Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 24 (2022)

Pseudorandomness of Expander Random Walks for Symmetric Functions and Permutation Branching Programs*

Louis Golowich†
Harvard University
Cambridge, MA USA
lgolowich@berkeley.edu

Salil Vadhan‡
Harvard University
Cambridge, MA USA
salil_vadhan@harvard.edu
salil.seas.harvard.edu

June 25, 2022

Abstract

We study the pseudorandomness of random walks on expander graphs against tests computed by symmetric functions and permutation branching programs. These questions are motivated by applications of expander walks in the coding theory and derandomization literatures. A line of prior work has shown that random walks on expanders with second largest eigenvalue λ fool symmetric functions up to a $O(\lambda)$ error in total variation distance, but only for the case where the vertices are labeled with symbols from a binary alphabet, and with a suboptimal dependence on the bias of the labeling. We generalize these results to labelings with an arbitrary alphabet, and for the case of binary labelings we achieve an optimal dependence on the labeling bias. We extend our analysis to unify it with and strengthen the expander-walk Chernoff bound. We then show that expander walks fool permutation branching programs up to a $O(\lambda)$ error in ℓ_2-distance, and we prove that much stronger bounds hold for programs with a certain structure. We also prove lower bounds to show that our results are tight. To prove our results for symmetric functions, we analyze the Fourier coefficients of the relevant distributions using linear-algebraic techniques. Our analysis for permutation branching programs is likewise linear-algebraic in nature, but also makes use of the recently introduced singular-value approximation notion for matrices (Ahmadinejad et al. 2021).

---

*Extended abstract to appear in the Computational Complexity Conference (CCC '22).
†Supported by Harvard College Herchel Smith Fellowship.
‡Supported by NSF grant CCF-1763299 and a Simons Investigator Award.

ISSN 1433-8092

Contents

1 Introduction

Random walks on expander graphs have numerous applications in computer science due to their pseudorandom properties (see e.g. [HLW06] for a survey). Typically, an expander random walk is used to provide a randomness-efficient means for generating a sequence of vertices $v_0, \dots, v_{t-1}$. In a given application, this expander walk will be used to "fool" certain desired test functions f, in the sense that the distribution of $f(v_0, \dots, v_{t-1})$ is approximately the same whether the vertices $v_0, \dots, v_{t-1}$ are sampled from a random walk on an expander, or independently and uniformly at random (which is equivalent to using a random walk on a complete graph with self loops). In this paper, we prove tight bounds on the extent to which expander graph random walks fool certain functions f of interest, namely, symmetric functions as well as functions computable by permutation branching programs. These results improve on a recent line of work [GK21, CPTS21, CMP+21]. Our results also yield further implications, including a strengthening of the expander-walk Chernoff bound [Gil98, Hea08].

An expander graph is a graph that is sparse but well connected. In this paper we consider regular λ-spectral expanders, which are constant-degree graphs for which all nontrivial eigenvalues of the random walk matrix have absolute value at most λ. Intuitively, the spectrum of an expander graph approximates that of the complete graph, so an expander provides a sparsification of the complete graph. Random walks on expander graphs therefore provide a derandomized approximation for random walks on complete graphs. A major aim of this paper is to obtain tight bounds on the error in this approximation.

Many explicit constructions of λ-spectral expanders are known for arbitrarily small $\lambda > 0$ (e.g. [Mar73, LPS88, RVW02, BATS11]). Random walks on such expanders have many applications, such as in randomness-efficient error reduction, error-correcting codes, and small-space derandomization (see the surveys [HLW06, Vad12, Gur04]). Randomness-efficient error reduction uses the ability of expander random walks to fool threshold functions, while Ta-Shma's recent breakthrough construction of ϵ-balanced codes [TS17] uses their ability to fool the parity function. Meanwhile, work on small-space derandomization starting from [INW94] uses the ability of expander walks to fool branching programs. In this paper, we prove new bounds on the extent to which expander walks fool symmetric functions (which include the threshold and parity functions), as well as (permutation) branching programs.

Specifically, we strengthen and generalize a result of Cohen et al. [CMP+21], which shows that a random walk on a sequence of λ-spectral expanders fools symmetric functions up to a $O(\lambda)$ error in total variation distance. Our result extends the result of Cohen et al. [CMP+21] to labelings of the vertices by symbols from an arbitrary alphabet and, in the binary case, achieves the optimal dependence on the bias of the labeling; the Cohen et al. [CMP+21] result only applies to binary labelings, and has a suboptimal dependence on the labeling bias. We also unite this total variation bound with a tail bound, which yields a strengthening of the expander-walk Chernoff bound. We furthermore show that expander random walks fool width-w permutation branching programs up to a $O(\lambda)$ error in ℓ_2-distance and a $O(\sqrt{w} \cdot \lambda)$ error in total variation distance, which extends a result of [AKM+20, HPV21] to walks of length > 2, and also strengthens the $O(w^4 \cdot \sqrt{\lambda})$ total variation bound of Cohen et al. [CPTS21]. For programs possessing a certain structure, we prove much stronger bounds. We also present several lower bounds that show our upper bounds to be tight.

1.1 Problem overview

For a sequence $\mathcal{G} = (G_1, \ldots, G_{t-1})$ of graphs on a shared vertex set V, let $\mathrm{RW}^t_{\mathcal{G}}$ denote the random variable taking values in V^t that is given by taking a length-t random walk on V, where the ith step is taken in the graph G_i. If all $G_i = G$ then we write $\mathrm{RW}^t_{\mathcal{G}} = \mathrm{RW}^t_G$.

For some fixed integer $d \geq 2$, we are given a labeling $\mathrm{val} : V \to [d] = \{0, \ldots, d-1\}$, which we extend to act on sequences componentwise, that is, $\mathrm{val}(v_0, \ldots, v_{t-1}) = (\mathrm{val}(v_0), \ldots, \mathrm{val}(v_{t-1}))$. We let the tuple $p = (p_0, \ldots, p_{d-1}) \in [0,1]^d$ specify the weights of the labels, so that p_b equals the fraction of vertices with label $b \in [d]$.

In this paper, we study the distribution of $\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ for a sequence $\mathcal{G}$ of λ-spectral expanders. In particular, letting J denote the complete graph with self-loops, we will compare the distributions of $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and $f(\mathrm{val}(\mathrm{RW}^t_J))$ for certain test functions f on $[d]^t$. Specifically, we study functions f that are either symmetric or computable by a permutation branching program.

Let $\Sigma : [d]^{[t]} \to [t+1]^{[d]}$ be the histogram function, so that $(\Sigma a)_b = |\{i \in [t] : a_i = b\}|$ denotes the number of copies of b in the sequence a. All symmetric functions factor through Σ, so to study symmetric functions we restrict attention to Σ.

1.2 Contributions

This section describes our main results. The theorem statements below are informal; the reader is referred to the later sections for precise statements with explicit constants.

1.2.1 Symmetric functions

A major objective of this paper is to study the extent to which expander walks fool symmetric functions. In our notation, for a sequence $\mathcal{G}$ of λ-spectral expanders, we would like to bound the distance between the distributions of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_J)$ as a function of λ, regardless of the choice of $\mathcal{G}$. Rather than directly comparing these distributions, in the following theorem we bound the change in $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ when one of the graphs G_u in the sequence $\mathcal{G}$ is changed, for the case of $d = 2$ possible labels. We then apply a hybrid argument by changing the graphs in $\mathcal{G}$ to J one at a time.

Thus the consideration of arbitrary expander sequences $\mathcal{G}$ is inherent in our proof. Yet as a side benefit, we are able to show fine-grained bounds on the distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})$ when $\mathcal{G}$ and $\mathcal{G}'$ only differ at a few steps. Such bounds are used in a follow-up work [Gol22] to prove a new Berry-Esseen theorem for expander walks.

In a slight abuse of notation below, we let G both denote a graph and its random walk matrix. We use $\|\cdot\|$ to denote the spectral norm of a matrix.

Theorem 1 (Informal statement of Theorem 18)**.** *Fix positive integers $u < t$. Let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$ be sequences of regular $1/100$-spectral expanders on a shared vertex set V such that $G_i = G'_i$ for all $i \neq u$. Fix a labeling $\mathrm{val} : V \to [2]$ that assigns each label $b \in [2]$ to p_b-fraction of the vertices. Then for every $c \geq 0$,*

$$\sum_{j \in [t+1] : |j - p_1 t| \geq c} \left| \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}) = (t-j, j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j, j)\right] \right|$$
$$= O\left(\frac{\|G'_u - G_u\| \cdot e^{-c^2/8t}}{t} \right).$$

Theorem 1 bounds the change in the distribution of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ when the graph at a single step in $\mathcal{G}$ is changed. A key point is that the bound decays linearly in t. That is, the longer the walk, the less effect changing one of the graphs has. By changing all the graphs to the complete graph with self loops J one step at a time, we obtain the following corollary.

Corollary 2 (Informal statement of Corollary 19)**.** *For all positive integers t and all $0 \leq \lambda \leq 1/100$, let $\mathcal{G} = (G_i)_{1\leq i\leq t-1}$ be a sequence of regular λ-spectral expanders on a shared vertex set V with labeling $\mathrm{val} : V \to [2]$. Then for every $c \geq 0$,*

$$\sum_{j\in[t+1]:|j-p_1t|\geq c} \left|\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j,j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{J}) = (t-j,j)\right]\right|$$
$$= O(\lambda \cdot e^{-c^2/8t}).$$

The bounds in Theorem 1 and Corollary 2 both provide unified bounds for two different notions of distance, namely total variation distance and tail bounds. Specifically, when $c = 0$ then the results above bound total variation distance, while as c grows large they provide tail bounds, as p_1t is the expected value of $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1$.

Both the total variation and tail bounds above are novel, to the best of our knowledge. Our tails bounds can be viewed as strengthening the expander-walk Chernoff bound [Gil98, Hea08], and indeed our proof of Theorem 1 draws on similar techniques as used in Healy's [Hea08] proof of the expander-walk Chernoff bound. Recall that for a sequence $\mathcal{G}$ of λ-spectral expanders with λ bounded away from 1, the expander-walk Chernoff bound states that

$$\sum_{j\in[t+1]:|j-p_1t|\geq c} \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j,j)\right] = O(e^{-\Omega(c^2/t)}),$$

that is, the tails of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ decay approximately as quickly as the tails of the binomial distribution as $c^2/t \to \infty$. Corollary 2 shows the stronger statement that as $\lambda \to 0$, the tails of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ converge to the tails of the binomial distribution $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{J})$, even when $c^2/t = O(1)$.

The $c = 0$ case of Corollary 2 shows a $O(\lambda)$ bound on the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{J})$. Equivalently, this result shows that every symmetric function $f : \{0,1\}^t \to \{0,1\}$ satisfies $|\mathbb{E}[f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))] - \mathbb{E}[f(\mathrm{val}(\mathrm{RW}^t_{J}))]| = O(\lambda)$, that is, random walks on λ-spectral expanders $O(\lambda)$-fool symmetric functions. This bound improves upon a line of prior work [GK21, CPTS21, CMP+21]. Guruswami and Kumar [GK21] initiated this line of work by showing a $O(\lambda)$ bound on the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{G})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{J})$ for the special case where G is the 2-vertex sticky random walk (see Section 5.1). Cohen et al. [CPTS21] then showed a $O(\lambda(\log(1/\lambda))^{3/2})$ bound on this total variation distance for arbitrary expanders G with a balanced labeling, that is, when $p_0 = p_1 = 1/2$. A follow-up paper of Cohen et al. [CMP+21] generalized to arbitrary p, and improved the total variation distance bound to $O(\lambda/\sqrt{\min(p)})$, where $\min(p) = \min\{p_0, p_1\}$. In contrast, the $c = 0$ case of Corollary 2 strengthens this bound to $O(\lambda)$ regardless of p. Our results also allow for sequences $\mathcal{G}$ of λ-spectral expanders with different graphs at different steps, whereas the prior work [CPTS21, CMP+21] assumed that the graph was the same at each step.

Theorem 1, Corollary 2, and all of the prior work [GK21, CPTS21, CMP+21] assumes a binary labeling $\mathrm{val} : V \to \{0,1\}$ on the expander graph's vertices. Jalan and Moshkovitz [JM21] asked whether these results generalize to labelings $\mathrm{val} : V \to [d]$ for $d > 2$. We provide an affirmative answer to this question in the following results, which generalizing the total variation distance bounds in Theorem 1 and Corollary 2 to arbitrary $d \geq 2$. Below, we let $\min(p) = \min_{b\in[d]} p_b$.

Theorem 3 (Informal statement of Theorem 20)**.** *For every integer $d \geq 2$ and every distribution $p \in [0,1]^d$ over the labels $[d]$, there exists a constant $\lambda_0 = \lambda_0(d,p) > 0$ such that the following holds. For all positive integers $u < t$, let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$ be sequences of λ_0-spectral expanders on a shared vertex set V, such that for all $i \neq u$ we have $G_i = G'_i$. Let* $\mathrm{val}: V \to [d]$ *be any labeling that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Then*

$$d_{TV}\left(\Sigma\, \mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}), \Sigma\, \mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})\right) = O\left(\left(\frac{d}{\min(p)}\right)^{O(d)} \cdot \frac{\|G'_u - G_u\|}{t}\right).$$

Corollary 4 (Informal statement of Corollary 21)**.** *For all integers $t \geq 1$ and $d \geq 2$, let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ be a sequence of λ-spectral expanders on a shared vertex set V with labeling* $\mathrm{val}: V \to [d]$ *that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Then*

$$d_{TV}\left(\Sigma\, \mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}), \Sigma\, \mathrm{val}(\mathrm{RW}^t_{J})\right) = O\left(\left(\frac{d}{\min(p)}\right)^{O(d)} \cdot \lambda\right).$$

In the results above, it is helpful to think of d and p as fixed, so that Corollary 4 gives a $O(\lambda)$ bound on total variation distance. When $d = 2$, Theorem 1 and Corollary 2 with $c = 0$ show that the factor of $(d/\min(p))^{O(d)}$ in the bounds above can be removed. We suspect that this $(d/\min(p))^{O(d)}$ dependence for $d > 2$ is not tight, and we leave the determination of the optimal dependence on d and p as an open question.

To show that the $O(\lambda)$ upper bounds on total variation distance described above are tight, we present the following lower bound.

Theorem 5 (Informal statement of Theorem 47)**.** *For every $p = (p_0, p_1)$ and $0 < \lambda < 1$, there exists a sufficiently large $t_0 = t_0(p, \lambda) \in \mathbb{N}$ and a λ-spectral expander G with vertex labeling* $\mathrm{val}: V \to [2]$ *that has label weights given by p, such that for every $t \geq t_0$,*

$$d_{TV}(\Sigma\, \mathrm{val}(\mathrm{RW}^t_G), \Sigma\, \mathrm{val}(\mathrm{RW}^t_J)) = \Omega(\lambda).$$

Theorem 5 generalizes a similar result of Guruswami and Kumar [GK21] for the special case of $p_0 = p_1 = 1/2$, and indeed our proof method is similar to theirs. Cohen et al. [CMP$^+$21] showed a similar $\Omega(\lambda)$ lower bound for all t but only when $p_0 = p_1 = 1/2$. Their result is incomparable to ours, as Theorem 5 considers all p but only sufficiently large t.

1.2.2 Permutation branching programs

This section describes our main results on the extent to which expander walks fool permutation branching programs. The reader is referred to Section 4 for background on permutation branching programs.

We first present a bound that makes no assumptions on the structure of the program.

Theorem 6 (Informal statement of Theorem 32)**.** *For integers $t \geq 1$, $w \geq 2$, and $d \geq 2$, let G be a λ-spectral expander with $\lambda < .1$, and assign some vertex labeling* $\mathrm{val}: V \to [d]$*. Let $B : [d]^t \to [w]$ be computed by a permutation branching program $\mathcal{B}$ of length t, width w, and degree d. Then*

$$d_{\ell_2}(B(\mathrm{val}(\mathrm{RW}^t_G)), B(\mathrm{val}(\mathrm{RW}^t_J))) = O(\lambda).$$

Note that the bound in Theorem 6 has no dependence on the width w of the branching program, but only bounds ℓ_2 rather than total variation distance. Applying the Cauchy-Schwartz inequality to this ℓ_2-bound gives the total variation bound

$$d_{\mathrm{TV}}(B(\mathrm{val}(\mathrm{RW}_G^t)), B(\mathrm{val}(\mathrm{RW}_J^t))) = O(\sqrt{w} \cdot \lambda).$$

This bound improves upon the work of Cohen et al. [CPTS21], who showed a $O(w^4 \cdot \sqrt{\lambda})$ bound on $d_{\mathrm{TV}}(B(\mathrm{val}(\mathrm{RW}_G^t)), B(\mathrm{val}(\mathrm{RW}_J^t)))$ for the special case where $d = 2$ and $p_0 = p_1 = 1/2$.

Theorem 6 is closely related to the analysis of the Impagliazzo-Nisan-Wigderson [INW94] pseudorandom generator studied by Hoza et al. [HPV21], which also uses expander walks to fool permutation branching programs. Both Theorem 6 and the results of Hoza et al. [HPV21] are also proven using similar matrix approximation notions. However, Hoza et al. [HPV21] consider many length-2 expander walks, whereas we consider a single longer walk.

Although we show in Section 5.3 that Theorem 6 is tight in general, much stronger bounds hold for certain permutation branching programs. Theorem 38 in Section 4.2 presents a class of such permutation branching programs $\mathcal{B}$ for which $B(\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t))$ approaches a uniform distribution exponentially quickly. For illustrative purposes here, we simply present one implication of this result.

Theorem 7 (Informal statement of Corollary 41)**.** *For integers $t \geq 1$, $w \geq 2$, and $d \geq 2$, let $\mathcal{G}$ be a sequence of λ-spectral expanders on a shared vertex set V with labeling* $\mathrm{val} : V \to [d]$*. Let $B^t : [d]^t \to [w]$ denote the sum modulo w, that is $B^t(a) = \sum_{i \in [t]} a_i \pmod{w}$. Then there exists a constant $c = c(d, w, p, \lambda) < 1$ such that*

$$d_{TV}(B^t(\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t)), B^t(\mathrm{val}(\mathrm{RW}_J^t))) \leq \sqrt{w} \cdot c^t.$$

That is, expander walks fool the small modular functions B^t, which are naturally computed by permutation branching programs, up to an exponentially small error. This result can be viewed as a generalization of the previously known fact that expander walks fool the parity function up to an exponentially small error, as can be recovered by letting $w = 2$ and $d = 2$ Theorem 7. This fact that expander walks are good parity samplers played a pivotal role in Ta-Shma's breakthrough construction of almost optimal ϵ-balanced codes [TS17].

For arbitrary $w \geq 2$, Guruswami and Kumar [GK21] showed that the total variation distance between $B^t(\mathrm{val}(\mathrm{RW}_G^t))$ and $B^t(\mathrm{val}(\mathrm{RW}_J^t))$ is exponentially small in t when G is the 2-vertex sticky random walk (see Section 5.1). Theorem 7 generalizes this exponential decay bound to arbitrary expander walks.

1.3 Proof overview for symmetric functions

In this section, we outline the proof of Theorem 1, which contains many of the key technical insights in our paper. In particular, the proof of Theorem 3 follows the same general argument, so for the exposition in this section we focus on Theorem 1. All of the proof details can be found in Section 3.1 and Appendix B.

As in Theorem 1, for some $u < t$ let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$ be sequences of $1/100$-spectral expanders that agree at all positions $i \neq u$, and again fix a vertex labeling $\mathrm{val} : V \to [2]$. Define $g \in [-1, 1]^{[t+1]} \subseteq [-1, 1]^{\mathbb{Z}}$ to be the difference between the probability mass functions of $(\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}'}^t))_1$ and $(\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t))_1$, that is,

$$g_j = \Pr\big[\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}'}^t) = (t - j, j)\big] - \Pr\big[\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t) = (t - j, j)\big].$$

In this notation, the $c = 0$ case of Theorem 1 states that g has ℓ_1-norm $\|g\|_1 = O(\|G'_u - G_u\|/t)$, which is bounded by $O(\lambda/t)$ if G'_u and G_u are λ-spectral expanders.

We first show that the ℓ_2-norm of g satisfies

$$\|g\| = O\left(\frac{\|G'_u - G_u\|}{t} \cdot \frac{1}{(p_0 p_1 t)^{1/4}}\right). \tag{1}$$

The proof of this bound is sketched below in Section 1.3.1. We will then explain in Section 1.3.2 how to go from this ℓ_2-bound to the desired ℓ_1-bound. We compare our techniques to those of prior work in Section 1.3.3, and in particular we draw connections with Healy's proof of the expander-walk Chernoff bound [Hea08].

1.3.1 Bounding the ℓ_2-distance $\|g\|$

In this section, we sketch the proof of the ℓ_2-bound (1) (which is the $r = 0$ case of Theorem 22). Because the Fourier transform preserves ℓ_2-norms, we will bound the ℓ_2-norm $\|\hat{g}\| = \|g\|$ of the Fourier transform $\hat{g}$ of g. Recall that here the Fourier transform is given by $\hat{g}(\theta) = \sum_{j \in \mathbb{Z}} e^{-i\theta j} g_j$, and has ℓ_2-norm $\|\hat{g}\| = \sqrt{\int_{\theta=-\pi}^{\pi} |\hat{g}(\theta)|^2 d\theta / 2\pi}$.

To motivate this shift to the Fourier basis, recall that the Fourier transform interchanges convolution and multiplication, so that addition of independent random variables translates to multiplication of the Fourier transforms of their probability density functions (i.e. multiplication of their *characteristic functions*). Such products can be easier to analyze than convolutions, so the Fourier transform is a natural tool for analyzing sums of independent random variables, as is exemplified in proofs of the central limit theorem. Theorem 1 and Corollary 2 intuitively show that the expander walk distribution $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1$ is close to the sum of independent variables, so it is also natural to analyze this distribution with the Fourier transform.

Whereas we apply the Fourier transform over the group $\mathbb{Z}$ to the random variable $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ (which is distributed over $\mathbb{Z}$), the prior work of Cohen et al. [CPTS21] and Cohen et al. [CMP+21] applied the Fourier transform over the group $(\mathbb{Z}/2)^t$ to the random variable $\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ (which is distributed over $\{0,1\}^t \cong (\mathbb{Z}/2)^t$). As described above, our approach seems well suited for symmetric functions, and it generalizes naturally to give Theorem 3 and Corollary 4 for alphabet sizes $d > 2$. In contrast, Cohen et al. [CPTS21] only consider $d = 2$, but they are able to apply their techniques to other classes of functions such as bounded-depth circuits, which we do not consider. More comparisons to prior techniques are provided in Section 1.3.3.

To begin, we express $\hat{g}(\theta)$ linear-algebraically. Specifically, let $\vec{1} = (1/\sqrt{|V|}, \ldots, 1/\sqrt{|V|})$ denote the uniform unit vector, and define the diagonal matrix $P_\theta = \mathrm{diag}(x_\theta) \in \mathbb{C}^{V \times V}$, where $x_\theta \in \mathbb{C}^V$ is the vector with $(x_\theta)_v = e^{-i\theta(\mathrm{val}(v) - p_1)}$. Then it can be verified that

$$e^{i\theta p_1 t} \cdot \hat{g}(\theta) = \vec{1}^\top \left(\prod_{i=u+1}^{t} G_i P_\theta\right)(G'_u - G_u)\left(\prod_{i=0}^{u-1} P_\theta G_i\right)\vec{1},$$

where the products above multiply from right-to-left, and we take $G_0 = G_t = J$. This equality can be seen by expanding the right hand side above as a sum over all length-t walks $v_0, \ldots, v_{t-1}$ on V (see Lemma 26). Therefore because $G'_u - G_u$ annihilates $\vec{1}$ from both sides, we have

$$|\hat{g}(\theta)| \leq \left\|\left(\vec{1}^\top \left(\prod_{i=u+1}^{t} G_i P_\theta\right)\right)^\perp\right\| \cdot \|G'_u - G_u\| \cdot \left\|\left(\left(\prod_{i=0}^{u-1} P_\theta G_i\right)\vec{1}\right)^\perp\right\|, \tag{2}$$

where the notation $x^\perp$ denotes the projection of a vector x onto the orthogonal complement of $\vec{1}$. We will also use $x^\parallel$ to denote the projection of x onto $\vec{1}$.

We bound the rightmost factor above by induction on u. Splitting off a factor of $P_\theta G_{u-1}$ gives

$$\begin{aligned}
&\left\| \left(\left(\prod_{i=0}^{u-1} P_\theta G_i \right) \vec{1} \right)^\perp \right\| \\
&\quad \le \|(P_\theta \vec{1})^\perp\| \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta G_i \right) \vec{1} \right)^\parallel \right\| + \|P_\theta\| \cdot \lambda(G_{u-1}) \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta G_i \right) \vec{1} \right)^\perp \right\| \qquad (3) \\
&\quad \le \|(P_\theta \vec{1})^\perp\| \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta G_i \right) \vec{1} \right)^\parallel \right\| + \frac{1}{100} \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta G_i \right) \vec{1} \right)^\perp \right\|,
\end{aligned}$$

where the last inequality follows because $\|P_\theta\| = 1$ and G_{u-1} is a 1/100-spectral expander. Thus if we can bound the first term on the right hand side of (3) by some $B(u)$ that decays less rapidly than 100^{-u} (i.e. $B(u) = \Theta(\beta^{-u})$ for $\beta < 100$), we can inductively bound the left hand side by $B(u) + B(u-1)/100 + B(u-2)/100^2 + \cdots = O(B(u))$. Specifically, we will show this bound for $B(u) = \Theta(\sqrt{p_0 p_1} \cdot \theta \cdot e^{-\Omega(p_0 p_1 (u-1)\theta^2)})$. Intuitively, it suffices to bound what happens to the component parallel to $\vec{1}$, because the component orthogonal to $\vec{1}$ is shrunk by a factor of 100 with each application of $P_\theta G_i$.

