The Expander Hitting Property When the Sets Are Arbitrarily Unbalanced

Amnon Ta-Shma
Department of Computer Science, Tel Aviv University, Israel

Ron Zadicario
Department of Computer Science, Tel Aviv University, Israel

Abstract

Numerous works have studied the probability that a length $t-1$ random walk on an expander is confined to a given rectangle $S_1 \times \ldots \times S_t$, providing both upper and lower bounds for this probability. However, when the densities of the sets S_i may depend on the walk length (e.g., when all set are equal and the density is $1-1/t$), the currently best known upper and lower bounds are very far from each other. We give an improved confinement lower bound that almost matches the upper bound.

We also study the more general question, of how well random walks fool various classes of test functions. Recently, Golowich and Vadhan proved that random walks on λ-expanders fool *Boolean, symmetric* functions up to a $O(\lambda)$ error in *total variation distance*, with *no dependence on the labeling bias*. Our techniques extend this result to cases not covered by it, e.g., to functions testing confinement to $S_1 \times \ldots \times S_t$, where each set S_i either has density ρ or $1-\rho$, for arbitrary ρ.

Technique-wise, we extend Beck's framework for analyzing what is often referred to as the "flow" of linear operators, reducing it to bounding the entries of a product of 2×2 matrices.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains; Theory of computation → Expander graphs and randomness extractors

Keywords and phrases Expander random walks, Expander hitting property

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2024.31

Category RANDOM

Related Version *Full Version*: https://eccc.weizmann.ac.il/report/2024/118 [14]

Funding *Amnon Ta-Shma*: The research leading to these results has received funding from the Israel Science Foundation (grant number 443/22)
Ron Zadicario: The research leading to these results has received funding from the Israel Science Foundation (grant number 443/22) and the Blavatnik Family Foundation.

Acknowledgements We thank the RANDOM 2024 anonymous referees for their helpful comments.

1 Introduction

Fix a set of vertices $V = [n]$ and t subsets $S_1, \ldots, S_t \subseteq V$. The hitting property of expander graphs [1] says that for a sufficiently good expander graph G on the set of vertices V, the probability that for all $i = 1, \ldots, t$ the i'th step of a random walk on G falls inside S_i is small, and therefore, with a good probability, the walk escapes the confinement $S_1 \times \ldots \times S_t$. Specifically,

▶ **Theorem 1** (Expander Hitting Property, based on [10]). *Let $G = (V, E)$ be a λ-expander. Then, for every sequence of subsets $S_1, \ldots, S_t \subseteq V$ such that S_i is of density $\rho_i = |S_i| / |V|$,*

$$\Pr_{(v_1,\ldots,v_t)\sim \mathsf{RW}_G^t} [\forall i \; v_i \in S_i] \le \sqrt{\rho_1 \rho_t} \cdot \prod_{i=1}^{t-1} \left((1-\lambda)\sqrt{\rho_i \rho_{i+1}} + \lambda\right). \tag{1}$$

We remark that a slightly weaker bound of $\prod_{i=1}^{t-1}\left(\sqrt{\rho_i\rho_{i+1}}+\lambda\right)$ appears in [10]. For the case where all densities ρ_i are the same ρ, a bound of $\rho\left((1-\lambda)\rho+\lambda\right)^{t-1}$ appears in [15], and of $\rho(\rho+\lambda)^{t-1}$ appears in [2]. The bound in the general case (Equation 1) follows by a similar proof, with a slightly more careful analysis. See Subsection 4.1.

However, on a conceptual level, one expects an expander random walk to mimic a truly random walk, each time choosing a vertex uniformly at random independent of all other choices. I.e., ideally, we would have liked a bound stating that the probability of an expander random walk being confined to $S_1 \times \ldots \times S_t$ is roughly the same as the probability of the same event with respect to a walk on the complete graph with self loops (which equals the product of the densities of the sets). Indeed, for the case in which all densities are equal, the following has been proven in [2]:

▶ **Theorem 2** ([2]). *Let $G=(V,E)$ be a λ-expander. For every sequence of subsets $S_1,\ldots,S_t \subseteq V$ such that S_i is of density ρ,*

- *If $\lambda < \rho/6$, then* $\Pr_{(v_1,\ldots,v_t)\sim\mathsf{RW}_G^t}[\forall i \; v_i \in S_i] \geq \rho\cdot(\rho-2\lambda)^{t-1}$.
- *If $\lambda < \rho^2/2$, then* $\Pr_{(v_1,\ldots,v_t)\sim\mathsf{RW}_G^t}[\forall i \; v_i \in S_i] \geq \rho\cdot(\rho-\lambda)^{t-1}$.

How tight are these bounds?

To get a feeling for the upper and lower bounds, let us look at the special case where all densities ρ_i are the same ρ. In this case, independent sampling gives the exact answer ρ^t. The upper bound (Theorem 1) is $\rho\mu^{t-1}$, where $\mu=\rho+(1-\rho)\lambda$, and

$$|\rho\mu^{t-1}-\rho^t| = \rho(\mu^{t-1}-\rho^{t-1}) = \rho(\mu-\rho)\sum_{j=0}^{t-2}\rho^j\mu^{t-2-j} \leq \rho(1-\rho)\lambda\sum_{j=0}^{t-2}\rho^j \leq \rho\cdot\lambda,$$

where the first equality is because $\mu \geq \rho$ and the second equality is using $a^k-b^k=(a-b)\sum_{j=0}^{k-1}a^jb^{k-1-j}$. We also use $\mu-\rho=(1-\rho)\lambda$. In particular the error term is at most λ, and tends to zero when λ tends to 0.

However, for the lower bound (Theorem 2), for any λ we have

$$\begin{aligned}\left|\rho^t-\rho(\rho-\lambda)^{t-1}\right| &= \rho(\rho^{t-1}-(\rho-\lambda)^{t-1}) = \rho\lambda\sum_{j=0}^{t-2}\rho^{t-2-j}(\rho-\lambda)^j \\ &\geq \rho\lambda\rho^{t-2}\sum_{j=0}^{t-2}(\rho-\lambda)^j \approx \rho^{t-1}\frac{\lambda}{\lambda+\frac{1}{t}}.\end{aligned}$$

Thus, when λ is some small constant, independent of t and $\rho=1-1/t$, the difference between independent sampling and the lower bound is $\rho^{t-1}\frac{\lambda}{\lambda+1/t} =\approx 1/e$. Therefore, even for arbitrarily small λ, if we let t grow to infinity and we let the density ρ depend on t, there is a constant gap between the independent sampling probability and the lower bound! Thus, a natural question is: can we find a better lower bound that matches the independent probability? In this work we prove:

▶ **Theorem 3** (New confinement lower-bound). *Let $G=(V,E)$ be a λ-expander, and let $S_1,\ldots,S_t \subseteq V$ be each of density ρ for some ρ.*[1]*. If $\lambda \leq \frac{\rho^2}{3}$, then*

$$\Pr_{(v_1,\ldots,v_t)\sim\mathsf{RW}_G^t}[\forall i \; v_i \in S_i] \geq \rho\cdot\left(\rho-\lambda(1-\rho^2)\right)^{t-1}.$$

[1] In fact, we prove the theorem under more general conditions, see Section 4

This bound is close to the independent sampling probability:

$$\begin{aligned}\left|\rho^t - \rho \cdot (\rho \ - \lambda(1-\rho^2))^{t-1}\right| &= \rho^t - \rho \cdot (\rho \ - \lambda(1-\rho^2))^{t-1} \\ &= \rho \cdot \lambda(1-\rho^2) \sum_{j=0}^{t-2} \rho^j (\rho - \lambda(1-\rho^2))^{t-2-j} \\ &\leq \lambda \cdot \rho(1-\rho^2) \cdot \sum_{j=0}^{\infty} \rho^j = \lambda \cdot \rho(1+\rho) \leq 2\rho\lambda.\end{aligned}$$

Therefore, for any λ, if we let t grow to infinity, and even if we let the density ρ depend on t, the distance between the independent probability (ρ^t) and the lower bound is at most 2λ (instead of an absolute constant before).

1.1 Further Results

Expander random walks are typically used as a randomness-efficient way of generating a uniform-like sequence of vertices $v_1, \ldots, v_t$. In most applications, the walk is used to "fool" a test function f. For example, we may think of the confinement problem when all sets S_i are the same set S, as taking an expander with $|V|$ vertices, which we label with 0 or 1 according to membership in S. We set f to be the AND function. We compare the probability that $f(x_1, \ldots, x_t)$ evaluates to 1 when $x_1, \ldots, x_t$ are the labels obtained from a random walk on the graph (which is the quantity we want to bound) with the probability that f evaluates to 1 when the labels are obtained from vertices chosen uniformly at random (which is a known quantity and equals the density S raised to the power of t). We wish to claim these two quantities are close to each other.

More generally, we say a test function $f : \mathbb{Z}_{d'}^t \to \mathbb{Z}_d$ is ε-fooled by expander random walks if for every λ-expander graph $G = (V, E)$ and every labeling $\mathsf{val} : V \to \mathbb{Z}_{d'}$, $d_{TV}\left(f(\mathsf{val}(\mathsf{RW}_G^t)),\ f(\mathsf{val}(\mathsf{Ind}_V^t))\right) \leq \varepsilon$. where

- RW_G^t is the distribution obtained by taking a length $t-1$ random walk on G. That is, we sample $v_1 \in V$ uniformly at random. Then, for $i = 2, \ldots, t$ sample v_i uniformly at random from the neighbours of v_{i-1}. $f(\mathsf{val}(\mathsf{RW}_G^t))$ is the distribution of $f(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$ when $(v_1, \ldots, v_t)$ is sampled from RW_G^t.
- Ind_V^t is the distribution obtained by sampling $v_1, \ldots, v_t \in V$ uniformly at random. Note that $\mathsf{Ind}_V^t = \mathsf{RW}_J^t$ where J is the complete graph on V with self loops. $f(\mathsf{val}(\mathsf{Ind}_V^t))$ is the distribution of $f(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$ when $(v_1, \ldots, v_t)$ is sampled from Ind_V^t.

Cohen et al. [4] proved that all Boolean *symmetric* functions f are fooled by expander random walks with up to a $O(\lambda/\sqrt{\rho_{\mathsf{min}}})$ error in total variation distance, where $\rho_{\mathsf{min}} = \min\{\rho_0, \rho_1\}$, and ρ_b is the *density* of b, i.e., that fraction of vertices with label b. Thus, even in the symmetric Boolean case, the error bound of [4] is $O(\lambda)$ only when ρ_{min} is bounded from below by some constant. When ρ_{min} is allowed to depend on t, the error bound of [4] may weaken as t increases.

A remarkable recent result of Golowich and Vadhan [8] significantly strengthened and extended the results of [4], and using new techniques managed to eliminate the dependence on the bias. That is, they prove that all symmetric *Boolean* functions are fooled by expander random walks with up to $O(\lambda)$ error in total variation distance, where the constant hidden in the Big-O notation is absolute and does not depend on ρ_{min}.

Notice that [8] implies that for confinement to a single set (which is a symmetric function) the difference between independent sampling and RW sampling is bounded by $O(\lambda)$, even when the density ρ may depend on t. Thus, it implies that Theorem 2, which gives constant difference for $\rho = 1 - 1/t$, is not tight. In this regard, Theorem 3 gives a bound that replaces the $O(\lambda)$ difference guaranteed by [8] with a more precise bound (that is in particular at most 2λ).

Let us now discuss whether the are functions for which the [8] bound does not guarantee an $O(\lambda)$ error, while our technique does.

