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

Sarah Miracle
University of St. Thomas, St. Paul, MN, USA
https://www.stthomas.edu/cisc/faculty/sarah-miracle.html
sarah.miracle@stthomas.edu

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

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

Abstract

In this paper, we address a conjecture of Fill [Fill03] about the spectral gap of a nearest-neighbor transposition Markov chain $\mathcal{M}_{\text{nn}}$ over biased permutations of $[n]$. Suppose we are given a set 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 [25] 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 [23], leading to an exponentially small lower bound on the gap. We make progress by introducing a new complementary 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 the decomposition given in [25] of a related Markov chain $\mathcal{M}_{\text{pp}}$ over k-class particle systems is $1/n^2$-orthogonal when 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 at most a constant.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains

Keywords and phrases Markov chains, Permutations, Decomposition, Spectral Gap, Iterated Decomposition

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.3

Category RANDOM

Related Version A full version of the paper is available at https://arxiv.org/abs/1910.05184.

Acknowledgements We wish to thank Dana Randall for several useful discussions about the decomposition method.

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, 30], card shuffling [2, 32], 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 a set of input probabilities $\mathcal{P} = \{p_{i,j}\}$ for all $1 \leq i, j \leq n$ 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 $p_{j,i}$, independent 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 biased settings [3, 4, 8, 10, 32].

A central question is under what conditions the spectral gap of $\mathcal{M}_{\text{nn}}$ is an inverse polynomial in n, which implies a polynomial time 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 $p_{i,j} \geq 1/2$ for all $i < j$. It is easy to see that this condition is necessary (see, e.g. [4]); however, it is not sufficient, as demonstrated in [4].

In 2003, Fill [13, 14] introduced the following monotonicity conditions: $p_{i,j+1} \geq p_{i,j}$ for all $1 \leq i < j \leq n-1$ and $p_{i-1,j} \geq p_{i,j}$ for all $2 \leq i < j \leq n$. He conjectured that if $\mathcal{P}$ is positively biased and monotone, then the spectral gap of $\mathcal{M}_{\text{nn}}$ is an inverse polynomial in n, and moreover that the smallest spectral gap is attained at the uniform distribution.

Despite significant work on this subject, Fill's conjecture remains mostly open after more than 15 years. In the uniform case, there is a clever path coupling argument that achieves tight bounds on the mixing time [32]. Various papers [3, 4] have identified different classes of $\mathcal{P}$ for which $\mathcal{M}_{\text{nn}}$ 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 $\mathcal{M}_{\text{nn}}$ 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 distributions 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 one in the list; in this case, $p_{i,j} = w_i/(w_i + w_j)$. The Markov chain $\mathcal{M}_{\text{nn}}$ was termed a "gladiator chain" in this case by Haddadan and Winkler [16].

A partition of [7]: $C_1 = \{1, 2, 3\}$, $C_2 = \{4, 5\}$, $C_3 = \{6\}$, $C_4 = \{7\}$ $3765241 \rightarrow 1432121$

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, $\mathcal{M}_{\text{nn}}$ can be seen as having dual duties: *whisking*,

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

which uniformly mixes particles of the same type[2], and *sifting*, which mixes particles of different types in together [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 $\mathcal{M}_{\text{pp}}$, 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, 28] when evaluating the spectral gap of $\mathcal{M}_{\text{nn}}$.

The *bias* towards having a particle of type i ahead of a particle of type j is $\overline{q}_{i,j} = \overline{p}_{i,j}/\overline{p}_{j,i}$. We say that the bias is *bounded* if there exists a constant $q > 1$ such that $\overline{q}_{i,j} \geq q$ for all $i < j$. After a series of papers [16, 25], it was shown that the spectral gap of $\mathcal{M}_{\text{pp}}$ is at least $\Omega(n^{-2(k-1)})$ for positively biased, monotone, and bounded distributions. 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 monotonicity. Weak monotonicity in the setting of k-classes is defined as follows.

▶ **Definition 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 on the spectral gap of $\Omega(n^{-2(k-1)})$, 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, 12, 15, 16, 18, 22, 23, 24, 26, 27]. Suppose $\mathcal{M}$ is a finite, ergodic Markov chain that is reversible and has stationary distribution π.