Letting $F = J + (1/10)(I - J)$ be the matrix that preserves $\vec{1}$ and scales its orthogonal complement by 1/10, then because by assumption all $i \ne u$ have $\lambda(G_i) \le 1/100$, it follows that $\|F^{-1} G_i F^{-1}\| \le 1$. Thus

$$\left\| \left(\left(\prod_{i=0}^{u-2} P_\theta G_i \right) \vec{1} \right)^\parallel \right\| = \left\| \vec{1}^\top \left(\prod_{i=0}^{u-2} F P_\theta F \cdot F^{-1} G_i F^{-1} \right) \vec{1} \right\| \le \|F P_\theta F\|^{u-1}.$$

Next via some technical calculations (Lemma 27), we show that for all $-\pi < \theta \le \pi$,

$$\|(P_\theta \vec{1})^\perp\| = \frac{\|x_\theta^\perp\|}{\sqrt{|V|}} = \Theta(\sqrt{p_0 p_1} \cdot \theta). \qquad (4)$$

For intuition, observe that if $p_0 p_1$ or θ equals 0, then all entries of x_θ are the same, so $x_\theta^\perp = 0$. Using (4), we also deduce (Lemma 28) that

$$\|F P_\theta F\| \le 1 - \Omega((\|(P_\theta \vec{1})^\perp\|^2) = e^{-\Omega(p_0 p_1 \theta^2)}.$$

Here for intuition, as F is a 1/10-spectral expander, we should expect $\|F P_\theta F\|$ to be close to $\|J P_\theta J\| = \|(P_\theta \vec{1})^\parallel\| = \sqrt{1 - \|(P_\theta \vec{1})^\perp\|^2} = 1 - \Omega(\|(P_\theta \vec{1})^\perp\|^2)$. Thus (3) becomes

$$\left\| \left(\left(\prod_{i=0}^{u-1} P_\theta G_i \right) \vec{1} \right)^\perp \right\| \le O\left(\sqrt{p_0 p_1} \cdot \theta \cdot e^{-\Omega(p_0 p_1 (u-1)\theta^2)} \right) + \frac{1}{100} \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta G_i \right) \vec{1} \right)^\perp \right\|.$$

Recursively applying this inequality to bound the last term on its right hand side then gives

$$\left\|\left(\left(\prod_{i=0}^{u-1} P_\theta G_i\right)\vec{1}\right)^\perp\right\| = O\left(\sqrt{p_0p_1}\cdot\theta\cdot e^{-\Omega(p_0p_1u\cdot\theta^2)}\right).$$

We now apply the above bound on $\|((\prod_{i=0}^{u-1} P_\theta G_i)\vec{1})^\perp\|$, along with an analogous bound on $\|(\vec{1}^\top(\prod_{i=u+1}^{t} G_iP_\theta))^\perp\|$, in (2) to give

$$|\hat{g}(\theta)| = O\left(p_0p_1\cdot\theta^2\cdot e^{-\Omega(p_0p_1t\cdot\theta^2)}\cdot\|G'_u - G_u\|\right).$$

We then obtain the desired ℓ_2-bound (1) by squaring and integrating this bound with the substitution $q = c\sqrt{p_0p_1t}\cdot\theta$ for a sufficiently small constant $c > 0$:

$$\begin{aligned}\|g\| = \|\hat{g}\| &= O\left(p_0p_1\cdot\|G'_u - G_u\|\cdot\sqrt{\int_{-\pi}^{\pi}\theta^4 e^{-\Omega(p_0p_1t\cdot\theta^2)}\frac{d\theta}{2\pi}}\right)\\ &= O\left(\frac{\|G'_u - G_u\|}{t\cdot(p_0p_1t)^{1/4}}\cdot\sqrt{\int_{-\infty}^{\infty} q^4e^{-q^2}\,dq}\right)\\ &= O\left(\frac{\|G'_u - G_u\|}{t\cdot(p_0p_1t)^{1/4}}\right).\end{aligned}$$

1.3.2 Going from an ℓ_2 to ℓ_1 bound

In this section, we show how to extend the techniques for bounding $\|g\|$ described above to bound $\|g\|_1$, and more generally to prove Theorem 1.

First observe that by the expander-walk Chernoff bound, $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})$ are mostly supported in an interval of length $\ell \approx O(\sqrt{t})$ about their mean. Applying the Cauchy-Schwartz inequality to (1) on this interval (which costs a factor of $\sqrt{\ell} \approx O(t^{1/4})$ to convert from ℓ_2 to ℓ_1), and the expander-walk Chernoff bound on the tails lying outside of the interval, yields a total variation bound of

$$\|g\|_1 = O\left(\frac{\|G'_u - G_u\|}{t}\cdot\left(\frac{\log(\|G'_u - G_u\|/t)}{p_0p_1}\right)^{1/4}\right).$$

However, the above ℓ_1-bound does not help us prove Theorem 1 when c^2/t is large. Furthermore, even to prove the $c = 0$ case Theorem 1, we need to remove the factor $(\log(\|G'_u - G_u\|/t)/p_0p_1)^{1/4}$ from the bound above.

To obtain these improvements, we first generalize (1) to bound the ℓ_2-norm of the vector $g^{(sr)} = (e^{sr(j-p_1t)}g_j)_{j\in[t+1]}$ for $s = \pm1$ and various values of $r \geq 0$. The proof of this bound on $\|g^{(sr)}\|$ for general r (Theorem 22) simply generalizes the argument presented in Section 1.3.1. The special case $r = 0$ recovers $g = g^{(0)}$, while when $r > 0$ then the sum of the elements of $g^{(sr)}$ equals the difference between the moment generating functions $\mathbb{E}[e^{sr((\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1-p_1t)}]$ and $\mathbb{E}[e^{sr((\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))_1-p_1t)}]$ that are used in the proofs of Chernoff bounds.

We then partition $[t+1] \subseteq \mathbb{Z}$ into intervals of length approximately $\sqrt{p_0p_1t}$, and we bound the ℓ_1-norm of g restricted to each interval by applying the Cauchy-Schwartz inequality with our

ℓ_2-bound on $g^{(sr)}$ for appropriately chosen s, r. Summing these bounds over all intervals lying at least some distance c from $p_1 t$ yields Theorem 1.

Intuitively, as $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})$ have standard deviation $\Theta(\sqrt{p_0 p_1 t})$, we would expect these distributions to be somewhat evenly distributed across an interval of length $\sqrt{p_0 p_1 t}$. This is the regime where Cauchy-Schwartz is tight. Appropriately choosing s, r allows us to "isolate" a given length-$\sqrt{p_0 p_1 t}$ interval, by ensuring that the components of $g^{(sr)}$ in that interval dominate components outside that interval.

1.3.3 Comparison with techniques in prior work

Our techniques described above to prove Theorem 1 are closely related to Healy's [Hea08] proof of the expander-walk Chernoff bound. In some sense, Healy's proof [Hea08] makes up "half" of our proof: Healy's proof bounds the moment-generating function $\mathbb{E}[e^{sr((\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1 - p_1 t)}]$, but does not bound the characteristic function $\mathbb{E}[e^{-i\theta((\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1 - p_1 t)}]$ as described in Section 1.3.1 (as the Fourier coefficient $\hat{g}(\theta)$ by definition equals the difference between the characteristic functions of $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1$ and $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))_1$). Intuitively, our proof combines the moment generating and characteristic function bounds, as in order to bound $\|g^{(sr)}\|$, we bound the difference $e^{i\theta p_1 t} \cdot \hat{g}^{(sr)}(\theta)$ between the generating functions $\mathbb{E}[e^{(sr - i\theta)((\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1 - p_1 t)}]$ and $\mathbb{E}[e^{(sr - i\theta)((\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))_1 - p_1 t)}]$.

Although Cohen et al. [CPTS21] and Cohen et al. [CMP+21] also studied the extent to which expander walks fool symmetric functions, their proofs are less similar to ours. Most notably, both of these papers use Fourier analysis over the group $(\mathbb{Z}/2\mathbb{Z})^t$ by viewing $\mathrm{val}(\mathrm{RW}^t_G)$ as a distribution on $(\mathbb{Z}/2\mathbb{Z})^t$. In contrast, we use Fourier analysis over $\mathbb{Z}^{d-1}$ by viewing $\Sigma\,\mathrm{val}(\mathrm{RW}^t_G)$ as a distribution on $\mathbb{Z}^d$ (or $\mathbb{Z}^{d-1}$, if we drop the first component). This explains why our results generalize more naturally to the case $d > 2$, which is not considered in [CPTS21, CMP+21]. We could also do our analysis using discrete Fourier analysis over $(\mathbb{Z}/m)^{d-1}$ instead, for any $m \geq t$, but then the modulus m is superfluous (as it is cleaner to avoid modular reduction) and only makes the notation more cumbersome.

1.4 Proof overview for permutation branching programs

In this section, we outline the proof of Theorem 6, which uses singular-value approximations as described below. We do not outline the proof of our other results for permutation branching programs, specifically Theorem 7, and instead refer the reader to Section 4.2, as this latter result uses techniques somwhat similar to those described above in Section 1.3.

For the purpose of this paper, a branching program $\mathcal{B}$ of length t, width w, and degree d is a program with w possible states of memory, that reads an input in $[d]^t$ and outputs a value in $[w]$. Thus each branching program $\mathcal{B}$ computes a specific function $B : [d]^t \to [w]$. More formally, $\mathcal{B}$ begins with an initial state $0 \in [w]$, and sequentially reads in t inputs $a_0, \dots, a_{t-1} \in [d]$. Upon receiving each input a_i, then $\mathcal{B}$ updates its state according to a specified function $B_i(a_i) : [w] \to [w]$. We restrict attention to permutation branching programs, meaning that all functions $B_i(a_i)$ are permutations.

1.4.1 Proof outline

We now describe the proof of Theorem 6. As in the theorem statement, for arbitrary integers $t \geq 1$, $w \geq 2$, and $d \geq 2$, let $\mathcal{B}$ be a permutation branching program of length t, width w, and degree

d that computes some function $B : [d]^t \to [w]$. Let G be a λ-spectral expander with $\lambda < .1$, and assign some vertex labeling $\mathrm{val} : G \to [d]$. We again let $g \in [-1,1]^{[w]}$ denote the difference between the distributions $B(\mathrm{val}(\mathrm{RW}^t_G))$ and $B(\mathrm{val}(\mathrm{RW}^t_J))$ of interest, that is,

$$g_j = \Pr\big[B(\mathrm{val}(\mathrm{RW}^t_G)) = j\big] - \Pr\big[B(\mathrm{val}(\mathrm{RW}^t_J)) = j\big].$$

In this notation, Theorem 6 states that $\|g\| = O(\lambda)$.

As in the proof of Theorem 1, we begin by expressing g linear-algebraically. Let $\tilde{P}$ be the operator on the vector space $\mathbb{R}^V \otimes \mathbb{R}^t \otimes \mathbb{R}^w$ given by

$$\tilde{P} = \sum_{v \in V, i \in [t]} \delta_v \delta_v^\top \otimes \delta_{i+1} \delta_i^\top \otimes B_i(\mathrm{val}(v)),$$

where $i+1$ is taken $\pmod t$ above, and by abuse of notation $B_i(\mathrm{val}(v)) \in \mathbb{R}^{w \times w}$ refers to the permutation matrix associated to the permutation $B_i(\mathrm{val}(v)) : [w] \to [w]$. Also for $W = G$ or J, let $\tilde{W} = W \otimes I \otimes I$. Then for every $j \in [w]$,

$$g = (\vec{1} \otimes \delta_0 \otimes I)^\top ((\tilde{G}\tilde{P})^t - (\tilde{J}\tilde{P})^t)(\vec{1} \otimes \delta_0 \otimes \delta_0). \tag{5}$$

This equality can again be seen by expanding the right hand side above as a sum over all length-t walks on V (see Lemma 36).

We will bound the right hand side using singular-value approximations [Ahm20, APSV21]. A matrix $W' \in \mathbb{C}^{N \times N}$ is a singular-value ϵ-approximation of another matrix $W \in \mathbb{C}^{N \times N}$, written $W' \overset{\mathrm{sv}}{\approx}_\epsilon W$, if for all $x, y \in \mathbb{C}^N$,

$$|x^*(W' - W)y| \le \frac{\epsilon}{2}(\|x\|^2 + \|y\|^2 - \|x^* W\|^2 - \|Wy\|^2),$$

where x^* denotes the conjugate transpose of x. The following properties were shown by Ahmadinejad et al. [Ahm20, APSV21] (see Section 2.3 for details):

1. $\tilde{G} \overset{\mathrm{sv}}{\approx}_\lambda \tilde{J}$.

Assume that $W' \overset{\mathrm{sv}}{\approx}_\epsilon W$. Then:

2. For every matrix X with spectral norm $\|X\| \le 1$, then $W'X \overset{\mathrm{sv}}{\approx}_\epsilon WX$.

3. If $\epsilon < .1$, then $(W')^t \overset{\mathrm{sv}}{\approx}_{\epsilon + 5\epsilon^2} W^t$. (Importantly, the bound $\epsilon + O(\epsilon^2)$ does not grow with t.)

4. $\|W' - W\| \le \epsilon$.

We now bound the right hand side of (5) using singular value approximations. Because by definition $\|\tilde{P}\| = 1$, property 1 and property 2 above imply that $\tilde{G}\tilde{P} \overset{\mathrm{sv}}{\approx}_\lambda \tilde{J}\tilde{P}$. Then property 3 implies that $(\tilde{G}\tilde{P})^t \overset{\mathrm{sv}}{\approx}_{\lambda + 5\lambda^2} (\tilde{J}\tilde{P})^t$, and property 4 then gives that $\|(\tilde{G}\tilde{P})^t - (\tilde{J}\tilde{P})^t\| \le \lambda + 5\lambda^2$, so $\|g\| \le \lambda + 5\lambda^2 = O(\lambda)$.

1.4.2 Comparison with techniques in prior work

The proof of Theorem 6 described above is closely related to the analysis of the Impagliazzo-Nisan-Wigderson (INW) [INW94] pseudorandom generator in Hoza et al. [HPV21]. Hoza et al. [HPV21] use unit-circle approximations [AKM+20] to show that length-2 walks on λ-spectral expanders fool permutation branching programs up to a $O(\lambda)$ ℓ_2-error; the INW generator they study recursively applies many such length-2 walks. We generalize this $O(\lambda)$ bound to walks of arbitrary length, and simplify the analysis by replacing the unit-circle approximations with singular-value approximations. We obtain these improvements because the unit-circle approximations, though similar in nature to singular-value approximations, do not satisfy property 2 described above. Although our results do not directly translate to an improved pseudorandom generator, it is an interesting question whether longer walks could somehow be used to improve the seed length.

As described in Section 1.2.2, Theorem 6 implies a $O(\sqrt{w} \cdot \lambda)$ total variation distance bound, which improves upon the $O(w^4 \cdot \sqrt{\lambda})$ total variation bound of Cohen et al. [CPTS21]. However, Cohen et al. [CPTS21] prove their result using bounds on the Fourier tails over $(\mathbb{Z}/2\mathbb{Z})^t$ of permutation branching programs with alphabet size $d = 2$, differing significantly from our proof using singular-value approximations, which generalizes readily to $d > 2$.

2 Preliminaries

This section describes the basic notation and problem setup that is used throughout the paper.

2.1 Notation

For $N \in \mathbb{N}$, let $[N] = \{0, \dots, N-1\}$. For the field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, let $\vec{1}_N = \begin{pmatrix} 1/\sqrt{N} & \dots & 1/\sqrt{N} \end{pmatrix}^\top \in \mathbb{F}^N$ denote the normalization of the all 1's vector. When the dimension N is clear from context, this vector will simply be denoted $\vec{1}$. The vector $\delta_i \in \mathbb{F}^N$ denotes the ith standard basis vector, which has a 1 in the ith component and 0s elsewhere. For a matrix $A \in \mathbb{F}^{N \times N}$, the spectral norm of A is defined to be $\|A\| = \max_{x \in \mathbb{F}^N \setminus \{0\}} \|Ax\|/\|x\|$. For $A \in \mathbb{C}^{N \times M}$, the conjugate transpose is denoted $A^* = \overline{A}^\top$.

A matrix $W \in [0,1]^{N \times N}$ is a random walk matrix on N vertices if the columns of W sum to 1, so that $W_{j,i}$ denotes the transition probability from vertex i to vertex j. The $N \times N$ identity matrix is denoted I, while the matrix $J = \vec{1}\vec{1}^\top$ refers to the $N \times N$ matrix with all entries $1/N$, where N will be understood from context. Thus J is the random walk matrix for the complete graph with self-loops. The random walk matrix for the directed cycle on N vertices is denoted C_N, so that $C_N(y_1, y_2, \dots, y_N) = (y_N, y_1, y_2, \dots, y_{N-1})$.

2.2 Distance between probability distributions

We will use total variation, ℓ_1, ℓ_2, and Kolmogorov distances between probability distributions, defined below.

Definition 8. Let Ω be a sample space with σ-algebra $\mathcal{A}$. The **total variation distance** between probability measures $\mu_1, \mu_2 : \mathcal{F} \to \mathbb{R}$ is

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = \sup_{A \in \mathcal{A}} |\mu_1(A) - \mu_2(A)|$$

Definition 9. Let Ω be a sample space with σ-algebra $\mathcal{A}$. The ℓ_1**-distance** between probability measures $\mu_1, \mu_2 : \mathcal{F} \to \mathbb{R}$ is equal to twice the total variation distance, that is,

$$d_{\ell_1}(\mu_1, \mu_2) = 2d_{\text{TV}}(\mu_1, \mu_2)$$

In particular, when Ω is countable and $\mathcal{A} = 2^\Omega$, then

$$d_{\ell_1}(\mu_1, \mu_2) = \sum_{a \in \Omega} |\mu_1(a) - \mu_2(a)|.$$

Definition 10. Let Ω be a countable sample space. The ℓ_2**-distance** between probability measures $\mu_1, \mu_2 : 2^\Omega \to \mathbb{R}$ is

$$d_{\ell_2}(\mu_1, \mu_2) = \sqrt{\sum_{a \in \Omega} (\mu_1(a) - \mu_2(a))^2}.$$

Definition 11. The **Kolmogorov distance** between real Borel probability measures $\mu_1, \mu_2 : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ is

$$d_{\text{Kol}}(\mu_1, \mu_2) = \sup_{a \in \mathbb{R}} |\mu_1(-\infty, a] - \mu_2(-\infty, a]|.$$

That is, the Kolmogorov distance between two probability distributions on $\mathbb{R}$ equals the sup-norm of the difference between their cumulative distribution functions.

2.3 Singular-value approximation

We will make use of singular-value spectral approximations for matrices:

Definition 12 ([Ahm20, APSV21])**.** A matrix $\tilde{W} \in \mathbb{C}^{N \times N}$ is a **singular-value** ϵ**-approximation** of another matrix $W \in \mathbb{C}^{N \times N}$, written $\tilde{W} \overset{\text{sv}}{\approx}_\epsilon W$, if for all $x, y \in \mathbb{C}^N$,

$$|x^*(\tilde{W} - W)y| \le \frac{\epsilon}{2}(\|x\|^2 + \|y\|^2 - \|x^* W\|^2 - \|Wy\|^2).$$

Some properties of this approximation notion are given below.

Lemma 13 ([Ahm20, APSV21])**.** *If $\tilde{W} \overset{sv}{\approx}_\epsilon W$ and X is any matrix with $\|X\| \le 1$, then $\tilde{W}X \overset{sv}{\approx}_\epsilon WX$ and $X\tilde{W} \overset{sv}{\approx}_\epsilon XW$.*

Lemma 14 ([Ahm20, APSV21])**.** *For $0 \le \epsilon < .1$ and $k \ge 1$, if $\tilde{W} \overset{sv}{\approx}_\epsilon W$, then $\tilde{W}^k \overset{sv}{\approx}_{\epsilon + 5\epsilon^2} W^k$.*

Lemma 15 ([Ahm20, APSV21])**.** *If $\tilde{W} \overset{sv}{\approx}_\epsilon W$, then $\tilde{W} \otimes I \overset{sv}{\approx}_\epsilon W \otimes I$, where I denotes the identity matrix over any complex vector space $\mathbb{C}^M$.*

Lemma 16 ([Ahm20, APSV21])**.** *If G is the random walk matrix for a λ-spectral-expander (see Section 2.4), then $G \overset{sv}{\approx}_\lambda J$.*

2.4 Problem setup

For a regular digraph $G = (V, E)$ on n vertices, the spectral expansion is defined as

$$\lambda(G) = \|G|_{\vec{1}^\perp}\| = \max_{x \perp \vec{1}} \frac{\|x^\top G\|}{\|x\|} = \max_{x, x' \perp \vec{1}} \frac{|x^\top G x'|}{\|x\|\|x'\|} = \max_{x' \perp \vec{1}} \frac{\|G x'\|}{\|x'\|},$$

where by abuse of notation G also denotes the random walk matrix of G. For some integer $d \geq 2$, let V have an associated labeling $\mathrm{val} : V \to [d]$, and let $p = (p_0, \ldots, p_{d-1})$ be the vector with $p_b = |\mathrm{val}^{-1}(b)|/n$ equal to the fraction of vertices in V with label b. Assume without loss of generality that all labels are used, so that $p_b > 0$ for all $b \in [d]$. For $t \in \mathbb{N}$, we extend the label function component-wise to $\mathrm{val} : V^t \to [d]^t$.

Given $t \in \mathbb{N}$ and a sequence of random walk matrices $\mathcal{W} = (W_1, \ldots, W_{t-1})$ on shared vertex set V, let $\mathrm{RW}^t_{\mathcal{W}}$ denote the probability distribution over V^t obtained by taking a t-step random walk on V, where the ith step is taken according to the transition probabilities in W_i. Formally, to sample $(v_0, \ldots, v_{t-1}) \sim \mathrm{RW}^t_W$, the initial vertex $v_0 \in V$ is chosen uniformly at random, and then for $1 \leq i \leq t-1$ the vertex v_i is sampled given v_{i-1} according to $\Pr[v_i = v] = (W_i)_{v, v_{i-1}}$. If all W_i equal some matrix W, we let $\mathrm{RW}^t_W = \mathrm{RW}^t_{\mathcal{W}}$.

Our goal in this paper is to study the probability distribution of $\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ for a sequence $\mathcal{G} = (G_1, \ldots, G_{t_1})$ of λ-spectral expanders. In particular, as J is the optimal (0-spectral) expander, it is natural to study the distance between the distributions $\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\mathrm{val}(\mathrm{RW}^t_J)$, for appropriate distance metrics. Intuitively, this distance measures the cost of replacing truly random steps in J with expander random walks; note that $\mathrm{val}(\mathrm{RW}^t_J)$ is simply t iid copies of the probability distribution p over $[d]$.

In particular, we will study the distance (i.e. total variation distance) between the distributions $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and $f(\mathrm{val}(\mathrm{RW}^t_J))$ for specific classes of functions f on $[d]^t$, namely symmetric functions and functions computable by permutation branching programs. We are interested in distance bounds that depend on G only through λ. Equivalently, this problem can be formulated as studying which functions f are fooled by all λ-spectral expanders G.

3 Symmetric functions

In this section, we address the following question: for an arbitrary symmetric function f on $[d]^t$, how similar are the distributions of $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))$, for certain graph sequences $\mathcal{G}, \mathcal{G}'$ of interest? In particular, to study how well symmetric functions are fooled by expander walks, we may let $\mathcal{G}$ be a sequence of λ-spectral expanders and $\mathcal{G}' = (J, \ldots, J)$ (though our results apply more generally). To measure the "similarity" of the distributions, we will prove both total variation distance bounds and tail bounds.

To begin, without loss of generality we restrict our attention from the class of all symmetric functions to the symmetric function Σ defined below.

Definition 17. For a string $a \in [d]^{[t]}$, let $\Sigma a \in [t+1]^{[d]}$ be the vector counting the number of occurences of each $b \in [d]$ in a. That is, for $b \in [d]$, let $(\Sigma a)_b = |\{i \in t : a_i = b\}|$.

Remark. The output of $\Sigma : [d]^t \to [t+1]^d$ is slightly redundant, as $\sum_{b \in [d]} (\Sigma a)_b = t$.

By definition, every symmetric function f on $[d]^t$ factors through Σ, in the sense that the value $f(a)$ can be computed given only f and Σa. Thus for every pair of sequences $\mathcal{W}, \mathcal{W}'$ of random walk matrices, the total variation distance between $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}}))$ and $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}'}))$ is bounded above by the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}'})$. Conversely, by definition there always exists a symmetric function $f : [d]^t \to \{0,1\}$ such that the total variation distance between $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}}))$ and $f(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}'}))$ equals the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}'})$, as if the subset $A \subseteq [t+1]^d$ maximizes the variation between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}'})$, then f may simply be defined as the indicator function for the event $\Sigma a \in A$. Thus to prove upper bounds on total variation distance that hold over all symmetric functions, we may restrict attention to the function $\Sigma : [d]^t \to [t+1]^d$.

Because Σ is supported inside $\mathbb{Z}^d$, we may study the difference between the tails as well as the total variation distance between $\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and $\Sigma(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))$. The following result provides a unified bound for these two objectives when $d = 2$.

Theorem 18. *Fix integers $t \geq 1$ and $1 \leq u \leq t-1$. Let $\mathcal{G} = (G_i)_{1\leq i\leq t-1}$ and $\mathcal{G}' = (G'_i)_{1\leq i\leq t-1}$ be sequences of regular graphs on a shared vertex set V such that for all $i \neq u$ we have $G_i = G'_i$ with $\lambda(G_i) = \lambda(G'_i) \leq 1/100$. Fix a labeling $\mathrm{val} : V \to [2]$ that assigns each label $b \in [2]$ to p_b-fraction of the vertices. Then for every $c \geq 0$,*

$$\begin{aligned}\sum_{j\in[t+1]:|j-p_1t|\geq c} &\left|\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}) = (t-j,j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j,j)\right]\right| \\ &\leq 4000 \cdot \frac{\|G'_u - G_u\| \cdot e^{-c^2/8t}}{t}.\end{aligned} \tag{6}$$

To interpret Theorem 18, it is useful to consider the following corollary. This corollary bounds the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and the binomial distribution, and simultaneously provides a tail bound that strengthens the expander-walk Chernoff bound.

