On the Spectral Expansion of Monotone Subsets of the Hypercube

Yumou Fei
Department of EECS, Massachusetts Institute of Technology, Cambridge, MA, USA

Renato Ferreira Pinto Jr.
Cheriton School of Computer Science, University of Waterloo, Canada

Abstract

We study the spectral gap of subgraphs of the hypercube induced by monotone subsets of vertices. For a monotone subset $A \subseteq \{0,1\}^n$ of density $\mu(A)$, the previous best lower bound on the spectral gap, due to Cohen [14], was $\gamma \gtrsim \mu(A)/n^2$, improving upon the earlier bound $\gamma \gtrsim \mu(A)^2/n^2$ established by Ding and Mossel [16]. In this paper, we prove the optimal lower bound $\gamma \gtrsim \mu(A)/n$. As a corollary, we improve the mixing time upper bound of the random walk on constant-density monotone sets from $O(n^3)$, as shown by Ding and Mossel, to $O(n^2)$. Along the way, we develop two new inequalities that may be of independent interest: (1) a directed L^2-Poincaré inequality on the hypercube, and (2) an "approximate" FKG inequality for monotone sets.

2012 ACM Subject Classification Mathematics of computing → Markov processes; Theory of computation → Random walks and Markov chains

Keywords and phrases Random walks, mixing time, FKG inequality, Poincaré inequality, directed isoperimetry

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2025.42

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/pdf/2505.02685`

Funding *Renato Ferreira Pinto Jr.*: Supported by an NSERC Canada Graduate Scholarship.

Acknowledgements We thank Lap Chi Lau for many helpful insights and discussions on spectral theory for directed graphs, and Esty Kelman for early discussions in the development of this project.

1 Introduction

Suppose $G = (V, E)$ is a good expander graph, so that a random walk on the vertices of G is *fast mixing*, i.e. converges quickly to its stationary distribution. When is a random walk on a subgraph of G also fast mixing? More precisely, what kinds of subgraph restrictions preserve good expansion?

This paper studies the case of the hypercube graph H_n, where vertices $x, y \in \{0,1\}^n$ are connected by an edge if and only if they differ in exactly one coordinate. Recall that the *lazy random walk* on H_n has mixing time $\Theta(n \log n)$. Given a subset $A \subseteq \{0,1\}^n$ of vertices, we consider the random walk on $\{0,1\}^n$ *censored* to A.

▶ **Definition 1** (Censored random walk, [16]). *Given $A \subseteq \{0,1\}^n$, the* random walk on $\{0,1\}^n$ censored to A *is defined as follows. On state $x \in A$, sample $i \in [n]$ uniformly at random and let y be the vertex obtained by flipping the i-th bit of x. Then*

1. *If $y \in A$, flip a coin and either stay at x or move to y (each with probability $1/2$).*
2. *If $y \notin A$, stay at x (in which case we call this a* censored *step).*

Without further guarantees on A, the censored random walk may mix well or extremely poorly even when A is large and connected, as the following two examples illustrate:

▶ **Example 2** (Subcube). Let $S \subset [n]$ be a set of indices, and let A be the subcube given by the vertices $x \in \{0,1\}^n$ satisfying $x_i = 0$ for all $i \in S$. Then the censored random walk is essentially a random walk on the smaller cube $\{0,1\}^{n'}$, where $n' := n - |S|$, except that only an $O(n'/n)$-fraction of the transitions are not censored. Thus the censored random walk has mixing time $O(\frac{n}{n'} \cdot n' \log n') = O(n \log n)$.

▶ **Example 3** (Middle slice bridge). Let $x^* \in \{0,1\}^n$ be an arbitrary vertex with Hamming weight $|x^*| = \lfloor n/2 \rfloor$, and consider the set $A := \{x \in \{0,1\}^n : |x| \neq \lfloor n/2 \rfloor\} \cup \{x^*\}$. A spectral argument shows that the mixing time of the censored random walk is exponential in n.

Thus, it is natural to ask: what properties of A ensure fast mixing? In [16], Ding & Mossel initiated the study of random walks censored to *monotone* sets[1] A and showed that, when A is not too small, monotonicity implies fast mixing. Concretely, letting μ denote the uniform distribution on $\{0,1\}^n$, they proved

▶ **Theorem 4** ([16, Corollary 1.2]). *Let $A \subseteq \{0,1\}^n$ be a non-empty monotone set. Then the random walk on $\{0,1\}^n$ censored to A has mixing time*

$$t_{\mathsf{mix}} \le 512 \cdot \left(\frac{n}{\mu(A)}\right)^2 \log(4 \cdot 2^n \mu(A)).$$

When the density $\mu(A)$ is a constant, the above implies a mixing time bound of $O(n^3)$. In particular, for the uncensored special case $A = H_n$, this result only yields an upper bound of $O(n^3)$ on the mixing time, versus the optimal $\Theta(n \log n)$; this suggests the potential for improving upon Theorem 4, and indeed [16] asked the following question.

▶ **Question 5** ([16, Question 1.1]). *Suppose $\mu(A) \ge \varepsilon$ for some constant $\varepsilon > 0$. Is it true that $t_{\mathsf{mix}} \le O_\varepsilon(n \log n)$?*

Our main result makes progress on this question by showing an $O(n^2)$ mixing time bound for monotone sets A of constant density.

▶ **Theorem 6.** *Let $A \subseteq \{0,1\}^n$ be a non-empty monotone set. Then the random walk on $\{0,1\}^n$ censored to A has mixing time*

$$t_{\mathsf{mix}} \le \frac{2n}{\mu(A)} \cdot \log(4 \cdot 2^n \mu(A)) \,.$$

1.1 Spectral gap

It is well-known that the mixing time of a Markov chain is related to the *spectral gap* of its generator (see e.g. [35]), or equivalently the *spectral expansion* of the underlying graph. The Poincaré inequality (see e.g. [39]) for the hypercube states that the spectral expansion of the (lazy) hypercube H_n is exactly $1/n$, which implies the (non-tight) mixing time bound $O(n^2)$ for the lazy random walk on H_n.

Thus, a natural question related to mixing under censoring is the *robustness* of the classical Poincaré inequality under vertex removal from H_n. Specifically, does the spectral expansion of H_n remain on the order of $1/n$ if only a small fraction of the vertices are removed, or does it exhibit a significant deviation? Example 3 demonstrates that the spectral

[1] A set A is called monotone if $x \in A$ implies $y \in A$ whenever $x \preceq y$, where the latter denotes the natural partial order on the hypercube: $x \preceq y$ if $x_i \le y_i$ for every $i \in [n]$.

expansion can shrink to exponentially small values if the removed set is arbitrary. Our goal is to show that when the removed set of vertices is monotone (and not too large), the spectral expansion remains at least on the order of $1/n$.

For the purpose of clearer comparison with the classical Poincaré inequality, we introduce the Dirichlet form of a function on A. Note that in the case $A = H_n$, the following definition is exactly the "influence" [39, Definition 2.27] of the function $f : \{0,1\}^n \to \mathbb{R}$.

▶ **Definition 7.** *Fix a monotone set $A \subseteq \{0,1\}^n$. For all $f : A \to \mathbb{R}$, we define*[2]

$$\mathcal{E}_A(f) := \frac{1}{4} \cdot \mathop{\mathbb{E}}_{x \in A}\left[\sum_{i=1}^{n} \left(f(x) - f(x^{\oplus i})\right)^2 \cdot \mathbb{1}\left[x^{\oplus i} \in A\right]\right].$$

Here $x^{\oplus i}$ denotes the binary string obtained by flipping the i-th bit of x.

We can now state our "robust" version of the Poincaré inequality.

▶ **Theorem 8.** *Let $A \subseteq \{0,1\}^n$ be a non-empty monotone set. We have for all $f : A \to \mathbb{R}$*

$$\mathrm{Var}_A[f] \le \frac{1}{1 - \sqrt{1 - \mu(A)}} \cdot \mathcal{E}_A(f).$$

Here $\mathrm{Var}_A[f]$ stands for the variance of $f(x)$ where x is a uniformly random element of A.

Note that in the case $A = H_n$, the above theorem recovers the Poincaré inequality on the hypercube. We remark that Theorem 6 follows directly from Theorem 8 due to standard Markov chain theory (e.g. [35, Theorem 12.4]), so the rest of the paper focuses mainly on proving Theorem 8.

Theorem 8 can also be stated as a lower bound on the spectral gap of H_A – the subgraph of H_n induced by A, with a self-loop added to vertex x for each edge $\{x, y\}$ of the hypercube with $x \in A$ and $y \notin A$ (which counts as 1 toward the degree of x). For convenience of reference, in this paper we *define* the spectral gap using the language of Theorem 8.

▶ **Definition 9.** *Let $A \subseteq \{0,1\}^n$ be a monotone set with at least 2 elements. We define*

$$\gamma(H_A) := \frac{1}{n} \cdot \inf_{f \notin \mathsf{const}_A} \frac{\mathcal{E}_A(f)}{\mathrm{Var}_A[f]},$$

where f ranges over all non-constant functions from A to $\mathbb{R}$.

