R, N\}$ uniformly at random.
- If $d = L$, find the largest j with $1 \leq j < i$ and $C(\sigma^t(j)) \geq C(\sigma^t(i))$ (if one exists). If $C(\sigma^t(j)) > C(\sigma^t(i))$, then with probability $1/2$, exchange $\sigma^t(i)$ and $\sigma^t(j)$ to obtain σ^{t+1}.
- If $d = R$, find the smallest j with $n \geq j > i$ and $C(\sigma^t(j)) \geq C(\sigma^t(i))$ (if one exists). If $C(\sigma^t(j)) > C(\sigma^t(i))$, then with probability
$$\frac{1}{2}\, q_{\sigma^t(j),\sigma^t(i)} \prod_{i<k<j} \left(q_{\sigma^t(j),\sigma^t(k)} q_{\sigma^t(k),\sigma^t(i)} \right),$$
exchange $\sigma^t(i)$ and $\sigma^t(j)$ to obtain σ^{t+1}.
- If $d = N$, find the largest j with $1 \leq j < i$ and $C(\sigma^t(j)) = C(\sigma^t(i))$. If such an element exists, then with probability $1/2$, exchange the elements $\sigma^t(i)$ and $\sigma^t(j)$ to obtain σ^{t+1}.
- With all remaining probability, $\sigma^{t+1} = \sigma^t$.

The Markov chain $\mathcal{M}_{\mathrm{T}}$ has the same stationary distribution as $\mathcal{M}_{\mathrm{nn}}$ and is a product of $k+1$ independent Markov chains [25]. The first k chains involve moves between particles in the same class and the ith is an unbiased nearest-neighbor Markov chain over permutations on c_i particles. This chain has spectral gap $\Theta(c_i^{-3})$. (This is an unpublished result of Diaconis. See, e. g., [35].)

Lemma 7.12 (Diaconis)**.** *The spectral gap of the unbiased nearest neighbor Markov chain over permutations on* $[n]$ *is* $(1-\cos(\pi/n))/(n-1) = \Theta(n^{-3})$.

However, the chain $\mathcal{M}_{\mathrm{T}}$ updates one of the first k chains only if direction N is selected, which happens with probability $c_i/(6n)$ for each particle class $1 \le i \le k$. Thus, the spectral gap of the slowed-down version of this chain is $\Theta(1/(nc_i^2))$. The final chain $\mathcal{M}_{\mathrm{pp}}$ which we analyzed in Section 7 allows only moves between different particle classes; it is updated when direction L or R is selected (i. e., with probability 2/3), so, by Theorem 1.3, it has spectral gap $\Omega(n^{-2})$. Therefore, by Corollary 3.1, the spectral gap of $\mathcal{M}_{\mathrm{T}}$ is $\Omega(n^{-3})$.

The final step is to relate the spectral gap of $\mathcal{M}_{\mathrm{nn}}$ to that of $\mathcal{M}_{\mathrm{T}}$ using the comparison theorem of [9], Theorem 7.11. Section 5 of [25] proves the following lemma where ζ is defined as in equation (7.2).

Lemma 7.13 ([25])**.** *If the probabilities* $\mathcal{P}$ *are weakly monotonic and form a bounded* k*-class for* $k \ge 2$*, then* $\zeta = O(n^4)$.

Combining Lemma 7.13 and Theorem 7.11 with our bound on the gap of $\mathcal{M}_{\mathrm{T}}$, we get $\gamma(\mathcal{M}_{\mathrm{nn}}) = \Omega(n^{-7})$. □

Finally, we may bound the mixing time of $\mathcal{M}_{\mathrm{nn}}$.

Theorem 7.14. *If the probability array* $\mathcal{P}$ *is weakly monotonic and forms a bounded* k*-class for* $k \ge 2$*, with* $|C_i| \ge 2N^*$*, then the mixing time* τ_{nn} *of* $\mathcal{M}_{nn}$ *satisfies* $\tau_{nn} = O(n^9 \ln(1/\epsilon))$.