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

Let $\Omega = \cup_{i=1}^{r} \Omega_i$ 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, 12, 15, 18, 22, 23, 26, 27]. 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 smaller 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, 12]; 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 [26], 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 $\mathcal{M}_{\text{pp}}$ 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

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

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

▶ **Theorem 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 5. There we also present several generalizations of Theorem 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. 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. However, we do not require 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 is very small even if $r(a,b)$ is far from 1 for 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 $\mathcal{M}_{\mathrm{pp}}$. We prove that this decomposition is $1/n^2$-orthogonal at each level of the decomposition. Thus, we may apply Theorem 2 iteratively k times to get a bound of $\Omega(n^{-2})$ on the spectral gap $\gamma(\mathcal{M}_{\mathrm{pp}})$. This bound is optimal up to constants. More formally, let $N^* = \max\left\{\frac{\log(6n^2)+\log((1+q)/(q-1))}{\log((1+q)/2)}, \frac{\log^2(14)}{\log^2(2q/(1+q))}\right\} =: C_q \log n$, and let c_i denote the number of particles of type i, for $1 \leq i \leq k$. We prove the following.

▶ **Theorem 3.** *If the probabilities $\mathcal{P}$ are weakly monotonic and bounded and $c_i \geq 2N^*$ for all $1 \leq i \leq k$, then the spectral gap γ of the chain $\mathcal{M}_{pp}$ satisfies $\gamma = \Omega(n^{-2})$.*

41311<u>2</u>413142

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.

The iterated decomposition works as follows. At the ith level of the decomposition, all particles of type less than i are in fixed positions, and particles of larger type are allowed to swap across these (this is the reason for allowing these non-adjacent transpositions); see Figure 2. This decomposition is designed to exploit the hypothesis that the movement of the particles of type i is nearly independent of the relative order of the particles of type bigger than i. The tool of ϵ-orthogonality allows us to make this intuition precise. We define the complementary restriction chains to contain the moves involving only particles of type bigger than i, and we define the restriction chains to contain moves between particles of type i and particles of larger type. We prove that this decomposition is $1/n^2$-orthogonal if the number of particles of type $i+1$ 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 many particles of type $i+1$ ensures that a typical configuration will not have a particle of type i after a particle of type at least $i+2$, 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, but not exponentially suppressed.

1223123112133413143444

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

We use Theorem 3 and comparison techniques to prove the following result for $\mathcal{M}_{\mathrm{nn}}$.

▶ **Theorem 4.** *If the probabilities $\mathcal{P}$ are weakly monotonic and form 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 bound on the spectral gap of $\mathcal{M}_{\text{nn}}$ 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 4 also leads to a bound of $O(n^9)$ on the mixing time of $\mathcal{M}_{\text{nn}}$ 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 2 and illustrate its use by applying it to the one-dimensional Ising model. In Section 4, we apply Theorem 2 to the biased permutation problem. In Section 5, 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. Appendix A provides even more detail concerning the proofs of the decomposition theorems, whereas the complete proofs appear in the full version of the paper.

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.

We will be interested in decomposing P into the part that performs restriction moves and the part that performs all other moves. 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 $P_1, P_2, \ldots, P_r$.

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 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 Complementary decomposition theorem

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 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 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 5.** *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 $\lambda = e^{-\beta}$, where $\beta > 0$ represents inverse temperature. We are interested in sampling from the Gibbs distribution given by $\pi(\sigma) = e^{-\beta H(\sigma)}/Z$, where the Hamiltonian $H(\sigma)$ is the number of edges whose endpoints have different spins and Z is the normalizing constant $\sum_{\sigma\in\Omega} e^{-\beta H(\sigma)}$, also called the *partition function*. (See [20] for background on the ferromagnetic Ising model.)

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

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

`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, $\mathcal{M}_{\text{gd}}$ 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 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 4.

Define $\lambda = e^{-\beta}$. 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 $Z_B = \sum_b w(b)$.

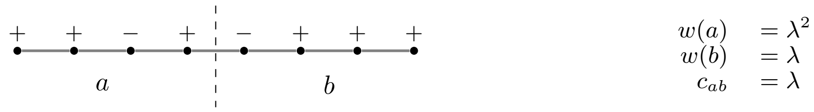


Figure 4 An example configuration of the one-dimensional Ising model.

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

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

Applying the complementary decomposition theorem (Theorem 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 Application to permutations

In this section, we apply the complementary decomposition theorem, Theorem 2, to the problem of sampling biased permutations. We give an overview of the proof of Theorem 4 which bounds the spectral gap of the following nearest-neighbor Markov chain over S_n, the permutations of $[n]$. The complete details are provided in the full version of the paper.

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 $q_{\sigma(i),\sigma(j)} = \frac{p_{\sigma(i),\sigma(j)}}{p_{\sigma(j),\sigma(i)}}$.

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 $1/2$. 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 $1 \leq i < j \leq k$. The constant q is called the *minimum bias*. We prove the following.