Now Theorem 8 can be stated as $\gamma(H_A) \ge \frac{1}{n}\left(1 - \sqrt{1 - \mu(A)}\right) \gtrsim \mu(A)/n$, for monotone sets A with $|A| \ge 2$.

1.2 Proof overview: previous work

We begin by briefly describing the proof of Theorem 4 in [16]. The proof in [16] also analyzes the spectral expansion of H_A, achieving the lower bound $\gamma(H_A) \gtrsim \mu(A)^2/n^2$.

By Cheeger's inequality, to obtain a lower bound on $\gamma(H_A)$, it suffices to lower bound the *bottleneck ratio*

$$\phi(H_A) := \min_{S \subseteq A} \frac{|E(S, A \setminus S)|}{\min\{|S|, |A \setminus S|\}},$$

[2] The energy functional $\mathcal{E}_A(\cdot)$ defined here differs from the standard Dirichlet form of the censored random walk on A by a factor of n. This normalization is chosen to better align with the notion of total influence in the analysis of Boolean functions.

where $E(S, A \setminus S)$ denotes the set of edges $\{x, y\}$ of H_n with $x \in S$ and $y \in A \setminus S$. By an isoperimetric inequality of the hypercube, there is good lower bound on the number of boundary edges connecting a vertex in S to a vertex in $\{0,1\}^n \setminus S$. However, it is not immediately clear how many of these edges actually lead to $A \setminus S$. The crucial observation of [16] is that if an edge from a vertex $x \in S$ goes "upward" – that is, if its other endpoint $y \in \{0,1\}^n \setminus S$ satisfies $y \succeq x$ – then by the monotonicity of A, we must have $y \in A \setminus S$.

Coincidentally, there is a "directed isopetrimetric inequality" [25, Theorem 2], developed by the property testing community, which provides a lower bound on exactly the number of such *upward* boundary edges. Specifically, it gives a lower bound on the number of edges connecting a vertex $x \in S$ to a vertex $y \in \{0,1\}^n \setminus S$ with $y \succeq x$ in terms of the *distance* of S to monotonicity. The precise notion of distance is less important than the key fact that, when $|A|$ is not too small, at least one of S and $A \setminus S$ must be far from any monotone set – i.e., it must have a large distance to monotonicity – due to the FKG inequality [22]. As a result, we can lower bound the number of upward boundary edges from either S or $A \setminus S$. Since both sets of upward edges are subsets $E(S, A \setminus S)$, we can thus arrive at a lower bound on $|E(S, A \setminus S)|$ and hence $\phi(H_A)$.

The work [16] actually achieves the optimal bound $\phi(H_A) \gtrsim \mu(A)/n$ on the bottleneck ratio. However, this only translates to a quadratically worse bound $\gamma(H_A) \gtrsim \mu(A)^2/n^2$ for the spectral expansion, due to the loss incurred by applying Cheeger's inequality. One natural idea is to avoid using Cheeger's inequality by directly bounding the spectral gap $\gamma(H_A)$. In this direction, [14] used a canonical path argument to show the bound $\gamma(H_A) \gtrsim \mu(A)/n^2$, which improves upon Theorem 4 by a factor of $\mu(A)$. However, this improvement is only effective when $\mu(A) \ll 1$ and the dependence on n remains suboptimal.

1.3 Proof overview: our work

At first glance, the proof in [16], as described in the previous subsection, appears to heavily depend on the discrete nature of the bottleneck ratio. In our view, a key conceptual contribution of this work is that the arguments in [16] can be adapted to the L^2 setting. While the discrete setting leads to the bottleneck ratio, in the L^2 setting, the corresponding arguments directly lead to the spectral expansion as stated in Theorem 8. For a full set of analogies, see Table 1.

Table 1 The analogies between the discrete and L^2 settings.

The discrete setting	The L^2 setting
subset $S \subseteq A$	function $f : A \to \mathbb{R}$
the complement set $A \setminus S$	the function $-f$
$\lvert E(S, A \setminus S)\rvert$	$\mathcal{E}_A(f)$
$\min\{\lvert S\rvert, \lvert A \setminus S\rvert\}$	$\mathrm{Var}_A[f]$
bottleneck ratio $\phi(H_A)$	spectral gap $\gamma(H_A)$
directed isoperimetric inequality [25]	directed L^2-Poincaré inequality (Theorem 12)
classical FKG inequality [22]	approximate FKG inequality (Theorem 14)

In contrast to [16], where the two main inequalities used in the proof – the directed isoperimetric inequality from [25] and the FKG inequality from [22] – are classical results, in our L^2 settings we have to formulate and prove new versions of these inequalities, which may be of interest on their own.

Directed Poincaré inequality

As indicated in Section 1.2, directed isoperimetric inequalities aim to lower bound the number of "upward boundary edges" from a set S to $\{0,1\}^n \setminus S$ in terms of the "distance" of S to monotonicity; see Section 1.6 for more background.

For our application, we require a directed isoperimetric inequality in the L^2 setting, which is the setting associated with the classical Poincaré inequality and the spectral gap. The first step is to define, for any $f : \{0,1\}^n \to \mathbb{R}$, its L^2-distance to monotonicity and its "upward boundary edges".

▶ **Definition 10** (Distance to monotonicity). *For a function $f : \{0,1\}^n \to \mathbb{R}$, we define*

$$\mathsf{dist}_2^{\mathsf{mono}}(f) := \inf_{g \in \mathsf{mono}} \sqrt{\mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\big(f(x) - g(x)\big)^2 \right]},$$

where g ranges over all monotone increasing functions from $\{0,1\}^n$ to $\mathbb{R}$.

▶ **Definition 11** (Upward boundary). *For all $f : \{0,1\}^n \to \mathbb{R}$, we define*

$$\mathcal{E}^-(f) := \frac{1}{4} \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\sum_{i=1}^n \min\left\{0, f(x^{i\to 1}) - f(x^{i \to 0})\right\}^2 \right].$$

Here for $x \in \{0,1\}^n$ and $b \in \{0,1\}$, $x^{i\to b}$ stands for the string $(x_1, \dots, x_{i-1}, b, x_{i+1}, \dots, x_n)$.

We are now ready to state our directed L^2-Poincaré inequality.

▶ **Theorem 12** (Directed Poincaré inequality). *For all functions $f : \{0,1\}^n \to \mathbb{R}$, we have*

$$\mathsf{dist}_2^{\mathsf{mono}}(f)^2 \le \mathcal{E}^-(f).$$

Approximate FKG inequality

The classical FKG inequality of [22] states that if $f, g : \{0,1\}^n \to \mathbb{R}$ are monotone increasing functions and x is a uniformly random element of $\{0,1\}^n$, then the random variables $f(x)$ and $g(x)$ are nonnegatively correlated. It is well-known that this statement holds for increasing functions over a broader class of partially ordered sets (posets). In our proof, we crucially need a lower bound on the correlation ratio of any two increasing functions $A \to \mathbb{R}$, where the set A is partially ordered by the natural partial order of the hypercube. However, it is easy to see that the FKG inequality does not generally hold on this poset. Thus, we seek an "approximate" version of the FKG inequality, where we are content with a correlation ratio bounded away from -1, rather than necessarily nonnegative.

▶ **Definition 13** (Approximate FKG ratio). *Fix a monotone set $A \subseteq \{0,1\}^n$ with at least 2 elements. We define the* approximate FKG ratio *of the poset A to be*

$$\delta(A) := \min\left\{0, \inf_{f,g \in \mathsf{mono}_A \setminus \mathsf{const}_A} \frac{\mathrm{Cov}_A[f,g]}{\sqrt{\mathrm{Var}_A[f] \cdot \mathrm{Var}_A[g]}}\right\},$$

where f and g range over all non-constant monotone increasing functions from A to $\mathbb{R}$. Here, $\mathrm{Cov}_A[f,g]$ *stands for the covariance of the random variable pair $(f(x), g(x))$ where x is a uniformly random element of A.*

▶ **Theorem 14** (Approximate FKG inequality). *For any monotone set $A \subseteq \{0,1\}^n$ with at least 2 elements, we have $\delta(A) \ge -\sqrt{1-\mu(A)}$.*

1.4 The case of small A: fast mixing requires good FKG ratio

The bound in Theorem 8 gives only a spectral gap bound $\gamma(H_A) \gtrsim \mu(A)/n$ for the random walk censored to A. The dependence on n is clearly optimal: even in the case $A = H_n$, the spectral gap is exactly $1/n$. The next example shows that the asymptotic dependence on $\mu(A)$ is also optimal.

▶ **Example 15** ([16, Example 1.3])**.** Assume $n/4 \le m \le n/2$ and consider $A = \{x \in \{0,1\}^n : x_1 = \cdots = x_m = 1\} \cup \{x \in \{0,1\}^n : x_{m+1} = \cdots = x_{2m} = 1\}$, the union of two subcubes. In this case, $\mu(A) \sim 2^{-m+1}$. Let $f : A \to \mathbb{R}$ be defined by $f(x) = 1$ if $x_1 = \cdots = x_m = 1$ and $f(x) = -1$ otherwise. Then $\mathcal{E}_A(f) \sim m \cdot 2^{-m}$ and $\mathrm{Var}_A[f] \sim 1$, so $\mathrm{Var}_A[f] \gtrsim \mu(A)^{-1} \cdot \mathcal{E}_A(f)$. Standard Markov chain theory (e.g. [35, Theorem 7.4]) shows that the mixing time of the random walk censored to A is exponentially large in n.