Proof. Finally, to get the mixing time of $\mathcal{M}_{\mathrm{nn}}$, we let $q_* = \max_{i<j} q_{i,j}$ then $\pi_* = \min_{x\in\Omega} \pi_{\mathrm{nn}}(x) \ge (q_*^{\binom{n}{2}} n!)^{-1}$ (see [4] and [25] for more details), so $\log(1/\epsilon\pi_*) = O(n^2 \ln \epsilon^{-1})$ since q is bounded from above by a positive constant (all $p_{i,j}$ are constant with respect to n). Applying Theorem 7.10(a) we have that the mixing time of $\mathcal{M}_{\mathrm{nn}}$ is $O(n^9 \ln \epsilon^{-1})$. This proves Theorem 7.14. □

References

[1] DAVID ALDOUS: Random walk on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII*, pp. 243–297. Springer, 1983. [doi:10.1007/BFb0068322] 2

[2] DAVE BAYER AND PERSI DIACONIS: Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992. [doi:10.1214/aoap/1177005705] 2

[3] ITAI BENJAMINI, NOAM BERGER, CHRISTOPHER HOFFMAN, AND ELCHANAN MOSSEL: Mixing times of the biased card shuffling and the asymmetric exclusion process. *Trans. AMS*, 357(8):3013–3029, 2005. [doi:10.1090/S0002-9947-05-03610-X] 2, 3

[4] PRATEEK BHAKTA, SARAH MIRACLE, DANA RANDALL, AND AMANDA PASCOE STREIB: Mixing times of Markov chains for self-organizing lists and biased permutations. *Random Struct. Algor.*, 61(4):638–665, 2022. Preliminary version in SODA'13. [doi:10.1002/rsa.21082] 2, 3, 4, 22, 23, 37

[5] FLAVIO CHIERICHETTI, ANIRBAN DASGUPTA, RAVI KUMAR, AND SILVIO LATTANZI: On reconstructing a hidden permutation. In *Proc. 18th Internat. Workshop on Randomization and Computation (RANDOM'14)*, pp. 604–617. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014. [doi:10.4230/LIPIcs.APPROX-RANDOM.2014.604] 2

[6] COLIN COOPER, MARTIN DYER, ALAN FRIEZE, AND RACHEL RUE: Mixing properties of the Swendsen–Wang process on the complete graph and narrow grids. *J. Math. Phys.*, 41(3):1499–1527, 2000. [doi:10.1063/1.533194] 4

[7] NICOLAS DESTAINVILLE: Bounding spectral gaps of Markov chains: A novel exact multi-decomposition technique. *J. Phys. A: Math. Theoret.*, 36(13):3647–3670, 2003. [doi:10.1088/0305-4470/36/13/301] 4, 5, 11

[8] PERSI DIACONIS AND LAURENT SALOFF-COSTE: Comparison techniques for random walk on finite groups. *Ann. Appl. Probab.*, 21(4):2131–2156, 1993. [doi:10.1214/AOP/1176989013] 2

[9] PERSI DIACONIS AND LAURENT SALOFF-COSTE: Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993. [doi:10.1214/AOAP/1177005359] 2, 3, 21, 36, 37

[10] PERSI DIACONIS AND MEHRDAD SHAHSHAHANI: Generating a random permutation with random transpositions. *Probab. Theory Relat. Fields*, 57(2):159–179, 1981. [doi:10.1007/BF00535487] 2

[11] JIAN DING, EYAL LUBETZKY, AND YUVAL PERES: Mixing time of critical Ising model on trees is polynomial in the height. *Comm. Math. Phys.*, 295(1):161–207, 2010. [doi:10.1007/s00220-009-0978-y] 4

[12] PAUL ERDŐS: On an elementary proof of some asymptotic formulas in the theory of partitions. *Ann. Math.*, 43(3):437–450, 1942. [doi:10.2307/1968802] 28