▶ **Theorem 4.** *If the probabilities $\mathcal{P}$ are weakly monotonic and form a bounded k-class with at least $2N^*$ particles in each class, then $\gamma(\mathcal{M}_{nn}) = \Omega(n^{-7})$.*

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. The complete analysis can be found in the full version of the paper. 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 $\mathcal{M}_{pp}$

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

- At time t, choose a position $1 \le i \le n$ and direction $d \in \{L, R\}$ uniformly at random.
- If $d = L$, find the largest j less than i with $\sigma^t(j) \ge \sigma^t(i)$ (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 $\sigma^t(j) \ge \sigma^t(i)$ (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 $\overline{q}_{\sigma(i),\sigma(j)} = \frac{\overline{p}_{\sigma(i),\sigma(j)}}{\overline{p}_{\sigma(j),\sigma(i)}}$.

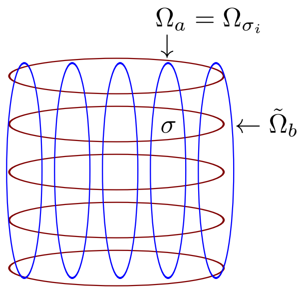
Recall the definition of weakly monotonic from Section 1. We will assume that property (2) holds. If instead property (3) holds, then as described in [25] we would modify $\mathcal{M}_{\text{pp}}$ to allow swaps between elements of different particle types across elements whose particle types are larger (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).

We prove the following bound on the spectral gap of $\mathcal{M}_{\text{pp}}$. Then, using comparison techniques [9, 28], we prove the bound on the spectral gap of $\mathcal{M}_{\text{nn}}$ given in Theorem 4. The details of this comparison argument can be found in the full version of the paper.

▶ **Theorem 3.** *If the probabilities $\mathcal{P}$ are weakly monotonic and bounded and $c_i \ge 2N^*$ for all $1 \le i \le k$, then the spectral gap γ of the chain $\mathcal{M}_{pp}$ satisfies $\gamma = \Omega(n^{-2})$.*

The proof of Theorem 3 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. For $i \ge 0$, let σ_i represent a fixed location of the particles of type $1, 2, \ldots, i$ (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 of type 3 or higher. We consider the chain $\mathcal{M}_{\sigma_i}$ whose state space Ω_{σ_i} is the set of all k-particle systems σ where the elements with type in $[i]$ are consistent with σ_i. The moves of $\mathcal{M}_{\sigma_i}$ are those moves from $\mathcal{M}_{\text{pp}}$ that do not involve an element of type at most 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 2.

Let A consist of all 2-particle systems with c_i particles of type i and $\sum_{j=i+1}^{k} c_j$ particles of type "_"; A is in bijection with staircase walks by mapping each i to a step right and each "_" to a step down, as in Figure 6. Let B consist of all $k-i$ particle systems with c_j particles of type j for all $i+1 \le j \le k$. Our goal is to show that the set of permutations σ consistent with σ_{i-1} on particles of type at most $i-1$ 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 of type less than i (see Figure 5), as those particles are in a fixed position for all of $\Omega_{\sigma_{i-1}}$.



σ_2 = 12_1_2_1_ _
σ_3 = 1231_2_13_
a = 3_ _3_
b = 454
σ = 1231425134

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

Next, define $b \in B$ to be the restriction of σ to elements of type bigger than 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.

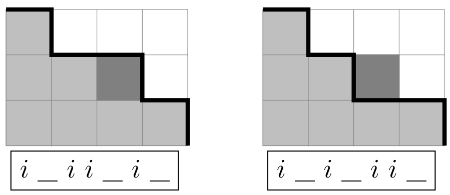


Figure 6 An exclusion process on staircase walks operates by adding or removing a square.

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 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 of type bigger than i, the complementary restriction chains can be seen as bounded exclusion processes on particles of type i with particles of type bigger than i. Bounded generalized biased exclusion processes operate on staircase walks as in Figure 6, where every square has a different bias but they are all bounded by some q. These processes were analyzed in [25]. We prove the following lemma in the full version of the paper.

▶ **Lemma 6.** *The complementary restrictions at each level of the induction are bounded generalized biased exclusion processes with spectral gap $\Omega(n^{-2})$.*

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, we show that the above decomposition is $1/n^2$-orthogonal by bounding $r(a, b)$. We define "good" a's to be those with fewer than $N^* = C_q \log n$ inversions and "good" b's to be those with fewer than N^* inversions involving particles of type $i+1$. Thus, as there are at least $2N^*$ particles of type $i+1$, then (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. 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 particles of type $i+1$ with particles of any higher type, 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 prove the following lemma in the full version of the paper.

▶ **Lemma 7.** *For any biased exclusion process with minimum bias constant $q > 1$, the total weight of staircase walks with area larger than N^* satisfies*

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

This allows us to prove the following.