Remarkably, Example 15 is also where the approximate FKG inequality fails badly – if we let A be the union of two subcubes as in Example 15 and consider the indicator functions of the two subcubes, it is easy to see that they are both increasing functions on the poset A but are very anti-correlated (i.e. the approximate FKG ratio $\delta(A)$ is very close to -1).

Our results actually reveal that, when A is a monotone set, torpid mixing happens *if and only if* the approximate FKG ratio of A is close to -1.

▶ **Theorem 16.** *Let $A \subseteq \{0,1\}^n$ be a monotone set with at least 2 elements. Then for all functions $f : A \to \mathbb{R}$, we have $(1 + \delta(A)) \cdot \mathrm{Var}_A[f] \le \mathcal{E}_A(f)$.*

▶ **Theorem 17.** *Let $A \subseteq \{0,1\}^n$ be a monotone set with at least 2 elements. Then for some non-constant function $f : A \to \mathbb{R}$, we have $(1 + \delta(A)) \cdot n \cdot \mathrm{Var}_A[f] \ge \mathcal{E}_A(f)$.*

The two theorems above imply that $(1 + \delta(A))/n \le \gamma(H_A) \le 1 + \delta(A)$, which means the approximate FKG ratio of A characterizes the spectral gap of H_A up to a factor of n.

▶ Remark 18. The case $A = H_n$ demonstrates that the lower bound $(1 + \delta(A))/n \le \gamma(H_A)$ is tight. Moreover, there are examples indicating that the upper bound $\gamma(H_A) \le 1 + \delta(A)$ is tight up to a constant factor – for instance, when $A = \{x \in \{0,1\}^n : |x| \le 1\}$.

1.5 Open problems

Our work leaves two avenues for potential improvement, roughly corresponding to two regimes in the size of the set A. We discuss each of these directions in turn.

Large A

When $\mu(A) \ge \varepsilon$ for some fixed constant $\varepsilon > 0$, we establish the tight asymptotic bound $\gamma(H_A) \gtrsim 1/n$. However, this only yields the mixing time bound $t_{\mathsf{mix}} = O_\varepsilon(n^2)$, which does not resolve Question 5. One way to establish an $O(n \log n)$ mixing time bound would be to prove a log-Sobolev inequality instead of an L^2-Poincaré inequality. It is plausible that our techniques could be further adapted to establish a log-Sobolev inequality, similar to how we extended the argument of [16] from the discrete setting to the L^2 setting. Are there analogous versions of the directed isoperimetric inequality and the approximate FKG inequality in the log-Sobolev setting? We leave these as open questions.

Small A

When no additional structure is imposed, the random walk censored to A may mix very slowly if $\mu(A) \ll 1$ (e.g. Example 15). In many problems of interest, however, A possesses some form of structure, and the goal is to obtain a good bound on the mixing time, aiming for an efficient approximate sampling algorithm. For example, when A is a halfspace, i.e. defined by $A = \{x \in \{0,1\}^n : a_1x_1 + \cdots + a_nx_n \geq b\}$ for nonnegative numbers $a_1, \ldots, a_n, b$, [38] proves a mixing time bound of $n^{9/2+o(1)}$ which yields a sampling algorithm for 0-1 knapsack solutions.

Our work (Theorems 16 and 17) reveals that the deciding factor for whether rapid mixing holds is the *approximate FKG ratio* of A rather than the size of A. However, we do not know how to leverage additional structure of A in a direct study of its approximate FKG ratio, and we leave the development of new tools for this purpose as a direction for future work.

1.6 Related work

Mixing time of censored random walks

Markov Chain Monte Carlo methods are the object of extensive study in mathematics, statistical physics and theoretical computer science, and the question of mixing time of random walks lies at the core of algorithms for approximate sampling and counting; see e.g. [29, 27, 37, 35]. In settings featuring combinatorial structure such as in sampling matchings, independent sets, or spanning forests of a graph, or their natural generalizations to the more algebraic setting of matroids, a *basis exchange* or *down-up* random walk is usually employed, and spectral arguments are used to bound the mixing time; see e.g. [30, 2, 15, 1].

Our work focuses on the setting where the set A may not enjoy such rich structure, and instead is only guaranteed to be monotone. As discussed above, our results improve upon the spectral gap and mixing time bounds shown by the previous works of [16, 14]. In the special case where the monotone set A is additionally promised to contain every $x \in \{0,1\}^n$ with Hamming weight at least $(\frac{1}{2} - \varepsilon)n$, i.e. the middle layers of the hypercube, one may expect the censored and uncensored random walks to behave similarly, and indeed in this case [36] gave the optimal $\Omega_\varepsilon(1/n)$ bound for the spectral gap and log-Sobolev constant of the censored random walk, which implies the optimal $O_\varepsilon(n \log n)$ mixing time bound.

Besides combinatorial or algebraic structure, one may also ask what *geometric* structure affords fast mixing. As mentioned above, [38] studied the censored random walk when the set A is a halfspace, i.e. $A = \{x \in \{0,1\}^n : a_1x_1 + \cdots + a_nx_n \leq b\}$ for nonnegative numbers $a_1, \ldots, a_n, b$, which corresponds to 0-1 knapsack solutions. Intuitively, such set A should not contain bottlenecks even if it is very small, and indeed [38] showed a mixing time bound of $n^{9/2+o(1)}$. By our Theorem 17, this also gives an inverse polynomial lower bound on the quantity $1 + \delta(A)$. On the other hand, a lower bound of $\widetilde{\Omega}(n^2)$ holds for the mixing time [38], and closing this gap is an interesting open problem[3].

[3] In a different line of investigation, a series of works has explored approaches for approximately sampling and counting knapsack solutions using dynamic programming [17, 26, 42, 24, 41, 18]; the best results in this direction are $\widetilde{O}(n^{5/2})$ and $\widetilde{O}(n^4)$-time algorithms for approximately sampling (depending on the model of computation) [17, 24, 41], and a (subquadratic) $\widetilde{O}(n^{3/2})$-time algorithm for approximately counting knapsack solutions [18]. However, these results do not directly say anything about the mixing time of the random walk on knapsack solutions.

Directed isoperimetric inequalities

Directed versions of isoperimetric inequalities such as the classical Poincaré inequality and Talagrand's inequality [43] have emerged over the last couple of decades as a key tool in the field of property testing. Since the introduction of the problem of monotonicity testing [25] and especially since the work of [11], directed isoperimetric inequalities have unlocked new results on the query complexity of testing monotonicity of Boolean functions over discrete domains such as the hypercube and hypergrid [31, 7, 40, 5, 6, 10], and, more recently, real-valued functions over the hypercube [8] and the continuous cube [20]. In recent work, [12] proved such an inequality toward testing monotonicity of *probability distributions* over the hypercube, using certain conditional samples from that distribution. To the best of our knowledge, the work of [16] was the first application of a directed isoperimetric inequality (i.e. the inequality of [25] for Boolean functions) outside of property testing, and the present work seems to be the first application of a real-valued directed isoperimetric inequality outside of property testing.

The specific Poincaré inequality we prove in Theorem 12 is most closely related to the following previous developments. The work of [19] introduced the systematic study of directed L^p-Poincaré inequalities for monotonicity testing, and proved an L^1 version of our Theorem 12; in the language of that paper, Theorem 12 is a directed (L^2, ℓ^2)-Poincaré inequality, whereas [19] proved an (L^1, ℓ^1) inequality. Another similar inequality for real-valued functions was proved by [8], who considered the Hamming distance as opposed to L^p distance. An (L^2, ℓ^2) inequality (i.e. the same flavor as ours) was proved for functions over the *continuous* cube $[0, 1]^n$ in [20], and our proof via the study of a dynamical system is directly inspired by theirs. Most recently, [12] proved a directed (L^1, ℓ^2)-Poincaré inequality[4] for real-valued functions on the hypercube, by extending the result of [31] for Boolean functions via a thresholding argument of [4]. There does not seem to be a trivial reduction between our inequality and the foregoing results.

Organization of the paper

It is clear that Theorems 14 and 16 together imply Theorem 8, and we recall that standard Markov chain theory ([35, Theorem 12.4]) derives Theorem 6 from Theorem 8.

Section 2 presents a proof of the approximate FKG inequality for large monotone sets (Theorem 14). Section 3 is where the heart of the argument of [16] is carried out in the L^2 setting. Section 3 demonstrates that the directed isoperimetric inequality of $\{0, 1\}^n$ (Theorem 12) implies the undirected isoperimetric inequality of the monotone subset A (Theorem 16), and why the approximate FKG ratio of A is important for this implication. Sections 4 and 5 are devoted to proving the directed L^2-Poincaré inequality (Theorem 12).

We defer the proof of Theorem 17, which is logically independent from the proof of the main result Theorem 6, to the full version of the paper.