A first candidate for such a problem is the confinement problem for $S_1 \times \ldots \times S_t$, where the sets S_i might be different, and are only guaranteed to all have the same density. Theorem 3 still guarantees the same bound, whereas [8] seems to not apply, because the function is not symmetric anymore. However, the Golowich-Vadhan result might be modified to cover this case as well, by using one fixed set, and adding corresponding permutation operators to the expanders, making them directed (which is still fine for [8]).[2]

However, using our techniques, we prove the following. Let $\mathbf{1}_S(i)$ equal 1 if $i \in S$ and 0 otherwise. Then, $\mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}$ equals one if the input is confined to $S_1 \times \ldots \times S_t$ and zero otherwise. We prove:

▶ **Theorem 4.** *Let $G = (V, E)$ be a λ-expander where $\lambda \leq 1/3$, and $t \geq 1$ an integer . Let $S_1, \ldots, S_t \subseteq V$ be a sequence of subsets such that the largest subset also has the maximal variance. Then,*

$$d_{TV}\left(\mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}(\mathsf{RW}_G^t),\ \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}(\mathsf{Ind}_V^t)\right) < 3\rho_{\mathsf{max}} \cdot \lambda,$$

where ρ_{max} is the density of the largest subset.

In particular,

▶ **Corollary 5.** *Let $G = (V, E)$ be a λ-expander where $\lambda \leq 1/3$, and $t \geq 1$ an integer. Let $S \subseteq V$ be a subset of density ρ, and suppose $S_1, \ldots, S_t \subseteq V$ are subsets such that for every i, $S_i = S$ or $S_i = \overline{S}$. Then,*

$$d_{TV}\left(\mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}(\mathsf{RW}_G^t),\ \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}(\mathsf{Ind}_V^t)\right) < 3\rho \cdot \lambda.$$

Notice that these functions are not symmetric, and therefore the results of [8] do not apply to them, while our techniques still work.

We also use similar techniques to analyze the extent to which the sum function modulo d is fooled by expander random walks on graphs with arbitrarily biased labelings, and prove that it is fooled with an $O(\sqrt{d} \cdot \lambda)$ error in total variation distance, with no dependence on the labeling bias. We prove:

▶ **Theorem 6.** *For integers $t \geq 1$, $d' \geq 2$, and $d \geq 2$ let $G = (V, E)$ be a λ-expander where $\lambda \leq 1/6$. Let $\mathsf{val} : V \to \mathbb{Z}_{d'}$ be any labeling. Then*

$$d_{TV}\left(\mathsf{Sum}_d\left(\mathsf{val}(\mathsf{RW}_G^t)\right), \mathsf{Sum}_d\left(\mathsf{val}(\mathsf{Ind}_V^t)\right)\right) \leq 5\sqrt{d} \cdot \lambda.$$

The $O(\sqrt{d} \cdot \lambda)$ error term also follows from the work of [8] on width-d permutation branching programs, using different techniques.

Additionally, we prove a bound on the bias of a labeling in terms of the density of the most frequent label, ρ_{max}. This is in contrast to previous bias-dependent result (e.g [8] for symmetric functions over $\mathbb{Z}_d$ with $d > 2$) where the total variation bound degrades with ρ_{min}, rather than $1 - \rho_{\mathsf{max}}$ (and notice that always $\rho_{\mathsf{min}} \leq 1 - \rho_{\mathsf{max}}$). This dependence is more

[2] We thank the anonymous referee for bringing this to our attention.

resilient as it can tolerate very rare labels, as long as the most common label is not too dominant. We think that this observation could potentially serve as an incentive to shift the bias dependence in previous works from the smallest label weight to the largest. Specifically, we prove,

▶ **Proposition 7.** *For a prime p, let $\mathsf{val} : [n] \to \mathbb{Z}_p$ be a labeling that assigns label $a \in \mathbb{Z}_p$ to ρ_a fraction of the vertices, and denote $\rho_{\mathsf{max}} = \max_a \rho_a$. Then, for every non-trivial character χ of $\mathbb{Z}_p$, $\mathsf{bias}_\chi(\mathsf{val}) \leq \sqrt{1 - \left(1 - \cos\frac{2\pi}{p}\right)(1 - \rho_{\mathsf{max}})}$, where $\mathsf{bias}_\chi(\mathsf{val}) \stackrel{\text{def}}{=} \left|\mathbb{E}_{i \in [n]}\chi(\mathsf{val}(i))\right|$.*

We also point out that our proofs apply even if the graph is different for each of the t steps, as long as it is a λ-expander at each step. The same property holds in previous works as well, e.g [8].

1.2 The Technique

We extend the techniques of Gillman [6], Healy [9] and Beck [3], that established a framework for analyzing what is often referred to as the "flow" of linear operators. The flow of a linear operator T from the linear subspace V_2 to the linear subspace V_1 is the quantity $\|\Pi_1 T \Pi_2\|$ where Π_i is the projection operator onto V_i. In our context, V_1 and V_2 will be either the line spanned by the all-ones vector (The "parallel space"), or its orthogonal complement (The "perpendicular space").

Let G also denote the transition matrix of a λ-expander graph, and let P denote the projection matrix on the set S. That is, P is the diagonal matrix satisfying $P[v, v] = 1$ if $v \in S$ and 0 otherwise. The probability that a length t random walk on G never escapes S can be expressed algebraically as $\mathbf{1}^T(PG)^{t-1}P\mathbf{1}$, where we denote $\mathbf{1} = \frac{1}{\sqrt{|V|}}(1, \ldots, 1)^T$.

One way to analyze this expression is to decompose the probability distribution at each of the t steps to its parallel and perpendicular components. The parallel component is identical to the independent sampling case, while the perpendicular component is shrunk by a factor of λ after each step on G. The above approach underlies many results in the field, and, in particular, the expander Chernoff bound [6, 9]. Beck [3] simplified the analysis by defining a 2×2 "flow" matrix for a linear operator T. The i, j'th entry of the flow matrix is the flow of T from V_j to V_i, where V_i and V_j are either the perpendicular space or the parallel space. This notation reduced the problem of bounding quantities like $\left|\mathbf{1}^T T \mathbf{1}\right|$ to bounding the $[0, 0]$ entry of a 2×2 matrix with non-negative entries. In this language, the expression $\mathbf{1}^T(PG)^{t-1}P\mathbf{1}$ is the flow of the operator $(PG)^{t-1}P$ from the parallel space to itself. For more details about the flow framework see Section 3.

[2] proved their confinement probability lower bound by giving simultaneous upper and lower bounds on flows between the perpendicular and parallel spaces. However, they did it explicitly and specifically for the confinement problem with equal density at each step, and obtained sub-optimal bounds. In this paper we analyze flows emerging from confinement problems (and additional problems) using the 2×2 flow matrix notation. As a result, we achieve simpler terms that are easier to follow and generalize to a broader setting of confinement problems with varying densities. These terms also indicate how to improve upon previous work (even when all densities are equal).

1.3 Summary and Discussion

As mentioned before, several total variation bounds in previous works depend on the labeling bias, namely on the weights ρ_b that are induced by a labeling. Cohen et al. [4] proved that all Boolean symmetric functions are fooled by expander random walks with up to a $O(\lambda/\sqrt{\rho_{\mathsf{min}}})$ error in total variation distance.

Recently, Golowich and Vadhan [8], significantly strengthened and extended these results using new techniques, and in some cases managed to eliminate the dependence on the bias. In particular, they prove that for the Boolean case, all symmetric functions are fooled by expander random walks with up to $O(\lambda)$ error in total variation distance, where the constant hidden in the Big-O notation is absolute.

For the non-Boolean case much less is known:

- For *symmetric* functions defined on $\mathbb{Z}_d^t$, Golowich and Vadhan prove an $O((\frac{d}{\rho_{\mathsf{min}}})^{O(d)} \cdot \lambda)$ total-variation bound where $\rho_{\mathsf{min}} = \min_a \rho_a$, and ρ_a is the density of label a. Notice that in this bound there is a dependence on ρ_{min}. It is an intriguing open problem whether the dependence on the bias is necessary.
- Golowich and Vadhan [8] also 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, a bound that does not depend on the bias of the labeling. Notice that this bias-independent bound also holds for non-symmetric functions, as long as they are computed by a low-width permutation branching program.

In this work we add another example where the error bound does not depend on the labeling bias. We show for the confinement problem, when the set of maximal density $\rho(S)$ is also of maximal variance (the variance is $\sqrt{\rho(S)(1-\rho(S))}$), the error bound is $O(\lambda)$ regardless of the densities. Note that this case is not symmetric. We also improve the lower bound for the symmetric case, as previously discussed.

There are many open problems left.

- First, and foremost, is it possible that all symmetric functions over Σ^t are $O_{|\Sigma|}(\lambda)$ fooled by random-walks? For $\Sigma = \{0, 1\}$ [8] gave an affirmative answer, but the general case is left open.
- What other non-symmetric functions are fooled by random-walks without a dependence on the bias? [8] showed all small-width permutation branching programs are such. We added the confinement test functions when all sets have the same variance. What other functions have this property?
- As alluded to by Proposition 7, we think that for many functions the parameter dominating the bias-dependent error is $1 - \rho_{\mathsf{max}}$ rather than ρ_{min}. For example, the bias-dependent bound for *any* confinement test function (Proposition 29) is $O(\frac{\lambda}{1-\rho_{\mathsf{max}}})$ where ρ_{max} is the density of the largest set. It would be interesting to examine previous results and see if the error terms can be correspondingly amended.

The paper is organized as follows: In Section 2 we give some preliminaries and background, and introduce our notations. In Section 3 we review Beck's flow framework [3] and extend it. In Section 4 we prove Theorem 3, and prove analogous lower bounds in the general setting of varying sets and densities. In Section 5 we study fooling confinement test functions, and in particular prove Theorem 4. In Section 6, we prove Theorem 6 using our techniques. The proof for Proposition 7 appears in the full version of this paper [14].

2 Preliminaries

Notation

For any positive integer d, let $\mathbb{Z}_d$ denote the group of integers modulo d, and $[d] = \{1, \ldots, d\}$. We define the ℓ_1-norm of a vector $x \in \mathbb{F}^n$ as $\|x\|_1 = \sum_i |x_i|$, and its ℓ_2-norm as $\|x\| = \sqrt{\sum_i |x_i|^2}$. For a field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, let $\mathbf{1}_n = (1/\sqrt{n}, \ldots, 1/\sqrt{n}) \in \mathbb{F}^n$ denote the normalized all-ones vector. When n is clear from context we simply write $\mathbf{1}$. For a matrix $M \in \mathbb{F}^{n\times n}$, the operator norm of M is given by $\max_{x\in\mathbb{F}^n\setminus\{0\}} \|Mx\| \,/\, \|x\|$. For $M \in \mathbb{C}^{n\times n}$, its conjugate transpose is denoted as M^*. For two real matrices $L, M \in \mathbb{R}^{n\times n}$, the notation $L \leq_{e.w} M$ stands for entry-wise inequality

A symmetric matrix $W \in [0,1]^{n\times n}$ is an undirected random walk matrix on n vertices if the columns and rows of W sum to 1, which implies that $W_{j,i} = W_{i,j}$ represents the transition probability between vertex i and j, or vice versa. In this context, I_n denotes the $n \times n$ identity matrix, and $J_n = \mathbf{1}_n \mathbf{1}_n^T$ represents a matrix with all entries being $1/n$. When the dimension is clear from the context, we use the notations I and J respectively. Notably, J_n is the random walk matrix for a complete graph on n vertices with self-loops. For a sequence of matrices $M_1, \ldots, M_t$, we denote $\prod_{i=1}^{t} M_i = M_t \cdot M_{t-1} \cdot \ldots \cdot M_1$.