▶ **Lemma 8.** *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 we have a $1/n^2$-orthogonal decomposition.*

We illustrate the main ideas of the proof of Lemma 8 here in the simplified $k = 3$ case (the details are deferred to the full version). 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

$$\pi(\sigma) = Z^{-1} \prod_{\substack{i<j:\\ \sigma(i)>\sigma(j)}} \overline{q}_{\sigma(i),\sigma(j)}, \tag{2}$$

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 8 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 $w(a, b) = 1$, 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.

To make the above statements precise is somewhat technical. Deferring details to the full version, we now give a bit more intuition. 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 $Z = \sum_{a,b} w(a)w(b)w(a, b) \approx Z_AZ_B$. 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, $r(a, b) = \frac{\pi(a,b)}{\pi(a)\pi(b)} \approx \frac{Z}{Z_AZ_B} \approx 1$. This allows us to show

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

We must also consider the case that either a or b is bad. In this case, we 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. We use the loose bound

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

The summation on the left is bounded by $\Pr(a \text{ bad}) + \Pr(b \text{ bad}) \leq 2\epsilon_1/(1-\epsilon_1)$. The summands in the summation on the right are products of conditional probabilities. We use the law of total probability to bound that summation by $\sum_{a \text{ bad}} w(a) + \sum_{b \text{ bad}} w(b)/(1-\epsilon_1) \leq \epsilon_1 + \epsilon_1/(1-\epsilon_1)$. Putting this all together, we have, for $\epsilon_1 \leq .225$,

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

This shows that the decomposition is $1/n^2$-orthogonal.

5 Generalizations and comparison with other theorems

In Sections 5.1 and 5.2, we present several generalizations of Theorem 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 5.3, we summarize the proofs, with more details given in Appendix A.

5.1 Generalized complementary decomposition theorems

Theorem 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 9.** *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 10 to prove Theorem 9, which in turn implies Theorem 2.

▶ **Theorem 10.** $\gamma(\mathcal{M}) \geq \min_{x \perp \sqrt{\pi}, \|x\|=1} \gamma_{\min}\|x^\perp\|^2 + \tilde{\gamma}_{\min}\|\tilde{x}^\perp\|^2$.

Here, $x^\perp$ and $\tilde{x}^\perp$ are orthogonal projections of a vector x onto the complement of the eigenspace of the top eigenvectors of certain matrices (defined in Section A.4) 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, $\|x^\perp\|^2 + \|\tilde{x}^\perp\|^2$ 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].

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

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 2.

5.2 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 [26] 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.

▶ **Theorem 11.** *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 17, in Section A.3. This bound allows us to rederive several known classical decomposition theorems.

▶ **Corollary 12.** *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 A.2). In Corollary 13 we show that Theorem 11 can be seen as a generalization of Theorem 1 of [18], except that it instead bounds the spectral gap.

▶ **Corollary 13.** $\gamma \geq \min\left\{\frac{\bar{\gamma}}{3}, \frac{\gamma_{\min}\bar{\gamma}}{3T+\bar{\gamma}}\right\}$.

In particular, if $T/\bar{\gamma}$ is a constant, then we get within a constant of the minimum gap as well. Theorem 11 produces slightly improved bounds over Corollary 13 when $T \approx \bar{\gamma} \ll \gamma_{\min}$.

5.3 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:

$$\lambda = \max_{x\perp v} \frac{\langle x, xA\rangle}{\|x\|^2} = \max_{x\perp v:\|x\|=1} \langle x, xA\rangle. \tag{3}$$

For a general reversible Markov chain with transition matrix P, we apply Equation 3 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. [29, 31]), and decompose the vector x into $x^{\perp} + x^{\parallel}$, where $x^{\parallel}$ 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 3, for any $x \perp v$, we need to bound

$$\langle x, xA\rangle = \langle x^{\perp}, x^{\perp}A\rangle + \langle x^{\parallel}, x^{\parallel}A\rangle + 2\langle x^{\perp}, x^{\parallel}A\rangle. \tag{4}$$

It is easy to bound $\langle x^{\perp}, x^{\perp}A\rangle$ and $\langle x^{\parallel}, x^{\parallel}A\rangle$ using ideas from other decomposition results [18, 23]. The term $\langle x^{\perp}, x^{\parallel}A\rangle$ 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 $\langle x^{\perp}, x^{\parallel}A\rangle \approx 0$ and $\gamma(\mathcal{M}) \approx \min\{\gamma(\mathcal{M}_1), \gamma(\mathcal{M}_2)\}$, or whether they are far from independent, in which case $\langle x^{\perp}, x^{\parallel}A\rangle$ is large and $\gamma(\mathcal{M}) = \Theta(\gamma_{\min}\bar{\gamma})$. The key to our decomposition proofs lies in our bounds on $\langle x^{\perp}, x^{\parallel}A\rangle$, which are different for our complementary decomposition theorems than they are for our classical decomposition theorems. More details are provided in the appendix.