2 Approximate FKG inequality

The goal of this section is to prove Theorem 14, where we need to lower bound the correlation ratio between two monotone increasing functions on A. We first note that the case where the functions take values in $\{0, 1\}$ is easy. Indeed, we have the following simple lemma.

[4] Compared to Theorem 12, that inequality uses the L^1 as opposed to L^2 distance, and takes a square-root inside the expectation operator in our definition of $\mathcal{E}^-(f)$ in Definition 11. By Jensen's inequality, the square root of the left- and right-hand sides of our inequality are respectively larger than the left- and right-hand sides of the inequality of [12], so the two results are not immediately comparable.

▶ **Lemma 19.** *Assume that $f, g : A \to \{0,1\}$ are monotone increasing functions. Then we have* $\mathop{\mathbb{E}}_{x\in A}[f(x)g(x)] \geq \mu(A) \cdot \mathop{\mathbb{E}}_{x\in A}[f(x)] \cdot \mathop{\mathbb{E}}_{x\in A}[g(x)]$.

Proof. Let $B = \{x \in A : f(x) = 1\}$ and $C = \{x \in A : g(x) = 1\}$. By the monotonicity of f and g, the sets B and C are both monotone subsets of the hypercube $\{0,1\}^n$. By the classical FKG inequality [22] we know that $\mu(B \cap C) \geq \mu(B) \cdot \mu(C)$. Therefore,

$$\mathop{\mathbb{E}}_{x\in A}[f(x)g(x)] = \frac{\mu(B\cap C)}{\mu(A)} \geq \mu(A) \cdot \frac{\mu(B)}{\mu(A)} \cdot \frac{\mu(C)}{\mu(A)} = \mu(A) \cdot \mathop{\mathbb{E}}_{x\in A}[f(x)] \cdot \mathop{\mathbb{E}}_{x\in A}[g(x)]\,. \qquad ◀$$

It is straightforward to deduce from Lemma 19 that for monotone increasing functions $f, g : A \to \{0,1\}$, the desired approximate FKG inequality

$$\mathrm{Cov}_A[f,g] \geq -\sqrt{1-\mu(A)} \cdot \sqrt{\mathrm{Var}_A[f] \cdot \mathrm{Var}_A[g]}$$

holds.

The main challenge in Theorem 14 lies in extending this idea to real-valued functions. In fact, the problem can be reduced to proving the following statement, which involves purely random variables rather than any structural property of the partially ordered set A.

▶ **Theorem 20.** *Let (X, Y) be a pair of real-valued random variables with bounded second moment. Suppose there is a constant $c \in [0,1)$ such that for all $a, b \in \mathbb{R}$,*

$$\mathbb{P}[X \geq a, Y \geq b] \geq c \cdot \mathbb{P}[X \geq a] \cdot \mathbb{P}[Y \geq b]\,, \tag{1}$$

then we must have

$$\mathrm{Cov}[X,Y] \geq -\sqrt{(1-c)\cdot \mathrm{Var}[X] \cdot \mathrm{Var}[Y]}. \tag{2}$$

Proof of Theorem 14 assuming Theorem 20. Let x be a uniformly random element of A and let $X = f(x)$ and $Y = g(x)$. Thus $\mathrm{Var}[X] = \mathrm{Var}_A[f]$, $\mathrm{Var}[Y] = \mathrm{Var}_A[g]$, and $\mathrm{Cov}[X,Y] = \mathrm{Cov}_A[f,g]$.

Now for each pair of $a, b \in \mathbb{R}$, if we define $f_a, g_b : A \to \mathbb{R}$ by

$$f_a(x) := \mathbb{1}[f(x) \geq a] \qquad \text{and} \qquad g_b(x) := \mathbb{1}[g(x) \geq b]\,,$$

since they are clearly monotone increasing 0/1-valued functions, we can apply Lemma 19 to f_a and g_b to deduce that $\mathbb{P}[X \geq a, Y \geq b] \geq \mu(A) \cdot \mathbb{P}[X \geq a] \cdot \mathbb{P}[Y \geq b]$. If $\mu(A) = 1$, then $A = \{0,1\}^n$ and the conclusion follows from the classical FKG inequality. If $\mu(A) < 1$, we apply Theorem 20 to the random variable pair (X, Y) with constant $c = \mu(A)$, which yields exactly the desired conclusion. ◀

The remainder of this section is devoted to proving Theorem 20, which is surprisingly nontrivial. To illustrate the complexity of this inequality, we note that equality in (2) holds for a wide range of joint distributions of (X, Y) beyond the case captured by Lemma 19, i.e. where X and Y take only two possible values.

▶ **Example 21.** Let (X, Y) follow a discrete distribution supported on the grid $\{0,2,3\}^2$. Specifically, let

$$\mathbb{P}[X=3,Y=3] = \mathbb{P}[X=3,Y=2] = \mathbb{P}[X=2,Y=3] = \frac{1}{5},$$

$$\mathbb{P}[X=0,Y=3] = \mathbb{P}[X=3,Y=0] = \frac{1}{15}, \qquad \text{and} \qquad \mathbb{P}[X=2,Y=2] = \frac{4}{15}.$$

Then (1) holds for $c = 45/49$ and all $a, b \in \mathbb{R}$. On the other hand, we have $\mathrm{Cov}[X,Y] = -8/45$ and $\mathrm{Var}[X] = \mathrm{Var}[Y] = 28/45$, so equality in (2) holds for $c = 45/49$ as well.

2.1 A symmetric model

A key challenge in Theorem 20 lies in its lack of "centrosymmetry" with respect to (X, Y). While the assumption (1) does not remain invariant under the substitution $X \mapsto -X$ and $Y \mapsto -Y$, the conclusion is unaffected by such substitutions. This raises an intriguing question: what is the "symmetric" information inherent in (1) that leads to the conclusion?

In this subsection, we present an approach that effectively extracts the "symmetric" information from (1). To this end, we first define two Borel measures on $[0,1]$ induced by X and Y.

▶ **Definition 22.** *Let $\varphi_X : \mathbb{R} \to [0,1]$ be the Borel-measurable map defined by $a \mapsto \mathbb{P}[X \geq a]$, and then for each Borel set $E \subseteq [0,1]$, let $\alpha(E)$ be the Lebesgue measure of the inverse image $\varphi_X^{-1}(E)$. The countable additivity of α easily follows from the countable additivity of the Lebesgue measure.*

The measure α is referred to as the push-forward of the Lebesgue measure on $\mathbb{R}$ by the map $a \mapsto \mathbb{P}[X \geq a]$. Similarly define the Borel measure β on $[0,1]$ to be the push-forward of the Lebesgue measure by the map $b \mapsto \mathbb{P}[Y \geq b]$.

The definition of push-forward measures naturally leads to the following "change of variable" formula, which is a standard fact in measure theory.

▶ **Proposition 23** ([9, Theorem 3.6.1])**.** *Suppose φ is a Borel-measurable map from $\mathbb{R}$ to $[0,1]$, and suppose λ is the push-forward of the Lebesgue measure under φ. Then for any Borel-measurable function $f : [0,1] \to \mathbb{R}^{\geq 0}$, we have*

$$\int_0^1 f(x)\,\mathrm{d}\lambda(x) = \int_{\mathbb{R}} f(\varphi(a))\,\mathrm{d}a.$$

We then define the reverse of a measure, which corresponds to the substitution $X \mapsto -X$.

▶ **Definition 24.** *If λ is a Borel measure on $[0,1]$, we let λ^{R} be the Borel measure on $[0,1]$ defined by $\lambda^{\mathsf{R}}(E) = \lambda(\{1 - x : x \in E\})$, for Borel subsets E of $[0,1]$.*

The following definition is the crucial tool in our proof of Theorem 20.

▶ **Definition 25.** *Fix a constant $c \in [0,1)$. We define the operator $K_c(\cdot,\cdot)$ by*

$$K_c(\lambda,\nu) := \int\int \min\left\{\sqrt{1-c}\cdot xy, \frac{(1-x)(1-y)}{\sqrt{1-c}}\right\} \mathrm{d}\lambda(x)\,\mathrm{d}\nu(y),$$

for Borel measures λ, ν on $[0,1]$. For $c = 0$, we omit the subscript and write $K := K_0$.

The next two propositions demonstrate that the operator $K_c(\cdot,\cdot)$ is able to capture the variances and covariance of X and Y. Proposition 27 is the key place where "symmetric" information is extracted from the condition (1).

▶ **Proposition 26.** *We have* $\mathrm{Var}[X] = K(\alpha, \alpha^{\mathsf{R}})$. *Similarly,* $\mathrm{Var}[Y] = K(\beta, \beta^{\mathsf{R}})$.