Corollary 19. *Fix an integer $t \geq 1$. Let $\lambda \leq 1/100$, and let $\mathcal{G} = (G_i)_{1\leq i\leq t-1}$ be a sequence of regular λ-spectral expanders on shared vertex set V. Fix a labeling $\mathrm{val} : V \to [2]$ that assigns each label $b \in [2]$ to p_b-fraction of the vertices. Then for every $c \geq 0$,*

$$\begin{aligned}\sum_{j\in[t+1]:|j-p_1t|\geq c} &\left|\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j,j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{J}) = (t-j,j)\right]\right| \\ &\leq 4000 \cdot \lambda \cdot e^{-c^2/8t}.\end{aligned}$$

Proof. For $0 \leq i \leq t-1$, define $\mathcal{G}^{(i)} = (G_1, \ldots, G_i, J, \ldots, J)$ to be the sequence consisting of the first i elements of $\mathcal{G}$ followed by $t-1-i$ copies of J. Then because $\|G_i - J\| = \lambda(G_i)$ for all i, applying Theorem 18 to each pair of sequences $\mathcal{G}^{(u)}, \mathcal{G}^{(u-1)}$ for $1 \leq u \leq t-1$ and summing the resulting bounds with the triangle inequality gives

$$\begin{aligned}\sum_{j\in[t+1]:|j-p_1t|\geq c} &\left|\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}^{(t-1)}}) = (t-j,j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}^{(0)}}) = (t-j,j)\right]\right| \\ &\leq \sum_{u=1}^{t-1} 4000 \cdot \frac{\lambda(G_u) \cdot e^{-c^2/8t}}{t} \\ &\leq 4000 \cdot \lambda \cdot e^{-c^2/8t}.\end{aligned}$$

□

When $c = 0$, Corollary 19 implies that for a sequence $\mathcal{G}$ of λ-spectral expanders, the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_J)$ is at most $O(\lambda)$, where the big-O constant does not depend on p. In contrast, the best previously known upper bound on this total variation distance, shown by Cohen et al. [CMP+21], was $O(\lambda/\sqrt{\min(p)})$, which weakens as $\min(p) = \min\{p_0, p_1\}$ falls. Cohen et al. [CMP+21] also assumed that all graphs in the sequence $\mathcal{G}$ are identical, whereas we allow the graph to change at each step.

When we consider $c > 0$, our bounds have further new implications. In particular, Corollary 19 strengthens the expander-walk Chernoff bound [Gil98, Hea08]. The expander-walk Chernoff bound shows that when λ is bounded away from 1 and $\mathcal{G}$ is a sequence of λ-spectral expanders, then

$$\sum_{j\in[t+1]:|j-p_1t|\geq c} \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j,j)\right] = \Pr\left[|(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{G}))_1 - p_1 t| \geq c\right] = O(e^{-\Omega(c^2/t)}),$$

that is, that the tails of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ decay approximately as quickly as the tails of the binomial distribution $\Sigma\,\mathrm{val}(\mathrm{RW}^t_J)$. In contrast, our tail bound in Corollary 19 shows the stronger statement that as $\lambda \to 0$, the tails of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ converge to the tails of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_J)$.

While Corollary 18 compares $\mathrm{RW}^t_{\mathcal{G}}$ with RW^t_J, Theorem 18 provides bounds for comparing $\mathrm{RW}^t_{\mathcal{G}}$ with $\mathrm{RW}^t_{\mathcal{G}'}$ for any two sequences $\mathcal{G}, \mathcal{G}'$ of $1/100$-spectral expanders. Thus if the sequences $\mathcal{G}$ and $\mathcal{G}'$ agree at all but a few steps, then Theorem 18 gives stronger bounds on the resulting distributions. More generally, as demonstrated in the proof of Corollary 19, Theorem 18 suggests that the total variation distance between $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})$ increases somewhat "smoothly" as the distance between the sequences $\mathcal{G}$ and $\mathcal{G}'$ increases.

Below, we present an analogue of Theorem 18 for arbitrary $d \geq 2$, although without a tail bound. The $d = 2$ case of the theorem below is simply the $c = 0$ case of Theorem 18. We prove the $d > 2$ case in Appendix B, as the proof closely resembles the proof of Theorem 18 given in Section 3.1 below.

Theorem 20. *Fix integers $t \geq 1$ and $1 \leq u \leq t-1$. For some integer $d \geq 2$, let V be a finite set of vertices with labeling $\mathrm{val} : V \to [d]$ that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Then there exist constants $\ell_{d,p}, c_{d,p} > 0$ depending only on d and p such that the following holds. Let $\mathcal{G} = (G_i)_{1\leq i\leq t-1}$ and $\mathcal{G}' = (G'_i)_{1\leq i\leq t-1}$ be sequences of regular graphs on the shared vertex set V such that for all $i \neq u$ we have $G_i = G'_i$ with $\lambda(G_i) = \lambda(G'_i) \leq \ell_{d,p}$. Then*

$$d_{TV}\left(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}), \Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})\right) \leq c_{d,p} \cdot \frac{\|G'_u - G_u\|}{t}.$$

In particular, for $d = 2$ and all $p = (p_0, p_1)$ we may take

$$\ell_{d,p} = 1/100 \qquad \text{and} \qquad c_{d,p} = 2000,$$

while for $d > 2$, letting $\min(p) = \min_{b\in[d]} p_b$, we may take

$$\ell_{d,p} = \min(p)/400 \qquad \text{and} \qquad c_{d,p} = \frac{2^{9d/4+13} d^{d/2+3}}{\min(p)^{(d+3)/4}}.$$

The following direct application of Theorem 20 follows by an argument that is analogous to the proof of Corollary 19.

Corollary 21. *Fix an integer $t \geq 1$. For some integer $d \geq 2$, let V be a finite set of vertices with labeling $\mathrm{val} : V \to [d]$ that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Define constants $\ell_{d,p}, c_{d,p}$ as in Theorem 20. Let $\lambda \leq \ell_{d,p}$, and let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ be a sequence of regular λ-spectral expanders on shared vertex set V. Then*

$$d_{TV}\left(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}), \Sigma\,\mathrm{val}(\mathrm{RW}^t_J)\right) \leq c_{d,p} \cdot \lambda.$$

Corollary 21 provides an affirmative answer to a question of Jalan and Moshkovitz [JM21], who asked whether the bound of Cohen et al. [CPTS21] extends to $d > 2$. However, we suspect that the dependence on p in Theorem 20 and Corollary 21 is not tight, and we leave the determination of the optimal dependence as an open question.

3.1 Proof of Theorem 18

In this section, we restrict attention to the case where $d = 2$, and we prove Theorem 18. Define

$$g = \left(\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}) = (t-j, j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j, j)\right]\right)_{j \in [t+1]} \in [-1, 1]^{[t+1]} \tag{7}$$

to denote the difference between the distributions of $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))_1$ and $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_1$.

3.1.1 Reduction to bounding an ℓ_2-norm

In this section, we show how to prove Theorem 18 given Theorem 22 below, which bounds the norm of the vector

$$g^{(sr)} = (e^{sr(j-p_1t)} g_j)_{j \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}} \tag{8}$$

for $s = \pm 1$ and $0 \leq r \leq 1/2$, where g_j is given by (7) for $j \in [t+1]$ and $g_j = 0$ for $j \notin [t+1]$. We will then prove Theorem 22 in the following sections.

Theorem 22. *As in Theorem 18, let $u < t$ be positive integers and let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$ be sequences of regular graphs on a shared vertex set V with labeling $\mathrm{val} : V \to [2]$, such that for all $i \neq u$ we have $G_i = G'_i$ with $\lambda(G_i) = \lambda(G'_i) \leq 1/100$. Let val assign each label $b \in [2]$ to p_b-fraction of the vertices. Then for every $s = \pm 1$ and $0 \leq r \leq 1/2$, defining $g^{(sr)}$ as in (8), we have*

$$\|g^{(sr)}\| \leq \|G'_u - G_u\| \cdot p_0 p_1 \cdot e^{2p_0p_1tr^2} \cdot \min\left\{44,\ \frac{22 \cdot r^2}{(p_0p_1t)^{1/4}} + \frac{70}{(p_0p_1t)^{5/4}}\right\}.$$

By reducing Theorem 18 to Theorem 22, we reduce our goal of showing an ℓ_1-bound to the problem of showing an ℓ_2-bound. Because the ℓ_2-norm is invariant under orthonormal basis changes, we can transition to the Fourier basis to prove Theorem 22, as will be shown in Section 3.1.2.

To prove Theorem 18, we will partition $[t+1]$ into intervals of length approximately $\sqrt{p_0p_1t}$, which equals the standard deviation of $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_J))_1$. We then bound the ℓ_1-norm of g on each such interval in terms of $\|g^{(sr)}\|$ for appropriately chosen $s = \pm 1$, $0 \leq r \leq 1/2$, as shown in Lemma 23 below. Intuitively, an appropriate selection of s, r allows the components of $g^{(sr)}$ in the specified interval to dominate components outside the interval, so within each interval we obtain asymptotically tight bounds.

Lemma 23. *For $k \geq 0$, let*

$$S_k^+ = \{j \in [t+1] : k\sqrt{p_0p_1t} \leq j - p_1t < (k+1)\sqrt{p_0p_1t}\}$$
$$S_k^- = \{j \in [t+1] : -(k+1)\sqrt{p_0p_1t} < j - p_1t \leq -k\sqrt{p_0p_1t}\}.$$

Then for $s = \pm 1$, $r \geq 0$, the following holds:

1. *For $j \in [t+1]$, $|g_j| \leq e^{-sr(j-p_1t)} \cdot \|g^{(sr)}\|$.*
2. *If $p_0p_1t \geq 1$, then $\|g_{S_k^s}\|_1 \leq \sqrt{2} \cdot (p_0p_1t)^{1/4} \cdot e^{-rk\sqrt{p_0p_1t}} \cdot \|g^{(sr)}\|$.*

Proof. 1. By definition $|g_j| = e^{-sr(j-p_1t)} \cdot |e^{sr(j-p_1t)}g_j| \leq e^{-sr(j-p_1t)} \cdot \|g^{(sr)}\|$.

2. By definition

$$\begin{aligned}\|g_{S_k^s}\|_1 &\leq e^{-rk\sqrt{p_0p_1t}} \cdot \sum_{j \in S_k^s} |e^{srj}g_j| \\ &\leq e^{-rk\sqrt{p_0p_1t}} \cdot \sqrt{|S_k^s|} \cdot \|g^{(sr)}\|,\end{aligned}$$

where the second inequality above holds by the Cauchy-Schwartz inequality. Then the desired result follows because $|S_k^s| \leq \sqrt{\sqrt{p_0p_1t}+1} \leq \sqrt{2} \cdot (p_0p_1t)^{1/4}$ as $p_0p_1t \geq 1$.

□

Because $\bigcup_{s=\pm1, 0 \leq k \leq \sqrt{t/p_1}} S_k^s = [t+1]$, Lemma 23 shows how to bound $\|g\|_1$ given bounds on $\|g^{(sr)}\|$. We prove Theorem 18 using this approach below. The theorem is restated for convenience.

Theorem 18. *Fix integers $t \geq 1$ and $1 \leq u \leq t-1$. Let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G_i')_{1 \leq i \leq t-1}$ be sequences of regular graphs on a shared vertex set V such that for all $i \neq u$ we have $G_i = G_i'$ with $\lambda(G_i) = \lambda(G_i') \leq 1/100$. Fix a labeling $\mathrm{val} : V \to [2]$ that assigns each label $b \in [2]$ to p_b-fraction of the vertices. Then for every $c \geq 0$,*

$$\begin{aligned}\sum_{j \in [t+1] : |j-p_1t| \geq c} &\left|\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}) = (t-j, j)\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}) = (t-j, j)\right]\right| \\ &\leq 4000 \cdot \frac{\|G_u' - G_u\| \cdot e^{-c^2/8t}}{t}.\end{aligned} \tag{6}$$

Proof. First consider the case where $p_0p_1t < 1$. Then by Lemma 23 and Theorem 22, for $j \in [t+1]$,

$$\begin{aligned}|g_j| &\leq e^{-sr(j-p_1t)} \cdot 44 \cdot \frac{\|G_u' - G_u\|}{t} \cdot p_0p_1t \cdot e^{2p_0p_1tr^2} \\ &\leq 44 \cdot \frac{\|G_u' - G_u\|}{t} \cdot e^{2r^2 - sr(j-p_1t)}.\end{aligned}$$

Thus setting $r = 1/2$ and $s = \mathrm{sgn}(j - p_1t)$ and then summing over j gives

$$\begin{aligned}\sum_{j \in [t+1] : |j-p_1t| \geq c} |g_j| &\leq \sum_{j \in [t+1] : |j-p_1t| \geq c} 44 \cdot \frac{\|G_u' - G_u\|}{t} \cdot e^{1/2 - |j-p_1t|/2} \\ &\leq 88 \cdot \frac{\|G_u' - G_u\|}{t} \cdot \frac{e^{1/2-c/2}}{1-e^{-1/2}}.\end{aligned}$$

When $c > t$, the left hand side above equals 0, so assume that $0 \le c \le t$. Then $-c/2 \le -c^2/2t$, so (6) holds by the above inequality because $88 \cdot e^{1/2}/(1-e^{-1/2}) \le 4000$.

Now assume that $p_0p_1t \ge 1$. Then by Lemma 23 and Theorem 22, for $k \ge 0$ and $s = \pm 1$,

$$\begin{aligned}\|g_{S_k^s}\|_1 &\le \sqrt{2}\cdot(p_0p_1t)^{1/4}\cdot e^{-rk\sqrt{p_0p_1t}}\cdot\|G'_u - G_u\|\cdot p_0p_1\cdot e^{2p_0p_1tr^2}\cdot\left(\frac{22\cdot r^2}{(p_0p_1t)^{1/4}}+\frac{70}{(p_0p_1t)^{5/4}}\right)\\ &\le \frac{\|G'_u-G_u\|}{t}\cdot e^{2r^2p_0p_1t-rk\sqrt{p_0p_1t}}\cdot\left(22\sqrt{2}\cdot r^2\cdot p_0p_1t+70\sqrt{2}\right).\end{aligned}$$

To make the bound above tight, we choose $0 \le r \le 1/2$ to minimize $2r^2p_0p_1t - rk\sqrt{p_0p_1t}$. That is, set

$$r = \begin{cases}\frac{k}{4\sqrt{p_0p_1t}}, & k \le 2\sqrt{p_0p_1t}\\ \frac{1}{2}, & k > 2\sqrt{p_0p_1t}.\end{cases}$$

Then

$$2r^2p_0p_1t - rk\sqrt{p_0p_1t} = \begin{cases}-\frac{k^2}{8}, & k \le 2\sqrt{p_0p_1t}\\ \frac{p_0p_1t}{2}-\frac{k\sqrt{p_0p_1t}}{2}, & k > 2\sqrt{p_0p_1t}.\end{cases}$$

For $k > 2\sqrt{p_0p_1t}$, we have

$$\begin{aligned}\frac{p_0p_1t}{2}-\frac{k\sqrt{p_0p_1t}}{2} &= -\frac{p_0p_1t}{2}-\frac{(k-2\sqrt{p_0p_1t})\sqrt{p_0p_1t}}{2}\\ &\le -(2\sqrt{p_0p_1t}+(k-2\sqrt{p_0p_1t}))\frac{\sqrt{p_0p_1t}}{4}\\ &= -\frac{k\sqrt{p_0p_1t}}{4}.\end{aligned}$$

Furthermore, because $p_0p_1t \ge 1$, by definition $r^2\cdot p_0p_1t \le k^2/16$. Thus

$$\begin{aligned}\|g_{S_k^s}\|_1 &\le \frac{\|G'_u-G_u\|}{t}\cdot\max\{e^{-k^2/8}, e^{-k\sqrt{p_0p_1t}/4}\}\cdot\left(22\sqrt{2}\cdot\frac{k^2}{16}+70\sqrt{2}\right)\\ &\le 70\sqrt{2}\cdot\frac{\|G'_u-G_u\|}{t}\cdot\max\{e^{-k^2/16}, e^{-k\sqrt{p_0p_1t}/8}\},\end{aligned}$$

where the second inequality above holds because $p_0p_1t \ge 1$ so that $e^{-k^2/16}\cdot\left(22\sqrt{2}\cdot\frac{k^2}{16}+70\sqrt{2}\right)$ and $e^{-k\sqrt{p_0p_1t}/8}\cdot\left(22\sqrt{2}\cdot\frac{k^2}{16}+70\sqrt{2}\right)$ are both maximized at $k=0$, and therefore are both bounded above by $70\sqrt{2}$. Therefore

$$\begin{aligned}\sum_{j\in[t+1]:|j-p_1t|\ge c}|g_j| &\le \sum_{\ell=0}^{\lfloor\sqrt{t/p_0p_1}\rfloor}\left(\left\|g_{S^+_{c/\sqrt{p_0p_1t}+\ell}}\right\|_1+\left\|g_{S^-_{c/\sqrt{p_0p_1t}+\ell}}\right\|_1\right)\\ &\le 140\sqrt{2}\cdot\frac{\|G'_u-G_u\|}{t}\cdot\sum_{\ell=0}^{\lfloor\sqrt{t/p_0p_1}\rfloor}\max\{e^{-(c/\sqrt{p_0p_1t}+\ell)^2/16}, e^{-c/8-\ell\sqrt{p_0p_1t}/8}\}\\ &\le 140\sqrt{2}\cdot\frac{\|G'_u-G_u\|}{t}\cdot\frac{\max\{e^{-c^2/16p_0p_1t}, e^{-c/8}\}}{1-e^{-1/16}}.\end{aligned}$$

When $c > t$, the left hand side above equals 0, so assume that $0 \leq c \leq t$. Then $-c/8 \leq -c^2/8t$, and also $-c^2/16p_0p_1t \leq -c^2/4t$, so the above inequality gives

$$\sum_{j \in [t+1]:|j-p_1t| \geq c} |g_j| \leq 140\sqrt{2} \cdot \frac{\|G'_u - G_u\|}{t} \cdot \frac{e^{-c^2/8t}}{1 - e^{-1/16}}$$
$$\leq 4000 \cdot \|G'_u - G_u\| \cdot \frac{e^{-c^2/8t}}{t}.$$

□

3.1.2 Reduction to bounding Fourier coefficients

In this section, we show how to prove Theorem 22 given bounds on the Fourier transform of $g^{(sr)}$. Specifically, we prove Theorem 22 assuming Theorem 25 below, which will be proven in Section 3.1.3.

We first introduce the Fourier transform for the group $\mathbb{Z}$.

Definition 24. Let $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ with ℓ^2 norm $\|f\| = \sqrt{\int_{-\pi}^{\pi} |f(\theta)|^2 d\theta/2\pi}$. Let $\ell^2(\mathbb{Z})$ and $\ell^2(S^1)$ denote the subspaces of $\mathbb{C}^{\mathbb{Z}}$ and $\mathbb{C}^{S^1}$ respectively containing all elements of finite ℓ^2 norm. Then the **Fourier transform for the group** $\mathbb{Z}$ is the map $\mathcal{F} : \ell^2(\mathbb{Z}) \to \ell^2(S^1)$ such the Fourier transform of $h \in \ell^2(\mathbb{Z})$, denoted $\mathcal{F}h = \hat{h} \in \ell^2(S^1)$, is given by

$$\hat{h}(\theta) = \sum_{j \in \mathbb{Z}} h_j e^{-i\theta j}.$$

The Fourier transform may also be expressed in terms of the **Fourier characters** $\chi_\theta = (e^{i\theta j})_{j \in \mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$, as $\hat{h}(\theta) = \chi_\theta^* h$.

It is well know that the Fourier transform preserves the ℓ^2 norm, so that $\|\hat{h}\| = \|h\|$. Below, we associate $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ with the interval $(-\pi, \pi]$, so that all $\theta \in S^1$ have $|\theta| \leq \pi$.

Theorem 25. *As in Theorem 18, let $u < t$ be positive integers and let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G'_i)_{1 \leq i \leq t-1}$ be sequences of regular graphs on a shared vertex set V with labeling* $\mathrm{val} : V \to [2]$, *such that for all $i \neq u$ we have $G_i = G'_i$ with $\lambda(G_i) = \lambda(G'_i) \leq 1/100$. Let* val *assign each label $b \in [2]$ to p_b-fraction of the vertices. Then for every $s = \pm 1$ and $0 \leq r \leq 1/2$, defining $g^{(sr)}$ as in (8), we have for all $-\pi < \theta \leq \pi$ that*

$$|\hat{g}^{(sr)}(\theta)| \leq 4 \cdot \|G'_u - G_u\| \cdot p_0p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right) \cdot e^{p_0p_1t(2r^2 - \theta^2/20)}.$$

To obtain the desired bound on $\|g^{(sr)}\|$ in Theorem 22, we square the inequality in Theorem 25 and then integrate over $\theta \in (-\pi, \pi]$, using the fact that $\|g^{(sr)}\|^2 = \|\hat{g}^{(sr)}\|^2 = \int_{-\pi}^{\pi} |\hat{g}^{(sr)}(\theta)|^2 d\theta/2\pi$ because the Fourier transform preserves ℓ^2-norms. This calculation is shown below.

Proof of Theorem 22. Because the Fourier transform preserves the ℓ^2-norm, by Theorem 25,

$$\begin{aligned}
\|g^{(sr)}\| &= \sqrt{\int_{-\pi}^{\pi} |\hat{g}^{(sr)}(\theta)|^2 \frac{d\theta}{2\pi}} \\
&\leq 4 \cdot \|G'_u - G_u\| \cdot p_0 p_1 \cdot e^{2p_0p_1tr^2} \cdot \sqrt{\int_{-\pi}^{\pi} \left(4r^2 + \frac{3}{2}\theta^2\right)^2 \cdot e^{-p_0p_1t\theta^2/10} \frac{d\theta}{2\pi}} \\
&\leq 4\sqrt{2} \cdot \|G'_u - G_u\| \cdot p_0p_1 \cdot e^{2p_0p_1tr^2} \\
&\quad \cdot \left(4r^2 \sqrt{\int_{-\pi}^{\pi} e^{-p_0p_1t\theta^2/10} \frac{d\theta}{2\pi}} + \sqrt{\int_{-\pi}^{\pi} \frac{9}{4}\theta^4 \cdot e^{-p_0p_1t\theta^2/10} \frac{d\theta}{2\pi}}\right),
\end{aligned} \tag{9}$$

where the final inequality above holds because all $a, b \geq 0$ satisfy $(a + b)^2 \leq 2(a^2 + b^2)$ and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. Applying the bound $e^{-p_0p_1t\theta^2/10} \leq 1$ in the right hand side of (9) gives

$$\begin{aligned}
\|g^{(sr)}\| &\leq 4\sqrt{2} \cdot \|G'_u - G_u\| \cdot p_0p_1 \cdot e^{2p_0p_1tr^2} \cdot \left(4r^2 + \frac{3\pi^2}{2\sqrt{5}}\right) \\
&\leq 44 \cdot \|G'_u - G_u\| \cdot p_0p_1 \cdot e^{2p_0p_1tr^2},
\end{aligned}$$

where the second inequality above holds because $0 \leq r \leq 1/2$. Alternatively, substituting $q = \sqrt{p_0p_1t/10} \cdot \theta$ in the integrals in the right hand side of (9) gives

$$\begin{aligned}
\|g^{(sr)}\| &\leq 4\sqrt{2} \cdot \|G'_u - G_u\| \cdot p_0p_1 \cdot e^{2p_0p_1tr^2} \\
&\quad \cdot \left(\frac{4r^2}{\sqrt{2\pi}} \left(\frac{10}{p_0p_1t}\right)^{1/4} \sqrt{\int_{-\infty}^{\infty} e^{-q^2} dq} + \frac{3}{2\sqrt{2\pi}} \left(\frac{10}{p_0p_1t}\right)^{5/4} \sqrt{\int_{-\infty}^{\infty} q^4 \cdot e^{-q^2} dq}\right).
\end{aligned}$$

Substituting the values $\int_{-\infty}^{\infty} e^{-q^2} dq = \sqrt{\pi}$ and $\int_{-\infty}^{\infty} q^4 \cdot e^{-q^2} dq = 3\sqrt{\pi}/4$ in the right hand side above yields

$$\|g^{(sr)}\| \leq \|G'_u - G_u\| \cdot p_0p_1 \cdot e^{2p_0p_1tr^2} \cdot \left(\frac{22 \cdot r^2}{(p_0p_1t)^{1/4}} + \frac{70}{(p_0p_1t)^{5/4}}\right).$$

The above inequalities together imply the desired result. □

3.1.3 Bounding Fourier coefficients

In this section we prove Theorem 25. Throughout this section, for convenience we extend the sequences G and G' to include 0th and tth components $G_0 = G'_0 = G_t = G'_t = J$. In the proofs below, these matrices will typically be applied to $\vec{1}$, and they therefore could be removed at the cost of more cumbersome notation.

First, we express the Fourier transform $\hat{g}^{(sr)}(\theta)$ of $g^{(sr)}$ linear-algebraically below.