References

1 D. Aldous. Random walk on finite groups and rapidly mixing Markov chains. *In Seminaire de Probabilites XVII*, pages 243–297, 1983.

2 D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2:295–313, 1992.

3 I. Benjamini, N. Berger, C. Hoffman, and E. Mossel. Mixing times of the biased card shuffling and the asymmetric exclusion process. *Trans. Amer. Math. Soc*, 357:3013–3029, 2005.

4 P. Bhakta, S. Miracle, D. Randall, and A. P. Streib. Mixing times of Markov chains for self-organizing lists and biased permutations. In *Proceedings of the 24th ACM/SIAM Symposium on Discrete Algorithms*, SODA '13, pages 1–15, 2013.

5 F. Chierichetti, A. Dasgupta, R. Kumar, and S. Lattanzi. On reconstructing a hidden permutation. *Leibniz International Proceedings in Informatics, LIPIcs*, 28:604–617, September 2014. doi:10.4230/LIPIcs.APPROX-RANDOM.2014.604.

6 C. Cooper, M. Dyer, A. Frieze, and R. Rue. Mixing properties of the Swendsen-Wang process on the complete graph and narrow grids. *Journal of Mathematical Physics*, 41(3):1499–1527, 2000. doi:10.1063/1.533194.

7 N. Destainville. Bounding spectral gaps of markov chains: a novel exact multi-decomposition technique. *Journal of Physics A: Mathematical and General*, 36(13):3647–3670, March 2003. doi:10.1088/0305-4470/36/13/301.

8 P. Diaconis and L. Saloff-Coste. Comparison techniques for random walks on finite groups. *The Annals of Applied Probability*, 21:2131–2156, 1993.

9 P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability*, 3:696–730, 1993.

10 P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 57:159–179, 1981.

11 J. Ding, E. Lubetzky, and Yuval Peres. Mixing time of critical Ising model on trees is polynomial in the height. *Communications in Mathematical Physics*, 295(1):161–207, April 2010.

12 P.L. Erdős, I. Miklós, and Z. Toroczkai. New classes of degree sequences with fast mixing swap Markov chain sampling. *Combinatorics Probability and Computing*, 27(2):186–207, 2018.

13 J. Fill. Background on the gap problem. *Unpublished manuscript*, 2003.

14 J. Fill. An interesting spectral gap problem. *Unpublished manuscript*, 2003.

15 R. Grone, K. Heinz Hoffmann, and P. Salamon. An interlacing theorem for reversible Markov chains. *Journal of Physics A: Mathematical and Theoretical*, 41(21):212002, May 2008. doi:10.1088/1751-8113/41/21/212002.

16 S. Haddadan and P. Winkler. Mixing of permutations by biased transposition. In *34th Symposium on Theoretical Aspects of Computer Science*, STACS '17, pages 41:1–41:13, 2017.

17 J. H. Hester and D. S. Hirschberg. Self-organizing linear search. *ACM Computing Surveys (CSUR)*, 17 Issue 3:295–311, 1985.

18 M. Jerrum, J. Son, P. Tetali, and E. Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.*, 14(4):1741–1765, November 2004. doi:10.1214/105051604000000639.

19 Donald K. *The Art of Computer Programming*, volume 2: Seminumerical Algorithms. Addison-Wesley, 1969.

20 D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

21 N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, pages 581–606, 2002.

22 Y. Mao and Liang H. Spectral gap for open Jackson networks. *Acta Mathematica Sinica, English Series*, 31:1879–1894, December 2015. doi:10.1007/s10114-015-4138-3.

23 R. Martin and D. Randall. Sampling adsorbing staircase walks using a new Markov chain decomposition method. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, pages 492–502, 2000.

24 S. Miracle, D. Randall, and A. P. Streib. Cluster algorithms for discrete models of colloids with bars. In *8th Workshop on Analytic Algorithmics and Combinatorics*, ANALCO, pages 135–150, 2011.

25 S. Miracle and A.P. Streib. Rapid mixing of k-class biased permutations. In *LATIN*, volume 10807 of *Lecture Notes in Computer Science*, pages 820–834. Springer, 2018.

26 N. Pillai and A. Smith. Elementary bounds on mixing times for decomposable Markov chains. *Stochastic Processes and their Applications*, 127(9):3068–3109, 2017. doi:10.1016/j.spa.2017.02.002.

27 D. Randall. Decomposition methods and sampling circuits in the Cartesian lattice. In *Proceedings of the 26th International Symposium on Mathematical Foundations of Computer Science*, MFCS '01, pages 74–86, London, UK, UK, 2001. Springer-Verlag. URL: http://dl.acm.org/citation.cfm?id=645730.668026.