Proof. Writing expected values as integrations of cumulative distribution functions (the "layer cake representation"), we have

$$\begin{aligned}
\mathrm{Var}\left[X\right] &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Big(\mathbb{P}\left[X \geq a, X \geq b\right] - \mathbb{P}\left[X \geq a\right]\mathbb{P}\left[X \geq b\right]\Big)\,\mathrm{d}a\,\mathrm{d}b \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\Big\{\mathbb{P}\left[X \geq a\right]\Big(1 - \mathbb{P}\left[X \geq b\right]\Big), \mathbb{P}\left[X \geq b\right]\Big(1 - \mathbb{P}\left[X \geq a\right]\Big)\Big\}\,\mathrm{d}a\,\mathrm{d}b \\
&= \int_0^1\int_0^1 \min\{x(1-y), y(1-x)\}\,\mathrm{d}\alpha(x)\,\mathrm{d}\alpha(y) \\
&= \int_0^1\int_0^1 \min\{xy, (1-x)(1-y)\}\,\mathrm{d}\alpha(x)\,\mathrm{d}\alpha^{\mathsf{R}}(y) = K(\alpha, \alpha^{\mathsf{R}}),
\end{aligned}$$

where the fourth equality above follows from Proposition 23. ◀

▶ **Proposition 27.** *Assuming* (1)*, we have* $\mathrm{Cov}\left[X,Y\right] \geq -\sqrt{1-c}\cdot K_c(\alpha,\beta)$*.*

Proof. In a similar way to the proof of Proposition 26, we have

$$\begin{aligned}
\mathrm{Cov}\left[X,Y\right] &= \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Big(\mathbb{P}\left[X \geq a, Y \geq b\right] - \mathbb{P}\left[X \geq a\right]\mathbb{P}\left[Y \geq b\right]\Big)\,\mathrm{d}a\,\mathrm{d}b. \qquad (3)
\end{aligned}$$

Note that on one hand, by (1) we have

$$\mathbb{P}\left[X \geq a, Y \geq b\right] - \mathbb{P}\left[X \geq a\right]\mathbb{P}\left[Y \geq b\right] \geq -(1-c)\mathbb{P}\left[X \geq a\right]\mathbb{P}\left[Y \geq b\right]. \qquad (4)$$

On the other hand, by union bound we have

$$\begin{aligned}
\mathbb{P}\left[X \geq a, Y \geq b\right] - \mathbb{P}\left[X \geq a\right]\mathbb{P}\left[Y \geq b\right] &\geq 1 - \mathbb{P}\left[X < a\right] - \mathbb{P}\left[Y < b\right] - \mathbb{P}\left[X \geq a\right]\mathbb{P}\left[Y \geq b\right] \\
&= -\Big(1 - \mathbb{P}\left[X \geq a\right]\Big)\Big(1 - \mathbb{P}\left[Y \geq b\right]\Big). \qquad (5)
\end{aligned}$$

Plugging (4) and (5) into (3), we have

$$\begin{aligned}
&\mathrm{Cov}\left[X,Y\right] \\
&\geq -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\Big\{(1-c)\mathbb{P}\left[X \geq a\right]\mathbb{P}\left[Y \geq b\right], \Big(1 - \mathbb{P}\left[X \geq a\right]\Big)\Big(1 - \mathbb{P}\left[Y \geq b\right]\Big)\Big\}\,\mathrm{d}a\,\mathrm{d}b \\
&= -\int_0^1\int_0^1 \min\{(1-c)xy, (1-x)(1-y)\}\,\mathrm{d}\alpha(x)\,\mathrm{d}\beta(y) = -\sqrt{1-c}\cdot K_c(\alpha,\beta),
\end{aligned}$$

where the first equality above follows from Proposition 23. ◀

We can now reduce Theorem 20 to the following more "symmetric" lemma, whose proof we defer to the full version.

▶ **Lemma 28.** *For any two Borel measures* α, β*, and any constant* $c \in [0,1)$ *we have*

$$K_c(\alpha,\beta)^2 \leq K(\alpha,\alpha^{\mathsf{R}}) \cdot K(\beta,\beta^{\mathsf{R}}).$$

Proof of Theorem 20 assuming Lemma 28. Using Proposition 27, Lemma 28 and Proposition 26 successively, we have

$$\begin{aligned}
\mathrm{Cov}\left[X,Y\right] &\geq -\sqrt{1-c}\cdot K_c(\alpha,\beta) \geq -\sqrt{1-c}\cdot\sqrt{K(\alpha,\alpha^{\mathsf{R}})K(\beta,\beta^{R})} \\
&= -\sqrt{(1-c)\cdot \mathrm{Var}\left[X\right]\cdot\mathrm{Var}\left[Y\right]}.
\end{aligned}$$

◀

3 From directed to undirected isoperimetry

In this section, we provide a proof of Theorem 16 assuming Theorem 12. Throughout the section, we fix a monotone set $A \subseteq \{0,1\}^n$ with at least 2 elements.

3.1 Domain extension

Since the target result, Theorem 16, focuses solely on the subset A of the hypercube, while Theorem 12 applies only to functions defined on the entire hypercube, we first introduce a simple method for extending function domains to the whole hypercube.

▶ **Definition 29.** *We define an operator T that extends any function $f : A \to \mathbb{R}$ to the function $T[f] : \{0,1\}^n \to \mathbb{R}$ defined by*

$$T[f](x) = \begin{cases} \min_{y \in A} f(y), & \text{if } x \notin A, \\ f(x), & \text{if } x \in A. \end{cases}$$

By defining the value of the function outside of the original domain A to be sufficiently small, the extension operator enjoys the following two useful properties that allow us to access the power of Theorem 12.

▶ **Proposition 30.** *For every function $f : A \to \mathbb{R}$ we have*

$$\mu(A) \cdot \mathcal{E}_A(f) = \mathcal{E}^-(T[f]) + \mathcal{E}^-(T[-f]).$$

Proof. See the full version. ◀

▶ **Proposition 31.** *For every function $f : A \to \mathbb{R}$, there exists a monotone increasing function $g : A \to \mathbb{R}$ such that $\|f - g\|_2 \leq \mu(A)^{-1/2} \cdot \mathsf{dist}_2^{\mathsf{mono}}(T[f])$, where the L^2-norm is the norm in the inner product space $L^2(A)$.*

Proof. Since the collection of all monotone increasing real-valued functions on $\{0,1\}^n$ form a closed set in the Euclidean space $\mathbb{R}^{\{0,1\}^n}$, there exists a monotone increasing function $\widetilde{g} : \{0,1\}^n \to \mathbb{R}$ such that $\|T[f] - \widetilde{g}\|_2 = \mathsf{dist}_2^{\mathsf{mono}}(T[f])$, where the L^2-norm is the norm in the space $L^2(\{0,1\}^n)$. Now note that the restriction $g := \widetilde{g}|_A$ is a monotone increasing function on A. Therefore,

$$\begin{aligned} \|f - g\|_2^2 &= \mathop{\mathbb{E}}_{x \in A}\left[(f(x) - g(x))^2\right] = \mathop{\mathbb{E}}_{x \in A}\left[\left(T[f](x) - \widetilde{g}(x)\right)^2\right] \\ &\leq \mu(A)^{-1} \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n}\left[\left(T[f](x) - \widetilde{g}(x)\right)^2\right] = \mu(A)^{-1} \cdot \mathsf{dist}_2^{\mathsf{mono}}(T[f])^2. \end{aligned}$$ ◀

3.2 Correlation analysis

In this subsection, we lay some groundwork about correlation of functions (or equivalently, random variables) that will help prove Theorem 16. We begin with the following natural definition of correlation ratios.

▶ **Definition 32.** *For non-constant functions $g, h : A \to \mathbb{R}$, we define*

$$\rho(g,h) := \frac{\mathrm{Cov}_A[g,h]}{\sqrt{\mathrm{Var}_A[g] \cdot \mathrm{Var}_A[h]}}.$$

The following triangle-inequality-type lemma is going to be important in the proof of Theorem 16. Conceptually, the lemma says that if functions g and h on A are not very correlated with each other (that is, $\rho(g, h)$ is bounded away from 1), then f cannot be very correlated with both g and h at the same time. In particular, we will later use the lemma in the case where g is a monotone increasing function and h is a monotone decreasing function, which cannot be very correlated if $\delta(A)$ is bounded away from -1.

▶ **Proposition 33.** *Consider three non-constant functions $f, g, h : A \to \mathbb{R}$. We have*

$$\max\{0, \rho(f, g)\}^2 + \max\{0, \rho(f, h)\}^2 \le 1 + \max\{0, \rho(g, h)\}.$$

Proof. We may without loss generality assume that $\mathrm{Var}_A[f] = \mathrm{Var}_A[g] = \mathrm{Var}_A[h] = 1$. In this case, $\mathrm{Cov}_A[f, g] = \rho(f, g)$, $\mathrm{Cov}_A[f, h] = \rho(f, h)$ and $\mathrm{Cov}_A[g, h] = \rho(g, h)$.

If $\rho(f, g) < 0$ then the conclusion trivially holds since $\max\{0, \rho(f, h)\}^2 \le 1$. Similarly if $\rho(f, h) < 0$, the conclusion is also trivial. In the following, we assume that $\rho(f, g) \ge 0$ and $\rho(f, h) \ge 0$.