We often use the decomposition $\mathbb{F}^n = \mathbf{V}_0 \oplus \mathbf{V}_1$ where $\mathbf{V}_0 = \mathsf{Span}\{\mathbf{1}\}$ is the subspace of $\mathbb{F}^n$ spanned of the all ones vector, and $\mathbf{V}_1 = \mathbf{V}_0^{\perp}$ is its orthogonal complement. We define Π_0 as the projection operator onto $\mathbf{V}_0$, noting that $\Pi_0 = J_n$, and Π_1 as the projection on $\mathbf{V}_1$, noting that $\Pi_1 = I_n - J_n$. For a vector $x \in \mathbb{F}^n$ we define $x^{\parallel} = \Pi_0 x$ and $x^{\perp} = \Pi_1 x$.

For two probability distributions p_1 and p_2 over a finite sample space Ω, their total variation distance is $d_{TV}(p_1, p_2) = \frac{1}{2} \cdot \sum_{s\in\Omega} |p_1(s) - p_2(s)|$.

The Information Theoretic XOR-Lemma

The characters of the group $\mathbb{Z}_d$ are the maps $\chi_b(a) = \omega_d^{b\cdot a}$ for $b = 0, \ldots, d-1$, where $\omega_d = e^{\frac{2\pi i}{d}}$. Let $\mathbb{C}^{\mathbb{Z}_d}$ denote the vector space of all complex valued function on $\mathbb{Z}_d$, equipped with the inner product $\langle h, g\rangle = \sum_{a\in\mathbb{Z}_d} h(a)\overline{g(a)}$.

The information theoretic XOR-Lemma [7] relates the total variation distance between two distributions over $\mathbb{Z}_d$ to the heaviest Fourier coefficient of their difference, also called the maximum bias.

▶ **Lemma 8** (Based on [7]). *For any two distributions p_1 p_2 over $\mathbb{Z}_d$: $d_{TV}(p_1, p_2) \leq \frac{\sqrt{d}}{2} \cdot \max_{b\in\mathbb{Z}_d} |\langle \chi_b, p_1 - p_2\rangle|$.*

The proof, based on [7], appears in the full version of this paper.

Expanders

For a regular, undirected graph $G = (V, E)$ on n vertices, the random walk matrix is the normalized adjacency matrix. The *spectral expansion* is defined as the second largest eigenvalue of the graph's random walk matrix in absolute value, namely $\lambda(G) = \max_{x,y\perp\mathbf{1}} \frac{|\langle x, Gy\rangle|}{\|x\|\cdot\|y\|} = \max_{x\perp\mathbf{1}} \frac{\|Gx\|}{\|x\|}$, where the maximum is over all non-zero $x, y \in \mathbb{R}^n$ which are orthogonal to the all-ones vector, and by abuse of notation G also denotes the random walk matrix of the graph G. We say G is a λ-expander if $\lambda(G) = \lambda$. For a λ-expander G, let $A = \frac{1}{\lambda}(G - J)$. Since the all-ones vector is an eigenvector of both G and J with eigenvalue 1, it follows that A is zero on the parallel space $\mathsf{Span}\{\mathbf{1}\}$. Additionally, $\|Ax\| = \left\|Ax^{\perp} + Ax^{\parallel}\right\| = \frac{1}{\lambda} \cdot \left\|Gx^{\perp}\right\| \leq \left\|x^{\perp}\right\| \leq \|x\|$. This implies a valuable decomposition $G = J + \lambda A$ where the symmetric "error matrix" A is zero on the parallel space, and $\|A\| \leq 1$. Another useful decomposition follows by setting $E = \frac{1}{\lambda}(G - (1-\lambda)\cdot J)$. One can easily verify that E acts like the identity on the parallel space, and that the orthogonal space is E-invariant. Thus, for every vector x we have

$$\|Ex\|^2 = \left\|Ex^{\parallel}\right\|^2 + \left\|Ex^{\perp}\right\|^2 \leq \left\|x^{\parallel}\right\|^2 + \frac{1}{\lambda}\left\|Gx^{\perp}\right\|^2 \leq \|x\|^2 .$$

This gives rise to the decomposition $G = (1-\lambda)J + \lambda E$ where the symmetric "error matrix" E satisfies $\|E\| \leq 1$.

3 Flow

Let $\mathbb{F}$ be either $\mathbb{C}$ or $\mathbb{R}$. We decompose $\mathbb{F}^n = \mathbf{V}_0 \oplus \mathbf{V}_1$ where $\mathbf{V}_0$ is the span of the all-ones vector $\mathsf{Span}\{\mathbf{1}\}$ (the "parallel" space) and $\mathbf{V}_1 = \mathbf{V}_0^{\perp}$ its orthogonal complement (the "orthogonal" space). Let Π_0 be the projection operator onto $\mathbf{V}_0$, and Π_1 the projection onto $\mathbf{V}_1$.

Throughout this work we study linear operators $T : \mathbb{F}^n \to \mathbb{F}^n$ by examining $\|\Pi_{b_1} T \Pi_{b_2}\|$ for $b_1, b_2 \in \{0,1\}$. Intuitively, this can be understood as the "flow of mass" from $\mathbf{V}_{b_2}$ to $\mathbf{V}_{b_1}$ under the linear operator T. To study the flow of a linear operator, we extend upon the techniques introduced by Gillman, Healy, and Beck, using the notation and claims of Beck [3]. These were used mostly in the context of the expander Chernoff bound [6, 9].

▶ **Definition 9** (The Flow Matrix). *Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be any linear operator. Then the flow matrix of T, denoted $\widetilde{T}$, is the 2×2 non-negative matrix defined by*

$$\widetilde{T} = \begin{pmatrix} \|\Pi_0 T \Pi_0\| & \|\Pi_0 T \Pi_1\| \\ \|\Pi_1 T \Pi_0\| & \|\Pi_1 T \Pi_1\| \end{pmatrix}$$

▶ **Example 10.** Let G be the random walk operator of a λ-expander graph. Then

$$\widetilde{G} \leq_{e.w} \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$$

where $\leq_{e.w}$ stands for entry-wise inequality.

To see this, apply the decomposition $G = J + \lambda A$ where $\|A\| \leq 1$ and A is zero on the parallel space. That is, $A\Pi_0 = \Pi_0 A = 0$. We then have

(i) $\|\Pi_0 G \Pi_0\| = \|\Pi_0 J \Pi_0\| = \|J\| = 1$,
(ii) $\|\Pi_0 G \Pi_1\| = \|\Pi_0 (J + \lambda A) \Pi_1\| = \|\Pi_0 J \Pi_1 + \lambda \Pi_0 A \Pi_1\| = 0$,
(iii) By symmetry $\|\Pi_1 G \Pi_0\| = 0$.
(iv) Finally, $\|\Pi_1 G \Pi_1\| = \lambda \|\Pi_1 A \Pi_1\| \leq \lambda$.

By submultiplicativity and subadditivity of the operator norm, we have the following submultiplicativity property of the flow operator:

▷ Claim 11 ([3]). For every linear operators $L, M : \mathbb{F}^n \to \mathbb{F}^n$, we have $\widetilde{L \cdot M} \leq_{e.w} \widetilde{L} \cdot \widetilde{M}$.

Proof. Let $i, j \in \{0,1\}$. Recall that $\Pi_0 = J$ and $\Pi_1 = I - J$, and thus $\Pi_0 + \Pi_1 = I$. We have

$$\begin{aligned} \widetilde{L \cdot M}[i,j] = \|\Pi_i L M \Pi_j\| = \|\Pi_i L (\Pi_0 + \Pi_1) M \Pi_j\| &\leq \|\Pi_i L \Pi_0 M \Pi_j\| + \|\Pi_i L \Pi_1 M \Pi_j\| \\ &\leq \|\Pi_i L \Pi_0\| \cdot \|\Pi_0 M \Pi_j\| + \|\Pi_i L \Pi_1\| \cdot \|\Pi_1 M \Pi_j\| \\ &= \widetilde{L}[i,0] \cdot \widetilde{M}[0,j] + \widetilde{L}[i,1] \cdot \widetilde{M}[1,j] = \widetilde{L} \cdot \widetilde{M}[i,j]. \end{aligned}$$ ◁

Typically, the primary technical tool utilized for analyzing flow matrices consists of the following bound, which generally hold for non-negative 2×2 matrices.

▶ **Lemma 12** ([3]). *If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \geq_{e.w} 0$ with $a \geq 1$ and $d < 1$, then*

$$A^t[0,0] \leq a \cdot \left(a + \frac{bc}{1-d} \right)^{t-1}$$

Proof. By induction on t. The base case $t = 1$ is clear. Assume for $1, \ldots, t-1$ and let us prove for t. We have the following recurrence relation

$$A^t[0,0] = A^{t-1}[0,0] \cdot A[0,0] + \sum_{j=0}^{t-2} A^j[0,0] \cdot A[0,1] \cdot A[1,1]^{t-2-j} \cdot A[1,0]$$

where j goes over the last time the path was at vertex 0 before taking the final step. As $A[i,j] \geq 0$ and $A[0,0] \geq 1$, we see that $A^{k_2}[0,0] \geq A^{k_1}[0,0]$ for all $k_2 \geq k_1$. Hence,

$$\begin{aligned} A^t[0,0] &= A^{t-1}[0,0] \cdot a + \sum_{j=0}^{t-2} A^j[0,0] \cdot bc \cdot d^{t-2-j} \\ &\leq A^{t-1}[0,0] \left(a + bc \sum_{j=0}^{\infty} d^j \right) \leq A^{t-1}[0,0] \left(a + \frac{bc}{1-d} \right) \end{aligned}$$

The proof is complete by applying the induction hypothesis. ◀

A simple way to generalize this lemma to the case where $A[0,0] > A[1,1]$ but not necessarily $A[0,0] > 1$ is as follows.

▶ **Lemma 13.** *If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \geq_{e.w} 0$ with $a > d$ then $A^t[0,0] \leq a \cdot \left(a + \frac{bc}{a-d} \right)^{t-1}$.*

Proof. Write $A = a \cdot \begin{pmatrix} 1 & \frac{b}{a} \\ \frac{c}{a} & \frac{d}{a} \end{pmatrix}$. Then, by the previous lemma

$$A^t[0,0] \leq a^t \cdot \left(1 + \frac{\frac{b}{a} \cdot \frac{c}{a}}{1 - \frac{d}{a}} \right)^{t-1} = a \cdot \left(a + \frac{bc}{a-d} \right)^{t-1}.$$ ◀

▶ Remark 14. Note that the lemma above is not tight when a is small. Indeed, $A^t[0,0]$ decreases with a, while the bound of Lemma 13 blows up when a approaches d. We do not try to optimize the bound for d close to a. Also, it would be nice to have a generalization of this lemma for the case of possibly different $A_1, \ldots, A_t$.