[13] PÉTER L. ERDŐS, ISTVÁN MIKLÓS, AND ZOLTÁN TOROCZKAI: New classes of degree sequences with fast mixing swap Markov chain sampling. *Combin. Probab. Comput.*, 27(2):186–207, 2018. [doi:10.1017/S0963548317000499] 4

[14] JAMES ALLEN FILL: An interesting spectral gap problem, from Jim Fill. Unpublished manuscript, 2003, posted on arXiv in 2025. [arXiv:2508.12557] 2, 22

[15] ROBERT GRONE, KARL HEINZ HOFFMANN, AND PETER SALAMON: An interlacing theorem for reversible Markov chains. *J. Phys. A: Math. Theoret.*, 41(21):212002, 2008. [doi:10.1088/1751-8113/41/21/212002] 4

[16] SHAHRZAD HADDADAN AND PETER WINKLER: Mixing of permutations by biased transpositions. *Theory Computing Sys.*, 63:1068–1088, 2018. Preliminary version in STACS'17. [doi:10.1007/s00224-018-9899-5] 3, 4, 22

[17] JAMES HERBERT HESTER AND DANIEL S. HIRSCHBERG: Self-organizing linear search. *ACM Computing Surveys (CSUR)*, 17(3):295–311, 1985. [doi:10.1145/5505.5507] 2

[18] MARK JERRUM, JUNG-BAE SON, PRASAD V. TETALI, AND ERIC VIGODA: Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.*, 14(4):1741–1765, 2004. [doi:10.1214/105051604000000639] 4, 5, 7, 10, 11, 12, 20, 21, 27

[19] DONALD KNUTH: *The Art of Computer Programming*. Volume 2: Seminumerical Algorithms. Addison-Wesley, 1969/1997. Link at ACM DL. 2

[20] DAVID A. LEVIN, YUVAL PERES, AND ELIZABETH L. WILMER: *Markov chains and mixing times*. American Mathematical Society, 2006. [doi:10.1090/mbk/058] 8, 13

[21] NEAL MADRAS AND DANA RANDALL: Markov chain decomposition for convergence rate analysis. *Ann. Appl. Probab.*, 12(2):581–606, 2002. [doi:10.1214/AOAP/1026915617] 4, 11

[22] YONG HUA MAO AND LIANG HUI XIA: Spectral gap for open Jackson networks. *Acta Math. Sinica, English Ser.*, 31(12):1879–1894, 2015. [doi:10.1007/S10114-015-4138-3] 4

[23] RUSSELL A. MARTIN AND DANA RANDALL: Sampling adsorbing staircase walks using a new Markov chain decomposition method. In *Proc. 41st FOCS*, pp. 492–502. IEEE Comp. Soc., 2000. [doi:10.1109/SFCS.2000.892137] 4, 7, 11, 12

[24] SARAH MIRACLE, DANA RANDALL, AND AMANDA PASCOE STREIB: Cluster algorithms for discrete models of colloids with bars. In *Proc. 8th Workshop on Analytic Algorithmics and Combinatorics (ANALCO'11)*, pp. 135–150. SIAM, 2011. [doi:10.1137/1.9781611973013.15] 4

[25] SARAH MIRACLE AND AMANDA PASCOE STREIB: Rapid mixing of k-class biased permutations. *SIAM J. Discr. Math.*, 38(1):702–725, 2024. Preliminary version in LATIN'18. [doi:10.1137/22M148063X] 3, 4, 6, 22, 23, 24, 25, 36, 37

[26] SARAH MIRACLE, AMANDA PASCOE STREIB, AND NOAH STREIB: Iterated decomposition of biased permutations via new bounds on the spectral gap of Markov chains. In *Proc. 24th Internat. Conf. on Randomization and Computation (RANDOM'20)*, pp. 3:1–21. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. [doi:10.4230/LIPICS.APPROX/RANDOM.2020.3] 1