Consider the matrix

$$B := \begin{bmatrix} 1 & \rho(f, g) & \rho(f, h) \\ \rho(f, g) & 1 & \rho(g, h) \\ \rho(f, h) & \rho(g, h) & 1 \end{bmatrix}.$$

For each vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$, we know $\boldsymbol{\lambda}^T B \boldsymbol{\lambda} = \mathrm{Var}_A[\lambda_1 f + \lambda_2 g + \lambda_3 h] \ge 0$. So B is a positive semi-definite matrix. This means $\det B \ge 0$, and we can expand it into

$$1 + 2\rho(f, g)\rho(f, h)\rho(g, h) \ge \rho(f, g)^2 + \rho(f, h)^2 + \rho(g, h)^2. \tag{6}$$

If $\rho(g, h) < 0$, then (6) implies $1 \ge \rho(f, g)^2 + \rho(f, h)^2$ and we arrive at the conclusion. In the following we assume $\rho(g, h) \ge 0$.

Expanding the Cauchy-Schwarz inequality $\mathrm{Var}_A[f] \cdot \mathrm{Var}_A[g + h] \ge \mathrm{Cov}_A[f, g + h]^2$ yields

$$2 + 2\rho(g, h) \ge (\rho(f, g) + \rho(f, h))^2 \ge 4\rho(f, g)\rho(f, h). \tag{7}$$

Multiplying both sides of (7) by $\rho(g, h)/2$ and then adding it to (6), we get the desired conclusion $1 + \rho(g, h) \ge \rho(f, g)^2 + \rho(f, h)^2$. ◀

The following definition serves to interpret correlation ratios in terms of L^2 distances.

▶ **Definition 34.** *For functions $f, g : A \to \mathbb{R}$, we define*

$$\tau(f, g) := \min_{a \in \mathbb{R}_{\ge 0}, b \in \mathbb{R}} \|f - (ag + b)\|_2,$$

where the L^2-norm is the norm in the inner product space $L^2(A)$.

▶ **Proposition 35.** *Consider two non-constant functions $f, g : A \to \mathbb{R}$. We have*

$$\tau(f, g)^2 = \left(1 - \max\{0, \rho(f, g)\}^2\right) \cdot \mathrm{Var}_A[f].$$

Proof. Note that

$$\begin{aligned} \tau(f, g)^2 &= \min_{a \in \mathbb{R}_{\ge 0}, b \in \mathbb{R}} \|f - (ag + b)\|_2^2 = \min_{a \in \mathbb{R}_{\ge 0}} \mathrm{Var}_A[f - ag] \\ &= \min_{a \in \mathbb{R}_{\ge 0}} \left(a^2 \cdot \mathrm{Var}_A[g] - 2a \cdot \mathrm{Cov}_A[f, g] + \mathrm{Var}_A[f]\right). \end{aligned} \tag{8}$$

If $\rho(f,g) < 0$, then $\mathrm{Cov}_A[f,g] < 0$, and the quadratic polynomial in the right hand side of (8) is minimized at $a = 0$. Therefore $\tau(f,g)^2 = \mathrm{Var}_A[f]$, as desired.

If $\rho(f,g) \geq 0$, then $\mathrm{Cov}_A[f,g] \geq 0$, and the quadratic polynomial in the right hand side of (8) is minimized at $a = \mathrm{Cov}_A[f,g]/\mathrm{Var}_A[g]$. Therefore (8) simplifies to

$$\tau(f,g)^2 = -\frac{\mathrm{Cov}_A[f,g]^2}{\mathrm{Var}_A[g]} + \mathrm{Var}_A[f] = (1-\rho(f,g)^2)\cdot \mathrm{Var}_A[f],$$

as desired. ◀

3.3 Proof of Theorem 16

We are now ready to prove Theorem 16 assuming Theorem 12.

Proof of Theorem 16 assuming Theorem 12. We have

$$\begin{aligned}
\mathcal{E}_A(f) &= \mu(A)^{-1}\cdot \mathcal{E}^-(T[f]) + \mu(A)^{-1}\cdot \mathcal{E}^-(T[-f]) && \text{(by Proposition 30)}\\
&\geq \mu(A)^{-1}\cdot \mathsf{dist}_2^{\mathsf{mono}}(T[f])^2 + \mu(A)^{-1}\cdot \mathsf{dist}_2^{\mathsf{mono}}(T[-f])^2 && \text{(by Theorem 12)}\\
&\geq \|f-g_0\|_2^2 + \|-f-h_0\|_2^2 && \text{(by Proposition 31)}, \quad (9)
\end{aligned}$$

for some monotone increasing functions $g_0, h_0 : A \to \mathbb{R}$. If g_0 is non-constant, we pick $g : A \to \mathbb{R}$ to be $g := g_0$. If g_0 is constant, we pick an arbitrary non-constant increasing function $g : A \to \mathbb{R}$. In either case, we trivially have

$$\|f-g_0\|_2^2 \geq \min_{a\in\mathbb{R}_{\geq 0}, b\in\mathbb{R}} \|f-(ag+b)\|_2^2 = \tau(f,g)^2.$$

Similarly we pick a non-constant increasing function $h : A \to \mathbb{R}$ such that $\|-f-h_0\|_2^2 \geq \tau(-f,h)^2$. We can then continue from (9) and have

$$\begin{aligned}
&\mathcal{E}_A(f) \geq \tau(f,g)^2 + \tau(-f,h)^2\\
&= \left(1-\max\{0,\rho(f,g)\}^2\right)\cdot \mathrm{Var}_A[f]\\
&\qquad + \left(1-\max\{0,\rho(-f,h)\}^2\right)\cdot \mathrm{Var}_A[f] && \text{(by Proposition 35)}\\
&= \left(2-\max\{0,\rho(f,g)\}^2 - \max\{0,\rho(f,-h)\}^2\right)\cdot \mathrm{Var}_A[f]\\
&\geq \left(1-\max\{0,\rho(g,-h)\}\right)\cdot \mathrm{Var}_A[f] && \text{(by Proposition 33)}\\
&= \left(1+\min\{0,\rho(g,h)\}\right)\cdot \mathrm{Var}_A[f] \geq (1+\delta(A))\cdot \mathrm{Var}_A[f] && \text{(by Definition 13).}
\end{aligned}$$

◀

4 Spectral theory and heat flow for directed graphs

In this section, as a first step toward proving our directed Poincaré inequality for the hypercube (Theorem 12), we first set up a framework that applies to the more general case of directed weighted graphs. Specifically, we revisit and extend the study of directed analogues of classical concepts from spectral graph theory such as the Laplacian operator, the Dirichlet energy, and the heat flow; define a directed notion of spectral gap for weighed directed graphs; and show that bounding this *dynamical spectral gap* suffices for proving a directed Poincaré inequality. Then, in the next section, Theorem 12 will follow as an application once we establish a bound on the directed spectral graph of the directed hypercube graph.

Prior work on spectral theory for directed graphs

There has been extensive prior work developing spectral graph theory beyond the classical setting of undirected graphs, toward capturing directed graphs and hypergraphs. Early work of [21, 13] associated a certain Hermitian matrix with each directed graph, and showed a Cheeger-type inequality based on the eigenvalues of that matrix, and many subsequent works have built upon that foundation; we refer to the recent thesis [45] for a thorough review, and here we mention two recent lines of work that are closest to our setting. One line of works [33, 34, 45] has developed a theory of *reweighted eigenvalues* capturing expansion properties of directed graphs and hypergraphs, proved Cheeger inequalities for these settings, and devised efficient algorithms for graph partitioning. Another line of works [47, 48, 23, 28] has pursued similar goals by analyzing a *nonlinear Laplacian* operator and the heat equation associated with it.

While our interest in a spectral theory for directed graphs is related to these previous works (and indeed we will build upon the approach of [47]), our focus is slightly different. In a nutshell, while prior works have focused on spectral characterizations of good expansion of a directed graph G as captured by directed versions of Cheeger inequalities (for edge conductance or vertex expansion) and mixing time of random walks, our focus will be on the quality of G as the "substrate" for a dynamical process; we will consider G a good directed spectral expander if it affords fast convergence for that process. In particular, our perspective allows for *directed acyclic graphs* to be considered good expanders, which is a stark departure from prior perspectives – as we briefly explain next.

Indeed, a central focus of prior works has been to establish Cheeger inequalities of the type $\vec{\lambda}_2 \lesssim \vec{\phi}(G) \lesssim \sqrt{\vec{\lambda}_2}$, where $\vec{\lambda}_2$ denotes a relevant second eigenvalue related to the directed weighted graph G, and the edge conductance $\vec{\phi}(G)$ of G is

$$\vec{\phi}(G) := \min_{\emptyset \neq S \subsetneq V} \frac{\min\{w(\delta^+(S)), w(\delta^+(V \setminus S))\}}{\min\{\mathrm{vol}_w(S), \mathrm{vol}_w(V \setminus S)\}},$$

where $\delta^+(S)$ denotes the outgoing edge boundary of S and $\mathrm{vol}_w(S)$ denotes the total weighted degree of all vertices in S. Now, if G is not strongly connected, then in general there exists a set S with positive volume but no outgoing edges, which makes $\vec{\phi}(G)$ and thus $\vec{\lambda}_2$ zero. In particular, this is the case for the directed hypercube graph which we are interested in, so if we hope to show a non-trivial directed Poincaré inequality via a spectral argument, such a quantity $\vec{\lambda}_2$ will not do.