28 D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics*, 41:1598–1615, 2000.

29 O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, pages 3–13, Washington, DC, USA, 2000. IEEE Computer Society. URL: http://dl.acm.org/citation.cfm?id=795666.796583.

30 R. Rivest. On self-organizing sequential search heuristics. *Commun. ACM*, 19(2):63–67, February 1976. doi:10.1145/359997.360000.

31 S. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012. doi:10.1561/0400000010.

32 D. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *The Annals of Applied Probability*, 1:274–325, 2004.

A Proofs of the decomposition theorems

In this section, we provide illustrate the main ideas of the proofs of the decomposition theorems. Some details are deferred to the full version of the paper. First, in Section A.1, we introduce some notation and terminology which will be useful for the proofs. Several parts are common to all of the proofs, so we present those parts in Section A.2. In Section A.3, we prove our classical decomposition theorem, Theorem 11. Finally, in Section A.4, we prove our complementary decomposition theorems, Theorems 2, 9, and 10.

A.1 Preliminaries

We first fix some notation and terminology. We write I_n to mean the $n \times n$ identity matrix. 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.

Define $\bar{P}$ to be the aggregated transition matrix on the state space $[r]$ defined by $\bar{P}(i,j) = \pi(\Omega_i)^{-1} \sum_{\sigma \in \Omega_i, \tau \in \Omega_j} \pi(\sigma) P(\sigma, \tau)$. Then $\bar{P}$ is the transition matrix of a reversible Markov chain $\bar{\mathcal{M}}$ with stationary distribution $\bar{\pi}$ defined by $\bar{\pi}(i) := \pi(\Omega_i)$. We call $\bar{\mathcal{M}}$ the *projection* chain.

It is useful to decompose the matrix P into the part that performs restriction moves and the part that performs all other moves. 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. Then $(\hat{P} + \tilde{P})(\sigma, \tau) = P(\sigma, \tau)$ 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 $P = \hat{P} + \tilde{P} - I_{|\Omega|}$.

Note that for any pair $\sigma, \tau \in \Omega$, the transitions (σ, τ) and (τ, σ) are either both nonzero in $\tilde{P}$ or both zero in $\tilde{P}$. Thus $\tilde{P}$ is itself the disjoint union of a set of ergodic, reversible Markov chains $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_{\tilde{r}}$ on state spaces $\tilde{\Omega}_1, \tilde{\Omega}_2, \ldots, \tilde{\Omega}_{\tilde{r}}$. We call these chains *complementary restrictions*.

In order to prove our decomposition results, we wish to apply Equation 3 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 3 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 $\tilde{A} \neq A(\tilde{P})$. This allows us to write $A = \hat{A} + \tilde{A} - I_{|\Omega|}$.

▶ **Proposition 14.** *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 $x^{\perp} = \sum_i a_i^{\perp} v_i$ 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 $v\tilde{A} = \sum_i a_i \mu_i v_i$.

A.2 Key ideas and lemmas for the proofs

We wish to apply Equation 3 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 $x^{\perp}$ as follows (note: this is similar to the vector decomposition used for the Zig Zag Product in [29]). 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 $x_i^{\perp}$, the part that is perpendicular to $\sqrt{\pi_i}$; recall that $\sqrt{\pi_i}$ is the top eigenvector of A_i. Finally, define $x^{\|}, x^{\perp} \in \mathbb{R}^{|\Omega|}$ by $x^{\|} = \sum_i e_i \otimes x_i^{\|}$ and $x^{\perp} = \sum_i e_i \otimes x_i^{\perp}$. Hence $x = \sum_i e_i \otimes x_i = x^{\|} + x^{\perp}$. Define $\tilde{x}^{\|}$ and $\tilde{x}^{\perp}$ analogously.

As described in Section 5.3, we will bound $\langle x, xA\rangle$ via Equation 4:

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

We need the following simple proposition.

▶ **Lemma 15.** *The following holds:* $x^{\|}A = x^{\|}\tilde{A}$.

Applying Lemma 15, Equation 4 becomes

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

For ease of notation, we define the following quantities:

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

Plugging these in, we have

$$\langle x, xA\rangle = \delta_4\|x^{\|}\|^2 + 2\langle x^{\perp}, x^{\|}\tilde{A}\rangle + (\delta_1 + \delta_2 - 1)\|x^{\perp}\|^2. \quad (5)$$

Bounding δ_1 and δ_4 is straightforward, and borrows many of the ideas from classical decomposition results. If $x^{\|}\tilde{A}$ were orthogonal to $x^{\perp}$, then doing so would be sufficient to proving a strong decomposition theorem. However, this is not true in general, so we must also bound $\langle x^{\perp}, x^{\|}\tilde{A}\rangle$. Our two types of theorems do so in different ways, which are presented in Sections A.3 and A.4.