Lemma 26. *Let $P_\theta^{(sr)} \in \mathbb{C}^{V \times V}$ be the matrix given by:*

$$P_\theta^{(sr)} = \sum_{v \in V} \delta_v \delta_v^* e^{(sr - i\theta)(\mathrm{val}(v) - p_1)} = \sum_{v \in \mathrm{val}^{-1}(0)} \delta_v \delta_v^* e^{-p_1(sr - i\theta)} + \sum_{v \in \mathrm{val}^{-1}(1)} \delta_v \delta_v^* e^{p_0(sr - i\theta)}.$$

Then

$$e^{i\theta p_1 t} \cdot \hat{g}^{(sr)}(\theta) = \vec{1}^* G_t P_\theta^{(sr)} \cdots G_{u+1} P_\theta^{(sr)} (G'_u - G_u) P_\theta^{(sr)} G_{u-1} \cdots P_\theta^{(sr)} G_0 \vec{1}$$
$$= \vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)} \right) (G'_u - G_u) \left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1}.$$

Proof. For $\mathcal{W} = \mathcal{G}$ or $\mathcal{G}'$, writing $W_i = \sum_{v,v' \in V} \delta_{v'} \delta_v^* (W_i)_{v',v}$, then the expression $\vec{1}^* \left(\prod_{i=1}^{t} (W_i P_\theta^{(sr)}) \right) \vec{1} = \sqrt{n} \vec{1}^* \left(\prod_{i=1}^{t} (W_i P_\theta^{(sr)}) \right) (1/\sqrt{n}) \vec{1}$ expands to give

$$\vec{1}^* \left(\prod_{i=1}^{t} (W_i P_\theta^{(sr)}) \right) \vec{1} = \sum_{(v_0,\ldots,v_{t-1}) \in V^{[t]}} \Pr\left[\mathrm{RW}_{\mathcal{W}}^t = (v_0, \ldots, v_{t-1})\right] \prod_{i' \in [t]} e^{(sr - i\theta)(\mathrm{val}(v_{i'}) - p_1)}$$
$$= \mathbb{E}\left[e^{(sr - i\theta)((\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{W}}^t))_1 - p_1 t)}\right].$$

The first equality above follows from an expansion analogous to the one described in detail in the proof of Lemma 36 in Section 4.1, to which the reader is referred for details; we omitted some intermediate steps here for readability. Now it follows that

$$e^{i\theta p_1 t} \cdot \hat{g}^{(sr)}(\theta) = \sum_{j \in \mathbb{Z}} \left(\Pr\left[\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}'}^t) = (t-j, j)\right] - \Pr\left[\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t) = (t-j, j)\right]\right) e^{(sr - i\theta)(j - p_1 t)}$$
$$= \mathbb{E}\left[e^{(sr - i\theta)((\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}'}^t))_1 - p_1 t)}\right] - \mathbb{E}\left[e^{(sr - i\theta)((\Sigma\, \mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t))_1 - p_1 t)}\right]$$
$$= \vec{1}^* \left(\prod_{i=1}^{t} (G'_i P_\theta^{(sr)}) \right) \vec{1} - \vec{1}^* \left(\prod_{i=1}^{t} (G_i P_\theta^{(sr)}) \right) \vec{1}$$
$$= \vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)} \right) (G'_u - G_u) \left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1},$$

where the final equality above holds because $G'_i = G_i$ for $i \neq u$ by assumption. □

Lemma 26 shows that in order to bound the Fourier transform of $g^{(sr)}$, it is sufficient to bound $\vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)} \right) (G'_u - G_u) \left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1}$. For this purpose, because the matrix $G'_u - G_u$ annihilates $\vec{1}$ from both sides, we will bound the components of $\left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)} \right) \vec{1}$ and $\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1}$ that are orthogonal to $\vec{1}$ (denoted by $^\perp$ below). To bound the orthogonal components of these vectors, we will apply the following two lemmas, which provide bounds for the matrix $P_\theta^{(sr)}$, and will be proven in Section 3.1.4.

Lemma 27. *For $-\pi < \theta \leq \pi$, $s = \pm 1$, and $0 \leq r \leq 1/2$, we have*

$$|\vec{1}^* P_\theta^{(sr)} \vec{1}| \leq 1 + p_0 p_1 \cdot r^2 - \frac{2}{\pi^2 e^{1/2}} \cdot p_0 p_1 \cdot \theta^2$$

and

$$\|(P_\theta^{(sr)} \vec{1})^\perp\| = \|(\vec{1}^* P_\theta^{(sr)})^\perp\| \leq \sqrt{e \cdot p_0 p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right)}.$$

Lemma 28. *Let $\rho = 1/10$ and $F = J + \rho(I - J)$. Then for every $s = \pm 1$ and $0 \le r \le 1/2$,*

$$\|FP_\theta^{(sr)}F\| \le 1 + 2p_0p_1 \cdot r^2 - \frac{p_0p_1}{20} \cdot \theta^2.$$

Lemma 29. *For $1 \le u \le t-1$,*

$$\left\| \left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\perp \right\| \le 2 \cdot \sqrt{p_0p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right)} \cdot e^{p_0p_1u(2r^2 - \theta^2/20)}.$$

Proof. By definition

$$\left\| \left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\perp \right\| \le \|(P_\theta^{(sr)}\vec{1})^\perp\| \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\parallel \right\| + \|P_\theta^{(sr)}\| \cdot \lambda(G_{u-1}) \cdot \left\| \left(\left(\prod_{i=0}^{u-2} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\perp \right\|.$$

The above inequality can be recursively applied to bound the term $\left\| \left(\left(\prod_{i=0}^{u-2} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\perp \right\|$ on its right hand side. Performing $u-1$ such recursive applications gives that

$$\left\| \left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\perp \right\| \le \|(P_\theta^{(sr)}\vec{1})^\perp\| \sum_{i=0}^{u-1} \left(\prod_{i'=i+1}^{u-1} \|P_\theta^{(sr)}\| \cdot \lambda(G_{i'}) \right) \left\| \left(\left(\prod_{i'=0}^{i-1} P_\theta^{(sr)} G_{i'} \right) \vec{1} \right)^\parallel \right\|.$$

Let $\rho = 1/10$ and $F = J + \rho(I - J)$. By assumption all $i' \le u-1$ have $\lambda(G_{i'}) \le \rho^2$. It follows that $\|F^{-1}G_{i'}F^{-1}\| \le 1$, as $F^{-1}G_{i'}F^{-1}$ preserves the vector $\vec{1}$ and the subspace $\vec{1}^\perp$, and the restriction $F^{-1}G_{i'}F^{-1}|_{\vec{1}^\perp}$ has spectral norm $\rho^{-1}\lambda(G_{i'})\rho^{-1} \le 1$. It also holds that $\|P_\theta^{(sr)}\| \le e^{1/2}$ because $0 \le r \le 1/2$. Thus

$$\begin{aligned}
&\left\| \left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i \right) \vec{1} \right)^\perp \right\| \\
&\quad \le \|(P_\theta^{(sr)}\vec{1})^\perp\| \cdot \sum_{i=0}^{u-1} (e^{1/2}\rho^2)^{u-1-i} \cdot \left| \vec{1}^* \left(\prod_{i'=0}^{i-1} FP_\theta^{(sr)}F \cdot F^{-1}G_{i'}F^{-1} \right) \vec{1} \right| \\
&\quad \le \|(P_\theta^{(sr)}\vec{1})^\perp\| \cdot \sum_{i=0}^{u-1} (e^{1/2}\rho^2)^{u-1-i} \cdot \|FP_\theta^{(sr)}F\|^i.
\end{aligned}$$

Applying Lemma 27 and Lemma 28 to bound $\|(P_\theta^{(sr)}\vec{1})^\perp\|$ and $\|FP_\theta^{(sr)}F\|$ respectively gives

$$\begin{aligned}
&\left\|\left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i\right)\vec{1}\right)^\perp\right\| \\
&\leq \sqrt{e\cdot p_0p_1\cdot\left(4r^2+\frac{3}{2}\theta^2\right)}\cdot\sum_{i=0}^{u-1}(e^{1/2}\rho^2)^{u-1-i}\cdot e^{i\cdot p_0p_1(2r^2-\theta^2/20)} \\
&= \sqrt{e\cdot p_0p_1\cdot\left(4r^2+\frac{3}{2}\theta^2\right)}\cdot e^{p_0p_1u(2r^2-\theta^2/20)}\cdot\sum_{i=0}^{u-1}\frac{(e^{1/2}\rho^2)^{u-1-i}}{e^{p_0p_1(u-i)(2r^2-\theta^2/20)}}.
\end{aligned}$$

Because $\rho = 1/10$ and $e^{p_0p_1(2r^2-\theta^2/20} \geq e^{-\pi^2/80}$,

$$\sum_{i=0}^{u-1}\frac{(e^{1/2}\rho^2)^{u-1-i}}{e^{p_0p_1(u-i)(2r^2-\theta^2/20)}} \leq \sum_{i=-\infty}^{u-1}\frac{(e^{1/2}/100)^{u-1-i}}{(e^{-\pi^2/80})^{u-i}} = \frac{e^{\pi^2/80}}{1-e^{1/2+\pi^2/80}/100} \leq \frac{2}{\sqrt{e}}.$$

Thus

$$\left\|\left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i\right)\vec{1}\right)^\perp\right\| \leq 2\cdot\sqrt{p_0p_1\cdot\left(4r^2+\frac{3}{2}\theta^2\right)}\cdot e^{p_0p_1u(2r^2-\theta^2/20)}.$$

□

We now apply the above lemmas to prove Theorem 25.

Proof of Theorem 25. By Lemma 26,

$$\begin{aligned}
|\hat{g}^{(sr)}(\theta)| &= \left|\vec{1}^*\left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)}\right)(G'_u - G_u)\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i\right)\vec{1}\right| \\
&\leq \left\|\left(\vec{1}^*\left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)}\right)\right)^\perp\right\| \|G'_u - G_u\| \left\|\left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i\right)\vec{1}\right)^\perp\right\|,
\end{aligned}$$

where the inequality above holds because $G'_u - G_u$ annihilates $\vec{1}$ from both sides. Lemma 29 implies that

$$\begin{aligned}
\left\|\left(\left(\prod_{i=0}^{u-1} P_\theta^{(sr)} G_i\right)\vec{1}\right)^\perp\right\| &\leq 2\cdot\sqrt{p_0p_1\cdot\left(4r^2+\frac{3}{2}\theta^2\right)}\cdot e^{p_0p_1u(2r^2-\theta^2/20)} \\
\left\|\left(\vec{1}^*\left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)}\right)\right)^\perp\right\| &\leq 2\cdot\sqrt{p_0p_1\cdot\left(4r^2+\frac{3}{2}\theta^2\right)}\cdot e^{p_0p_1(t-u)(2r^2-\theta^2/20)},
\end{aligned}$$

where the second equality above holds because $P_\theta^{(sr)}$ is diagonal and therefore symmetric, so we may apply Lemma 29 to bound the norm of the transpose of $\left(\vec{1}^*\left(\prod_{i=u+1}^{t} G_i P_\theta^{(sr)}\right)\right)^\perp$. Now combining the above inequalities gives

$$|\hat{g}^{(sr)}(\theta)| \leq 4 \cdot \|G'_u - G_u\| \cdot p_0 p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right) \cdot e^{p_0 p_1 t (2r^2 - \theta^2/20)}.$$

□

3.1.4 Bounds for the matrix $P_\theta^{(sr)}$

In this section, we prove Lemma 27 and Lemma 28, thereby bounding the quantities $|\vec{1}^* P_\theta^{(sr)} \vec{1}|$, $\|(P_\theta^{(sr)}\vec{1})^\perp\|$, and $\|F P_\theta^{(sr)} F\|$. We first bound the former two quantities in the proof of Lemma 27 below by deriving explicit expressions for these quantities, and then simplifying the resulting expressions with Taylor approximations.

Proof of Lemma 27. By definition $\vec{1}^* P_\theta^{(sr)} \vec{1} = p_0 e^{-p_1(sr - i\theta)} + p_1 e^{p_0(sr-i\theta)}$, so by Lemma 54,

$$\begin{aligned}
|\vec{1}^* P_\theta^{(sr)} \vec{1}| &= (p_0 e^{-p_1 sr} + p_1 e^{p_0 sr}) \left| \frac{p_0 e^{-p_1 sr}}{p_0 e^{-p_1 sr} + p_1 e^{p_0 sr}} + \frac{p_1 e^{p_0 sr}}{p_0 e^{-p_1 sr} + p_1 e^{p_0 sr}} e^{-i\theta} \right| \\
&\leq (p_0 e^{-p_1 sr} + p_1 e^{p_0 sr}) \left(1 - \frac{p_0 p_1 e^{(p_0 - p_1) sr}}{(p_0 e^{-p_1 sr} + p_1 e^{p_0 sr})^2} \cdot \frac{2}{\pi^2}\theta^2\right) \\
&= p_0 e^{-p_1 sr} + p_1 e^{p_0 sr} - \frac{2}{\pi^2} \cdot \frac{e^{(p_0 - p_1) sr}}{p_0 e^{-p_1 sr} + p_1 e^{p_0 sr}} \cdot p_0 p_1 \cdot \theta^2 \\
&\leq 1 + p_0 p_1 \cdot r^2 - \frac{2}{\pi^2 e^{1/2}} \cdot p_0 p_1 \cdot \theta^2,
\end{aligned}$$

where the final inequality above holds because $0 \leq r \leq 1/2$, so that $e^{-p_1 sr} \leq 1 - p_1 sr + (p_1 r)^2$, $e^{p_0 sr} \leq 1 + p_0 sr + (p_0 r)^2$, and $e^{(p_0-p_1)sr}/(p_0 e^{-p_1 sr} + p_1 e^{p_0 sr}) = 1/(p_0 e^{-p_0 sr} + p_1 e^{p_1 sr}) \geq e^{-1/2}$.

We now bound $\|(P_\theta^{(sr)}\vec{1})^\perp\|$. By definition

$$\begin{aligned}
(P_\theta^{(sr)}\vec{1})^\perp &= P_\theta^{(sr)}\vec{1} - \vec{1}\vec{1}^* P_\theta^{(sr)}\vec{1} \\
&= \frac{1}{\sqrt{n}} \sum_{v \in V} \delta_v e^{(sr - i\theta)(\mathrm{val}(v) - p_1)} - \frac{1}{\sqrt{n}} \sum_{v \in V} \delta_v (p_0 e^{-p_1(sr - i\theta)} + p_1 e^{p_0(sr - i\theta)}) \\
&= \frac{e^{-p_1(sr - i\theta)}(1 - e^{sr - i\theta})}{\sqrt{n}} \left(p_1 \sum_{v \in \mathrm{val}^{-1}(0)} \delta_v - p_0 \sum_{v \in \mathrm{val}^{-1}(1)} \delta_v\right),
\end{aligned}$$

so

$$\|(P_\theta^{(sr)}\vec{1})^\perp\| = e^{-p_1 sr} \cdot |1 - e^{sr - i\theta}| \cdot \sqrt{p_0 n \frac{p_1^2}{n} + p_1 n \frac{p_0^2}{n}} = \sqrt{p_0 p_1} \cdot e^{-p_1 sr} \cdot |1 - e^{sr - i\theta}|.$$

By analogous reasoning $\|(\vec{1}^* P_\theta^{(sr)})^\perp\| = \sqrt{p_0 p_1} \cdot e^{-p_1 sr} \cdot |1 - e^{sr - i\theta}|$. Now because $r \leq 1/2$,

$$\begin{aligned}
|1 - e^{sr - i\theta}|^2 &= (1 - e^{sr - i\theta})(1 - e^{sr + i\theta}) \\
&= 1 + e^{2sr} - 2e^{sr}\cos\theta \\
&\leq 1 + (1 + 2sr + (2r)^2) - 2(1 + sr)\left(1 - \frac{\theta^2}{2}\right) \qquad (10) \\
&\leq 4r^2 + \frac{3}{2}\theta^2,
\end{aligned}$$

so because $e^{-p_1 sr} \leq 1/2$,

$$\|(P_\theta^{(sr)}\vec{1})^\perp\| = \|(\vec{1}^* P_\theta^{(sr)})^\perp\| \leq \sqrt{e \cdot p_0 p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right)}.$$

□

We prove Lemma 28 below. The main idea in the proof is to bound $\|FP_\theta^{(sr)}F\|$ by the spectral norm of a 2×2 symmetric real matrix, which we in turn bound using Lemma 56.

Proof of Lemma 28. By definition

$$\|FP_\theta^{(sr)}F\| = \sup_{x,y \in \mathbb{C}^V : \|x\|=\|y\|=1} |x^* F P_\theta^{(sr)} F y|.$$

Decomposing $x = x^\parallel + x^\perp$ and $y = y^\parallel + y^\perp$ with $x^\parallel, y^\parallel \in \operatorname{span}\{\vec{1}\}$ and $x^\perp, y^\perp \in \vec{1}^\perp$ gives

$$\begin{aligned}|x^* F P_\theta^{(sr)} F y| &\leq |x^{\parallel *} P_\theta^{(sr)} y^\parallel| + |x^{\parallel *} P_\theta^{(sr)} \rho y^\perp| + |x^\perp \rho P_\theta^{(sr)} y^\parallel| + |x^\perp \rho P_\theta^{(sr)} \rho y^\perp| \\ &\leq \begin{pmatrix} \|x^\parallel\| & \|x^\perp\| \end{pmatrix} \begin{pmatrix} |\vec{1}^* P_\theta^{(sr)} \vec{1}| & \rho\|(\vec{1}^* P_\theta^{(sr)})^\perp\| \\ \rho\|(P_\theta^{(sr)}\vec{1})^\perp\| & \rho^2 \|P_\theta^{(sr)}\| \end{pmatrix} \begin{pmatrix} \|y^\parallel\| \\ \|y^\perp\| \end{pmatrix}\end{aligned}$$

Thus

$$\|FP_\theta^{(sr)}F\| \leq \left\| \begin{pmatrix} |\vec{1}^* P_\theta^{(sr)} \vec{1}| & \rho\|(\vec{1}^* P_\theta^{(sr)})^\perp\| \\ \rho\|(P_\theta^{(sr)}\vec{1})^\perp\| & \rho^2 \|P_\theta^{(sr)}\| \end{pmatrix} \right\|. \tag{11}$$

The lower right entry of the matrix on the right hand side above is at most $\rho^2 e^{1/2}$ because $\|P_\theta^{(sr)}\| = \max\{e^{-p_1 sr}, e^{p_0 sr}\} \leq e^{1/2}$ as $r \leq 1/2$. Applying Lemma 27 to bound the other three entries of this matrix gives that

$$\|FP_\theta^{(sr)}F\| \leq \left\| \begin{pmatrix} 1 + p_0p_1 \cdot r^2 - \frac{2}{\pi^2 e^{1/2}} \cdot p_0p_1 \cdot \theta^2 & \rho\sqrt{e \cdot p_0p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right)} \\ \rho\sqrt{e \cdot p_0p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right)} & \rho^2 e^{1/2} \end{pmatrix} \right\|.$$

Because $\rho = 1/10$, $p_0p_1 \leq 1/4$, and $|\theta| \leq \pi$,

$$\left(1 + p_0p_1 \cdot r^2 - \frac{2}{\pi^2 e^{1/2}} \cdot p_0p_1 \cdot \theta^2\right) - \rho^2 e^{1/2} \geq 1 - \frac{1}{2e^{1/2}} - \frac{e^{1/2}}{100} \geq \frac{2}{3},$$

so Lemma 56 implies that

$$\begin{aligned}\|FP_\theta^{(sr)}F\| &\leq \left(1 + p_0p_1 \cdot r^2 - \frac{2}{\pi^2 e^{1/2}} \cdot p_0p_1 \cdot \theta^2\right) + \frac{\frac{1}{100} \cdot e \cdot p_0p_1 \cdot \left(4r^2 + \frac{3}{2}\theta^2\right)}{2/3} \\ &\leq 1 + 2p_0p_1 \cdot r^2 - \frac{p_0p_1}{20} \cdot \theta^2.\end{aligned}$$

□

4 Permutation branching programs

We now turn to the problem of bounding the distance between the distributions of $f(\mathrm{val}(\mathrm{RW}_G^t))$ and $f(\mathrm{val}(\mathrm{RW}_J^t))$ for functions f on $[d]^t$ that can be computed by permutation branching programs, defined below.

Definition 30. A **permutation branching program** $\mathcal{B}$ of length t, width w, and degree d is a collection of functions $B_i : [d] \times [w] \to [w]$ for $i \in [t]$ such that for $b \in [d]$, each restriction $B_i(b) = B_i|_{\{b\}\times[w]} : [w] \to [w]$ is a permutation. The program is said to **compute** the function $B : [d]^t \to [w]$ defined by[1]

$$B(a) = (B_{t-1}(a_{t-1}) \circ \cdots \circ B_0(a_0))(0).$$

To demonstrate the definition above, we show in the example below that the function $\Sigma : [d]^t \to [t+1]^d$ from Definition 17 can be computed by a permutation branching program.

Example 31. Let $w = (t+1)^d$, and associate the set $[w]$ with $[t+1]^{[d]}$, such that the initial state $0 \in [w]$ of a width-w permutation branching program corresponds to $(0, \ldots, 0) \in [t+1]^{[d]}$. Define a permutation branching program $\mathcal{B}$ of length t, width w, and degree d so that for $i \in [t]$, $b \in [d]$, and $j \in [t+1]^{d-1}$,

$$B_i(b, j) = j + \delta_b,$$

where the addition $j + \delta_b$ is performed (mod $t+1$) in each component. Then $\mathcal{B}$ by definition computes the function Σ.

For a function B computed by a permutation branching program $\mathcal{B}$ such as in Example 31, our goal is to bound the difference between the distributions $B(\mathrm{val}(\mathrm{RW}_G^t))$ and $B(\mathrm{val}(\mathrm{RW}_J^t))$. To do so, we will first expess the computation of B linear-algebraically. For this purpose, we take the two related approaches presented in Section 4.1 and Section 4.2 respectively, each of which proves useful in certain settings.

We show in Section 5.3 that the our bounds in Section 4.1 below are tight in general, in the sense that no tighter bound holds for *all* permutation branching programs. Yet for the specific case of $B = \Sigma$, the bounds in Section 4.1 and Section 4.2 below are not tight. Hence we needed more specialzed methods in Section 3 to prove tight bounds for this case.

Remark. The labeling $\mathrm{val} : V \to [d]$ is not strictly necessary for our study of permutation branching programs. Specifically, we could consider degree-$|V|$ instead of degree-d permutation branching programs, and feed the vertices $(v_0, \ldots, v_{t-1}) \sim \mathrm{RW}_G^t$ directly as inputs to the program. While this simplified setup, which is equivalent to letting $\mathrm{val} : V \to V$ be the identity map, would be sufficient for our results in Section 4.1, the labeling val is useful in other results, particularly those regarding symmetric functions in Section 3. We therefore include the labeling $\mathrm{val} : V \to [d]$ throughout for consistency.

4.1 General total variation distance bound

In this section, we show the following bound for an arbitrary permutation branching program.

[1] Without loss of generality the initial state is assumed to be $0 \in [w]$.

Theorem 32. *Let $t \geq 1$, $w \geq 2$, $d \geq 2$ be integers and let $G = (V, E)$ be a regular λ-spectral expander with labeling* val $: V \to [d]$*. If $B : [d]^t \to [w]$ is computed by a permutation branching program $\mathcal{B}$ of length t, width w, and degree d, and if $\lambda < .1$, then*

$$d_{\ell_2}(B(\mathrm{val}(\mathrm{RW}_G^t)), B(\mathrm{val}(\mathrm{RW}_J^t))) \leq \lambda + 5\lambda^2.$$

Below, we implicitly let t, w, d, G, val be given as in Theorem 32. The Cauchy-Schwartz inequality implies the following corollary.

Corollary 33. *If $B : [d]^t \to [w]$ is computed by a permutation branching program $\mathcal{B}$, and $\lambda < .1$, then*

$$d_{TV}(B(\mathrm{val}(\mathrm{RW}_G^t)), B(\mathrm{val}(\mathrm{RW}_J^t))) \leq \frac{\sqrt{w}}{2}(\lambda + 5\lambda^2).$$

Although we have defined a permutation branching program $\mathcal{B}$ to compute a function $B : [d]^t \to [w]$, in the literature (see e.g. Hoza et al. [HPV21]) it is sometimes customary to designate a single state as an accepting state, which without loss of generality may be $0 \in [w]$. Then $\mathcal{B}$ is associated with the boolean-valued function $f : [d]^t \to \{0, 1\}$ given by $f(a) = \mathbb{1}_{B(a)=0}$. Theorem 32 then implies that the total variation distance, which equals the absolute difference in expectation, between $f(\mathrm{val}(\mathrm{RW}_G^t))$ and $f(\mathrm{val}(\mathrm{RW}_J^t))$ is $O(\lambda)$, as stated below.

Corollary 34. *For a permutation branching program $\mathcal{B}$ that computes B, let $f(a) = \mathbb{1}_{B(a)=0}$ be the boolean-valued function associated with $\mathcal{B}$. If $\lambda < .1$, then*

$$d_{TV}(f(\mathrm{val}(\mathrm{RW}_G^t)), f(\mathrm{val}(\mathrm{RW}_J^t))) \leq \lambda + 5\lambda^2.$$

Proof. Because the image of f is a subset of $\{0, 1\}$, by definition

$$\begin{aligned} d_{\mathrm{TV}}(f(\mathrm{val}(\mathrm{RW}_G^t)), f(\mathrm{val}(\mathrm{RW}_J^t))) &= |\Pr[f(\mathrm{val}(\mathrm{RW}_G^t)) = 1] - \Pr[f(\mathrm{val}(\mathrm{RW}_J^t)) = 1]| \\ &\leq d_{\ell_2}(B(\mathrm{val}(\mathrm{RW}_G^t)), B(\mathrm{val}(\mathrm{RW}_J^t))) \\ &\leq \lambda + 5\lambda^2, \end{aligned}$$

where the latter inequality above holds by Theorem 32. □

For comparison, Cohen et al. [CPTS21] showed that if $B : \{0, 1\}^t \to [w]$ is computed by some width-w, degree-2 permutation branching program $\mathcal{B}$ and the vertex label weights of G are $p_0 = p_1 = 1/2$, then $d_{\mathrm{TV}}(B(\mathrm{val}(\mathrm{RW}_G^t)), B(\mathrm{val}(\mathrm{RW}_J^t))) = O(w^4\sqrt{\lambda})$. Corollary 33 improves this total variation distance bound to $O(\sqrt{w} \cdot \lambda)$, while for the special case of programs with a single accepting state, Corollary 34 shows the even stronger total variation distance bound of $O(\lambda)$.

In Section 5.3, we show that Theorem 32 and Corollary 34 are tight, in the sense that there exist λ-spectral expanders G and permutation branching programs $\mathcal{B}$ that yield lower bounds of $\Omega(\lambda)$ to match the $O(\lambda)$ upper bounds above.