Our approach

The type of spectral theory for directed graphs we study in this section was first developed by [47] in the context of network analysis. In that work, [47] defined a nonlinear Laplacian operator acting on real-valued functions defined on the vertices of a directed graph, showed that this operator induces a dynamical process that is a directed analogue of the classical heat flow on graphs, proved that this operator has nontrivial eigenvalues, and established a Cheeger inequality for this setting. While [47] focused on the implications of directed spectral theory for graph partitioning and related problems in network analysis, we focus on the dynamical properties of the heat flow on directed graphs – namely its convergence to a *monotone* limit, and its connection to the directed Poincaré inequality.

Let us briefly motivate and preview the main ideas in our argument. In classical spectral graph theory, given a graph $G = (V, E)$, the following four concepts play a central role:

1. The Laplacian operator $\mathcal{L}$, which acts on function f by outputting another function $\mathcal{L}f$.
2. The Dirichlet energy functional $\mathcal{E}$, which associates with each f an energy $\mathcal{E}(f) \geq 0$ measuring the "local variance" of f along edges of G.
3. The heat flow semigroup S_t, which captures a dynamical process which starts at some initial state f and has its rate of change governed by the Laplacian: $\frac{\mathrm{d}}{\mathrm{d}t} S_t f = \mathcal{L} S_t f$. The heat flow informally "sends mass" along each edge of G from the vertex with higher f-value to the vertex with lower f-value, causing the system to converge to an equilibrium state.
4. The Poincaré inequality, which states that $\mathrm{Var}[f] \leq \frac{1}{\lambda} \mathcal{E}(f)$. The best constant λ is called the *spectral gap* of G.

The appearance of the Poincaré inequality above hints at the relevance of this theory to our goal of proving a directed Poincaré inequality, and we mentioned above that our strategy toward this goal will be to define a directed version of the spectral gap. As a motivation for this strategy, we recall that the spectral gap ties together essentially all of the elements of the list above; indeed, as summarized in [46, Theorem 2.18], the following are equivalent given a constant $c \geq 0$:

1. Poincaré inequality: $\mathrm{Var}[f] \leq c\mathcal{E}(f)$ for all f.
2. Variance decay: $\mathrm{Var}[S_t f] \leq e^{-2t/c} \mathrm{Var}[f]$ for all f, t.
3. Energy decay: $\mathcal{E}(S_t f) \leq e^{-2t/c} \mathcal{E}(f)$ for all f, t.

To prove a directed Poincaré inequality, we replace the Laplacian operator $\mathcal{L}$ with an operator $\mathcal{L}^-$ which, intuitively, only "sends mass" from vertex u to vertex v if (u, v) is a directed edge and $f(u) > f(v)$, i.e. f violates monotonicity along edge (u, v); the dynamical system induced by $\mathcal{L}^-$ is the *directed heat flow* on G, which was studied by [47]. The directed heat flow is precisely the *gradient system* for the directed energy functional $\mathcal{E}^-(f)$, i.e. the upward boundary from Definition 11. Thus, this system intuitively "corrects" the local violations of monotonicity as quickly as possible, and indeed it converges to a monotone function as $t \to \infty$.

This directed theory cannot fully analogize the classical situation above; for example, variance decay fails to hold, because non-constant monotone functions are (non-unique!) stationary solutions. Instead, we will *define* the *dynamical spectral gap* of G as the best constant characterizing the (directed) energy decay, as in Item 3 above, and then show that (the directed version of) Item 3 implies (the directed version of) Item 1.

As mentioned in the introduction, recent work of [20] also proved a directed Poincaré inequality – for functions defined on the continuous cube $[0, 1]^n$ – using a dynamical argument. Indeed, that work also took as its starting point the connections between the heat flow and the Poincaré inequality, and showed that the natural directed version of the heat flow in continuous space enjoys exponential energy decay, which implies a directed Poincaré inequality for that setting. Our proof is conceptually similar to the proof of [20], but our techniques differ in at least two ways: 1) [20] required analytical arguments from the theory of partial differential equations (PDEs), while we are able to study our dynamical process as an ordinary differential equation (ODE) thanks to the finite-dimensional nature of our problem; and 2) [20] used tools from optimal transport theory to tensorize their one-dimensional result, while we obtain a multidimensional inequality directly by studying the directed heat flow as a gradient system. In this last regard, our proof also bears resemblance to, and is inspired by prior work of [32] on the so-called *Paulsen problem* from operator theory, where a "movement decay" property of a suitable dynamical system was used to bound the distance between the initial and equilibrium states of that system.

The rest of this section is organized as follows. Sections 4.1–4.3 present the directed versions of the Laplacian operator, the Dirichlet energy functional, and the heat flow, respectively. These subsections are largely an alternative exposition of the ideas in [47], but with a different emphasis tailored to our goals[5]. In Section 4.4, we define the dynamical spectral gap, which mediates a directed Poincaré inequality for each directed weighted graph.

Notation

Given a discrete set V, we denote by $L^2(V)$ the Hilbert space obtained by endowing the set of $V \to \mathbb{R}$ functions with the inner product $\langle f, g \rangle := \mathop{\mathbb{E}}_{x \in V} [f(x)g(x)]$, where the expectation is taken with respect to the uniform distribution over V. This inner product induces the norm $\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\mathbb{E}[f^2]}$.

For each $u \in V$, we write $e_u \in L^2(V)$ for the standard basis vector given by $e_u(v) := \mathbb{1}[u = v]$.

4.1 The directed Laplacian

Let $G = (V, w)$ be a directed weighted graph, where $w : V \times V \to [0, +\infty)$ is function specifying the weights of edges in G. By convention, we say that $(u, v) \in V \times V$ is an edge in G when $w(u, v) > 0$. We say G is *undirected* is w is a symmetric function.

▶ **Definition 36** (Directed Laplacian [47]). *The* directed Laplacian operator *of G is the operator $\mathcal{L}^- = \mathcal{L}^-_G : L^2(V) \to L^2(V)$ given by*

$$\mathcal{L}^- f := \frac{1}{2} \sum_{u,v \in V} w(u, v) \left(f(u) - f(v)\right)^+ (e_v - e_u) \tag{10}$$

for each $f \in L^2(V)$. In this paper, we use the notation $x^+ := \max\{x, 0\}$, for $x \in \mathbb{R}$.

Given $f \in L^2(V)$, we say an edge (u, v) is f*-monotone* if $f(u) \le f(v)$, and we say it is f*-antimonotone* if $f(u) > f(v)$. We say f *is monotone* if every edge (u, v) of G is f-monotone. If we think of f as the distribution of "mass" over the vertices V and of $\mathcal{L}^- \boldsymbol{f}$ as the rate of change of $\boldsymbol{f} = \boldsymbol{f}(t)$ over time t, then Definition 36 posits that mass flows along the f-antimonotone edges, from the heavier vertex to the lighter one. When G is undirected, this process is the standard heat flow on G, and indeed Definition 36 recovers the standard graph Laplacian in this case:

▶ **Observation 37.** *If G is undirected, then $\mathcal{L}^-_G$ is (half of) the standard (unnormalized) Laplacian operator $\mathcal{L}_G$ of G. Indeed, we can see the action of $\mathcal{L}^-_G$ on $f \in L^2(V)$ as follows: 1) remove the f-monotone edges (u, v) from G; 2) view the resulting graph G' as undirected; and 3) apply the standard Laplacian operator $\mathcal{L}_{G'}$ to f.*

▶ Remark 38. In spectral graph theory, one typically defines the Laplacian operator of an undirected graph as $-\mathcal{L}_G$ in our notation, i.e. by replacing $e_v - e_u$ with $e_u - e_v$ in Definition 36. Our notation follows instead the tradition from probability theory (see e.g. [3, 46]), which has the advantages 1) that $\mathcal{L}^-$ itself, rather than $-\mathcal{L}^-$, will be the generator of the heat semigroup – our main object of interest; and 2) of consistency with the analytic setting, where we have the Laplacian operator Δ for smooth functions in Euclidean space.

[5] In particular, [47] defined both normalized and unnormalized versions of their Laplacian operator, and focused on the normalized one. We study a single definition for weighted graphs.

We note that unlike the standard Laplacian operator, the operator $\mathcal{L}^-$ is nonlinear and not self-adjoint in general. Instead, one may think of $\mathcal{L}^-$ as a "piecewise linear" operator on $L^2(V)$, which is in particular Lipschitz continuous. In the full version we show

▶ **Lemma 39.** *The operator $\mathcal{L}^- : L^2(V) \to L^2(V)$ is Lipschitz continuous.*

4.2 The energy functional

As in the case of the standard Laplacian operator, the directed Laplacian naturally induces an *energy functional* (or Dirichlet form). The following definition corresponds to the Rayleigh quotient defined in [47].

▶ **Definition 40** (Energy functional)**.** *The* directed Dirichlet energy *functional $\mathcal{E}^- : L^2(V) \to \mathbb{R}$ is given by $\mathcal{E}^-(f) := -\langle f, \mathcal{L}^- f\rangle$.*

The directed Dirichlet energy measures the local violations of monotonicity along edges of G, and it is indeed always non-negative, as shown in the following proposition, which is similar to Lemma 4.3 of [47] for the normalized nonlinear Laplacian and is proved in the full version.