▶ **Lemma 15.** *If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \geq_{e.w} 0$ with $a > d$. Then for all $t \geq 2$,*

$$A^t[0,0] - (A[0,0])^t \leq \frac{abc}{a-d} \cdot \sum_{k=0}^{t-2} a^k \left(a + \frac{bc}{a-d} \right)^{t-k-2}$$

Proof. For every integer $k \geq 1$ $x^k - y^k = (x-y) \cdot \sum_{i=0}^{k-1} x^i y^{k-i-1}$. Using this and Lemma 13, we see that

$$\begin{aligned} A^t[0,0] - (A[0,0])^t &\leq a \cdot \left(a + \frac{bc}{a-d} \right)^{t-1} - a^t = a \left(\left(a + \frac{bc}{a-d} \right)^{t-1} - a^{t-1} \right) \\ &= a \cdot \frac{bc}{a-d} \cdot \sum_{k=0}^{t-2} a^k \left(a + \frac{bc}{a-d} \right)^{t-k-2} \end{aligned}$$ ◀

▶ **Lemma 16.** *For an integer $t \geq 1$, Let $M_1, \ldots, M_t$ be a sequence of $n \times n$ matrices. Then*

$$\left| \mathbf{1}^T \left(\prod_{i=1}^{t} M_i \right) \mathbf{1} - \mathbf{1}^T \left(\prod_{i=1}^{t} \Pi_0 M_i \right) \mathbf{1} \right| \leq \left(\prod_{i=1}^{t} \widetilde{M_i} \right) [0,0] - \left(\prod_{i=1}^{t} \widetilde{M_i}[0,0] \right).$$

Proof. Writing $M_i = \Pi_0 M_i + \Pi_1 M_i$ we have

$$\mathbf{1}^T \left(\prod_{i=1}^{t} M_i\right) \mathbf{1} = \sum_{b\in\{0,1\}^t} \mathbf{1}^T \prod_{i=1}^{t} (\Pi_{b_i} M_i)\mathbf{1} = \sum_{b\in\{0,1\}^{t-1}} \mathbf{1}^T \Pi_0 M_t \left(\prod_{i=1}^{t-1} (\Pi_{b_i} M_i)\right) \cdot \mathbf{1}$$

Since $\mathbf{1}^T\Pi_1 = 0$. To complete the proof let $LHS = \left|\mathbf{1}^T \prod_{i=1}^{t} M_i \mathbf{1} - \mathbf{1}^T \left(\prod_{i=1}^{t} (\Pi_0 M_i)\right) \mathbf{1}\right|$. Then,

$$\begin{aligned}
LHS &= \left| \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} \mathbf{1}^T \Pi_0 M_t \left(\prod_{i=1}^{t-1} (\Pi_{b_i} M_i)\right) \cdot \mathbf{1} \right| \\
&\leq \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} \left| \mathbf{1}^T \Pi_0 M_t \left(\prod_{i=1}^{t-1} (\Pi_{b_i} M_i)\right) \cdot \Pi_0 \mathbf{1} \right| \leq \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} \left\| \Pi_0 M_t \left(\prod_{i=1}^{t-1} (\Pi_{b_i} M_i)\right) \cdot \Pi_0 \right\| \\
&= \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} \left\| \Pi_0 M_t \Pi_{b_{t-1}} \left(\prod_{i=2}^{t-1} \left(\Pi_{b_i} M_i \Pi_{b_{i-1}}\right)\right) \cdot \Pi_{b_1} M_1 \Pi_0 \right\| \\
&\leq \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} \left\| \Pi_0 M_t \Pi_{b_{t-1}} \right\| \cdot \prod_{i=2}^{t-1} \left\| \Pi_{b_i} M_i \Pi_{b_{i-1}} \right\| \left\| \Pi_{b_1} M_1 \Pi_0 \right\| \\
&= \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} \widetilde{M_t}[0, b_{t-1}] \left(\prod_{i=2}^{t-1} \widetilde{M_i}[b_i, b_{i-1}]\right) \widetilde{M_1}[b_1, 0] = \left(\prod_{i=1}^{t} \widetilde{M_i}\right)[0,0] - \prod_{i=1}^{t} \widetilde{M_i}[0,0].
\end{aligned}$$ ◀

▶ **Lemma 17.** *Let $A_1, \ldots, A_t$ be a sequence of non-negative 2×2 matrices such that for all i, $A_i \leq_{e.w} A$ for some 2×2 matrix A. Then*

$$\left(\prod_{i=1}^{t} A_i\right)[0,0] - \prod_{i=1}^{t} (A_i[0,0]) \leq A^t[0,0] - (A[0,0])^t.$$

Proof. We have

$$\begin{aligned}
\left(\prod_{i=1}^{t} A_i\right)[0,0] - \prod_{i=1}^{t} (A_i[0,0]) &= \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^t}} A_t[0, b_{t-1}] \left(\prod_{i=2}^{t-1} A_i[b_i, b_{i-1}]\right) A_1[b_1, 0] \\
&\leq \sum_{\substack{b\in\{0,1\}^{t-1}\\ b\neq 0^{t-1}}} A[0, b_{t-1}] \left(\prod_{i=2}^{t-1} A[b_{i+1}, b_i]\right) A[b_1, 0] = A^t[0,0] - (A[0,0])^t.
\end{aligned}$$ ◀

We now proceed to establish techniques for proving flow lower bounds. While these concepts were introduced specifically for the confinement problem with the same set density in [2], we extend them to general linear operators and use the flow matrix notation.

▶ **Lemma 18** (Flow Progress). *For linear operators $T_1, \dots, T_t$,*

$$\widetilde{\prod_{i=1}^{t} T_i}[0,0] \geq \widetilde{T_t}[0,0] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[0,0] - \widetilde{T_t}[0,1] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[1,0] \tag{2}$$

$$\widetilde{\prod_{i=1}^{t} T_i}[1,0] \leq \widetilde{T_t}[1,0] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[0,0] + \widetilde{T_t}[1,1] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[1,0] \tag{3}$$

Proof. We have

$$\begin{aligned}
\widetilde{\prod_{i=1}^{t} T_i}[0,0] &= \left\| \Pi_0 \prod_{i=1}^{t} T_i \Pi_0 \right\| = \left\| \Pi_0 \left(T_t \Pi_0 + T_t \Pi_1 \right) \prod_{i=1}^{t-1} T_i \Pi_0 \right\| \\
&\geq \|\Pi_0 T_t \Pi_0\| \cdot \left\| \Pi_0 \prod_{i=1}^{t-1} T_i \Pi_0 \right\| - \|\Pi_0 T_t \Pi_1\| \cdot \left\| \Pi_1 \prod_{i=1}^{t-1} T_i \Pi_0 \right\| \\
&= \widetilde{T_t}[0,0] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[0,0] - \widetilde{T_t}[0,1] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[1,0]
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\prod_{i=1}^{t} T_i}[1,0] &= \left\| \Pi_1 \prod_{i=1}^{t} T_i \Pi_0 \right\| = \left\| \Pi_1 \left(T_t \Pi_0 + T_t \Pi_1 \right) \prod_{i=1}^{t-1} T_i \Pi_0 \right\| \\
&\leq \|\Pi_1 T_t \Pi_0\| \cdot \left\| \Pi_0 \prod_{i=1}^{t-1} T_i \Pi_0 \right\| + \|\Pi_1 T_t \Pi_1\| \cdot \left\| \Pi_1 \prod_{i=1}^{t-1} T_i \Pi_0 \right\| \\
&= \widetilde{T_t}[1,0] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[0,0] + \widetilde{T_t}[1,1] \cdot \widetilde{\prod_{i=1}^{t-1} T_i}[1,0]
\end{aligned}$$

◀

▶ **Definition 19** (Flow sequence). *For a sequence of linear operators $T_1, \dots, T_t$, the flow sequence is defined recursively such that $c_1 = \frac{\widetilde{T_1}[0,0]}{\widetilde{T_1}[1,0]}$ and for $k \geq 1$*

$$c_{k+1} = \frac{\widetilde{T_{k+1}}[0,0] \cdot c_k - \widetilde{T_{k+1}}[0,1]}{\widetilde{T_{k+1}}[1,0] \cdot c_k + \widetilde{T_{k+1}}[1,1]}$$

The constants c_i emerge from recursively dividing Equation 2 of Lemma 18 by Equation 3, as demonstrated by the following lemmas. Therefore, from an intuitive perspective, the constants c_i in the definition above can be thought of as a lower bound on the ratio between the mass preserved inside the parallel space after the i-th step and the mass lost to its orthogonal complement.

We remark that the smaller $\widetilde{T_i}[0,1], \widetilde{T_i}[1,1]$ are taken relative to $\widetilde{T_i}[0,0]$ and $\widetilde{T_i}[1,0]$, the larger sequence elements will become. In all of our use cases, each operator T_i includes a step on a λ-expander graph G. Thus, as we shall later see, we can make $\widetilde{T_i}[0,1]$ and $\widetilde{T_i}[1,1]$ smaller by taking the the expansion parameter λ smaller, and hence the sequence elements larger. Specifically, in all instances considered in this work, the constants c_i are strictly positive. Therefore, for the remainder of this section, we proceed with the assumption that the provided linear operators $T_1, \dots, T_t$ are such that their corresponding flow sequence elements are positive.

▶ **Lemma 20.** *Let $T_1, \dots, T_t$ be linear operators with a positive flow sequence. Then, for all $k = 1, \dots, t$ it holds that $\widetilde{\prod_{i=1}^{k} T_i}[0,0] \geq c_k \cdot \widetilde{\prod_{i=1}^{k} T_i}[1,0]$.*

Proof. By induction on k. For $k = 1$ the claim holds by definition. For the induction step, assume that $\widetilde{\prod_{i=1}^{k} T_i}[0,0] \geq c_k \cdot \widetilde{\prod_{i=1}^{k} T_i}[1,0]$. Plugging the induction hypothesis into Equation 2 see that

$$\begin{aligned}
\widetilde{\prod_{i=1}^{k+1} T_i}[0,0] &\geq \widetilde{T_{k+1}}[0,0] \cdot \widetilde{\prod_{i=1}^{k} T_i}[0,0] - \widetilde{T_{k+1}}[0,1] \cdot \widetilde{\prod_{i=1}^{k} T_i}[1,0] \\
&\geq \left(\widetilde{T_{k+1}}[0,0] - \frac{\widetilde{T_{k+1}}[0,1]}{c_k} \right) \widetilde{\prod_{i=1}^{t-1} T_i}[0,0].
\end{aligned}$$

Similarly, plugging the induction hypothesis into Equation 3,

$$\begin{aligned}
\widetilde{\prod_{i=1}^{k+1} T_i}[1,0] &\leq \widetilde{T_{k+1}}[1,0] \cdot \widetilde{\prod_{i=1}^{k} T_i}[0,0] + \widetilde{T_{k+1}}[1,1] \cdot \widetilde{\prod_{i=1}^{k} T_i}[1,0] \\
&\leq \left(\widetilde{T_{k+1}}[1,0] + \frac{\widetilde{T_{k+1}}[1,1]}{c_k} \right) \widetilde{\prod_{i=1}^{k} T_i}[0,0].
\end{aligned}$$

Combining these we obtain

$$\begin{aligned}
\widetilde{\prod_{i=1}^{k+1} T_i}[0,0] &\geq \frac{\left(\widetilde{T_{k+1}}[0,0] - \frac{\widetilde{T_{k+1}}[0,1]}{c_k} \right)}{\left(\widetilde{T_{k+1}}[1,0] + \frac{\widetilde{T_{k+1}}[1,1]}{c_k} \right)} \prod_{i=1}^{k} T_i[0,0] \\
&= \left(\frac{\widetilde{T_{k+1}}[0,0] \cdot c_k - \widetilde{T_{k+1}}[0,1]}{\widetilde{T_{k+1}}[1,0] \cdot c_k + \widetilde{T_{k+1}}[1,1]} \right) \widetilde{\prod_{i=1}^{k} T_i}[0,0] = c_{k+1} \cdot \widetilde{\prod_{i=1}^{k} T_i}[0,0]
\end{aligned}$$ ◀

Hence we have the following corollary

▶ **Corollary 21.** *For all $k = 1, \dots, t$ we have*

$$\widetilde{\prod_{i=1}^{k} T_i}[0,0] \geq \widetilde{T_1}[0,0] \cdot \prod_{i=2}^{k} \left(\widetilde{T_i}[0,0] - \frac{\widetilde{T_i}[0,1]}{c_{i-1}} \right)$$

Proof. By induction on k. For $k = 1$ the product on the right hand side is empty and the equality trivially holds. For the induction step, Using Equation 2, the previous claim, and the induction hypothesis,

$$\begin{aligned}
\widetilde{\prod_{i=1}^{k+1} T_i}[0,0] &\geq \widetilde{T_{k+1}}[0,0] \cdot \widetilde{\prod_{i=1}^{k} T_i}[0,0] - \widetilde{T_{k+1}}[0,1] \cdot \widetilde{\prod_{i=1}^{k} T_i}[1,0] \\
&\geq \left(\widetilde{T_{k+1}}[0,0] - \frac{\widetilde{T_{k+1}}[0,1]}{c_k} \right) \widetilde{\prod_{i=1}^{k} T_i}[0,0] \geq \widetilde{T_1}[0,0] \cdot \prod_{i=2}^{k+1} \left(\widetilde{T_i}[0,0] - \frac{\widetilde{T_i}[0,1]}{c_{i-1}} \right).
\end{aligned}$$ ◀

4 Expander Hitting Property Revised

We use the following notations. For a set $S_i \subseteq [n]$ we define its density as $\rho_i = |S_i|/n$ and its variance as $\sigma_i = \sqrt{\rho_i(1-\rho_i)}$. We let P_i be the projection matrix on the set S_i. That is, P_i is the diagonal matrix satisfying $P_i[v,v] = 1$ if $v \in S_i$ and 0 otherwise. G is the random walk operator of the graph G.