4.1.1 Proof of Theorem 32

To prove Theorem 32, for a random walk matrix $W \in \mathbb{R}^{V \times V}$, we show below how the distribution of $B(\mathrm{val}(\mathrm{RW}_W^t))$ can be computed by alternating applications of two matrices. We then apply these matrix expressions for B to bound the distance between the distributions of $B(\mathrm{val}(\mathrm{RW}_G^t))$ and $B(\mathrm{val}(\mathrm{RW}_J^t))$.

For the definition below, in a slight abuse of notation we let $B_i(b) \in \mathbb{R}^{[w] \times [w]}$ denote the permutation matrix for the permutation $B_i(b) : [w] \to [w]$, so that $(B_i(b))_{j',j} = \mathbb{1}_{B_i(b,j)=j'}$.

Definition 35. For a permutation branching program $\mathcal{B}$, define the **global increment matrix** $\tilde{P}$ **of** $\mathcal{B}$ to be the random walk matrix for the directed graph with vertex set $V \times [t] \times [w]$ such that each vertex (v, i, j) has a single outgoing edge to the vertex $(v, i+1, B_i(\mathrm{val}(v), j))$, where $i+1$ is taken $\pmod{t}$. Note that this digraph is 1-regular, so $\tilde{P}$ is a permutation matrix. More explicitly, $\tilde{P}$ is the linear operator on the vector space $\mathbb{R}^V \otimes \mathbb{R}^t \otimes \mathbb{R}^w$ given by

$$\tilde{P} = \sum_{v \in V, i \in [t]} \delta_v \delta_v^\top \otimes \delta_{i+1} \delta_i^\top \otimes B_i(\mathrm{val}(v)),$$

where $i+1$ is again taken $\pmod{t}$ above, so that $\delta_t = \delta_0 \in \mathbb{R}^t$.

For a random walk matrix $W \in \mathbb{R}^{V \times V}$ such as G or J, we let $\tilde{W} = W \otimes I$ for the identity matrix I of appropriate dimension, so that in this section, $\tilde{W} = W \otimes (I \otimes I)$ is the random walk matrix on vertex set $V \times [t] \times [w]$ that applies W to the first component and ignores the latter two components.

The lemma below shows that the distribution of $B(\mathrm{val}(\mathrm{RW}_W^t))$ can be computed by alternative applications of $\tilde{P}$ and $\tilde{W}$.

Lemma 36. *Let $\tilde{P}$ be the global increment matrix of a permutation branching program $\mathcal{B}$. For a random walk matrix $W \in \mathbb{R}^{V \times V}$ and for every $j \in [w]$,*

$$\Pr\big[B(\mathrm{val}(\mathrm{RW}_W^t)) = j\big] = (\vec{1} \otimes \delta_0 \otimes \delta_j)^\top \tilde{P} (\tilde{W}\tilde{P})^{t-1} (\vec{1} \otimes \delta_0 \otimes \delta_0). \tag{12}$$

Proof. Recall that $n = |V|$, so that here $\vec{1} = (1/\sqrt{n}, \ldots, 1/\sqrt{n}) \in \mathbb{R}^V$. Decomposing $W = \sum_{(v,v') \in [w]^2} W_{v',v} \delta_{v'} \delta_v^\top$ and then expanding the right hand side of (12) gives that

$$\begin{aligned}
&(\vec{1} \otimes \delta_0 \otimes \delta_j)^\top \tilde{P} (\tilde{W}\tilde{P})^{t-1} (\vec{1} \otimes \delta_0 \otimes \delta_0) \\
&= (\sqrt{n}\vec{1} \otimes \delta_0 \otimes \delta_j)^\top \tilde{P} \left(\left(\sum_{(v,v') \in [w]^2} W_{v',v} \delta_{v'} \delta_v^\top \otimes I \otimes I \right) \tilde{P} \right)^{t-1} \left(\frac{1}{\sqrt{n}} \vec{1} \otimes \delta_0 \otimes \delta_0 \right) \\
&= \sum_{(v_0, \ldots, v_{t-1}) \in V^t} (\sqrt{n}\vec{1} \otimes \delta_0 \otimes \delta_j)^\top \tilde{P} \left(\prod_{i \in [t-1]} (W_{v_{i+1}, v_i} \delta_{v_{i+1}} \delta_{v_i}^\top \otimes I \otimes I) \tilde{P} \right) \left(\frac{1}{\sqrt{n}} \vec{1} \otimes \delta_0 \otimes \delta_0 \right) \\
&= \sum_{(v_0, \ldots, v_{t-1}) \in V^t} (\delta_{v_{t-1}}^\top \otimes \delta_{t-1}^\top \otimes \delta_j^\top B_{t-1}(\mathrm{val}(v_{t-1}))) \\
&\qquad \cdot \left(\prod_{i \in [t-1]} W_{v_{i+1}, v_i} \delta_{v_{i+1}} \delta_{v_i}^\top \otimes \delta_{i+1} \delta_i^\top \otimes B_i(\mathrm{val}(v_i)) \right) \left(\frac{1}{n} \delta_{v_0} \otimes \delta_0 \otimes \delta_0 \right) \\
&= \sum_{(v_0, \ldots, v_{t-1}) \in V^t} \left(\prod_{i \in [t-1]} W_{v_{i+1}, v_i} \right) \frac{1}{n} \cdot \delta_j^\top (B_{t-1}(\mathrm{val}(v_{t-1})) \circ \cdots \circ B_0(\mathrm{val}(v_0))) \delta_0 \\
&= \sum_{(v_0, \ldots, v_{t-1}) \in V^t} \Pr\big[\mathrm{RW}_W^t = (v_0, \ldots, v_{t-1})\big] \cdot \delta_j^\top (B_{t-1}(\mathrm{val}(v_{t-1})) \circ \cdots \circ B_0(\mathrm{val}(v_0))) \delta_0 \\
&= \sum_{(v_0, \ldots, v_{t-1}) \in V^t} \Pr\big[\mathrm{RW}_W^t = (v_0, \ldots, v_{t-1})\big] \cdot \mathbb{1}_{B(\mathrm{val}(v_0, \ldots, v_{t-1})) = j} \\
&= \Pr\big[B(\mathrm{val}(\mathrm{RW}_W^t)) = j\big],
\end{aligned}$$

where the second and third equalities above hold by the definition of $\tilde{P}$. Note that the products over $i \in [t-1]$ in the expressions above are expanded from right to left. $\square$

Remark. The proof of Lemma 36 also works when the matrix W changes at each step in the random walk. That is, for a sequence $\mathcal{W} = (W_i)_{1 \leq i \leq t-1}$ of random walk matrices on shared vertex set V, then an exactly analogous proof implies that

$$\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})) = j\right] = (\vec{1} \otimes \delta_0 \otimes \delta_j)^\top \left(\prod_{i=1}^{t-1} \tilde{P}\tilde{W}_i\right) \tilde{P}(\vec{1} \otimes \delta_0 \otimes \delta_0), \tag{13}$$

where the product above expands from right-to-left. We presented the proof for the special case where all W_i are equal to simplify notation, but the generalization (13) will be useful in Section 4.2.

By Lemma 36, to bound the difference between the distributions of $B(\mathrm{val}(\mathrm{RW}^t_W))$ for $W = G, J$, it is sufficient to bound the difference between the matrices $\tilde{P}(\tilde{W}\tilde{P})^{t-1}$ for $W = G, J$. Below, we derive such a bound using singular-value approximations.

Proposition 37. *For the global increment matrix $\tilde{P}$ of any permutation branching program $\mathcal{B}$, if $\lambda < .1$ then*

$$\tilde{P}(\tilde{G}\tilde{P})^{t-1} \overset{sv}{\approx}_{\lambda+5\lambda^2} \tilde{P}(\tilde{J}\tilde{P})^{t-1}.$$

Proof. By Lemma 16, $G \overset{\mathrm{sv}}{\approx}_\lambda J$, so by Lemma 15, $\tilde{G} \overset{\mathrm{sv}}{\approx}_\lambda \tilde{J}$. Then because $\tilde{P}$ is a permutation matrix so that $\|\tilde{P}\| = 1$, Lemma 13 implies that $\tilde{G}\tilde{P} \overset{\mathrm{sv}}{\approx}_\lambda \tilde{J}\tilde{P}$, and then Lemma 14 gives that $(\tilde{G}\tilde{P})^{t-1} \overset{\mathrm{sv}}{\approx}_{\lambda+5\lambda^2} (\tilde{J}\tilde{P})^{t-1}$. One more application of Lemma 13 then gives the desired result. $\square$

We are now ready to prove Theorem 32.

Proof of Theorem 32. Let $g \in [-1, 1]^{[w]}$ denote the difference between the distributions of $B(\mathrm{val}(\mathrm{RW}^t_G))$ and $B(\mathrm{val}(\mathrm{RW}^t_J))$, so that $g_j = \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_G)) = j\right] - \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = j\right]$. Then by Lemma 36,

$$g = (\vec{1} \otimes \delta_0 \otimes I)^\top (\tilde{P}(\tilde{G}\tilde{P})^{t-1} - \tilde{P}(\tilde{J}\tilde{P})^{t-1})(\vec{1} \otimes \delta_0 \otimes \delta_0),$$

so by Proposition 37 and by the definition of singular-value approximations,

$$\|g\| \leq \|\tilde{P}(\tilde{G}\tilde{P})^{t-1} - \tilde{P}(\tilde{J}\tilde{P})^{t-1}\| \leq \lambda + 5\lambda^2.$$

$\square$

4.2 Program-structure-dependent total variation distance bound

Although we show that the bounds of Section 4.1, and in particular Theorem 32, are tight over the class of all permutation branching programs, smaller classes of permutation branching programs may have special structure resulting in a smaller total variation distance between the distributions of $B(\mathrm{val}(\mathrm{RW}^t_G))$ and $B(\mathrm{val}(\mathrm{RW}^t_J))$. In this section, we formalize this idea, to show that for certain permutation branching programs, this total variation distance decays exponentially in t. In particular, we show the following.

Theorem 38. *Fix integers $t \geq 1$, $w \geq 2$, and $d \geq 2$. For some $0 \leq \lambda \leq 1$, let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ be a sequence of regular λ-spectral expanders on a shared vertex set V. Fix a labeling $\mathrm{val} : V \to [d]$, and for $b \in [d]$, let p_b denote the fraction of vertices with label b. Let $\mathcal{B} = (B_i)_{i \in [t]}$ be a permutation branching program of length t, width w, and degree d that computes B. For $i \in [t]$, let $B_i^p = \sum_{b \in [d]} p_b B_i(b) \in \mathbb{R}^{w \times w}$. Then*

$$d_{\ell_2}(B(\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t)), \mathrm{Unif}([w])) \leq \prod_{i \in [t]} \left(1 - \left(1 - \sqrt{\lambda}\right)^2 (1 - \lambda(B_i^p))\right).$$

Below, we implicitly let the variables $t, w, d, \lambda, \mathcal{G}, \mathrm{val}, p, B_i^p$ be given as in Theorem 38.

Corollary 39. *For every permutation branching program $\mathcal{B}$ that computes B,*

$$d_{\ell_2}(B(\mathrm{val}(\mathrm{RW}_J^t)), \mathrm{Unif}([w])) \leq \prod_{i \in [t]} \lambda(B_i^p).$$

Proof. Because J is the random walk matrix for the unweighted complete graph with self-loops, which is undirected, the desired result follows by letting $\mathcal{G} = (J, \ldots, J)$ in the first inequality in Theorem 38, as $\lambda(J) = 0$. □

The bounds in Theorem 38 and Corollary 39 are most useful when $\lambda(B_i^p) < 1$ for many $i \in [t]$, as then these bounds can be used to show that $d_{\ell_2}(B(\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t)), \mathrm{Unif}([w]))$ decays exponentially in t. The following lemma characterizes when we have $\lambda(B_i^p) < 1$.

Lemma 40. *Let $\mathcal{B}$ be a permutation branching program, and let $\mathcal{B}$ be normalized so that $B_i(0) = I$ is the identity for every $i \in [t]$. Then $\lambda(B_i^p) \leq 1$, with equality iff there exists a set $\emptyset \subsetneq S \subsetneq [w]$ such that all permutations $B_i(b)$ for $b \in [d]$ preserve the set S. That is, $\lambda(B_i^p) < 1$ iff the subgroup of the symmetric group generated by $B_i(b)$ for $b \in [d]$ acts transitively on $[w]$.*

Proof. For every $y \in \mathbb{R}^w$, because each $B_i(b)$ is a permutation matrix, Jensen's inequality implies that

$$\|B_i^p y\| = \left\| \sum_{b \in [d]} p_b B_i(b) y \right\| \leq \sum_{b \in [d]} p_b \|B_i(b) y\| = \|y\|,$$

where the inequality above is an equality iff all $B_i(b)y$ are equal, as by assumption all p_b are nonzero. Therefore $\lambda(B_i^p) = \|B_i^p|_{\vec{1}^\perp}\| \leq 1$, with equality iff there exists some nonzero $y \in \vec{1}^\perp$ such that all $b \in [d]$ yield the same value of $B_i(b)y$. As all $B_i(b)$ are permutation matrices and therefore preserve $\vec{1}$, it follows that $\lambda(B_i^p) = 1$ iff there exists some $y \in \mathbb{R}^w \setminus \mathrm{span}\{\vec{1}\}$ such that all $b \in [d]$ yield the same value of $B_i(b)y$.

Now given $y \in \mathbb{R}^w \setminus \mathrm{span}\{\vec{1}\}$ such that all $b \in [d]$ yield the same value of $B_i(b)y$, choose $\emptyset \subsetneq S \subsetneq [w]$ to be the support of y_0 (or of any y_j), so that all $B_i(b)$ must preserve S because they preserve y. Conversely, given $\emptyset \subsetneq S \subsetneq [w]$ such that all $B_i(b)$ for $b \in [d]$ preserve S, then choose $y = \mathbb{1}_S \in \mathbb{R}^w \setminus \mathrm{span}\{\vec{1}\}$. As all $B_i(b)$ preserve S, they must also preserve the complement of S, and thus they preserve y, as desired. □

For example, if $B_i(0) = I$ and $B_i(1) = C_w$ is the directed cycle, then the subgroup generated by $B_i(1)$ alone already acts transitively on $[w]$, so Lemma 40 implies that $\lambda(B_i^p) < 1$ regardless of the choice of $B_i(b)$ for $b \geq 2$. Thus we have the following corollary.

Corollary 41. *For any fixed d, w, and p, define $B^t : [d]^t \to [w]$ to be the sum modulo w, that is $B^t(a) = \sum_{i\in[t]} a_i \pmod w$. Then there exists a constant $c = c(d, w, p, \lambda) < 1$ such that*

$$d_{TV}(B^t(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})), B^t(\mathrm{val}(\mathrm{RW}^t_J))) \le \sqrt{w} \cdot c^t.$$

Proof. Letting $B^t_i(b) = C^b_w$ for all $t \in \mathbb{N}$, $i \in [t]$ and $b \in [d]$, then the permutation branching program $(B^t_i)_{i\in t}$ by definition computes B^t. Because the group generated by the cycle permutation acts transitively on $[w]$, Lemma 40 implies that $\lambda((B^t_i)^p) = \lambda(\sum_{b\in[d]} p_b C^b_w) < 1$, so by Theorem 38 and Corollary 39, there exists a constant $c = c(d, w, p, \lambda) < 1$ such that

$$d_{\ell_2}(B^t(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})), \mathrm{Unif}([w])) \le c^t \text{ and } d_{\ell_2}(B^t(\mathrm{val}(\mathrm{RW}^t_J)), \mathrm{Unif}([w])) \le c^t. \tag{14}$$

Thus the total variation distance between $B^t(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ and $B^t(\mathrm{val}(\mathrm{RW}^t_J))$ is

$$d_{\mathrm{TV}}(B^t(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})), B^t(\mathrm{val}(\mathrm{RW}^t_J))) \le \frac{\sqrt{w}}{2} d_{\ell_2}(B^t(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})), B^t(\mathrm{val}(\mathrm{RW}^t_J))) \le \sqrt{w} \cdot c^t,$$

where the first inequality above holds by the Cauchy-Schwartz inequality as $d_{\mathrm{TV}} = \frac{1}{2} d_{\ell_1}$, and the second inequality holds by applying the triangle inequality with (14). ☐

Thus for any fixed d, w, p, λ, the sum function modulo w is fooled by a random walk on λ-spectral expanders up to a total variation distance error that is exponentially small in t. In the special case of $w = 2$ and $d = 2$, Corollary 41 states that expander walks fool the parity function up to an exponentially small error. This fact was previously known, and is a key part of Ta-Shma's breakthrough construction of almost optimal ϵ-balanced codes [TS17]. Guruswami and Kumar [GK21] also showed an analagous result to Corollary 41 for the special case of the sticky random walk on $n = 2$ vertices when $d = 2$ and $p = (1/2, 1/2)$.

4.2.1 Proof of Theorem 38

To prove Theorem 38, we will decompose the global increment matrix $\tilde{P}$ used in Section 4.1 into a sequence of local increment matrices, for which we can more easily take advantage of spectral properties.

Definition 42. For a permutation branching program $\mathcal{B}$, define the **local increment matrices** $\tilde{P}_0, \ldots, \tilde{P}_{t-1}$ **of** $\mathcal{B}$ to be linear operators on the vector space $\mathbb{R}^V \otimes \mathbb{R}^w$ given by

$$\tilde{P}_i = \sum_{v\in V} \delta_v \delta_v^\top \otimes B_i(\mathrm{val}(v)).$$

We can relate the $\tilde{P}_i$'s to the global increment matrix $\tilde{P}$ of Definition 35 as follows: Up to a reordering of basis elements, we have

$$\tilde{P} = \sum_{i\in[t]} \delta_{i+1}\delta_i^\top \otimes \tilde{P}_i, \tag{15}$$

so that the local increment matrices assemble to form the global increment matrix of $\mathcal{B}$. The following lemma is therefore a consequence of Lemma 36.

Lemma 43. *Let $\tilde{P}_0, \ldots, \tilde{P}_{t-1}$ be the local increment matrices of a permutation branching program $\mathcal{B}$, and let $\mathcal{W} = (W_i)_{1 \le i \le t-1}$ be a sequence of random walk matrices with shared vertex set V. Then for every $j \in [w]$,*

$$\Pr\big[B(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})) = j\big] = (\vec{1} \otimes \delta_j)^\top \left(\prod_{i=1}^{t-1} \tilde{P}_i \tilde{W}_i \right) \tilde{P}_0 (\vec{1} \otimes \delta_0), \tag{16}$$

where the product above expands from right to left, that is, $\prod_{i=1}^{t-1} \tilde{P}_i \tilde{W}_i = \tilde{P}_{t-1}\tilde{W}_{t-1} \cdots \tilde{P}_1 \tilde{W}_1$.

Proof. Recall that here $\tilde{W}_i = W_i \otimes I$. By (13),

$$\begin{aligned}
\Pr\big[B(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})) = j\big] &= (\vec{1} \otimes \delta_0 \otimes \delta_j)^\top \left(\prod_{i=1}^{t-1} \tilde{P}(W_i \otimes I \otimes I) \right) \tilde{P}(\vec{1} \otimes \delta_0 \otimes \delta_0) \\
&= (\delta_0 \otimes \vec{1} \otimes \delta_j)^\top \left(\prod_{i=1}^{t-1} \tilde{P}(I \otimes \tilde{W}_i) \right) \tilde{P}(\delta_0 \otimes \vec{1} \otimes \delta_0).
\end{aligned}$$

Now by (15), $\tilde{P} = \sum_{i \in [t]} \delta_{i+1}\delta_i^\top \otimes \tilde{P}_i$. Therefore for $i \in [t]$, the ith application (counting from right to left) of $\tilde{P}$ in the right hand side above takes as input a vector in $\delta_i \otimes \mathbb{R}^V \otimes \mathbb{R}^w$ and outputs a vector in $\delta_{i+1} \otimes \mathbb{R}^V \otimes \mathbb{R}^w$, where $i+1$ is taken $\pmod t$. Therefore the ith application of $\tilde{P}$ acts as $\delta_{i+1}\delta_i^\top \otimes \tilde{P}_i$, with all other terms in the sum in (15) vanishing. Thus

$$\begin{aligned}
\Pr\big[B(\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})) = j\big] &= (\delta_0 \otimes \vec{1} \otimes \delta_j)^\top \left(\prod_{i=1}^{t-1} \delta_{i+1}\delta_i^\top \otimes \tilde{P}_i \tilde{W}_i \right) (\delta_1 \delta_0^\top \otimes \tilde{P}_0)(\delta_0 \otimes \vec{1} \otimes \delta_0) \\
&= (\vec{1} \otimes \delta_j)^\top \left(\prod_{i=1}^{t-1} \tilde{P}_i \tilde{W}_i \right) \tilde{P}_0 (\vec{1} \otimes \delta_0).
\end{aligned}$$

□

Below we apply Lemma 43 to bound the distance from $B(\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))$ to the uniform distribution on $[w]$. However, we first need the following lemma, which intuitively says that if X is a good expander, then $\tilde{X}\tilde{P}_i\tilde{X}$ has similar expansion to $\tilde{J}\tilde{P}_i\tilde{J}$.

Lemma 44. *Let $(\tilde{P}_i)_{i \in [t]}$ be the local increment matrices of a permutation branching program $\mathcal{B}$. Then for every matrix $X \in \mathbb{C}^{V \times V}$ such that $\vec{1}^* X = \vec{1}^*$, $X\vec{1} = \vec{1}$, and $\lambda(X) = \|X|_{\vec{1}^\perp}\| \le 1$, letting $\tilde{X} = X \otimes I$, it holds that*

$$\|\tilde{X}\tilde{P}_i\tilde{X}|_{(\vec{1}\otimes\vec{1})^\perp}\| \le 1 - (1 - \lambda(X))^2(1 - \lambda(B_i^p)).$$

Proof. Because X has spectral expansion $\lambda(X)$, the matrix $E = (X - (1 - \lambda(X))J)/\lambda(X)$ has

$\|E\| = 1$ (see e.g. Lemma 4.19 of Vadhan [Vad12]), so

$$\begin{aligned}
&\|\tilde{X}\tilde{P}_i\tilde{X}|_{(\vec{1}\otimes\vec{1})^\perp}\| \\
&\quad = \|((1-\lambda(X))\tilde{J} + \lambda(X)\tilde{E}) \cdot \tilde{P}_i \cdot ((1-\lambda(X))\tilde{J} + \lambda(X)\tilde{E})|_{(\vec{1}\otimes\vec{1})^\perp}\| \\
&\quad \leq (1-\lambda(X))^2 \cdot \|\tilde{J}\tilde{P}_i\tilde{J}|_{(\vec{1}\otimes\vec{1})^\perp}\| \\
&\qquad + \lambda(X)\left((1-\lambda(X))\cdot\|\tilde{J}\tilde{P}_i\tilde{E}\| + (1-\lambda(X))\cdot\|\tilde{E}\tilde{P}_i\tilde{J}\| + \lambda(X)\cdot\|\tilde{E}\tilde{P}_i\tilde{E}\|\right) \\
&\quad \leq (1-\lambda(X))^2 \cdot \|\tilde{J}\tilde{P}_i\tilde{J}|_{(\vec{1}\otimes\vec{1})^\perp}\| + \lambda(X)\cdot(2-\lambda(X)) \\
&\quad = 1 - (1-\lambda(X))^2 \cdot (1 - \|\tilde{J}\tilde{P}_i\tilde{J}|_{(\vec{1}\otimes\vec{1})^\perp}\|).
\end{aligned}$$

Now by definition

$$\tilde{J}\tilde{P}_i\tilde{J} = \sum_{v\in V} \vec{1}\vec{1}^\top \delta_v\delta_v^\top \vec{1}\vec{1}^\top \otimes B_i(\mathrm{val}(v)) = \vec{1}\vec{1}^\top \otimes \frac{1}{n}\sum_{v\in V} B_i(\mathrm{val}(v)) = J \otimes B_i^p.$$

so the desired inequality follows because

$$\|J \otimes B_i^p|_{(\vec{1}\otimes\vec{1})^\perp}\| = \lambda(J\otimes B_p^p) = \max\{\lambda(J), \lambda(B_i^p)\} = \lambda(B_i^p)\}.$$

□

We apply Lemma 44 to prove Theorem 38 below with $X = J + \sqrt{\lambda}(I - J)$.