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 16.** *With the above notation,*
1. $\delta_1 \le \lambda_{\max}$.
2. $\delta_4 \le \bar{\lambda}$.

A.3 Classical decomposition theorems

In this section, we will prove Theorem 17, which is a generalization of Theorem 11.

▶ **Theorem 17.** *Let* $\rho = \sqrt{(1-\delta_2)/\bar{\gamma}}$. *Then* $\gamma(\mathcal{M}) \ge \min_{p^2+q^2=1} \gamma_{\min}q^2 + \bar{\gamma}\,(q\rho - p)^2$.

With the technology developed in Section A.2, there is one critical piece remaining to prove Theorem 17, which is to bound the cross terms generated by applying the matrix $\tilde{A}$ to $x^{\|}$.

▶ **Lemma 18.** *With the above notation,* $|\langle x^\perp, x^\parallel \tilde{A}\rangle| \le \sqrt{(1-\delta_4)(1-\delta_2)}\|x^\parallel\|\|x^\perp\|$.

The proof appears in the full version.

Finally, we are ready to prove Theorem 17.

Proof of Theorem 17. Let $x \in \mathbb{R}^{|\Omega|}$ with $x \perp \sqrt{\pi}$ and $\|x\| = 1$. By Equation 3, $\gamma(\mathcal{M}) \ge 1 - \langle x, xA\rangle$. Applying Lemma 15 and the definitions of δ_1, δ_4, and δ_2, Equation 4 becomes

$$\begin{aligned}\langle x, xA\rangle &= \langle x^\parallel, x^\parallel \tilde{A}\rangle + 2\langle x^\perp, x^\parallel \tilde{A}\rangle + \langle x^\perp, x^\perp(\hat{A} + \tilde{A} - I_{|\Omega|})\rangle \\ &= \delta_4\|x^\parallel\|^2 + 2\langle x^\perp, x^\parallel \tilde{A}\rangle + (\delta_1 + \delta_2 - 1)\|x^\perp\|^2. \end{aligned} \tag{6}$$

Applying Lemma 18, we have

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

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

$$\gamma(\mathcal{M}) \ge \min_{x\perp\sqrt{\pi}:\|x\|=1} (1-\delta_1)\|x^\perp\|^2 + \left(\sqrt{1-\delta_2}\|x^\perp\| - \sqrt{1-\delta_4}\|x^\parallel\|\right)^2. \tag{7}$$

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

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

Using a bit of calculus one may show that the expression on the right is minimized when $(1-\delta_1)$ is minimized, when $(1-\delta_2)$ is maximized and when $(1-\delta_4)$ is minimized. By Lemma 16, $(1-\delta_1) \ge \gamma_{\min}$ and $(1-\delta_2) \ge \bar{\gamma}$. Therefore, we have

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

The statement of Theorem 17 is admittedly technical. However, from it we may derive several corollaries, as listed in Section 5. It is simple to show that Theorem 11 follows from Theorem 17 by noticing that $\delta_2 \ge 1 - 2T$.

We do not currently have a comparison between our Theorem 11 and Corollary 2 of [18], which requires a pointwise bound of π_i^j. However, see Remark 19 which shows that their result would not be sufficient for our application to permutations.

▶ **Remark 19.** It is worth pointing out that the decomposition of $\mathcal{M}_{\sigma_{i-1}}$ that we described in Section 4 does not satisfy the regularity conditions of [18] needed to obtain a better bound. For any $\upsilon \in \Omega_j$, define

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

We need to bound $\pi_j^{j'}(\upsilon)/\pi_j(\upsilon)$ for any j, j', and $\upsilon \in \Omega_j$. For example, let $\sigma_2 =$ 12_11111_2_1_ _ and $\sigma_3 =$ 12311111_231_ _. Notice that the two permutations $\upsilon_1 = 12311111423156$ and $\upsilon_2 = 12311111523146$ 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 υ_1) or 5 (in the case of υ_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, $\pi_j^{j'}(\upsilon)/\pi_j(\upsilon)$ cannot be uniformly bounded to within $1 \pm \eta$ unless η is exponentially large.

A.4 Complementary decomposition theorems

Next we use the technology developed in Section A.3 to prove Theorem 10. 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 $\tilde{\delta}_1 = \frac{\langle \tilde{x}^\perp, \tilde{x}^\perp \tilde{A}\rangle}{\|\tilde{x}^\perp\|^2}$. Now we can make an explicit statement about the gap of $\mathcal{M}$; notice the equality in Equation 8.