4.1 Confinement Probability Upper-bounds

We begin with the hitting property for sets with possibly different densities. In [10] the authors give the bound $\prod_{j=1}^{t-1}(\sqrt{\rho_j\rho_{j+1}}+\lambda)$, which corresponds to $\|P_tG\dots GP_1\|$ rather than $\mathbf{1}^TP_tG\dots GP_1\mathbf{1}$. However, we observe that this loss is not necessary.

▶ **Proposition 22** (Expander Hitting Property). *Let $G=(V,E)$ be a λ-expander. Then, for every sequence of subsets $S_1,\dots,S_t \subseteq V$ such that S_i is of density ρ_i,*

$$\Pr_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t}[\forall i\ v_i \in S_i] \le \sqrt{\rho_1\rho_t}\cdot\prod_{i=1}^{t-1}\left((1-\lambda)\sqrt{\rho_i\rho_{i+1}}+\lambda\right).$$

Proof. First note that for all i, $\|P_iJP_{i+1}\| = \sqrt{\rho_i\rho_{i+1}}$. Indeed,

$$\|P_iJP_{i+1}\| = \left\|P_i\mathbf{1}(P_{i+1}\mathbf{1})^T\right\| = \|P_i\mathbf{1}\|\cdot\|P_{i+1}\mathbf{1}\| = \sqrt{\rho_i\rho_{i+1}}$$

Decomposing $G=(1-\lambda)J+\lambda E$ with $\|E\|\le 1$, we find that

$$\begin{aligned}\|P_iGP_{i+1}\| &= \|(1-\lambda)\cdot P_iJP_{i+1}+\lambda\cdot P_iEP_{i+1}\|\\ &\le (1-\lambda)\cdot\|P_iJP_{i+1}\|+\lambda \le (1-\lambda)\sqrt{\rho_i\rho_{i+1}}+\lambda.\end{aligned}$$

Let $u=(1/n,\dots,1/n)\in\mathbb{R}^n$ be the uniform vector. Expressing the probability linear-algebraically we obtain

$$\begin{aligned}\Pr_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t}[\forall i\ v_i\in S_i] &= \mathbf{1}^TP_t\left(\prod_{i=1}^{t-1}GP_i\right)\mathbf{1} = \mathbf{1}^TP_t\prod_{i=1}^{t-1}(P_{i+1}GP_i)P_1\mathbf{1}\\ &= \left\|P_t\prod_{i=1}^{t-1}(P_{i+1}GP_i)P_1u\right\|_1 \le \sqrt{\rho_t\cdot n}\cdot\left\|\prod_{i=1}^{t-1}(P_{i+1}GP_i)P_1u\right\|\\ &\le \sqrt{\rho_t\cdot n}\cdot\left\|\prod_{i=1}^{t-1}(P_{i+1}GP_i)\right\|\cdot\|P_1u\| = \sqrt{\rho_t\cdot n}\cdot\left\|\prod_{i=1}^{t-1}(P_{i+1}GP_i)\right\|_2\cdot\sqrt{\frac{\rho_1}{n}}\\ &\le \sqrt{\rho_t\rho_1}\cdot\prod_{i=1}^{t-1}\|(P_{i+1}GP_i)\| \le \sqrt{\rho_1\rho_t}\cdot\prod_{i=1}^{t-1}\left((1-\lambda)\sqrt{\rho_i\rho_{i+1}}+\lambda\right),\end{aligned}$$

where we use $P_i^2 = P_i$, and the first inequality is Cauchy-Schwartz, noting that after multiplying by P_t, the resulting vector has at most $\rho_t\cdot n$ non-zero entries. ◀

4.2 Confinement Probability Lower-bounds

As explained in the introduction, previous lower bounds do not give an $O(\lambda)$ bound on the error term comparing with the independent sampling case. In this section, we give a tighter lower bound that, in particular, is $O(\lambda)$-close to the probability of the same confinement event but with independently chosen samples.

Expressing the probability linear-algebraically we find that

$$\mathop{\mathbf{Pr}}_{(v_1,\dots,v_t)\sim\mathsf{RW}_G^t}[\forall i\ v_i \in S_i] = \mathbf{1}^T \prod_{i=2}^{t} (P_iG)\, P_1\mathbf{1} = \left\| \Pi_0 \prod_{i=1}^{t} (P_iG)\, \Pi_0 \right\| = \widetilde{\prod_{i=1}^{t} (P_iG)}[0,0].$$

Therefore, we see that this quantity is applicable to bounds via the lower-bound part of the flow framework. Consider the sequence of linear operators $P_1G,\dots,P_tG$ with corresponding flow sequence $c_1,\dots c_t$. It follows from Corollary 21 that

$$\widetilde{\prod_{i=1}^{t} (P_iG)}[0,0] \geq \widetilde{P_1G}[0,0] \cdot \prod_{i=2}^{k+1} \left(\widetilde{P_iG}[0,0] - \frac{\widetilde{P_iG}[0,1]}{c_{i-1}} \right)$$

Hence, our next objective is to bound the entries of $\widetilde{P_iG}$, and find lower bounds on the constants $c_1,\dots c_t$.

▶ **Lemma 23.** *For all $i = 1,\dots,t$ we have $\widetilde{P_iG} \leq_{e.w} \begin{pmatrix} \rho_i & \lambda\sigma_i \\ \sigma_i & \lambda \end{pmatrix}$ where the first column holds with equality.*

Proof. First, observe that

$$\widetilde{P_iG}[0,0] = \|\Pi_0 P_i G \Pi_0\| = \left\|\mathbf{1}\mathbf{1}^T P_i G \mathbf{1}\mathbf{1}^T\right\| = \left\|\mathbf{1}\mathbf{1}^T P_i \mathbf{1}\mathbf{1}^T\right\| = \left|\mathbf{1}^T P_i \mathbf{1}\right| = \rho_i$$

Following the discussion about the norm of rank-one matrices, we see that for $b \in \{0,1\}$,

$$\|\Pi_b P_i G \Pi_0\| = \left\|\Pi_b P_i G \mathbf{1}\mathbf{1}^T\right\| = \|\Pi_b P_i \mathbf{1}\| \cdot \|\mathbf{1}\| = \|\Pi_b P_i \mathbf{1}\|.$$

Using this , we find that

$$\widetilde{P_iG}[0,0]^2 + \widetilde{P_iG}[1,0]^2 = \|\Pi_0 P_i \mathbf{1}\|^2 + \|\Pi_1 P_i \mathbf{1}\|^2 = \|P_i \mathbf{1}\|^2 = \rho_i$$

hence $\widetilde{P_iG}[1,0] = \sqrt{\rho_i(1-\rho_i)} = \sigma_i$.

Now, let us write $G = J + \lambda A$ where $\|A\| \leq 1$ and A is zero on the parallel space. Then

$$\begin{aligned} \widetilde{P_iG}[0,1] &= \|\Pi_0 P_i G \Pi_1\| = \|\Pi_0 P_i\, (J + \lambda A)\, \Pi_1\| = \lambda\, \|\Pi_0 P_i A \Pi_1\| = \lambda\, \|\Pi_0 P_i \Pi_1 A \Pi_1\| \\ &\leq \lambda\, \|\Pi_0 P_i \Pi_1\| = \lambda\sigma_i \end{aligned}$$

where we have used that $\Pi_1 A = A$ in the last equality. In the inequality we observe that $\widetilde{P_iG}[1,0] = \|\Pi_1 P_i G \Pi_0\| = \|\Pi_1 P_i \Pi_0\| = \|\Pi_0 P_i \Pi_1\|$. Hence we substitute $\|\Pi_0 P_i \Pi_1\| = \sigma_i$.

For the last entry we have $\widetilde{P_iG}[1,1] = \|\Pi_1 P_i G \Pi_1\| = \lambda\, \|\Pi_1 P_i A \Pi_1\| \leq \lambda$. ◀

By definition of flow sequence (Definition 19) and the previous lemma, we obtain:

▶ **Corollary 24** (Flow sequence lower-bound)**.** *Let $G = (V,E)$ be a λ-expander, and let $S_1,\dots,S_t \subseteq V$ be a sequence of subsets such that S_i is of density ρ_i and variance σ_i. Let $c_1,\dots,c_t$ be the flow sequence of the linear operators $P_1G,\dots,P_tG$. Then $c_1 = \frac{\rho_1}{\sigma_1}$ and: $c_{i+1} \geq \frac{c_i \cdot \rho_{i+1} - \lambda \cdot \sigma_{i+1}}{c_i \cdot \sigma_{i+1} + \lambda}$.*

▶ **Corollary 25.** *Let $G = (V,E)$ be a λ-expander, and let $S_1,\dots,S_t \subseteq V$ be a sequence of subsets such that S_i is of density ρ_i. Let $c_1,\dots,c_t$ be the flow sequence of the linear operators $P_1G,\dots,P_tG$. Suppose that λ is sufficiently small so that $c_i > 0$ for all i. Then, $\mathop{\mathbf{Pr}}_{(v_1,\dots,v_t)\sim\mathsf{RW}_G^t}[\forall i\ v_i \in S_i] \geq \rho_1 \cdot \prod_{i=2}^{t} \left(\rho_i - \frac{\sigma_i}{c_{i-1}}\lambda\right)$.*

Next, we demonstrate how distinct conditions imposed on λ lead to varying bounds on the flow sequence, consequently leading to corresponding confinement probability lower bounds.

▶ **Lemma 26.** *Let $G = (V, E)$ be a λ-expander, and let $S_1, \dots, S_t \subseteq V$ be a sequence of subsets each of density ρ_i. If for all i, $\lambda < \frac{1}{6} \cdot \sigma_i\sigma_{i+1} \cdot \frac{1+\rho_{i+1}}{1-\rho_{i+1}}$, then*

$$\Pr_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t}[\forall i\ v_i \in S_i] \geq \rho_1 \cdot \prod_{i=2}^{t}\left(\rho_i - 2\cdot\frac{\sigma_i}{\sigma_{i-1}}\lambda\right).$$

Proof. By Corollary 25, it suffices to prove that under our assumption on λ, we have $c_i \geq \sigma_i/2$ for all i.