Proof of Theorem 38. Let $h \in [-1,1]^{[w]}$ denote the difference between the distribution of $B(\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t))$ and the uniform distribution on $[w]$, so that $h_j = \Pr\big[B(\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t)) = j\big] - 1/w$. Let $F = J + \sqrt{\lambda}(I - J)$ be the matrix that preserves $\vec{1}$ and scales $\vec{1}^\perp$ by a factor of $\sqrt{\lambda}$. Also for notational convenience let $G_0 = J$. Then by Lemma 43,

$$\begin{aligned}
h &= (\vec{1}\otimes I)^\top \left(\prod_{i=1}^{t-1} \tilde{P}_i\tilde{G}_i\right) \tilde{P}_0(\vec{1}\otimes\delta_0) - \frac{1}{\sqrt{w}}\vec{1} \\
&= (\vec{1}\otimes I)^\top \left(\prod_{i=1}^{t-1} \tilde{P}_i\tilde{G}_i\right) \tilde{P}_0\left(\vec{1}\otimes\left(\delta_0 - \frac{1}{\sqrt{w}}\vec{1}\right)\right) \\
&= (\vec{1}\otimes I)^\top \left(\prod_{i=0}^{t-1} \tilde{F}\tilde{P}_i\tilde{F}\cdot\tilde{F}^{-1}\tilde{G}_i\tilde{F}^{-1}\right) \left(\vec{1}\otimes\left(\delta_0 - \frac{1}{\sqrt{w}}\vec{1}\right)\right),
\end{aligned}$$

where the second equality above holds because by definition all $\tilde{P}_i$ and $\tilde{G}_i$ preserve the vector $\vec{1}\otimes\vec{1}$, and the third equality holds because $\tilde{F}$ and all $\tilde{G}_i$ preserve the subspace $\vec{1}\otimes\mathbb{R}^w$. Thus because $\delta_0 - (1/\sqrt{w})\vec{1} \in \vec{1}^\perp$ is by definition the orthogonal projection of δ_0 onto $\vec{1}^\perp$ so that $\|\delta_0 - (1/\sqrt{w})\vec{1}\| \leq \|\delta_0\| = 1$, and the matrices $\tilde{G}$ and $\tilde{P}_i$ preserve the subspace $(\vec{1}\otimes\vec{1})^\perp$ because they preserve $\vec{1}\otimes\vec{1}$

from both sides, it follows that

$$\begin{aligned}\|h\| &\le \left\| \prod_{i=0}^{t-1} \tilde{F}\tilde{P}_i\tilde{F} \cdot \tilde{F}^{-1}\tilde{G}_i\tilde{F}^{-1} \Big|_{(\vec{1}\otimes\vec{1})^\perp} \right\| \\ &\le \prod_{i=0}^{t-1} \|\tilde{F}\tilde{P}_i\tilde{F}|_{(\vec{1}\otimes\vec{1})^\perp}\| \cdot \|\tilde{F}^{-1}\tilde{G}_i\tilde{F}^{-1}\| \\ &\le \prod_{i=0}^{t-1} \left(1 - \left(1-\sqrt{\lambda}\right)^2 (1-\lambda(B_i^p))\right),\end{aligned}$$

where the final inequality above holds by Lemma 44 and because $\|\tilde{F}^{-1}\tilde{G}_i\tilde{F}^{-1}\| = \|F^{-1}G_iF^{-1}\| \le 1$ as $F^{-1}G_iF^{-1}$ preserves the subspaces $\mathrm{span}\{\vec{1}\}$ and $\vec{1}^\perp$, and has norm at most $\sqrt{\lambda}^{-1} \cdot \lambda \cdot \sqrt{\lambda}^{-1} = 1$ on $\vec{1}^\perp$ because $F^{-1} = J + \sqrt{\lambda}^{-1}(I-J)$. □

5 Lower bounds

In this section, we show lower bounds to match our upper bounds in previous sections. We prove these lower bounds using the sticky random walk, which is a particularly simple λ-spectral expander.

5.1 Sticky random walk

In this section, we consider the case where $d = 2$, and introduce the λ-sticky, p-biased random walk, which is a random walk with spectral expansion λ and label weights p. The sticky random walk can be thought if as a "canonical" λ-spectral expander, and will be useful for proving lower bounds to match our upper bounds from previous sections. The special case of the sticky walk on $|V| = 2$ vertices with $p_0 = p_1 = 1/2$ was studied extensively by Guruswami and Kumar [GK21]. To prove lower bounds that match our more general upper bounds in Section 3 and Section 4.1, we generalize the sticky walk to arbitrary p.

Definition 45. Fix a vertex set $V = V_0 \sqcup V_1$ with labeling $\mathrm{val} : V \to \{0,1\}$ given by $\mathrm{val}(v) = b$ for $v \in V_b$, so that $p_0 = |V_0|/|V|$ and $p_1 = |V_1|/|V|$. For subsets $A, B \subseteq V$, let $J_{A,B} \in \mathbb{R}^{A\times B}$ denote the matrix with all entries equal to $1/|A|$. For $0 \le \lambda \le 1$, define the **λ-sticky, p-biased random walk matrix** $G_{\lambda,p} \in \mathbb{R}^{V\times V}$ by

$$G_{\lambda,p} = \begin{pmatrix} (p_0 + p_1\lambda)J_{V_0,V_0} & (p_0 - p_0\lambda)J_{V_0,V_1} \\ (p_1 - p_1\lambda)J_{V_1,V_0} & (p_1 + p_0\lambda)J_{V_1,V_1} \end{pmatrix}.$$

That is, $G_{\lambda,p}$ treats all vertices within V_b identically for each $b = 0, 1$, and if (v, v') represents a 1-step random walk on $G_{\lambda,p}$, then the transition probabilities are

$$\begin{aligned}\Pr\left[v' \in V_0 | v \in V_0\right] &= p_0 + p_1\lambda = (1-\lambda)p_0 + \lambda \\ \Pr\left[v' \in V_0 | v \in V_1\right] &= p_0 - p_0\lambda = (1-\lambda)p_0 \\ \Pr\left[v' \in V_1 | v \in V_0\right] &= p_1 - p_1\lambda = (1-\lambda)p_1 \\ \Pr\left[v' \in V_1 | v \in V_1\right] &= p_1 + p_0\lambda = (1-\lambda)p_1 + \lambda.\end{aligned}$$

We show below that the λ-sticky random walk is indeed a λ-spectral expander.

Lemma 46. *$\lambda(G_{\lambda,p}) = \lambda$.*

Proof. By definition

$$G_{\lambda,p} = (1-\lambda)J_{V,V} + \lambda W$$

for

$$W = \begin{pmatrix} J_{V_0,V_0} & 0 \\ 0 & J_{V_1,V_1} \end{pmatrix}.$$

We have $\|W\| = 1$, as W acts as J on the orthogonal subspaces $\mathbb{R}^{V_0}$ and $\mathbb{R}^{V_1}$ of $\mathbb{R}^V$. Thus $\lambda(G_{\lambda,p}) \leq \lambda$. The opposite inequality follows from the fact that $p_1 \mathbb{1}_{V_0} - p_0 \mathbb{1}_{V_1} \in \vec{1}^{\perp}$ is an eigenvector of $G_{\lambda,p}$ with eigenvalue λ, as is evident from the decomposition of $G_{\lambda,p}$ above. $\square$

5.2 Lower bound for symmetric functions

In this section, we use the sticky random walk $G_{\lambda,p}$ to show that the the $c = 0$ case of Corollary 19, or equivalently the $d = 2$ case of Corollary 21, is asymptotically tight. Specifically, we show the following.

Theorem 47. *For every $p = (p_0, p_1)$, $0 < \lambda < 1$, and $t \geq 10^{10}/(\lambda^2(1-\lambda)^4(p_0p_1)^3)$, then*

$$d_{TV}(\Sigma \operatorname{val}(\mathrm{RW}^t_{G_{\lambda,p}}), \Sigma \operatorname{val}(\mathrm{RW}^t_J)) \geq \frac{\lambda}{20}.$$

Theorem 47 implies that for every λ, p with $d = 2$, taking $G = G_{\lambda,p}$ and letting t be sufficiently large ensures that the total variation distance between $\Sigma \operatorname{val}(\mathrm{RW}^t_G)$ and $\Sigma \operatorname{val}(\mathrm{RW}^t_J)$ is $\Omega(\lambda)$. Thus Corollary 21, which showed an upper bound of $O(\lambda)$ for this total variation distance for all λ-spectral-expanders, is tight for large t.

Theorem 47 generalizes a similar result of Guruswami and Kumar [GK21] for the special case of $p_0 = p_1 = 1/2$, and indeed our proof method is similar to theirs. However, whereas Guruswami and Kumar [GK21] derived their lower bound by proving a special case of the CLT, we instead cite a similar but more general CLT result of Kloeckner [Klo19].

Cohen et al. [CMP+21] also showed a similar lower bound that is incomparable to Theorem 47. Specifically, for the case of $p_0 = p_1 = 1/2$, Cohen et al. [CMP+21] presented a λ-spectral expander G for which the total variation distance between $\Sigma \operatorname{val}(\mathrm{RW}^t_G)$ and $\Sigma \operatorname{val}(\mathrm{RW}^t_J)$ is $\Omega(\lambda)$ for all t. In contrast, Theorem 47 considers arbitrary p, but only sufficiently large t.

The main idea to prove Theorem 47 is that by the Markov chain CLT, as $t \to \infty$ then $((\Sigma \operatorname{val}(\mathrm{RW}^t_{G_{\lambda,p}}))_1 - p_1 t)/\sqrt{p_0 p_1 t}$ converges in distribution (that is, in Kolmogorov distance) to a normal distribution with variance $(1+\lambda)/(1-\lambda)$. In contrast, the CLT implies that the normalized binomial distribution $((\Sigma \operatorname{val}(\mathrm{RW}^t_J))_1 - p_1 t)/\sqrt{p_0 p_1 t}$ converges to a normal distribution with variance 1. Theorem 47 then follows because the distance between these two normals is $\Omega(\lambda)$.

The Markov chain CLT relies on the following asymptotic notion of variance.

Definition 48. For a sequence $X = (X^t)^{t \in \mathbb{N}}$ of probability distributions over $\mathbb{R}$, the **asymptotic variance** of X is

$$\sigma^2(X) = \lim_{t \to \infty} \frac{1}{t} \operatorname{Var}(X^t).$$

When $G = (V, E)$ is a graph with labeling $\mathrm{val} : V \to [2]$, we write $\sigma^2(\Sigma\,\mathrm{val}(\mathrm{RW}_G^t))$ to denote the asymptotic variance $\sigma^2((\Sigma\,\mathrm{val}(\mathrm{RW}_G^t))_0) = \sigma^2((\Sigma\,\mathrm{val}(\mathrm{RW}_G^t))_1)$. The following formula for this expression is well known; it for instance is a special case of the definition of asymptotic variance in Kloeckner [Klo19].

Lemma 49. *Let $G = (V, E)$ be a regular graph with labeling $\mathrm{val} : V \to [2]$ that assigns each label $b \in [2]$ to p_b-fraction of the n vertices. Viewing val and $\mathrm{val} - p_1 : V \to \mathbb{R}$ as vectors in $\mathbb{R}^V$, then*

$$\sigma^2(\Sigma\,\mathrm{val}(\mathrm{RW}_G^t)) = p_0 p_1 + 2\sum_{i=1}^{\infty} \frac{1}{n}(\mathrm{val} - p_1)^\top G^i\,\mathrm{val}\,.$$

Letting $\sigma^2 = \sigma^2(\Sigma\,\mathrm{val}(\mathrm{RW}_{G_{\lambda,p}}^t))$, then the Markov chain CLT implies that $((\Sigma\,\mathrm{val}(\mathrm{RW}_{G_{\lambda,p}}^t))_1 - p_1 t)/(\sigma\sqrt{t})$ converges in distribution to the standard normal $\mathcal{N}(0, 1)$ as $t \to \infty$. We specifically apply a Berry-Esseen bound of Kloeckner [Klo19], which quantifies the rate of convergence to the normal. As Kloeckner [Klo19] considers a significantly more general class of Markov chains than is needed here, we simply state a direct consequence of their result as it applies to the sticky random walk.

Theorem 50 (Follows from Theorem C of [Klo19])**.** *Let $\sigma^2 = \sigma^2(\Sigma\,\mathrm{val}(\mathrm{RW}_{G_{\lambda,p}}^t))$. Then*

$$d_{Kol}\left(\frac{1}{\sigma\sqrt{t}}\left((\Sigma\,\mathrm{val}(\mathrm{RW}_{G_{\lambda,p}}^t))_1 - p_1 t\right), \mathcal{N}(0, 1)\right) \le \frac{1000 \cdot \max\{1, 1/\sigma^3\}}{(1-\lambda)^2\sqrt{t}}.$$

To apply Theorem 50 to prove Theorem 47, we first compute the asymptotic variance for the sticky random walk.

Lemma 51. *Define $G_{\lambda,p}$ and val as in Definition 45. Then*

$$\sigma^2(\Sigma\,\mathrm{val}(\mathrm{RW}_{G_{\lambda,p}}^t)) = p_0 p_1 \cdot \frac{1+\lambda}{1-\lambda}.$$

Proof. View val and $\mathrm{val} - p_1 : V \to \mathbb{R}$ as vectors in $\mathbb{R}^V$. Then $\mathrm{val} - p_1 \in \vec{1}^\perp$ is an eigenvalue of $G_{\lambda,p} = (1-\lambda)J_{V,V} + \lambda \begin{pmatrix} J_{V_0,V_0} & 0 \\ 0 & J_{V_1,V_1} \end{pmatrix}$ with eigenvalue λ, so by Lemma 49,

$$\begin{aligned}
\sigma^2(\Sigma\,\mathrm{val}(\mathrm{RW}_{G_{\lambda,p}}^t)) &= p_0 p_1 + 2\sum_{i=1}^{\infty} \frac{1}{n}(\mathrm{val} - p_1)^\top G^i(\mathrm{val} - p_1) \\
&= p_0 p_1 + 2\sum_{i=1}^{\infty} \lambda^i \cdot \frac{\|\mathrm{val} - p_1\|^2}{n} \\
&= p_0 p_1 + 2 \cdot \frac{\lambda}{1-\lambda} \cdot p_0 p_1 \\
&= p_0 p_1 \cdot \frac{1+\lambda}{1-\lambda}.
\end{aligned}$$

□

We now prove Theorem 47 using Lemma 51.

Proof of Theorem 47. By Theorem 50 and Lemma 51,

$$d_{\text{Kol}}\left(\frac{1}{\sqrt{p_0p_1t}}\left((\Sigma\,\text{val}(\text{RW}^t_{G_{\lambda,p}}))_1 - p_1t\right),\ \mathcal{N}\left(0, \frac{1+\lambda}{1-\lambda}\right)\right) \leq \frac{1000}{(1-\lambda)^2(p_0p_1)^{3/2}\sqrt{t}}.$$

Furthermore, becuase $G_{0,p} = J$ by definition, Theorem 50 also implies that

$$d_{\text{Kol}}\left(\frac{1}{\sqrt{p_0p_1t}}\left((\Sigma\,\text{val}(\text{RW}^t_J))_1 - p_1t\right), \mathcal{N}(0,1)\right) \leq \frac{1000}{(p_0p_1)^{3/2}\sqrt{t}}.$$

Now by definition

$$\begin{aligned}
d_{\text{Kol}}\left(\mathcal{N}\left(0, \frac{1+\lambda}{1-\lambda}\right), \mathcal{N}(0,1)\right) &\geq \Pr[\mathcal{N}(0,1) \leq 1] - \Pr\left[\mathcal{N}\left(0, \frac{1+\lambda}{1-\lambda}\right) \leq 1\right] \\
&= \Pr[\mathcal{N}(0,1) \leq 1] - \Pr\left[\mathcal{N}(0,1) \leq \sqrt{\frac{1-\lambda}{1+\lambda}}\right] \\
&= \int_{\sqrt{\frac{1-\lambda}{1+\lambda}}}^{1} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\
&\geq \left(1 - \sqrt{\frac{1-\lambda}{1+\lambda}}\right) \frac{1}{\sqrt{2\pi}} e^{-1/2} \\
&\geq \frac{\lambda}{2\sqrt{2\pi e}},
\end{aligned}$$

where the final inequality above holds because $\sqrt{1-a} \leq 1 - a/2$ for $0 \leq a \leq 1$. Now by definition total variation distance dominates Kolmogorov distance, and both are invariant under scaling and translation, so applying the triangle inequality with the three inequalities above gives that

$$\begin{aligned}
&d_{\text{TV}}(\Sigma\,\text{val}(\text{RW}^t_{G_{\lambda,p}}), \Sigma\,\text{val}(\text{RW}^t_J)) \\
&\quad = d_{\text{TV}}\left(\frac{1}{\sqrt{p_0p_1t}}\left((\Sigma\,\text{val}(\text{RW}^t_{G_{\lambda,p}}))_1 - p_1t\right), \frac{1}{\sqrt{p_0p_1t}}\left((\Sigma\,\text{val}(\text{RW}^t_J))_1 - p_1t\right)\right) \\
&\quad \geq d_{\text{Kol}}\left(\frac{1}{\sqrt{p_0p_1t}}\left((\Sigma\,\text{val}(\text{RW}^t_{G_{\lambda,p}}))_1 - p_1t\right), \frac{1}{\sqrt{p_0p_1t}}\left((\Sigma\,\text{val}(\text{RW}^t_J))_1 - p_1t\right)\right) \\
&\quad \geq \frac{\lambda}{2\sqrt{2\pi e}} - \frac{1000}{(1-\lambda)^2(p_0p_1)^{3/2}\sqrt{t}} - \frac{1000}{(p_0p_1)^{3/2}\sqrt{t}} \\
&\quad \geq \frac{\lambda}{2\sqrt{2\pi e}} - \frac{2000}{(1-\lambda)^2(p_0p_1)^{3/2}\sqrt{t}},
\end{aligned}$$

Then the desired result follows because the right hand side above is at least $\lambda/20$ when

$$t \geq \frac{10^{10}}{\lambda^2(1-\lambda)^4(p_0p_1)^3}.$$

□

5.3 Lower bound for general permutation branching programs

In this section, we use the sticky random walk $G_{\lambda,p}$ to show that Theorem 32 and Corollary 34 are tight. In particular, in Proposition 53 below, we show below that for every $p = (p_0, p_1)$ there exists a degree-2 permutation branching program $\mathcal{B}$ For which $|\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_{G_{\lambda,p}})) = 0\right] - \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = 0\right]| = \Omega(\lambda)$. This lower bound matches the upper bound of $O(\lambda)$ from Theorem 32 and Corollary 34. Note that while we restrict to the $d = 2$ case in this section, the results extend naturally to the case where $d > 2$, as we may group all labels in $[d]$ into two sets, which can be renamed with labels 0 and 1 respectively, and the lower bounds still hold.

Before presenting Proposition 53, we first present the following basic result showing that if p_0, p_1 are bounded away from 0, then there exists a degree-2 permutation branching program $\mathcal{B}$ of any length $t \geq 2$ and any width $w \geq 2$ such that $|\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_{G_{\lambda,p}})) = 0\right] - \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = 0\right]| = \Omega(\lambda)$.

Proposition 52. *Let $G = G_{\lambda,p}$ be the sticky random walk for some $0 \leq \lambda \leq 1$ and $p = (p_0, p_1)$. For every $t \geq 2$ and $w \geq 2$, there exists a permutation branching program $\mathcal{B}$ of length t, width w, and degree $d = 2$ such that*

$$\left|\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_G)) = 0\right] - \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = 0\right]\right| = 2p_0p_1\lambda.$$

Proof. Define $\mathcal{B}$ to be the permutation branching program of length $t \geq 2$, width $w \geq 2$, and degree $d = 2$ with

$$B_i(b,j) = \begin{cases} j + b \pmod{w}, & i = 0 \\ j + b - 1 \pmod{w}, & i = 1 \\ j, & \text{otherwise.} \end{cases}$$

That is, $B_0(1) = C_w$ is the forwards cycle permutation, $B_1(0) = C_w^{-1}$ is the backwards cycle permutation, and all other $B_i(b)$ are the identity permutation. Then for an input $a \in \{0,1\}^t$, it holds that $B(a) = 0$ iff $a_0 + a_1 = 1$, or equivalently, $B(a) = 0$ iff $a_0 \neq a_1$. Thus

$$\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_G)) = 0\right] = \Pr\left[(\mathrm{RW}^t_G)_0 \neq (\mathrm{RW}^t_G)_1\right] = p_0(p_1 - p_1\lambda) + p_1(p_0 - p_0\lambda) = 2p_0p_1(1-\lambda)$$
$$\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = 0\right] = \Pr\left[(\mathrm{RW}^t_J)_0 \neq (\mathrm{RW}^t_J)_1\right] = p_0p_1 + p_1p_0 = 2p_0p_1.$$

Thus $\left|\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_G)) = 0\right] - \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = 0\right]\right| = 2p_0p_1\lambda$. □

The lower bound in Proposition 52 of $\Omega(p_0p_1\lambda)$ fails to meet the upper bounds of $O(\lambda)$ in Theorem 32 and Corollary 34 when p_0p_1 is small. The following result addresses this issue by improving the lower bound to $\Omega(\lambda)$, although it requires the permutation branching program length and width to be $\Omega(1/p_0p_1)$, whereas Proposition 52 only required length and width ≥ 2.

Proposition 53. *Let $G = G_{\lambda,p}$ be the sticky random walk for some $0 \leq \lambda \leq 1$ and $p = (p_0, p_1)$. There exists a permutation branching program $\mathcal{B}$ of degree $d = 2$, length $t = \lfloor 1/\min\{p_0, p_1\} \rfloor + 1$, and width $w = t + 1$ such that*

$$\left|\Pr\left[B(\mathrm{val}(\mathrm{RW}^t_G)) = 0\right] - \Pr\left[B(\mathrm{val}(\mathrm{RW}^t_J)) = 0\right]\right| \geq \frac{\lambda}{2e^2}.$$

Proof. Assume without loss of generality that $p_1 \leq 1/2$, as otherwise the labeling val could be replaced with $1 - \text{val}$. Let $\mathcal{B}$ be the length-t, width-w, degree-2 permutation branching program given by $B_i(b, j) = j + b$, so that $\mathcal{B}$ computes the sum function $B : \{0,1\}^t \to [w]$ defined by $B(a) = (\Sigma a)_1 = \sum_{i \in [t]} a_i$. Then

$$\begin{aligned}\Pr\big[B(\text{val}(\text{RW}_G^t)) = 0\big] &= \Pr_{(v_0,\dots,v_{t-1}) \sim \text{RW}_G^t}[v_0, \dots, v_{t-1} \in V_0] \\ &= p_0(p_0 + p_1\lambda)^{t-1},\end{aligned}$$

while

$$\begin{aligned}\Pr\big[B(\text{val}(\text{RW}_J^t)) = 0\big] &= \Pr_{(v_0,\dots,v_{t-1}) \sim \text{RW}_J^t}[v_0, \dots, v_{t-1} \in V_0] \\ &= p_0^t.\end{aligned}$$

Thus

$$\begin{aligned}&\big|\Pr\big[B(\text{val}(\text{RW}_G^t)) = 0\big] - \Pr\big[B(\text{val}(\text{RW}_J^t)) = 0\big]\big| \\ &= p_0(p_0 + p_1\lambda)^{t-1} - p_0^t \\ &= p_0^t \cdot \left(\left(1 + \frac{p_1}{p_0}\lambda\right)^{t-1} - 1\right) \\ &\geq p_0^t \cdot \left(1 + (t-1)\frac{p_1}{p_0}\lambda - 1\right) \\ &= (1 - p_1)^{t-1} \cdot (t-1)p_1\lambda \\ &\geq e^{-2(t-1)p_1} \cdot (t-1)p_1\lambda \\ &\geq e^{-2} \cdot \frac{1}{2}\lambda,\end{aligned}$$

where the second inequality above holds because $\log(1 - p_1) \geq -2p_1$ as $p_1 \leq 1/2$, and the third inequality holds because $1/2 \leq (t-1)p_1 \leq 1$ as $p_1 \leq 1/2$ and $t - 1 = \lfloor 1/p_1 \rfloor$. ☐

The $\Omega(\lambda)$ lower bounds above match the $O(\lambda)$ upper bounds in Theorem 32 and Corollary 34. However, for permutation branching programs of large width w, it remains an open problem to resolve the gap between these $\Omega(\lambda)$ lower bounds and our $O(\sqrt{w} \cdot \lambda)$ upper bound on total variation distance from Corollary 33.

6 Acknowledgments

S.V. thanks Dean Doron, Raghu Meka, Omer Reingold, and Avishay Tal, conversations with whom inspired some of this research. We also thank the anonymous CCC reviewers for helpful comments that have improved the presentation.

References

[Ahm20] AmirMahdi Ahmadinejad. *Computing stationary distributions: perron vectors, random walks, and ride-sharing competition.* PhD thesis, Stanford University, Stanford, California, 2020.

[AKM+20] AmirMahdi Ahmadinejad, Jonathan Kelner, Jack Murtagh, John Peebles, Aaron Sidford, and Salil Vadhan. High-precision Estimation of Random Walks in Small Space. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1295–1306, November 2020. ISSN: 2575-8454.

[APSV21] AmirMahdi Ahmadinejad, John Peebles, Aaron Sidford, and Salil Vadhan. Personal Communication, 2021.

[BATS11] Avraham Ben-Aroya and Amnon Ta-Shma. A Combinatorial Construction of Almost-Ramanujan Graphs Using the Zig-Zag Product. *SIAM Journal on Computing*, 40(2):267–290, January 2011.

[CMP+21] Gil Cohen, Dor Minzer, Shir Peleg, Aaron Potechin, and Amnon Ta-Shma. Expander Random Walks: The General Case and Limitations. *Electronic Colloquium on Computational Complexity*, 2021.

[CPTS21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a Fourier-analytic approach. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1643–1655, New York, NY, USA, June 2021. Association for Computing Machinery.

[Gil98] David Gillman. A Chernoff Bound for Random Walks on Expander Graphs. *SIAM Journal on Computing*, 27(4):1203–1220, August 1998.

[GK21] Venkatesan Guruswami and Vinayak M. Kumar. Pseudobinomiality of the Sticky Random Walk. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48:1–48:19, Dagstuhl, Germany, 2021. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Gol22] Louis Golowich. A Berry-Esseen Theorem for Expander Walks. *Forthcoming*, 2022.

[Gur04] Venkatesan Guruswami. Guest column: error-correcting codes and expander graphs. *ACM SIGACT News*, 35(3):25–41, September 2004.

[Hea08] Alexander D. Healy. Randomness-Efficient Sampling within NC1. *computational complexity*, 17(1):3–37, April 2008.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[HPV21] William M. Hoza, Edward Pyne, and Salil Vadhan. Pseudorandom Generators for Unbounded-Width Permutation Branching Programs. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of Computing*, STOC '94, pages 356–364, New York, NY, USA, May 1994. Association for Computing Machinery.

[JM21] Akhil Jalan and Dana Moshkovitz. Near-Optimal Cayley Expanders for Abelian Groups. *arXiv:2105.01149 [cs]*, May 2021. arXiv: 2105.01149 version: 1.

[Klo19] Benoît Kloeckner. Effective Berry–Esseen and concentration bounds for Markov chains with a spectral gap. *The Annals of Applied Probability*, 29(3):1778–1807, June 2019.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, September 1988.

[Mar73] Grigorii Aleksandrovich Margulis. Explicit constructions of expanders. *Problemy Peredachi Informatsii*, 9(4):71–80, 1973. Publisher: Russian Academy of Sciences, Branch of Informatics, Computer Equipment and . . . .