▶ **Theorem 10.**

$$\gamma(\mathcal{M}) = \min_{x \perp \sqrt{\pi}, \|x\|=1} (1-\delta_1)\|x^\perp\|^2 + (1-\tilde{\delta}_1)\|\tilde{x}^\perp\|^2. \tag{8}$$

In particular,

$$\gamma(\mathcal{M}) \geq \min_{x \perp \sqrt{\pi}, \|x\|=1} \gamma_{\min}\|x^\perp\|^2 + \tilde{\gamma}_{\min}\|\tilde{x}^\perp\|^2. \tag{9}$$

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

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

On the other hand, from Equation 6, we have

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

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

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

Applying Equation 3, we get Equation 8. To get Equation 9, we apply Lemma 16, which yields $1 - \delta_1 \geq 1 - \lambda_{\max} = \gamma_{\min}$. An analogous statement to Lemma 16 holds for $\tilde{\delta}_1$, and shows $1 - \tilde{\delta}_1 \geq \tilde{\gamma}_{\min}$. ◀

It remains to prove Theorem 9. By Theorem 10, if $\gamma_{\min}$ and $\tilde{\gamma}_{\min}$ are not too small, it suffices to show that $\|x^\perp\|^2$ and $\|\tilde{x}^\perp\|^2$ cannot both be small. To this end, we further decompose $x^\perp$ and $\tilde{x}^\parallel$ based on the eigenvectors of $\tilde{A}$. Define $S = \{i : \mu_i = 1\}$ and vectors $x_{11} = \sum_{i \in S} a_i^\parallel v_i$ and $x_{12} = \sum_{i \notin S} a_i^\parallel v_i$. Similarly, let $x_{21} = \sum_{i \in S} a_i^\perp v_i$ and $x_{22} = \sum_{i \notin S} a_i^\perp v_i$. Notice $\tilde{x}^\parallel = x_{11} + x_{21}$ and $\tilde{x}^\perp = x_{12} + x_{22}$, so that the vectors in each row (respectively, column) of the following table sum to the vector in its row (respectively, column) label.

	$\tilde{x}^\parallel$	$\tilde{x}^\perp$
$x^\parallel$	x_{11}	x_{12}
$x^\perp$	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 9 is that if $\|x_{11}\|$ is small, then $\|x^\perp\|^2 + \|\tilde{x}^\perp\|^2$ is large. The following lemma states that ϵ-orthogonality is sufficient to guarantee $\|x_{11}\|$ is small.

▶ **Lemma 20.** *Let ϵ be as defined in Equation 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,

$$\langle x^{\|}, \sqrt{\tilde{\pi}_j}\rangle = \sum_i \sum_{\sigma\in\Omega_i} x^{\|}(\sigma)\sqrt{\tilde{\pi}_j(\sigma)} = \sum_i \alpha_i \sum_{\sigma\in\Omega_i\cap\tilde{\Omega}_j} \sqrt{\pi_i(\sigma)\tilde{\pi}_j(\sigma)} = \sum_{i\in[r]} \alpha_i \frac{\pi(\Omega_i\cap\tilde{\Omega}_j)}{\sqrt{\pi(\Omega_i)\pi(\tilde{\Omega}_j)}}. \tag{10}$$

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

$$0 = \langle x^{\|}, \sqrt{\pi}\rangle = \sum_i \alpha_i \sum_{\sigma\in\Omega_i} \sqrt{\pi_i(\sigma)\pi(\sigma)} = \sum_i \alpha_i \sum_{\sigma\in\Omega_i} \frac{\sqrt{\pi(\sigma)}}{\sqrt{\pi(\Omega_i)}}\sqrt{\pi(\sigma)} = \sum_i \alpha_i\sqrt{\pi(\Omega_i)}, \tag{11}$$

and this final term is equal to $\sum_i \alpha_i\sqrt{\bar{\pi}_i} = \langle\alpha, \sqrt{\bar{\pi}}\rangle$. Multiplying Equation 11 by $\pi(\tilde{\Omega}_j)$ and subtracting it from Equation 10, 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

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

By the Cauchy-Schwartz 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,

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

To prove Theorem 9 from Theorem 10, we must show that if $\|x_{11}\|^2 \leq \epsilon$, then $\|x^{\perp}\|^2 + \|\tilde{x}^{\perp}\|^2 \geq (1-\sqrt{\epsilon})^2$. 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 $\|x^{\perp}\|^2 + \|\tilde{x}^{\perp}\|^2$ 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 $\|\tilde{x}^{\perp}\|^2 < \|x_{12}\|^2 + \|x_{22}\|^2$. The proof is deferred to the full version of this paper.