For $i = 1$, we clearly have $c_1 = \frac{\rho_1}{\sigma_1} = \frac{\sigma_1}{1-\rho_1} > \sigma_1$. Now, assume that $c_i \geq \sigma_i/2$. Using Corollary 24, we find that

$$\begin{aligned} c_{i+1} - \frac{\sigma_{i+1}}{2} &\geq \frac{c_i\rho_{i+1} - \lambda\sigma_{i+1}}{c_i\sigma_{i+1} + \lambda} - \frac{\sigma_{i+1}}{2} = \frac{c_i\left(2\rho_{i+1} - \sigma_{i+1}^2\right) - 3\lambda\sigma_{i+1}}{2\left(c_i\sigma_{i+1} + \lambda\right)} \\ &= \frac{c_i\rho_{i+1}\left(1 + \rho_{i+1}\right) - 3\lambda\sigma_{i+1}}{2\left(c_i\sigma_{i+1} + \lambda\right)} \end{aligned}$$

Therefore it suffices to show $3\lambda\sigma_{i+1} \leq c_i\rho_{i+1}\left(1 + \rho_{i+1}\right)$. Indeed, using our assumption on λ and the induction hypothesis,

$$\lambda < \frac{1}{6}\cdot\sigma_i\sigma_{i+1}\cdot\frac{1+\rho_{i+1}}{1-\rho_{i+1}} \leq \frac{c_i}{3}\cdot\frac{\sigma_{i+1}}{1-\rho_{i+1}}(1+\rho_{i+1}) = \frac{c_i}{3}\cdot\frac{\rho_{i+1}}{\sigma_{i+1}}(1+\rho_{i+1}). \qquad ◀$$

The first part of Theorem 2 follows as a special case of the lemma above, in which all sets have the same density. Indeed, in this case, our assumption on λ becomes $\lambda < \frac{1}{6}\cdot\sigma^2\cdot\frac{1+\rho}{1-\rho} = \frac{1}{6}\rho(1+\rho)$.

▶ **Lemma 27.** *Let $G = (V, E)$ be a λ-expander, and let $S_1, \dots, S_t \subseteq V$ be a sequence of subsets each of density ρ_i. If for all i, $\lambda < \frac{1}{2}\cdot\frac{\sigma_i}{\sigma_{i+1}}\cdot\rho_{i+1}^2$, then*

$$\Pr_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t}[\forall i\ v_i \in S_i] \geq \rho_1 \cdot \prod_{i=2}^{t}\left(\rho_i - \frac{\sigma_i}{\sigma_{i-1}}\lambda\right).$$

Proof. By Corollary 25 it suffices to prove that under our assumption on λ, we have $c_i \geq \sigma_i$ for all i. The proof is by induction on i. For $i = 1$ we clearly have $c_1 = \sqrt{\frac{\rho_1}{1-\rho_1}} = \frac{\sigma_1}{1-\rho_1} > \sigma_1$. Assume that $c_i \geq \sigma_i$. By Corollary 24, we have

$$\begin{aligned} \frac{c_{i+1}}{\sigma_{i+1}} &\geq \frac{c_i\rho_{i+1} - \lambda\sigma_{i+1}}{c_i\sigma_{i+1}^2 + \lambda\sigma_{i+1}} = \frac{c_i\rho_{i+1} + c_i\sigma_{i+1}^2 + \lambda\sigma_{i+1} - c_i\sigma_{i+1}^2 - 2\lambda\sigma_{i+1}}{c_i\sigma_{i+1}^2 + \lambda\sigma_{i+1}} \\ &= 1 + \frac{c_i\left(\rho_{i+1} - \sigma_{i+1}^2\right) - 2\lambda\sigma_{i+1}}{c_i\sigma_{i+1}^2 + \lambda\sigma_{i+1}} = 1 + \frac{c_i\rho_{i+1}^2 - 2\lambda\sigma_{i+1}}{c_i\sigma_{i+1}^2 + \lambda\sigma_{i+1}} \end{aligned}$$

Therefore it suffices to show $2\lambda\sigma_{i+1} \leq c_i\rho_{i+1}^2$. Indeed, using our assumption on λ and the induction hypothesis, $\lambda < \frac{1}{2}\cdot\frac{\sigma_i}{\sigma_{i+1}}\cdot\rho_{i+1}^2 \leq \frac{c_i\rho_{i+1}^2}{2\sigma_{i+1}}$. ◀

The second part of Theorem 2 follows as a special case of the lemma above, in which all sets have the same density. In that case our assumption on λ becomes $\lambda < \frac{\rho^2}{2}$.

The following lemma refines the bound given in [2] and also allows for arbitrary densities with decreasing variances.

▶ **Lemma 28.** *Let $G = (V, E)$ be a λ-expander, and let $S_1, \dots, S_t \subseteq V$ be a sequence of subsets, each of density ρ_i and variance σ_i. Suppose that $\sigma_1 \geq \dots \geq \sigma_t$. If $\lambda \leq \frac{\sigma_i}{\sigma_{i-1}} \cdot \frac{\rho_{i-1}\rho_i}{4}$ for all i, then $\mathop{\mathbf{Pr}}_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t}[\forall i\ v_i \in S_i] \geq \rho_1 \cdot \prod_{i=2}^t \left(\rho_i - \lambda(1-\rho_{i-1}^2)\right)$.*

Proof. Using our assumption that $\sigma_i \geq \sigma_{i+1}$ for all $i = 1, \dots, t$, it suffices to prove that $\sigma_i \leq (1-\rho_i^2)c_i$ for all i. Indeed, in that case, by Corollary 25 we obtain

$$\mathop{\mathbf{Pr}}_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t}[\forall i\ v_i \in S_i] \geq \rho_1 \cdot \prod_{i=2}^t \left(\rho_i - \frac{\sigma_i}{c_{i-1}}\lambda\right) \geq \rho_1 \cdot \prod_{i=2}^t \left(\rho_i - \frac{\sigma_{i-1}}{c_{i-1}}\lambda\right)$$
$$\geq \rho_1 \cdot \prod_{i=2}^t \left(\rho_i - (1-\rho_{i-1}^2)\cdot\lambda\right)$$

Now, we prove by induction on i that $\sigma_i \leq (1-\rho_i^2)c_i$. For $i = 1$ we clearly have

$$(1-\rho_1^2)c_1 = (1-\rho_1^2)\sqrt{\frac{\rho_1}{1-\rho_1}} = (1+\rho_1)\sqrt{\rho_1(1-\rho_1)} > \sigma_1.$$

Assume that $\sigma_{i-1} \leq (1-\rho_{i-1}^2)c_{i-1}$. Then, by Corollary 24

$$\frac{1-\rho_i^2}{\sigma_i} \cdot c_i \geq \frac{1-\rho_i^2}{\sigma_i} \cdot \frac{c_{i-1}\rho_i - \lambda\sigma_i}{c_{i-1}\sigma_i + \lambda} = \frac{c_{i-1}\sigma_i(1+\rho_i) - \lambda(1-\rho_i^2)}{c_{i-1}\sigma_i + \lambda}$$
$$= \frac{c_{i-1}\sigma_i + \lambda + c_{i-1}\sigma_i\rho_i - \lambda(2-\rho_i^2)}{c_{i-1}\sigma + \lambda} = 1 + \frac{c_{i-1}\sigma_i\rho_i - \lambda(2-\rho_i^2)}{c_{i-1}\sigma_i + \lambda}$$

where the second equality uses the identity $\rho(1-\rho^2)/\sigma = \sigma(1+\rho)$. Thus, it remains to prove that $c_{i-1}\sigma_i\rho_i \geq \lambda(2-\rho_i^2)$. Indeed, on the one hand, by our induction hypothesis

$$c_{i-1}\sigma_i\rho_i \geq \frac{\sigma_{i-1}\sigma_i\rho_i}{(1-\rho_{i-1}^2)} = \frac{\sigma_i}{\sigma_{i-1}} \cdot \frac{\rho_{i-1}\rho_i}{1+\rho_{i-1}}.$$

using the identity $\rho/\sigma = \sigma/(1-\rho)$.

On the other hand, our assumption on λ implies that

$$\lambda \leq \frac{\sigma_i}{\sigma_{i-1}} \cdot \frac{\rho_{i-1}\rho_i}{4} \leq \frac{\sigma_i}{\sigma_{i-1}} \cdot \frac{\rho_{i-1}\rho_i}{(1+\rho_{i-1})(2-\rho_i^2)}$$

and the proof is complete. ◀

When all subsets have the same density ρ, we observe that in fact $(1+\rho)(2-\rho^2) \leq 3$. Therefore, Theorem 3 follows.

5 Fooling Non-Symmetric Confinement Functions

The class of t-wise confinement functions $\mathsf{Conf}_n^{\otimes t} \subseteq \{f : [n]^t \to \{0,1\}\}$ is defined as $\mathsf{Conf}_n^{\otimes t} = \{\mathbf{1}_{S_1} \otimes \dots \otimes \mathbf{1}_{S_t} \mid S_1, \dots, S_t \subseteq [n]\}$ where $\mathbf{1}_S(i)$ equals 1 if $i \in S$ and 0 otherwise. This class of functions is sometimes referred to as cut-tensors or cut-functions. Generally, confinement functions are not symmetric, hence a density-independent total variation bound for this class is not implied by the previous work of [8]. Nevertheless, we show that the class of confinement functions where the sets have equal variances, is $O(\lambda)$-fooled by expander random walks regardless of the densities.

We begin with a density-dependent bound which holds for all confinement functions.

▶ **Proposition 29.** *For $t \geq 1$, let $G = (V, E)$ be a λ-expander, and let $S_1, \ldots, S_t$ be a sequence of subsets such that S_i is of density ρ_i. Let $\rho_{\mathsf{max}} = \max_i \rho_i$. Then,*

$$d_{TV}\left(\mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}(\mathsf{RW}_G^t),\ \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t}(\mathsf{Ind}_V^t)\right) \leq \left(1 + \frac{1-\rho_{\mathsf{max}}^{t-1}}{1-\rho_{\mathsf{max}}}\right) \cdot \lambda.$$

Proof. First, observe that we may assume $t \geq 2$, as for $t = 1$ the distributions are identical and claim trivially holds. The decomposition $G = J + \lambda A$ with $\|A\| \leq 1$ implies that $\|G - J\| \leq \lambda$. Also, recall that $\|JP_iJ\| = \rho_i$. Let *LHS* be left-hand size of the inequality in the proposition. Expressing *LHS* linear-algebraically, we see that $LHS = \left|\mathbf{1}^T \prod_{i=1}^t (P_iG)\,\mathbf{1} - \mathbf{1}^T \prod_{i=1}^t (P_iJ)\,\mathbf{1}\right|$ and

$$\begin{aligned}
LHS &\leq \left\|\prod_{i=1}^t (P_iG) - \prod_{i=1}^t (P_iJ)\right\| \leq \sum_{k=1}^t \left\|\left(\prod_{j=k+1}^t (P_jG)\right) P_k(G-J)\left(\prod_{j=1}^{k-1}(P_jJ)\right)\right\| \\
&\leq \sum_{k=1}^t \|(G-J)\| \left\|\left(\prod_{j=1}^{k-1}(P_jJ)\right)\right\| \leq \sum_{k=1}^t \lambda \cdot \prod_{j=1}^{k-2} \rho_j \leq \lambda\left(1 + \sum_{\ell=0}^{t-2} \rho_{\mathsf{max}}^{\ell}\right) \\
&= \lambda \cdot \left(1 + \frac{1-\rho_{\mathsf{max}}^{t-1}}{1-\rho_{\mathsf{max}}}\right)
\end{aligned}$$

◀

Note that, in particular, the proof implies a $t\lambda$ bound for all confinement functions. A similar hybrid idea was used in [12] to derive a generalization of the expander mixing lemma for length-t random walks.[3] Proposition 29 shows that when the all the sets are small, say, of density which is bounded from above by some constant α, the corresponding confinement function is $O_\alpha(\lambda)$-fooled.

We proceed with the main result for this section.

Proof of Theorem 4. First, observe that we may assume $t \geq 2$, as for $t = 1$ the distributions are identical and claim trivially holds. Let *LHS* be left-hand side of the inequality in the proposition. Let $n = |V|$ and identify V with $[n]$ arbitrarily. Let us denote by ρ_i the density of S_i, and by σ_i its variance, so that $\rho_{\mathsf{max}} = \max_i \rho_i$. Further denote $\sigma_{\mathsf{max}} = \max_i \sigma_i$.