[27] NATESH S. PILLAI AND AARON SMITH: Elementary bounds on mixing times for decomposable Markov chains. *Stoch. Proc. Appl.*, 127(9):3068–3109, 2017. [doi:10.1016/J.SPA.2017.02.002] 4, 11

[28] DANA RANDALL: Decomposition methods and sampling circuits in the Cartesian lattice. In *Proc. Internat. Symp. Math. Foundations of Comp. Sci. (MFCS'01)*, pp. 74–86. Springer, 2001. [doi:10.1007/3-540-44683-4_8] 4

[29] DANA RANDALL: Mixing. In *Proc. 44th FOCS*, pp. 4–15. IEEE Comp. Soc., 2003. [doi:10.1109/SFCS.2003.1238175] 35

[30] DANA RANDALL AND PRASAD TETALI: Analyzing Glauber dynamics by comparison of Markov chains. *J. Math. Phys.*, 41:1598–1615, 2000. Preliminary version in LATIN'98. [doi:10.1063/1.533199] 3, 21

[31] OMER REINGOLD, SALIL VADHAN, AND AVI WIGDERSON: Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Ann. Math.*, 155(1):157–187, 2002. Preliminary version in FOCS'00. [doi:10.2307/3062153] 12, 13

[32] RONALD RIVEST: On self-organizing sequential search heuristics. *Comm. ACM*, 19(2):63–67, 1976. [doi:10.1145/359997.360000] 2

[33] ALISTAIR SINCLAIR: *Algorithms for Random Generation and Counting: A Markov Chain Approach*. Progr. in Theoret. Comp. Sci. Birkhäuser, 1993. [doi:10.1007/978-1-4612-0323-0] 35

[34] SALIL VADHAN: Pseudorandomness. *Found. Trends Theor. Comp. Sci.*, 7(1–3):1–336, 2012. [doi:10.1561/0400000010] 12

[35] DAVID BRUCE WILSON: Mixing times of lozenge tiling and card shuffling Markov chains. *Ann. Appl. Probab.*, 1(1):274–325, 2004. [doi:10.1214/aoap/1075828054] 2, 37

AUTHORS

Sarah Miracle
Associate Professor
Computer and Information Science Department
University of St. Thomas
St. Paul, MN, USA
sarah.miracle@stthomas.edu

Amanda Pascoe Streib
Research Staff Member
Center for Computing Sciences
Bowie, MD, USA
ampasco@super.org

Noah Streib
Research Staff Member
Center for Computing Sciences
Bowie, MD, USA
nsstrei@super.org

ABOUT THE AUTHORS

SARAH MIRACLE is an Associate Professor in the Department of Computer and Information Science at the University of Saint Thomas in Minnesota. She received a Ph. D. from the Georgia Institute of Technology where she was advised by Dana Randall. Her research interests are in randomized algorithms, sampling, and Markov chains. In her spare time, she enjoys swimming, mountain biking, and skate skiing.

AMANDA PASCOE STREIB is an applied mathematician working at the Center for Computing Sciences. She received a Ph. D. in Algorithms, Combinatorics, and Optimization from the Georgia Institute of Technology and spent a year at the National Institute of Standards and Technology as a National Research Council postdoctoral fellow. She credits her Ph. D. advisor, Dana Randall, for teaching her how to craft an excellent talk and to distill complex mathematics into high-level ideas. Amanda's research interests are in the design and analysis of discrete algorithms, with an eye toward real-world applications. Outside of work, Amanda keeps busy and active with her two children.

NOAH STREIB is a researcher at the Center for Computing Sciences in Bowie, MD. After growing up in Ohio and attending Oberlin College, he received a Ph. D. in Algorithms, Combinatorics, and Optimization from the Georgia Institute of Technology in 2012 under the advisement of Tom Trotter. His research interests lie mainly in discrete mathematics and computer science. In his free time, Noah enjoys soccer, ultimate Frisbee, and playing with his kids.