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy Waves, the Zig-Zag Graph Product, and New Constant-Degree Expanders. *The Annals of Mathematics*, 155(1):157, January 2002.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, Montreal Canada, June 2017. ACM.

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, December 2012.

A Technical lemmas

The following technical lemmas are used at various points in the paper.

Lemma 54. *Let $p = (p_0, p_1) \in (0, 1)^2$ with $p_0 + p_1 = 1$. For all $-\pi \le \theta \le \pi$ with $\theta \ne 0$,*

$$\frac{p_0 p_1}{2} \le \frac{1 - |p_0 + p_1 e^{i\theta}|}{|1 - e^{i\theta}|^2} \le p_0 p_1. \tag{17}$$

Furthermore,

$$p_0 p_1 \frac{2}{\pi^2} \theta^2 \le 1 - |p_0 + p_1 e^{i\theta}| \le p_0 p_1 \theta^2. \tag{18}$$

Proof. For the first inequality, by definition

$$\begin{aligned}
\frac{1 - |p_0 + p_1 e^{i\theta}|}{|1 - e^{i\theta}|^2} &= \frac{1 - \sqrt{(p_0 + p_1 \cos\theta)^2 + (p_1 \sin\theta)^2}}{(1 - \cos\theta)^2 + (\sin\theta)^2} \\
&= \frac{1 - \sqrt{(p_0 + p_1)^2 - 2p_0 p_1 (1 - \cos\theta)}}{2 - 2\cos\theta} \\
&= \frac{1 - (1 - 2p_0 p_1 (1 - \cos\theta))}{(2 - 2\cos\theta)(1 + \sqrt{1 - 2p_0 p_1 (1 - \cos\theta)})} \\
&= \frac{p_0 p_1}{1 + \sqrt{1 - 2p_0 p_1 (1 - \cos\theta)}}.
\end{aligned}$$

The denominator of the right hand side above is always at least 1, which implies the desired upper bound of p_0p_1. The right hand side above is minimized by letting $\theta \to 0$, which gives the desired lower bound of $p_0p_1/2$.

For the second inequality, because $|1-e^{i\theta}|^2 = 2(1-\cos\theta)$, and for all $-\pi \le \theta \le \pi$ it holds that $\frac{2}{\pi^2}\theta^2 \le 1-\cos\theta \le \frac{1}{2}\theta^2$, it follows that

$$\frac{4}{\pi^2}\theta^2 \le |1-e^{i\theta}|^2 \le \theta^2.$$

Multiplying this inequality with (17) gives (18). □

Lemma 55. *For all $a, b > 0$ and $\rho > 0$,*

$$ab \le \frac{1}{2}\left(\rho a^2 + \frac{1}{\rho}b^2\right).$$

Proof. The result follows from the inequality

$$\rho a^2 + \frac{1}{\rho}b^2 - 2ab = \left(\sqrt{\rho}a - \frac{1}{\sqrt{\rho}}b\right)^2 \ge 0.$$

□

Lemma 56. *For all $a, b, c \in \mathbb{R}$ such that $a + c \ge 0$,*

$$\left\|\begin{pmatrix} a & b \\ b & c \end{pmatrix}\right\| = \frac{a+c+\sqrt{(a-c)^2+4b^2}}{2}.$$

In particular, if it also holds that $a > c$, then

$$\left\|\begin{pmatrix} a & b \\ b & c \end{pmatrix}\right\| \le a + \frac{b^2}{a-c}.$$

Proof. Let $M = \left\|\begin{pmatrix} a & b \\ b & c \end{pmatrix}\right\|$. By the spectral theorem, M has an orthonormal eigenbasis. Let $\mu_1 \ge \mu_2$ be the eigenvalues of M. Then $\|M\| = \max\{|\mu_1|, |\mu_2|\}$, and as $\mu_1 + \mu_2 = \operatorname{Tr} M = a + c \ge 0$, it follows that $\|M\| = \mu_1$. The eigenvalues μ_1, μ_2 are roots of the characteristic polynomial

$$p_M(\mu) = \mu^2 - (a+c)\mu + (ac - b^2),$$

so

$$\|M\| = \mu_1 = \frac{a+c+\sqrt{(a-c)^2+4b^2}}{2},$$

which proves the first part of the lemma. For the second part, if $a > c$, then because $\sqrt{u}$ is a concave function,

$$\begin{aligned}\sqrt{(a-c)^2+4b^2} &\le \sqrt{(a-c)^2} + \left(\frac{\mathrm{d}}{\mathrm{d}u}\sqrt{u}\Big|_{u=(a-c)^2}\right)\cdot 4b^2 \\ &= a - c + \frac{2b^2}{a-c}.\end{aligned}$$

Thus the above expression for $\|M\|$ is bounded by

$$\|M\| \le a + \frac{b^2}{a-c}.$$

□

Lemma 57. *Let $\beta \in (1/2)\mathbb{Z}$. If $\beta \ge 0$, then*

$$\int_0^\infty q^\beta e^{-q^2} dq \le (\beta+1)^{\beta/2}.$$

If $\beta \ge 2$, then

$$\int_0^\infty q^\beta e^{-q^2} dq \le \left(\frac{\beta-1}{2}\right)^{\beta/2}.$$

Proof. We show the result by induction. For the base case, the inequalities in the lemma statement can be directly verified for all eight possible values $0 \le \beta < 4$. For the inductive step, for some $\beta \ge 4$, assume that the bounds in the lemma statement hold for $\beta - 2$. Integrating by parts gives that

$$\begin{aligned}\int_0^\infty q^\beta e^{-q^2} dq &= \int_0^\infty q^{\beta-1} \cdot (q e^{-q^2}) dq \\ &= \frac{\beta-1}{2} \cdot \int_0^\infty q^{\beta-2} e^{-q^2} dq.\end{aligned}$$

Thus the desired result follows by our bounds on $\int_0^\infty q^{\beta-2} e^{-q^2} dq$ from the inductive hypothesis. □

B Proof of Theorem 20 for $d > 2$

In this section, we prove the $d > 2$ case of Theorem 20. The proof we present in this section generalizes the proof in Section 3.1 for the $d = 2$ case, and will therefore follow the same general outline. In fact, the proof here also applies for the $d = 2$ case, but gives a worse dependence on p than was achieved in Section 3.1; we also do not show a tail bound for the $d > 2$ case as was shown in Theorem 18 when $d = 2$. The precise result we will prove in this section is (re)stated below.

Theorem 58 ($d > 2$ case of Theorem 20)**.** *Fix integers $t \ge 1$ and $1 \le u \le t-1$. For some integer $d \ge 2$, let V be a finite set of vertices with labeling $\mathrm{val} : V \to [d]$ that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Let $\mathcal{G} = (G_i)_{1 \le i \le t-1}$ and $\mathcal{G}' = (G'_i)_{1 \le i \le t-1}$ be sequences of regular graphs on the shared vertex set V such that for all $i \ne u$ we have $G_i = G'_i$ with $\lambda(G_i) = \lambda(G'_i) \le \min(p)/400$. Then*

$$d_{TV}\left(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}), \Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})\right) \le \frac{2^{9d/4+13} d^{d/2+3}}{\min(p)^{(d+3)/4}} \cdot \frac{\|G'_u - G_u\|}{t}.$$

To begin, recall that for $\mathcal{W} = \mathcal{G}$ or $\mathcal{G}'$, the random variable $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})$ takes values in $[t+1]^{[d]}$, where $(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}}))_b$ denotes the number of steps in RW^t_W that land on a vertex with label b. Note that the $b = 0$ component (for instance) of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})$ is redundant, as the sum of all of the components equals t. Therefore we will often restrict to components $1 \le b \le d-1$ of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{W}})$, which are supported inside $[t+1]^{d-1} \subseteq \mathbb{Z}^{d-1}$. That is, let

$$S = \mathbb{Z}^{d-1} \cong \left\{ j \in \mathbb{Z}^d : \sum_{b\in[d]} j_b = t \right\},$$

where the isomorphism above is given by $(j_1,\dots,j_{d-1}) \mapsto (t - \sum_{b=1}^{d-1} j_b, j_1, \dots, j_{d-1})$.

Define

$$g = \left(\Pr\left[(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'}))_{[d]\setminus\{0\}} = j\right] - \Pr\left[(\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}}))_{[d]\setminus\{0\}} = j\right]\right)_{j\in S} \in [-1,1]^S \tag{19}$$

to denote the difference between the distributions of $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}'})$ and $\Sigma\,\mathrm{val}(\mathrm{RW}^t_{\mathcal{G}})$ when restricting to components in $[d] \setminus \{0\}$. In this notation, our goal is to show that if d and p are fixed, then $\|g\|_1 = O(\|G'_u - G_u\|/t)$.

B.1 Reduction to bounding an ℓ_2-norm

In this section, we show how to prove the $d > 2$ case of Theorem 20 given Theorem 59 below, which bounds the norm of the vector

$$g_b^{(sr)} = (e^{sr(j_b - p_b t)} g_j)_{j\in S} \in \mathbb{R}^S \tag{20}$$

for $1 \le b \le d-1$, $s = \pm 1$, and $0 \le r \le 1/2$, where g is given by (19). We will prove Theorem 59 in the following sections.

Theorem 59. *As in Theorem 20, let $u < t$ and $d \ge 2$ be positive integers, and let V be a set of vertices with labeling $\mathrm{val} : V \to [d]$ that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Let $\mathcal{G} = (G_i)_{1\le i\le t-1}$ and $\mathcal{G}' = (G'_i)_{1\le i\le t-1}$ be sequences of regular graphs on the shared vertex set V, such that for all $i \ne u$ we have $G_i = G'_i$ with $\lambda(G_i) = \lambda(G'_i) \le \min(p)/400$. Then for every $1 \le b \le d-1$, $s = \pm 1$, and $0 \le r \le 1/2$, defining $g_b^{(sr)}$ as in (20), we have*

$$\|g_b^{(sr)}\| \le 2^{d/2+9}(d-1)^{d/4+1} \cdot .\|G'_u - G_u\| \cdot e^{2p_b t r^2} \cdot \left(\frac{p_b r^2}{(\min(p)t)^{(d-1)/4}} + \frac{1}{(\min(p)t)^{(d+3)/4}} \right).$$

As in the $d = 2$ case from Section 3.1, reducing the desired ℓ_1-norm bound on g to an ℓ_2-norm bound on $g_b^{(sr)}$ will allow us to take the Fourier transform of $g_b^{(sr)}$, and then bound each Fourier coefficient separately, as described in Section B.2.

To prove the $d > 2$ case of Theorem 20, we will partition S into blocks, and then bound the ℓ_1-norm of g within each block in terms of $\|g_b^{(sr)}\|$ for appropriately chosen $s = \pm 1$, $0 \le r \le 1/2$, and $1 \le b \le d-1$. Intuitively, an appropriate selection of s, r, b allows the components of $g_b^{(sr)}$ inside the specified block to dominate components outside the block.

To begin, with $S = \mathbb{Z}^{d-1}$, then for integers $1 \le b \le d-1$ and $k \ge 0$, define

$$S^+_{b,k} = \{j \in S : \|j - pt\|_\infty < (k+1)\sqrt{t},\ j_b - p_b t \ge k\sqrt{t}\}$$
$$S^-_{b,k} = \{j \in S : \|j - pt\|_\infty < (k+1)\sqrt{t},\ j_b - p_b t \le -k\sqrt{t}\}.$$

Note that above $j-pt$ is viewed as a vector in $\mathbb{R}^{d-1}$, so that $j=(j_1,\dots,j_{d-1})$ and $p=(p_1,\dots,p_{d-1})$ with p_0 being ignored. By definition S equals the union over all b,k,s of $S^s_{b,k}$, and furthermore,

$$\mathrm{supp}(g) \subseteq [t+1]^{d-1} \subseteq \bigcup_{1\le b\le d-1,\ 0\le k\le\sqrt{t},\ s=\pm1} S^s_{b,k} \subseteq S.$$

Therefore letting $g_A=(g_j)_{j\in A}\in[-1,1]^A$ denote the restriction of g to a subset $A\subseteq S$, then

$$\|g\|_1 \le \sum_{1\le b\le d-1,\ 0\le k\le\sqrt{t},\ s=\pm1} \|g_{S^s_{b,k}}\|_1. \tag{21}$$

Thus we will focus on bounding the $g_{S^s_{b,k}}$ separately. We proceed analagously as in Section 3.1.

Lemma 60. *For $1\le b\le d-1$, $0\le k\le\sqrt{t}$, $s=\pm1$ and $r\ge0$,*

$$\|g_{S^s_{b,k}}\|_1 \le 2^{d-1}(k+1)^{(d-2)/2}t^{(d-1)/4}e^{-rk\sqrt{t}}\cdot\|g_b^{(sr)}\|.$$

Proof. By the definition of $g_b^{(sr)}$ and the Cauchy-Schwartz inequality,

$$\begin{aligned}\|g_{S^s_{b,k}}\|_1 &\le e^{-rk\sqrt{t}}\cdot\|(g_b^{(sr)})_{S^s_{b,k}}\|_1\\ &\le e^{-rk\sqrt{t}}\sqrt{|S^s_{b,k}|}\cdot\|g_b^{(sr)}\|\\ &\le e^{-rk\sqrt{t}}\sqrt{(\sqrt{t}+1)(2(k+1)\sqrt{t}+1)^{d-2}}\cdot\|g_b^{(sr)}\|\\ &\le 2^{d-1}(k+1)^{(d-2)/2}t^{(d-1)/4}e^{-rk\sqrt{t}}\cdot\|g_b^{(sr)}\|.\end{aligned}$$

□

We now apply the bound in (21) along with Theorem 59 and Lemma 60 to prove the $d>2$ case of Theorem 20.

Proof of Theorem 58 ($d>2$ case of Theorem 20). By Lemma 60 and Theorem 59,

$$\begin{aligned}\|g_{S^s_{b,k}}\|_1 &\le 2^{d-1}(k+1)^{(d-2)/2}t^{(d-1)/4}e^{-rk\sqrt{t}}\\ &\quad\cdot 2^{d/2+9}(d-1)^{d/4+1}\cdot.\|G'_u-G_u\|\cdot e^{2p_btr^2}\cdot\left(\frac{p_br^2}{(\min(p)t)^{(d-1)/4}}+\frac{1}{(\min(p)t)^{(d+3)/4}}\right)\\ &\le 2^{3d/2+8}(d-1)^{d/4+1}\cdot\|G'_u-G_u\|\cdot(k+1)^{(d-2)/2}\\ &\quad\cdot e^{2p_btr^2-rk\sqrt{t}}\cdot\left(\frac{p_br^2}{\min(p)^{(d-1)/4}}+\frac{1}{\min(p)^{(d+3)/4}\cdot t}\right).\end{aligned}$$

Thus setting

$$r=\begin{cases}\frac{k}{4p_b\sqrt{t}}, & k\le 2p_b\sqrt{t}\\ \frac{1}{2}, & k>2p_b\sqrt{t}\end{cases}$$

gives that if $k\le 2p_b\sqrt{t}$, then $2r^2p_bt-rk\sqrt{t}=-k^2/(8p_b)$, while if $k>2p_b\sqrt{t}$, then $p_bt\le k\sqrt{t}/2$, so $2r^2p_bt-rk\sqrt{t}=p_bt/2-k\sqrt{t}/2\le -k\sqrt{t}/4\le -k^2/4$, where this final inequality holds because $k\le\sqrt{t}$ by assumption. Thus in either case,

$$2r^2p_bt-rk\sqrt{t}\le\max\left\{-\frac{k^2}{8p_b},-\frac{k^2}{4}\right\}\le-\frac{k^2}{8}.$$

Furthermore, when $k \leq 2p_b\sqrt{t}$ then $p_b r^2 = k^2/(16p_b t)$, while when $k > 2p_b\sqrt{t}$, then $k^2/(16p_b t) > 4p_b^2 t/(16p_b t) = p_b/4 = p_b r^2$. Thus in either case,

$$p_b r^2 \leq \frac{k^2}{16p_b t} \leq \frac{k^2}{16\min(p)t}.$$

Combining the above inequalities gives that

$$\begin{aligned}\|g_{S^s_{b,k}}\|_1 &\leq 2^{3d/2+8}(d-1)^{d/4+1} \cdot \|G'_u - G_u\| \cdot (k+1)^{(d-2)/2} \\ &\quad \cdot e^{-k^2/8} \cdot \left(\frac{k^2/(16\min(p)t)}{\min(p)^{(d-1)/4}} + \frac{1}{\min(p)^{(d+3)/4} \cdot t}\right) \\ &= 2^{3d/2+8}(d-1)^{d/4+1} \cdot \|G'_u - G_u\| \cdot (k+1)^{(d-2)/2} \cdot e^{-k^2/8} \cdot \frac{k^2/16+1}{\min(p)^{(d+3)/4} \cdot t} \\ &\leq 2^{3d/2+8}(d-1)^{d/4+1} \cdot \|G'_u - G_u\| \cdot (k+1)^{(d+2)/2} \cdot e^{-k^2/8} \cdot \frac{1}{\min(p)^{(d+3)/4} \cdot t},\end{aligned}$$

where the final inequality above holds because $k^2/16+1 \leq (k+1)^2$ for all $k \geq 0$. Therefore by (21),

$$\begin{aligned}\|g\|_1 &\leq \frac{2^{3d/2+8}(d-1)^{d/4+1} \cdot \|G'_u - G_u\|}{\min(p)^{(d+3)/4} \cdot t} \cdot \sum_{1 \leq b \leq d-1,\ 0 \leq k \leq \sqrt{t},\ s=\pm 1} (k+1)^{(d+2)/2} \cdot e^{-k^2/8} \\ &= \frac{2^{3d/2+9}(d-1)^{d/4+2} \cdot \|G'_u - G_u\|}{\min(p)^{(d+3)/4} \cdot t} \cdot \sum_{k=0}^{\lfloor\sqrt{t}\rfloor}(k+1)^{(d+2)/2} \cdot e^{-k^2/8}.\end{aligned}$$

The sum above can be bounded as

$$\begin{aligned}\sum_{k=0}^{\lfloor\sqrt{t}\rfloor}(k+1)^{(d+2)/2}e^{-k^2/8} &\leq 1 + 2^{(d+2)/2}\sum_{k=1}^{\lfloor\sqrt{t}\rfloor} k^{(d-2)/2}e^{-k^2/8} \\ &\leq 1 + 2^{(d+2)/2}\left(\int_{k=0}^{\infty} k^{(d+2)/2}e^{-k^2/8}dk + 2\sup_{k \geq 0} k^{(d+2)/2}e^{-k^2/8}\right) \\ &\leq 1 + 2^{(d+2)/2}\left(\sqrt{8}^{(d+4)/2}(d/4)^{(d+2)/4} + 2\left(\frac{2(d+2)}{e}\right)^{(d+2)/4}\right) \\ &\leq 2^{3d/4+5}d^{d/4+1/2},\end{aligned}$$

where the second and third inequalities above hold because $(k+1)^{(d+2)/2}e^{-k^2/8}$ is increasing for $0 < k < \sqrt{2(d+2)}$ and decreasing for $k > \sqrt{2(d+2)}$, and the third inequality also uses the fact that with $\ell = k/\sqrt{8}$, then $\int_{k=0}^{\infty} k^{(d+2)/2}e^{-k^2/8}dk = \sqrt{8}^{(d+4)/2}\int_{\ell=0}^{\infty}\ell^{(d+2)/2}e^{-\ell^2}d\ell$, and this latter integral can be bounded using Lemma 57. Thus

$$\begin{aligned}\|g\|_1 &\leq \frac{2^{3d/2+9}(d-1)^{d/4+2} \cdot \|G'_u - G_u\|}{\min(p)^{(d+3)/4} \cdot t} \cdot 2^{3d/4+5}d^{d/4+1/2} \\ &\leq \frac{2^{9d/4+14}d^{d/2+3}}{\min(p)^{(d+3)/4}} \cdot \frac{\|G'_u - G_u\|}{t}.\end{aligned}$$

□

B.2 Reduction to bounding Fourier coefficients

In this section, we show how to prove Theorem 59 given bounds on the Fourier transform of $g_b^{(sr)}$. Specifically, we prove Theorem 59 assuming Theorem 62 below, which will be proven in Section B.3.

We first introduce the Fourier transform for the group $\mathbb{Z}^{d-1}$.

Definition 61. Let $(S^1)^{d-1} = (\mathbb{R}/2\pi\mathbb{Z})^{d-1}$ with ℓ^2 norm $\|f\| = \sqrt{\int_{(-\pi,\pi]^{d-1}} |f(\theta)|^2 d\theta/(2\pi)^{d-1}}$. Let $\ell^2(\mathbb{Z}^{d-1})$ and $\ell^2((S^1)^{d-1})$ denote the subspaces of $\mathbb{C}^{\mathbb{Z}^{d-1}}$ and $\mathbb{C}^{(S^1)^{d-1}}$ respectively containing all elements of finite ℓ^2 norm. Then the **Fourier transform for the group** $\mathbb{Z}^{d-1}$ is the map $\mathcal{F} : \ell^2(\mathbb{Z}^{d-1}) \to \ell^2((S^1)^{d-1})$ such the Fourier transform of $h \in \ell^2(\mathbb{Z}^{d-1})$, denoted $\mathcal{F}h = \hat{h} \in \ell^2((S^1)^{d-1})$, is given by

$$\hat{h}(\theta) = \sum_{j \in \mathbb{Z}^{d-1}} h_j e^{-i\theta \cdot j}.$$

The Fourier transform may also be expressed in terms of the **Fourier characters** $\chi_\theta = (e^{i\theta \cdot j})_{j \in \mathbb{Z}^{d-1}} \in \mathbb{C}^{\mathbb{Z}^{d-1}}$, as $\hat{h}(\theta) = \chi_\theta^* h$.

It is well know that the Fourier transform preserves the ℓ^2 norm, so that $\|\hat{h}\| = \|h\|$. Below, we associate $(S^1)^{d-1} = (\mathbb{R}/2\pi\mathbb{Z})^{d-1}$ with the space $(-\pi,\pi]^{d-1}$, so that all $\theta \in (S^1)^{d-1}$ have $\|\theta\|_\infty \leq \pi$.

Theorem 62. *As in Theorem 20, let $u < t$ and $d \geq 2$ be positive integers, and let V be a set of vertices with labeling $\mathrm{val} : V \to [d]$ that assigns each label $b \in [d]$ to p_b-fraction of the vertices. Let $\mathcal{G} = (G_i)_{1 \leq i \leq t-1}$ and $\mathcal{G}' = (G_i')_{1 \leq i \leq t-1}$ be sequences of regular graphs on the shared vertex set V, such that for all $i \neq u$ we have $G_i = G_i'$ with $\lambda(G_i) = \lambda(G_i') \leq \min(p)/400$. Then for every $1 \leq b \leq d-1$, $s = \pm 1$, and $0 \leq r \leq 1/2$, defining $g_b^{(sr)}$ as in (20), we have for all $\theta \in (-\pi,\pi]^{d-1}$ that*

$$|\hat{g}_b^{(sr)}(\theta)| \leq 4 \cdot \|G_u' - G_u\| \cdot \left(4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2\right) \cdot e^{t(2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40)}.$$

To obtain the desired bound on $\|g_b^{(sr)}\|$ in Theorem 59, we square the inequality in Theorem 62 and then integrate over $\theta \in (-\pi,\pi]^{d-1}$, using the fact that $\|g_b^{(sr)}\|^2 = \|\hat{g}_b^{(sr)}\|^2 = \int_{(-\pi,\pi]^{d-1}} |\hat{g}_b^{(sr)}(\theta)|^2 d\theta/(2\pi)^{d-1}$ because the Fourier transform preserves ℓ^2-norms. This calculation is shown below.

Proof of Theorem 59. Because the Fourier transform preserves the ℓ^2-norm, by Theorem 62,

$$
\begin{aligned}
\|g_b^{(sr)}\| &= \sqrt{\int_{(-\pi,\pi]^{d-1}} |\hat{g}_b^{(sr)}(\theta)|^2 \frac{d\theta}{(2\pi)^{d-1}}} \\
&\leq 4 \cdot \|G'_u - G_u\| \cdot e^{2p_b t r^2} \cdot \sqrt{\int_{(-\pi,\pi]^{d-1}} \left(4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2\right)^2 \cdot e^{-\min(p)t\cdot\|\theta\|_\infty^2/20} \frac{d\theta}{(2\pi)^{d-1}}} \\
&\leq 4\sqrt{2} \cdot \|G'_u - G_u\| \cdot e^{2p_b t r^2} \\
&\quad \cdot \left(4p_b r^2 \sqrt{\int_{(-\pi,\pi]^{d-1}} e^{-\min(p)t\cdot\|\theta\|_\infty^2/20} \frac{d\theta}{(2\pi)^{d-1}}} \right. \\
&\quad \left. + \sqrt{\int_{(-\pi,\pi]^{d-1}} \frac{9}{4}\|\theta\|_\infty^4 e^{-\min(p)t\cdot\|\theta\|_\infty^2/20} \frac{d\theta}{(2\pi)^{d-1}}}\right) \\
&= 4\sqrt{2} \cdot \|G'_u - G_u\| \cdot e^{2p_b t r^2} \\
&\quad \cdot \left(4p_b r^2 \sqrt{\int_0^\pi e^{-\min(p)t\cdot\eta^2/20} \cdot 2(d-1)(2\eta)^{d-2} \frac{d\eta}{(2\pi)^{d-1}}} \right. \\
&\quad \left. + \sqrt{\int_0^\pi \frac{9}{4}\eta^4 e^{-\min(p)t\cdot\eta^2/20} \cdot 2(d-1)(2\eta)^{d-2} \frac{d\eta}{(2\pi)^{d-1}}}\right),
\end{aligned}
\tag{22}
$$

where the second inequality above holds because all $a, b \geq 0$ satisfy $(a+b)^2 \leq 2(a^2+b^2)$ and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, and the second equality above holds because the $(d-2)$-dimensional region $\{\theta \in \mathbb{R}^{d-1} : \|\theta\|_\infty = \eta\}$ has volume $2(d-1)(2\eta)^{d-2}$. Substituting $q = \sqrt{\min(p)t/20} \cdot \eta$ in the integrals in the right hand side of (22) gives

$$
\begin{aligned}
\|g^{(sr)}\| \leq{} & 4\sqrt{2} \cdot \|G'_u - G_u\| \cdot e^{2p_b t r^2} \\
& \cdot \left(4p_b r^2 \sqrt{\frac{d-1}{\pi^{d-1}} \cdot \left(\frac{20}{\min(p)t}\right)^{(d-1)/2} \cdot \int_0^\infty q^{d-2} e^{-q^2} dq} \right. \\
& \left. + \frac{3}{2}\sqrt{\frac{d-1}{\pi^{d-1}} \cdot \left(\frac{20}{\min(p)t}\right)^{(d+3)/2} \cdot \int_0^\infty q^{d+2} e^{-q^2} \cdot dq}\right).
\end{aligned}
$$

Applying Lemma 57 to the two integrals on the right hand side above impies that both of these integrals are bounded above by $(d-1)^{(d+2)/2}$. Therefore

$$
\begin{aligned}
\|g^{(sr)}\| &\leq \frac{16\sqrt{2} \cdot 20^{(d+3)/4} \cdot (d-1)^{(d+4)/4}}{\pi^{(d-1)/2}} \\
&\quad \cdot .\|G'_u - G_u\| \cdot e^{2p_b t r^2} \cdot \left(\frac{p_b r^2}{(\min(p)t)^{(d-1)/4}} + \frac{1}{(\min(p)t)^{(d+3)/4}}\right) \\
&\leq 2^{d/2+9}(d-1)^{d/4+1} \cdot .\|G'_u - G_u\| \cdot e^{2p_b t r^2} \cdot \left(\frac{p_b r^2}{(\min(p)t)^{(d-1)/4}} + \frac{1}{(\min(p)t)^{(d+3)/4}}\right).
\end{aligned}
$$

□

B.3 Bounding Fourier coefficients

In this section we prove Theorem 62. Throughout this section, for convenience we extend the sequences G and G' to include 0th and tth components $G_0 = G'_0 = G_t = G'_t = J$. In the proofs below, these matrices will typically be applied to $\vec{1}$, and they therefore could be removed at the cost of more cumbersome notation.