We consider two cases according to the relationship between ρ_{max} and λ. Assume first that $\rho_{\mathsf{max}} \leq 2\lambda$. Applying the upper-bound Proposition 22 we find that

$$\begin{aligned}
LHS &= \left|\Pr_{(v_1,\ldots,v_t)\sim\mathsf{RW}_G^t}[\forall i\ v_i \in S_i] - \prod_{i=1}^t \rho_i\right| \leq \sqrt{\rho_1\rho_t} \cdot \prod_{i=1}^{t-1}\left(\lambda + (1-\lambda)\sqrt{\rho_i\rho_{i+1}}\right) \\
&\leq \rho_{\mathsf{max}}\left(\lambda + \rho_{\mathsf{max}}(1-\lambda)\right)^{t-1} \leq \rho_{\mathsf{max}}\left(3\lambda\right)^{t-1} \leq 3 \cdot \rho_{\mathsf{max}} \cdot \lambda
\end{aligned}$$

For the first inequality we observe that both terms inside the absolute value are non-negative, hence the magnitude of their difference is bounded by the maximal one. Additionally, the upper-bound provided by the hitting property as presented in Proposition 22 applies to both terms. Then, we bound $\lambda + 2\lambda(1-\lambda) \leq 3\lambda$. The last inequality holds under our assumption that $\lambda \leq 1/3$ and $t \geq 2$.

[3] In fact their result is more general, as it goes beyond random walk on expander graphs. Their "splittable-mixing lemma" holds for what they call "λ-splittable structures", which are subsets of $[n]^t$ that admit certain high-dimensional expansion properties.

Now, Assume that $\rho_{\mathsf{max}} \geq 2\lambda$. For $i = 1, \ldots, t$, let P_i be the $n \times n$ projection matrix on the set S_i. That is, P_i is the diagonal matrix satisfying $P_i[v, v] = 1$ if $v \in S_i$ and 0 otherwise. We have the following entry-wise bounds on the flow matrices: $\widetilde{P_i} \leq_{e.w} \begin{pmatrix} \rho_i & \sigma_i \\ \sigma_i & 1 \end{pmatrix}$ where all entries except for the right bottom are equality. To see this, consider that

$$\|\Pi_0 P_i \Pi_0\| = \|J P_i J\| = \left\|\mathbf{1}\mathbf{1}^T P_i \mathbf{1}\mathbf{1}^T\right\| = \left|\mathbf{1}^T P_i \mathbf{1}\right| = \rho_i.$$

Moreover, for $b \in \{0, 1\}$ we have

$$\|\Pi_b P_i \Pi_0\| = \left\|\Pi_b P_i \mathbf{1}\mathbf{1}^T\right\| = \|\Pi_b P_i \mathbf{1}\| \cdot \|\mathbf{1}\| = \|\Pi_b P_i \mathbf{1}\|.$$

Since $\|P_i \mathbf{1}\| = \sqrt{\sum_{i \in S_i} \frac{1}{n}} = \sqrt{\rho_i}$, we have

$$\|\Pi_1 P_i \Pi_0\| = \|\Pi_1 P_i \mathbf{1}\| = \sqrt{\|P_i \mathbf{1}\|^2 - \|\Pi_0 P_i \mathbf{1}\|^2} = \sqrt{\rho_i(1 - \rho_i)} = \sigma_i.$$

By symmetry we also have $\|\Pi_1 P \Pi_0\| = \sigma_i$. Finally, we bound $\|\Pi_1 P_i \Pi_1\| \leq \|\Pi_1\|^2 \|P_i\| \leq 1$.

Let $\sigma_{\mathsf{max}} = \max_i \sigma_i$, and recall that by assumption $\sigma_{\mathsf{max}} = \sqrt{\rho_{\mathsf{max}}(1 - \rho_{\mathsf{max}})}$. Through utilizing the submultiplicativity of the flow operator (Claim 11) and Example 10, we find that $\widetilde{GP_i} \leq_{e.w} \begin{pmatrix} \rho_i & \sigma_i \\ \lambda\sigma_i & \lambda \end{pmatrix} \leq_{e.w} A$ for $A \overset{\mathrm{def}}{=} \begin{pmatrix} \rho_{\mathsf{max}} & \sigma_{\mathsf{max}} \\ \lambda\sigma_{\mathsf{max}} & \lambda \end{pmatrix}$. Now, expressing the total variation distance linear algebraically, we have

$$\begin{aligned}
LHS &= \left|\Pr_{(v_1,\ldots,v_t)\sim \mathsf{RW}_G^t}[\forall i \; v_i \in S_i] - \prod_{i=1}^{t} \rho_i\right| = \left|\mathbf{1}^T \prod_{i=2}^{t} (P_i G)\, P_1 \mathbf{1} - \mathbf{1}^T \prod_{i=2}^{t} (P_i J)\, P_1 \mathbf{1}\right| \\
&= \left|\mathbf{1}^T \prod_{i=1}^{t} (G P_i)\, \mathbf{1} - \mathbf{1}^T \prod_{i=1}^{t} (J P_i)\, \mathbf{1}\right| = \left|\mathbf{1}^T \prod_{i=1}^{t} (G P_i)\, \mathbf{1} - \mathbf{1}^T \prod_{i=1}^{t} (\Pi_0 G P_i)\, \mathbf{1}\right| \\
&\leq \left(\prod_{i=1}^{t} \widetilde{GP_i}\right)[0,0] - \left(\prod_{i=1}^{t} \widetilde{GP_i}[0,0]\right) \\
&\leq A^t[0,0] - (A[0,0])^t
\end{aligned}$$

Where the first inequality is by Lemma 16, and the second is by Lemma 17 Using Lemma 15 we obtain the bound

$$\begin{aligned}
A^t[0,0] - (A[0,0])^t &\leq \frac{\lambda\rho_{\mathsf{max}}\sigma_{\mathsf{max}}^2}{\rho_{\mathsf{max}} - \lambda} \cdot \sum_{k=0}^{t-2} \rho_{\mathsf{max}}^k \left(\rho_{\mathsf{max}} + \frac{\lambda\sigma_{\mathsf{max}}^2}{\rho_{\mathsf{max}} - \lambda}\right)^{t-k-2} \\
&= \frac{\lambda\rho_{\mathsf{max}}^2(1 - \rho_{\mathsf{max}})}{\rho_{\mathsf{max}} - \lambda} \cdot \sum_{k=0}^{t-2} \rho_{\mathsf{max}}^k \left(\rho_{\mathsf{max}} + \frac{\lambda\rho_{\mathsf{max}}(1 - \rho_{\mathsf{max}})}{\rho_{\mathsf{max}} - \lambda}\right)^{t-k-2} \\
&\leq 2\rho_{\mathsf{max}} \cdot \lambda \cdot (1 - \rho_{\mathsf{max}}) \sum_{k=0}^{t-2} \rho_{\mathsf{max}}^k \left(\rho_{\mathsf{max}} + \rho_{\mathsf{max}}(1 - \rho_{\mathsf{max}})\right)^{t-k-2} \\
&\leq 2\rho_{\mathsf{max}} \cdot \lambda \cdot (1 - \rho_{\mathsf{max}}) \sum_{k=0}^{\infty} \rho_{\mathsf{max}}^k = 2\rho_{\mathsf{max}} \cdot \lambda,
\end{aligned}$$

where in the second inequality we have used that $\rho_{\mathsf{max}}/(\rho_{\mathsf{max}} - \lambda) \leq 2$ and $\lambda/(\rho_{\mathsf{max}} - \lambda) \leq 1$ under our assumption that $\rho_{\mathsf{max}} \geq 2\lambda$. ◀

6 Fooling The Sum Function modulo d

For integers $d \geq 2$, $d' \geq 2$ and $t \geq 1$, define the function $\mathsf{Sum}_{d',d} : \mathbb{Z}_{d'}^t \to \mathbb{Z}_d$ as $\mathsf{Sum}_{d',d}(a_1, \ldots, a_t) = \sum_{i=1}^t a_i \bmod d$. Given the insignificance of d' within this context, we will simplify our notation by omitting it, using only Sum_d.

In this section we use the flow framework to prove a bias-independent $O(\sqrt{d} \cdot \lambda)$ total variation bound for Sum_d. We also prove a bias amplification result (Lemma 33) from which an $O(\sqrt{d} \cdot c^t)$ total variation bound can be derived using Lemma 8, where $c < 1$ is a parameter that depends on the bias of the labeling and λ. [8] obtains a similar bound, which they derive from their results on permutation branching programs.

▶ Remark 30. While characters of $\mathbb{Z}_d$ are formally defined on values in $\mathbb{Z}_d$, throughout this section, we simplify notation by using $\chi(a)$ for an arbitrary integer a and character χ of $\mathbb{Z}_d$, to mean $\chi(a \bmod d)$.

6.1 Bias Amplification

We begin our discussion with the Boolean case.

▶ **Definition 31** (bias over $\mathbb{Z}_2$). *The bias of a labeling* $\mathsf{val} : [n] \to \{0, 1\}$ *is defined as* $\mathsf{bias}(\mathsf{val}) \overset{\text{def}}{=} \left|\mathbb{E}_{i \in [n]}(-1)^{\mathsf{val}(i)}\right|$.

The distribution over $\{0, 1\}$ obtained by sampling a uniformly random element of $[n]$ and outputting its label is $\mathsf{bias}(\mathsf{val})$-biased. However, the distribution obtained by taking t uniformly random samples from $[n]$ and computing the parity of the corresponding labels is only $\mathsf{bias}(\mathsf{val})^t$-biased. That is, the bias decreases exponentially with t. To see this, note that

$$\left|\mathop{\mathbb{E}}_{(i_1,\ldots,i_t)\in[n]^t}(-1)^{\sum_{j=1}^t \mathsf{val}(i_j)}\right| = \left|\prod_{j=1}^t \mathop{\mathbb{E}}_{i_j\in[n]}(-1)^{\mathsf{val}(i_j)}\right| = \mathsf{bias}(\mathsf{val})^t.$$

It has been observed in [13] that this bias reducing construction can be derandomized by taking length-$(t-1)$ expander random walks on $[n]$ rather than independent samples. In this case, it is shown that the bias of the resulting distribution is at most $(\mathsf{bias}(\mathsf{val}) + \lambda)^{\lfloor t/2 \rfloor}$, where λ is the expansion parameter of the graph. In [13], this property is called *parity sampling*, and it follows that expander random walks are *good parity samplers*. This observation is a key part of the breakthrough construction of almost optimal ε-balanced codes [13].

In the context of the sum function modulo d, we allow labelings with a larger alphabet size. It is therefore natural to ask whether the bias amplification phenomenon extends to $\mathbb{Z}_d$ where $d > 2$. Observe that the bias of a labeling $\mathsf{val} : [n] \to \{0, 1\}$ is simply the inner product of the distribution induced by the labeling with the non-trivial character of $\mathbb{Z}_2$. This notion extends naturally to characters of $\mathbb{Z}_d$ as follows.

▶ **Definition 32** (bias over $\mathbb{Z}_d$). *For integers $d \geq 2$ and $d' \geq 2$, the bias of a labeling* $\mathsf{val} : [n] \to \mathbb{Z}_{d'}$ *with respect to a character χ of $\mathbb{Z}_d$ is defined as* $\mathsf{bias}_\chi(\mathsf{val}) \overset{\text{def}}{=} \left|\mathbb{E}_{i\in[n]}\chi(\mathsf{val}(i))\right|$.

The same argument as before shows that for any character χ of $\mathbb{Z}_d$, taking t independent samples from $[n]$ and outputting the sum of their labels modulo d yields a distribution on $\mathbb{Z}_d$ with bias $\mathsf{bias}_\chi(\mathsf{val})^t$ with respect to the character χ. Moreover, we prove that replacing the independent samples by length $t-1$ random walk on a λ-expander graph obtains a distribution on $\mathbb{Z}_d$ with bias at most $(\mathsf{bias}_\chi(\mathsf{val}) + \lambda)^{\lfloor t/2 \rfloor}$ with respect to the character χ. (In fact, the bound is slightly better, as here it may be larger than 1). In other words, expander random walks are good *character samplers*. This fact has also been independently observed in [11].

▶ **Lemma 33** (Bias Amplification). *For integers $t \geq 2$, $d \geq 2$, and $d' \geq 2$, let $G = (V, E)$ be a λ-expander, and let $\mathsf{val} : V \to \mathbb{Z}_{d'}$ be any labeling. Let χ be a character of $\mathbb{Z}_d$. Then*

$$\left| \mathop{\mathbb{E}}_{(v_1,\dots,v_t)\sim \mathsf{RW}_G^t} \chi\left(\sum_{i=1}^{t} \mathsf{val}(v_i)\right)\right| \leq \left((1-\lambda)^2 \cdot \mathsf{bias}_\chi(\mathsf{val}) + 2\lambda(1-\lambda) + \lambda^2\right)^{\lfloor \frac{t}{2} \rfloor}, \tag{4}$$

where $\mathsf{bias}_\chi(\mathsf{val}) = \left|\mathop{\mathbb{E}}_{i\in[n]} \chi(\mathsf{val}(i))\right|$ *is the bias of* val *with respect to* χ.

The proof for Lemma 33 appears in the full version of this paper [14]

6.2 Bias-independent bound using the flow framework

When the bias is bounded by a constant, the bias amplification property implies that the distributions $\mathsf{Sum}_d(\mathsf{val}(\mathsf{Ind}_V^t))$ and $\mathsf{Sum}_d(\mathsf{val}(\mathsf{RW}_G^t))$ are highly unbiased, with bias which is exponentially small in t. Applying the triangle inequality to the XOR-Lemma (Lemma 8), we see that the total variation distance between these distributions is bounded by the sum of the biases of each distribution with respect to a worst-case non-trivial character. As such, it is decreasing exponentially fast with t as well. However, this argument hinges on the assumption that the given labeling is balanced. If we have, say, $\mathsf{bias}_\chi(\mathsf{val}) = 1 - 1/t$ for some non-trivial character χ of $\mathbb{Z}_d$, the bias amplification argument is insufficient for an effective total variation bound. This constitutes the primary reason why earlier works such as [5] and [4], which rely heavily on the bias-amplification property, result in total variation bounds that are bias-dependent.

Next, we use the flow framework to obtain an $O(\sqrt{d} \cdot \lambda)$ bias independent bound, similar to that of [8]. The proof in this case is arguably simpler than the more general case in [8], which applies for all small-width permutation branching programs.

Proof of Theorem 6. First, observe that we may assume $t \geq 2$, as for $t = 1$ the distributions are identical, and the claim trivially holds. Let *LHS* be left-hand side of the inequality in the theorem. Let $n = |V|$ and identify V with $[n]$ arbitrarily. Observe that in order to obtain a total variation bound, it suffices to bound the maximum bias of the difference between the two distributions. Indeed, by the XOR-Lemma (Lemma 8)

$$LHS \leq \frac{\sqrt{d}}{2} \cdot \max_{\chi \in \widehat{\mathbb{Z}_d}} \left|\left\langle \chi, \mathsf{Sum}_d\left(\mathsf{val}(\mathsf{RW}_G^t)\right) - \mathsf{Sum}_d\left(\mathsf{val}(\mathsf{Ind}_V^t)\right)\right\rangle\right|$$

We fix a character χ of $\mathbb{Z}_d$ that attains the maximum. Let $\mu = \mathsf{bias}_\chi(\mathsf{val})$ be the bias of the labeling val with respect to χ. We consider two cases according to the relation between μ and λ. To begin, let us assume that $\mu \leq 3\lambda$. In that case,

$$\begin{aligned} LHS &\leq \frac{\sqrt{d}}{2}\left| \mathop{\mathbb{E}}_{v=(v_1,\dots,v_t)\sim \mathsf{RW}_G^t} \chi\left(\sum_{i=1}^{t} \mathsf{val}(v_i)\right) - \mathop{\mathbb{E}}_{v\sim \mathsf{Ind}_V^t} \chi\left(\sum_{i=1}^{t} \mathsf{val}(v_i)\right)\right| \\ &\leq \sqrt{d} \cdot \left((1-\lambda)^2 \cdot \mu + 2\lambda(1-\lambda) + \lambda^2\right)^{\lfloor \frac{t}{2} \rfloor} \\ &\leq \sqrt{d} \cdot \left(3(1-\lambda)^2 \cdot \lambda + 2\lambda(1-\lambda) + \lambda^2\right)^{\lfloor \frac{t}{2} \rfloor} \leq \sqrt{d} \cdot (5\lambda)^{\lfloor \frac{t}{2} \rfloor} \leq 5\sqrt{d} \cdot \lambda, \end{aligned} \tag{5}$$

where the second inequality is implied by the triangle inequality and the bias amplification property established in the previous subsection. The last inequality holds under our assumption that $t \geq 2$ and $\lambda \leq 1/6$.

Now, let us assume that $\mu \geq 3\lambda$. Instead of applying the triangle inequality on the second line of Equation 5, we express it linear algebraically. We then give entry-wise bounds for the flow matrices of the involved linear operators. Let P be the $n \times n$ diagonal matrix $P = \operatorname{diag}(\chi(\mathsf{val}(1)), \ldots, \chi(\mathsf{val}(n)))$. We have the following entry-wise bounds on the flow matrix: $\widetilde{P} \leq_{e.w} \begin{pmatrix} \mu & \sqrt{1-\mu^2} \\ \sqrt{1-\mu^2} & 1 \end{pmatrix}$ where all entries except for the right bottom are equality. To see this, note that

$$\widetilde{P}[0,0] = \|\Pi_0 P \Pi_0\| = \|JPJ\| = \left\|\mathbf{1}\mathbf{1}^T P \mathbf{1}\mathbf{1}^T\right\| = \left|\mathbf{1}^T P \mathbf{1}\right| = \mu.$$

Moreover, we see that for $b \in \{0,1\}$,

$$\|\Pi_b P \Pi_0\| = \left\|\Pi_b P \mathbf{1}\mathbf{1}^T\right\| = \|\Pi_b P \mathbf{1}\| \cdot \left\|\mathbf{1}^T\right\| = \|\Pi_b P \mathbf{1}\|.$$

Now, since P is unitary,

$$\widetilde{P}[0,0]^2 + \widetilde{P}[1,0]^2 = \|\Pi_0 P \mathbf{1}\|^2 + \|\Pi_1 P \mathbf{1}\|^2 = \|P\mathbf{1}\|^2 = \|\mathbf{1}\| = 1$$

By symmetry we conclude that $\widetilde{P}[1,0] = \widetilde{P}[0,1] = \sqrt{1-\mu^2}$. Finally, we bound $\widetilde{P}[1,1] = \|\Pi_1 P \Pi_1\| \leq \|\Pi_1\|^2 \|P\| \leq 1$. By submultiplicativity of the flow operator (Claim 11) and Example 10, We see that $\widetilde{GP} \leq_{e.w} A$ for $A \overset{\text{def}}{=} \begin{pmatrix} \mu & \sqrt{1-\mu^2} \\ \lambda\sqrt{1-\mu^2} & \lambda \end{pmatrix}$. Now, Let us pick up Equation 5 after the first inequality. Expressing the bias linear-algebraically,

$$\left| \mathop{\mathbb{E}}_{v \sim \mathsf{RW}_G^t} \chi\left(\sum_{i=1}^t \mathsf{val}(v_i)\right) - \mathop{\mathbb{E}}_{v \sim \mathsf{Ind}_V^t} \chi\left(\sum_{i=1}^t \mathsf{val}(v_i)\right) \right| = \left|\mathbf{1}^T (PG)^{t-1} P \mathbf{1} - \mathbf{1}^T (PJ)^{t-1} P \mathbf{1}\right|$$

and,

$$\begin{aligned} \left|\mathbf{1}^T (PG)^{t-1} P \mathbf{1} - \mathbf{1}^T (PJ)^{t-1} P \mathbf{1}\right| &= \left|\mathbf{1}^T (GP)^t \mathbf{1} - \mathbf{1}^T (JP)^t \mathbf{1}\right| \\ &= \left|\mathbf{1}^T (GP)^t \mathbf{1} - \mathbf{1}^T (\Pi_0 GP)^t \mathbf{1}\right| && (\Pi_0 G = J) \\ &\leq (\widetilde{GP})^t[0,0] - (\widetilde{GP}[0,0])^t && (\textit{Lemma } 16) \\ &\leq A^t[0,0] - (A[0,0])^t. && (\textit{Lemma } 17) \end{aligned}$$

Applying Lemma 13 we obtain the bound

$$\begin{aligned} A^t[0,0] - (A[0,0])^t &\leq \frac{\lambda\mu(1-\mu^2)}{\mu-\lambda} \sum_{k=0}^{t-2} \mu^k \left(\mu + \frac{\lambda(1-\mu^2)}{\mu-\lambda}\right)^{t-k-2} \\ &\leq \frac{3}{2} \cdot \lambda(1-\mu^2) \sum_{k=0}^{t-1} \mu^k \left(\mu + \frac{1}{2}\cdot(1-\mu^2)\right)^{t-k-1} \leq \frac{3}{2}\cdot\lambda\cdot(1+\mu)(1-\mu)\sum_{k=0}^{\infty}\mu^k \\ &\leq 3\cdot\lambda. \end{aligned}$$

where the second inequality holds as our assumption $\mu \geq 3\lambda$ implies that $\mu/(\mu-\lambda) \leq 3/2$ and $\lambda/(\mu-\lambda) \leq 1/2$. In the third inequality we have used that $\mu + \frac{1}{2}\cdot(1-\mu^2) \leq 1$. Overall, we have $d_{TV}\left(\mathsf{Sum}_d\left(\mathsf{val}(\mathsf{RW}_G^t)\right), \mathsf{Sum}_d\left(\mathsf{val}(\mathsf{Ind}_V^t)\right)\right) < 5\sqrt{d}\cdot\lambda$ in all cases, and the proof is complete. ◀

References

1 Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 132–140, 1987.

2 Noga Alon, Uriel Feige, Avi Wigderson, and David Zuckerman. Derandomized graph products. *Computational Complexity*, 5:60–75, 1995.

3 C. Beck. Chernoff bounds for expander walks. `https://www.ias.edu/video/csdm/2015/0310-ChristopherBeck`, 2015. A recording of a lecture, given at the Institute for Advanced Study, Princeton, USA.

4 Gil Cohen, Dor Minzer, Shir Peleg, Aaron Potechin, and Amnon Ta-Shma. Expander random walks: The general case and limitations. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

5 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*, pages 1643–1655, 2021.

6 David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

7 Oded Goldreich. Three XOR-lemmas – An exposition. *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation: In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, pages 248–272, 2011.

8 Louis Golowich and Salil Vadhan. Pseudorandomness of expander random walks for symmetric functions and permutation branching programs. In *37th Computational Complexity Conference (CCC 2022)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

9 Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17:3–37, 2008.

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

11 Akhil Jalan and Dana Moshkovitz. Near-optimal cayley expanders for abelian groups. *arXiv preprint*, 2021. `arXiv:2105.01149`.

12 Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. Near-linear time decoding of ta-shma's codes via splittable regularity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1527–1536, 2021.

13 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, 2017.

14 Amnon Ta-Shma and Ron Zadiario. The expander hitting property when the sets are arbitrarily unbalanced. *Electronic Colloquium on Computational Complexity (ECCC)*, TR24-118, 2024. URL: `https://eccc.weizmann.ac.il/report/2024/118`.

15 Salil P Vadhan et al. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.