First, we express the Fourier transform $\hat{g}_b^{(sr)}(\theta)$ of $g_b^{(sr)}$ linear-algebraically below.

Lemma 63. *For $\theta \in (-\pi,\pi]^{[d]\setminus\{0\}}$, $1 \le b \le d-1$, $s = \pm 1$, and $0 \le r \le 1/2$, let $P_{\theta,b}^{(sr)} \in \mathbb{C}^{V\times V}$ be the matrix given by:*

$$P_{\theta,b}^{(sr)} = \sum_{v\in V} \delta_v\delta_v^* e^{(sr\delta_b - i\theta)\cdot\delta_{\mathrm{val}(v)} - p_b sr} = \sum_{v\in \mathrm{val}^{-1}(b)} \delta_v\delta_v^* e^{(1-p_b)sr - i\theta_b} + \sum_{b'\in[d]\setminus\{b\}} \delta_v\delta_v^* e^{-p_b sr - i\theta_{b'}},$$

where we extend θ to a tuple in $(-\pi,\pi]^{[d]}$ by letting $\theta_0 = 0$. Then

$$\begin{aligned}\hat{g}_b^{(sr)}(\theta) &= \vec{1}^* G_t P_{\theta,b}^{(sr)} \cdots G_{u+1} P_{\theta,b}^{(sr)} (G'_u - G_u) P_{\theta,b}^{(sr)} G_{u-1} \cdots P_{\theta,b}^{(sr)} G_0 \vec{1} \\ &= \vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_{\theta,b}^{(sr)}\right) (G'_u - G_u) \left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right) \vec{1}.\end{aligned}$$

Proof. The proof is analogous to that of Lemma 26. For $\mathcal{W} = \mathcal{G}$ or $\mathcal{G}'$, writing $W_i = \sum_{v,v'\in V} \delta_{v'}\delta_v^*(W_i)_{v',v}$, then $\vec{1}^*\left(\prod_{i=1}^t (W_i P_{\theta,b}^{(sr)})\right)\vec{1} = \sqrt{n}\vec{1}^*\left(\prod_{i=1}^t (W_i P_{\theta,b}^{(sr)})\right)(1/\sqrt{n})\vec{1}$ expands to give

$$\begin{aligned}\vec{1}^*\left(\prod_{i=1}^t (W_i P_{\theta,b}^{(sr)})\right)\vec{1} &= \sum_{(v_0,\dots,v_{t-1})\in V^{[t]}} \Pr\left[\mathrm{RW}_{\mathcal{W}}^t = (v_0,\dots,v_{t-1})\right] \prod_{i'\in[t]} e^{(sr\delta_b - i\theta)\cdot\delta_{\mathrm{val}(v_{i'})} - p_b sr} \\ &= \mathbb{E}\left[e^{(sr\delta_b - i\theta)\cdot\Sigma\,\mathrm{val}(\mathrm{RW}_{\mathcal{W}}^t) - p_b srt}\right].\end{aligned}$$

The first equality above follows from an expansion analogous to the one described in detail in the proof of Lemma 36, to which the reader is referred for details; we omitted some intermediate steps to avoid redundancy. Therefore

$$\begin{aligned}\hat{g}_b^{(sr)}(\theta) &= \sum_{j\in S}\left(\Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}_{\mathcal{G}'}^t) = j\right] - \Pr\left[\Sigma\,\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t) = j\right]\right) e^{(sr\delta_b - i\theta)\cdot j - p_b srt} \\ &= \mathbb{E}\left[e^{(sr\delta_b - i\theta)\cdot\Sigma\,\mathrm{val}(\mathrm{RW}_{\mathcal{G}'}^t) - p_b srt}\right] - \mathbb{E}\left[e^{(sr\delta_b - i\theta)\cdot\Sigma\,\mathrm{val}(\mathrm{RW}_{\mathcal{G}}^t) - p_b srt}\right] \\ &= \vec{1}^*\left(\prod_{i=1}^t (G'_i P_{\theta,b}^{(sr)})\right)\vec{1} - \vec{1}^*\left(\prod_{i=1}^t (G_i P_{\theta,b}^{(sr)})\right)\vec{1} \\ &= \vec{1}^*\left(\prod_{i=u+1}^t G_i P_{\theta,b}^{(sr)}\right)(G'_u - G_u)\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right)\vec{1},\end{aligned}$$

where the final equality above holds because $G'_i = G_i$ for $i \ne u$ by assumption. □

Lemma 63 shows that in order to bound the Fourier transform of $g_b^{(sr)}$, it is sufficient to bound $\vec{1}^*\left(\prod_{i=u+1}^t G_i P_{\theta,b}^{(sr)}\right)(G'_u - G_u)\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right)\vec{1}$. For this purpose, because the matrix

$G'_u - G_u$ annihilates $\vec{1}$ from both sides, we will bound the components of $\left(\prod_{i=u+1}^{t} G_i P_{\theta,b}^{(sr)}\right)\vec{1}$ and $\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right)\vec{1}$ that are orthogonal to $\vec{1}$ (denoted by $^\perp$ below). To bound the orthogonal components of these vectors, we will apply the following two lemmas, which provide bounds for the matrix $P_{\theta,b}^{(sr)}$, and will be proven in Section B.4.

Lemma 64. *For $\theta \in (-\pi, \pi]^{d-1}$, $1 \le b \le d-1$, $s = \pm 1$, and $0 \le r \le 1/2$, we have*

$$|\vec{1}^* P_{\theta,b}^{(sr)} \vec{1}| \le 1 + p_b r^2 - \frac{3\min(p)}{4\pi^2 e^{1/2}} \|\theta\|_\infty^2$$

and

$$\|(P_{\theta,b}^{(sr)}\vec{1})^\perp\| = \|(\vec{1}^* P_{\theta,b}^{(sr)})^\perp\| \le \sqrt{4ep_b r^2 + \frac{3e}{2}\|\theta\|_\infty^2}.$$

Lemma 65. *Let $\rho = \sqrt{\min(p)}/20$ and $F = J + \rho(I - J)$. Then for every $1 \le b \le d-1$, $s = \pm 1$ and $0 \le r \le 1/2$,*

$$\|F P_{\theta,b}^{(sr)} F\| \le 1 + 2p_b \cdot r^2 - \frac{\min(p)}{40} \cdot \|\theta\|_\infty^2.$$

Lemma 66. *For $1 \le u \le t-1$,*

$$\left\|\left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right)\vec{1}\right)^\perp\right\| \le 2 \cdot \sqrt{4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2} \cdot e^{u(2p_b \cdot r^2 - \min(p)\cdot\|\theta\|_\infty^2/40)}.$$

Proof. The proof is similar to that of Lemma 29. By definition

$$\left\|\left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right)\vec{1}\right)^\perp\right\| \le \|(P_{\theta,b}^{(sr)}\vec{1})^\perp\| \cdot \left\|\left(\left(\prod_{i=0}^{u-2} P_{\theta,b}^{(sr)} G_i\right)\vec{1}\right)^\parallel\right\| + \|P_{\theta,b}^{(sr)}\| \cdot \lambda(G_{u-1}) \cdot \left\|\left(\left(\prod_{i=0}^{u-2} P_{\theta,b}^{(sr)} G_i\right)\vec{1}\right)^\perp\right\|.$$

The above inequality can be recursively applied to bound the term $\left\|\left(\left(\prod_{i=0}^{u-2} P_{\theta,b}^{(sr)} G_i\right)\vec{1}\right)^\perp\right\|$ on its right hand side. Performing $u-1$ such recursive applications gives that

$$\left\|\left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i\right)\vec{1}\right)^\perp\right\| \le \|(P_{\theta,b}^{(sr)}\vec{1})^\perp\| \sum_{i=0}^{u-1}\left(\prod_{i'=i+1}^{u-1}\|P_{\theta,b}^{(sr)}\| \cdot \lambda(G_{i'})\right)\left\|\left(\left(\prod_{i'=0}^{i-1} P_{\theta,b}^{(sr)} G_{i'}\right)\vec{1}\right)^\parallel\right\|.$$

Let $\rho = \sqrt{\min(p)}/20$ and $F = J + \rho(I - J)$. By assumption all $i' \le u-1$ have $\lambda(G_{i'}) \le \rho^2$. It follows that $\|F^{-1}G_{i'}F^{-1}\| \le 1$, as $F^{-1}G_{i'}F^{-1}$ preserves the vector $\vec{1}$ and the subspace $\vec{1}^\perp$, and the restriction $F^{-1}G_{i'}F^{-1}|_{\vec{1}^\perp}$ has spectral norm $\rho^{-1}\lambda(G_{i'})\rho^{-1} \le 1$. It also holds that $\|P_{\theta,b}^{(sr)}\| \le e^{1/2}$

because $0 \le r \le 1/2$. Thus

$$\begin{aligned}
&\left\| \left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i \right) \vec{1} \right)^{\perp} \right\| \\
&\quad \le \|(P_{\theta,b}^{(sr)} \vec{1})^{\perp}\| \cdot \sum_{i=0}^{u-1} (e^{1/2}\rho^2)^{u-1-i} \cdot \left| \vec{1}^* \left(\prod_{i'=0}^{i-1} F P_{\theta,b}^{(sr)} F \cdot F^{-1} G_{i'} F^{-1} \right) \vec{1} \right| \\
&\quad \le \|(P_{\theta,b}^{(sr)} \vec{1})^{\perp}\| \cdot \sum_{i=0}^{u-1} (e^{1/2}\rho^2)^{u-1-i} \cdot \|F P_{\theta,b}^{(sr)} F\|^i.
\end{aligned}$$

Applying Lemma 64 and Lemma 65 to bound $\|(P_{\theta,b}^{(sr)} \vec{1})^{\perp}\|$ and $\|F P_{\theta,b}^{(sr)} F\|$ respectively gives

$$\begin{aligned}
&\left\| \left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i \right) \vec{1} \right)^{\perp} \right\| \\
&\quad \le \sqrt{4ep_b r^2 + \frac{3e}{2}\|\theta\|_\infty^2} \cdot \sum_{i=0}^{u-1} (e^{1/2}\rho^2)^{u-1-i} \cdot e^{i(2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40)} \\
&\quad = \sqrt{4ep_b r^2 + \frac{3e}{2}\|\theta\|_\infty^2} \cdot e^{u(2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40)} \cdot \sum_{i=0}^{u-1} \frac{(e^{1/2}\rho^2)^{u-1-i}}{e^{(u-i)(2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40)}}.
\end{aligned}$$

Because $\rho = \sqrt{\min(p)}/20$ and $e^{2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40} \ge e^{-\pi^2/80}$,

$$\begin{aligned}
\sum_{i=0}^{u-1} \frac{(e^{1/2}\rho^2)^{u-1-i}}{e^{(u-i)(2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40)}} &\le \sum_{i=-\infty}^{u-1} \frac{(e^{1/2} \cdot \min(p)/400)^{u-1-i}}{(e^{-\pi^2/80})^{u-i}} \\
&= \frac{e^{\pi^2/80}}{1 - e^{1/2+\pi^2/80} \cdot \min(p)/400} \\
&\le \frac{2}{\sqrt{e}}.
\end{aligned}$$

Thus

$$\left\| \left(\left(\prod_{i=0}^{u-1} P_{\theta}^{(sr)} G_i \right) \vec{1} \right)^{\perp} \right\| \le 2 \cdot \sqrt{4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2} \cdot e^{u(2p_b \cdot r^2 - \min(p) \cdot \|\theta\|_\infty^2/40)}.$$

□

We now apply the above lemmas to prove Theorem 62.

Proof of Theorem 62. By Lemma 63,

$$\begin{aligned}
|\hat{g}_b^{(sr)}(\theta)| &= \left| \vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_{\theta,b}^{(sr)} \right) (G'_u - G_u) \left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i \right) \vec{1} \right| \\
&\le \left\| \left(\vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_{\theta,b}^{(sr)} \right) \right)^{\perp} \right\| \|G'_u - G_u\| \left\| \left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i \right) \vec{1} \right)^{\perp} \right\|,
\end{aligned}$$

where the inequality above holds because $G'_u - G_u$ annihilates $\vec{1}$ from both sides. Lemma 66 implies that

$$\left\| \left(\left(\prod_{i=0}^{u-1} P_{\theta,b}^{(sr)} G_i \right) \vec{1} \right)^{\perp} \right\| \leq 2 \cdot \sqrt{4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2} \cdot e^{u(2p_b \cdot r^2 - \min(p)\cdot\|\theta\|_\infty^2/40)}$$
$$\left\| \left(\vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_{\theta,b}^{(sr)} \right) \right)^{\perp} \right\| \leq 2 \cdot \sqrt{4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2} \cdot e^{(t-u)(2p_b \cdot r^2 - \min(p)\cdot\|\theta\|_\infty^2/40)},$$

where the second equality above holds because $P_{\theta,b}^{(sr)}$ is diagonal and therefore symmetric, so we may apply Lemma 29 to bound the norm of the transpose of $\left(\vec{1}^* \left(\prod_{i=u+1}^{t} G_i P_{\theta,b}^{(sr)}\right)\right)^{\perp}$. Now combining the above inequalities gives

$$|\hat{g}_b^{(sr)}(\theta)| \leq 4 \cdot \|G'_u - G_u\| \cdot \left(4p_b r^2 + \frac{3}{2}\|\theta\|_\infty^2\right) \cdot e^{t(2p_b \cdot r^2 - \min(p)\cdot\|\theta\|_\infty^2/40)}.$$

□

B.4 Bounds for the matrix $P_{\theta,b}^{(sr)}$

In this section, we prove Lemma 64 and Lemma 65, thereby bounding the quantities $|\vec{1}^* P_{\theta,b}^{(sr)} \vec{1}|$, $\|(P_{\theta,b}^{(sr)} \vec{1})^{\perp}\|$, and $\|F P_{\theta,b}^{(sr)} F\|$. The proofs are similar to those in Section 3.1.4.

Proof of Lemma 64. For the first inequality in the lemma statement, by definition $\vec{1}^* P_{\theta,b}^{(sr)} \vec{1} = e^{-p_b sr}(p \cdot e^{sr\delta_b - i\theta})$. We will first show that

$$|p \cdot e^{sr\delta_b - i\theta}| \leq 1 + p_b(e^{sr} - 1) - \frac{3\min(p)}{4\pi^2}\|\theta\|_\infty^2. \tag{23}$$

Let $\bar{b} = \arg\max_{b' \in [d]} |\theta_{b'}|$, so that $|\theta_{\bar{b}}| = \|\theta\|_\infty$. If $\bar{b} \neq b$, then

$$\begin{aligned}
|p \cdot e^{sr\delta_b - i\theta}| &\leq (p_0 + p_{\bar{b}}) \left| \frac{p_0}{p_0 + p_{\bar{b}}} + \frac{p_{\bar{b}}}{p_0 + p_{\bar{b}}} e^{-i\theta_{\bar{b}}} \right| + \sum_{b' \in [d] \setminus \{0, \bar{b}\}} p_{b'} |e^{(sr\delta_b - i\theta)\cdot\delta_{b'}}| \\
&\leq (p_0 + p_{\bar{b}}) \left(1 - \frac{2}{\pi^2} \cdot \frac{p_0 p_{\bar{b}}}{(p_0 + p_{\bar{b}})^2} \theta_{\bar{b}}^2\right) + 1 - p_0 - p_{\bar{b}} + p_b(e^{sr} - 1) \\
&\leq 1 + p_b(e^{sr} - 1) - \frac{\min(p)}{\pi^2}\|\theta\|_\infty^2,
\end{aligned}$$

where the first inequality above holds by the triangle inequality, the second inequality holds by Lemma 54, and the third inequality holds because $p_0 p_{\bar{b}}/(p_0 + p_{\bar{b}}) = p_0/(p_0/p_{\bar{b}} + 1) = p_{\bar{b}}/(1 + p_{\bar{b}}/p_0)$ is minimized when $p_0 = p_{\bar{b}} = \min(p)$ are both minimized.

If instead $\bar{b} = b$, then

$$\begin{aligned}|p \cdot e^{sr\delta_b - i\theta}| &\leq (p_0 + p_b e^{sr}) \left| \frac{p_0}{p_0 + e^{sr} p_b} + \frac{p_b e^{sr}}{p_0 + p_b e^{sr}} e^{-i\theta_b} \right| + \sum_{b' \in [d] \setminus \{0,b\}} p_{b'} |e^{-i\theta_{b'}}| \\ &\leq (p_0 + p_b e^{sr}) \left(1 - \frac{2}{\pi^2} \cdot \frac{p_0 p_b e^{sr}}{(p_0 + p_b e^{sr})^2} \theta_b^2 \right) + 1 - p_0 - p_b \\ &\leq 1 + p_b(e^{sr} - 1) - \frac{3\min(p)}{4\pi^2} \|\theta\|_\infty^2,\end{aligned}$$

where the first inequality above holds by the triangle inequality, the second inequality holds by Lemma 54, and the third equality holds because $p_0 p_b e^{sr}/(p_0 + p_b e^{sr})$ is minimized when $p_0 = p_b = \min(p)$ are both minimized and when $sr = -1/2$ is minimized, so that $p_0 p_b e^{sr}/(p_0 + p_b e^{sr}) \geq \min(p) e^{-1/2}/(1 + e^{-1/2}) \geq 3\min(p)/8$.

Thus (23) holds. It follows that

$$\begin{aligned}|\vec{1}^* P_{\theta,b}^{(sr)} \vec{1}| &= e^{-p_b sr} |p \cdot e^{sr\delta_b - i\theta}| \\ &\leq (1 - p_b) e^{-p_b sr} + p_b e^{(1-p_b)sr} - e^{-p_b sr} \frac{3\min(p)}{4\pi^2} \|\theta\|_\infty^2 \\ &\leq 1 + p_b(1 - p_b) r^2 - \frac{3\min(p)}{4\pi^2 e^{1/2}} \|\theta\|_\infty^2 \\ &\leq 1 + p_b r^2 - \frac{3\min(p)}{4\pi^2 e^{1/2}} \|\theta\|_\infty^2,\end{aligned}$$

where the second inequality above holds because $e^a \leq 1 + a + a^2$ when $|a| \leq 1/2$, and because $e^{-p_b sr} \geq e^{-1/2}$ as $r \leq 1/2$.

For the second inequality in the lemma statement, the norms of $(P_{\theta,b}^{(sr)} \vec{1})^\perp$ and $(\vec{1}^* P_{\theta,b}^{(sr)})^\perp$ are equal because these vectors are by definition transposes of each other. Thus it suffices to bound the norm of the former. Because $\|(P_{\theta,b}^{(sr)} \vec{1})^\perp\|$ equals the shortest distance from the vector $P_{\theta,b}^{(sr)} \vec{1}$ to the subspace $\text{span}\{\vec{1}\}$, we have

$$\begin{aligned}\|(P_{\theta,b}^{(sr)} \vec{1})^\perp\| &\leq \|P_{\theta,b}^{(sr)} \vec{1} - e^{-p_b sr} \vec{1}\| \\ &= \left\| \frac{1}{\sqrt{n}} \sum_{v \in V} \delta_v e^{(sr\delta_b - i\theta) \cdot \delta_{\text{val}(v)} - p_b sr} - \frac{1}{\sqrt{n}} \sum_{v \in V} \delta_v e^{-p_b sr} \right\| \\ &= \sqrt{e^{-2p_b sr} \sum_{b' \in [d]} p_{b'} |e^{(sr\delta_b - i\theta) \cdot \delta_{b'}} - 1|^2} \\ &\leq \sqrt{e \left(p_b \left(4r^2 + \frac{3}{2} \theta_b^2 \right) + \sum_{b' \in [d] \setminus \{b\}} p_{b'} \frac{3}{2} \theta_{b'}^2 \right)} \\ &\leq \sqrt{4e p_b r^2 + \frac{3e}{2} \|\theta\|_\infty^2},\end{aligned}$$

where the second inequality above follows by (10) and because $e^{-2p_b sr} \leq e$ as $r \leq 1/2$. □

Proof of Lemma 65. By definition

$$\|FP_{\theta,b}^{(sr)}F\| = \sup_{x,y\in\mathbb{C}^V:\|x\|=\|y\|=1} |x^* F P_{\theta,b}^{(sr)} F y|.$$

Decomposing $x = x^{\|} + x^{\perp}$ and $y = y^{\|} + y^{\perp}$ with $x^{\|}, y^{\|} \in \text{span}\{\vec{1}\}$ and $x^{\perp}, y^{\perp} \in \vec{1}^{\perp}$ gives

$$\begin{aligned}|x^* F P_{\theta,b}^{(sr)} F y| &\leq |x^{\|*} P_{\theta,b}^{(sr)} y^{\|}| + |x^{\|*} P_{\theta,b}^{(sr)} \rho y^{\perp}| + |x^{\perp} \rho P_{\theta,b}^{(sr)} y^{\|}| + |x^{\perp} \rho P_{\theta,b}^{(sr)} \rho y^{\perp}| \\ &\leq \begin{pmatrix} \|x^{\|}\| & \|x^{\perp}\| \end{pmatrix} \begin{pmatrix} |\vec{1}^* P_{\theta,b}^{(sr)} \vec{1}| & \rho\|(\vec{1}^* P_{\theta,b}^{(sr)})^{\perp}\| \\ \rho\|(P_{\theta,b}^{(sr)}\vec{1})^{\perp}\| & \rho^2 \|P_{\theta,b}^{(sr)}\| \end{pmatrix} \begin{pmatrix} \|y^{\|}\| \\ \|y^{\perp}\| \end{pmatrix}\end{aligned}$$

Thus

$$\|FP_{\theta,b}^{(sr)}F\| \leq \left\| \begin{pmatrix} |\vec{1}^* P_{\theta,b}^{(sr)} \vec{1}| & \rho\|(\vec{1}^* P_{\theta,b}^{(sr)})^{\perp}\| \\ \rho\|(P_{\theta,b}^{(sr)}\vec{1})^{\perp}\| & \rho^2 \|P_{\theta,b}^{(sr)}\| \end{pmatrix} \right\|. \tag{24}$$

The lower right entry of the matrix on the right hand side above is at most $\rho^2 e^{1/2}$ because $\|P_{\theta,b}^{(sr)}\| = \max\{e^{-p_b sr}, e^{(1-p_b)sr}\} \leq e^{1/2}$ as $r \leq 1/2$. Applying Lemma 64 to bound the other three entries of this matrix gives that

$$\|FP_{\theta,b}^{(sr)}F\| \leq \left\| \begin{pmatrix} 1 + p_b r^2 - \frac{3\min(p)}{4\pi^2 e^{1/2}}\|\theta\|_\infty^2 & \rho\sqrt{4ep_b r^2 + \frac{3e}{2}\|\theta\|_\infty^2} \\ \rho\sqrt{4ep_b r^2 + \frac{3e}{2}\|\theta\|_\infty^2} & \rho^2 e^{1/2} \end{pmatrix} \right\|.$$

Because $\rho = \sqrt{\min(p)}/20$, $\min(p) \leq 1/2$, and $\|\theta\|_\infty \leq \pi$,

$$\left(1 + p_b r^2 - \frac{3\min(p)}{4\pi^2 e^{1/2}}\|\theta\|_\infty^2\right) - \rho^2 e^{1/2} \geq 1 - \frac{3}{8e^{1/2}} - \frac{e^{1/2}}{800} \geq \frac{3}{4},$$

so Lemma 56 implies that

$$\begin{aligned}\|FP_{\theta,b}^{(sr)}F\| &\leq \left(1 + p_b r^2 - \frac{3\min(p)}{4\pi^2 e^{1/2}}\|\theta\|_\infty^2\right) + \frac{\frac{\min(p)}{400}\cdot(4ep_b r^2 + \frac{3e}{2}\|\theta\|_\infty^2)}{3/4} \\ &\leq 1 + 2p_b \cdot r^2 - \frac{\min(p)}{40}\cdot\|\theta\|_\infty^2.\end{aligned}$$

□

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il