▶ **Proposition 41** (Energy functional measures local violations)**.** *For each $f \in L^2(V)$, it holds that*

$$\mathcal{E}^-(f) = \frac{1}{2} \mathop{\mathbb{E}}_{u \in V} \left[\sum_{v \in V} w(u,v) \left((f(u) - f(v))^+ \right)^2 \right] .$$

Observe that since $\mathcal{L}^-$ is "piecewise linear", the energy functional $\mathcal{E}^-$ is a "piecewise quadratic" functional on $L^2(V)$. Furthermore, at each point $f \in L^2(V)$ the Laplacian $\mathcal{L}^- f$ points at the direction opposite to the gradient of $\mathcal{E}^-$. Using the appropriate analytic formalism, we show the following lemma in the full version.

▶ **Lemma 42.** *For any $f \in L^2(V)$, we have $\mathcal{L}^- f = -\frac{1}{2}\nabla \mathcal{E}^-(f)$.*

4.3 Directed heat flow

As previewed in the Section 4.1, the directed Laplacian operator can be thought of as the rate of mass transfer along f-antimonotone edges of G in a dynamical process. Let us make this notion precise.

Given any $f \in L^2(V)$, we define the *directed heat flow* on G with *initial state* f as the dynamical system given by the initial value problem (IVP)

$$\boldsymbol{f}'(t) = \mathcal{L}^- \boldsymbol{f}(t) \quad \text{for all } t \geq 0 , \qquad \boldsymbol{f}(0) = f . \tag{11}$$

Since the operator $\mathcal{L}^-$ is Lipschitz, a standard existence and uniqueness theorem for ordinary differential equations (ODEs) implies that this IVP has a unique solution $\boldsymbol{f} : [0, +\infty) \to L^2(V)$; see e.g. [44, Corollary 2.6]. This can also be shown using the theory of maximal monotone operators, as done by [28] for the heat flow on hypergraphs, and by [48] in a study that generalizes both the directed graph setting of [47] and the hypergraph setting of [28].

Moreover, the directed heat flow enjoys the following semigroup structure. Define the operator family $(P_t)_{t \geq 0}$, with $P_t : L^2(V) \to L^2(V)$ for each $t \geq 0$, as follows: for each $f \in L^2(V)$, let $\boldsymbol{f} : [0, +\infty) \to L^2(V)$ be the solution to the IVP (11), and let $P_t f := \boldsymbol{f}(t)$. Then it immediately follows that P_t satisfies the properties of a semigroup, namely

1. P_0 is the identity operator.
2. $P_s P_t = P_{s+t}$ for all $s, t \geq 0$.
3. $\lim_{t \to 0} P_t f = f$ for all $f \in L^2(V)$ (follows from the differentiability of the solution $\boldsymbol{f}(t)$).

We call P_t the *directed heat semigroup* operator. Note that P_t is a nonlinear operator.

▶ **Observation 43** (Monotone functions are stationary solutions). *If $f \in L^2(V)$ is monotone, then $\mathcal{L}^- f = 0$ and hence $P_t f = f$ for all $t \geq 0$.*

Since monotone functions are stationary solutions to the directed heat flow (Observation 43), while any non-monotone f has non-zero Laplacian $\mathcal{L}^- f$ (see Proposition 41), it is natural to expect that $P_t f$ always converges to a monotone function as $t \to \infty$. This is indeed the case, because the directed heat flow is a *gradient system* for the *convex* energy functional $\mathcal{E}^-$.

▶ **Proposition 44** (Directed heat flow is gradient system). *For all $f \in L^2(V)$ and $t \geq 0$, we have $\frac{\mathrm{d}}{\mathrm{d}t} P_t f = -\frac{1}{2} \nabla \mathcal{E}^-(P_t f)$.*

Proof. By the definition of P_t we have $\frac{\mathrm{d}}{\mathrm{d}t} P_t f = \mathcal{L}^- P_t f$. The claim follows from Lemma 42. ◀

Using basic facts about gradient systems, in the full version we show

▶ **Corollary 45** (Convergence to monotone equilibrium). *For every $f \in L^2(V)$, there exists a (unique) monotone $f^* \in L^2(V)$ such that $P_t f \to f^*$ as $t \to \infty$.*

In light of Corollary 45, we may define the following limit operator.

▶ **Definition 46** (Monotone equilibrium). *We define the operator $P_\infty : L^2(V) \to L^2(V)$ by $P_\infty f := \lim_{t \to \infty} P_t f$ for each $f \in L^2(V)$, and call $P_\infty f$ the* monotone equilibrium *of f.*

4.4 Dynamical spectral gap

In this subsection, we associate with the directed Laplacian operator $\mathcal{L}^- = \mathcal{L}^-_G$ a quantity $\lambda^- = \lambda^-(G)$, the *dynamical spectral gap* of G, as a natural directed generalization of the classical (undirected) case. In particular, $\lambda^-(G)$ characterizes the rate of energy decay in the directed heat flow as $P_t f$ converges to its monotone equilibrium, and this also implies a directed Poincaré inequality linking the distance between the initial and equilibrium states to the energy of the initial state (i.e. its violations of monotonicity). We defer to the full version a thorough discussion of the motivation behind our definition. Here, we directly give the following definition, which accomplishes the goals described in the beginning of this section.

▶ **Definition 47.** *The* dynamical spectral gap *of G is the quantity $\lambda^-(G) \in [0, +\infty]$ given by*

$$\lambda^-(G) := \inf \left\{ \frac{\|\mathcal{L}^- f\|_2^2}{\mathcal{E}^-(f)} \;\middle|\; f \in L^2(V)\,,\ \mathcal{E}^-(f) > 0 \right\}.$$

In the full version, we show that $\lambda^-(G)$ characterizes the rate of exponential decay of $\mathcal{E}^-(P_t f)$, and this suffices to obtain, via a calculus argument, our directed Poincaré inequality:

▶ **Theorem 48** (Directed Poincaré inequality). *For all $f \in L^2(V)$, it holds that*

$$\|f - P_\infty f\|_2^2 \leq \frac{1}{\lambda^-(G)} \mathcal{E}^-(f)\,.$$

5 The dynamical spectral gap of the directed hypercube

Let H_n denote the unweighted directed hypercube in dimension n, i.e. $H_n = (\{0,1\}^n, w)$ where the weight function w is as follows: for each $x, y \in \{0,1\}^n$, $w(x,y) = 1$ if $\|x-y\|_1 = 1$ with $x \preceq y$, and $w(x,y) = 0$ otherwise. For simplicity of notation, in this section we also let $V := \{0,1\}^n$.

This section studies the spectral gap of H_n endowed with directed Laplacian operator $\mathcal{L}^- = \mathcal{L}^-_{H_n}$ and associated directed Dirichlet energy functional $\mathcal{E}^-$. We show

▶ **Theorem 49** (Dynamical spectral gap of the directed hypercube)**.** *H_n satisfies $\lambda^-(H_n) = 1$.*

Here we sketch the main idea behind the proof of Theorem 49, which is given in the full version. The main point is that $\mathcal{L}^-$ and $\mathcal{E}^-$ enjoy a useful coordinate-wise decomposition: we write $\mathcal{L}^- = \sum_{i=1}^n \mathcal{L}^{(i)}$, where each $\mathcal{L}^{(i)} : L^2(V) \to L^2(V)$ is given by

$$(\mathcal{L}^{(i)} f)(x) := \frac{1}{2}\left(f(x^{\oplus i}) - f(x)\right) \mathbb{1}\left[f(x^{i \to 0}) > f(x^{i \to 1})\right]$$

for each $f \in L^2(V)$ and $x \in \{0,1\}^n$. It is straightforward to check that this decomposition agrees with Definition 36. Similarly, from Proposition 41 we also obtain the decomposition $\mathcal{E}^- = \sum_{i=1}^n \mathcal{E}^{(i)}$, where each $\mathcal{E}^{(i)} : L^2(V) \to \mathbb{R}$ is given by

$$\mathcal{E}^{(i)}(f) := \frac{1}{4} \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n}\left[\left((f(x^{i \to 1}) - f(x^{i \to 0}))^-\right)^2\right]$$

for each $f \in L^2(V)$. (The extra factor of $1/2$ compared to Proposition 41 appears because the summation above counts each edge of H_n twice.)

Now, the upper bound of Theorem 49 is easy and attained by anti-dictator functions, so the main point is to show the lower bound $\lambda^-(H_n) \geq 1$. To this end, we fix any function f, consider the fraction in Definition 47, and expand the operators $\mathcal{L}^-$ and $\mathcal{E}^-$ according to the coordinate-wise decompositions above. The definitions of $\mathcal{L}^{(i)}$ and $\mathcal{E}^{(i)}$ readily yield $\left\|\mathcal{L}^{(i)} f\right\|_2^2 = \mathcal{E}^{(i)}(f)$, so the main step is to control the correlation terms $\left\langle \mathcal{L}^{(i)} f, \mathcal{L}^{(j)} f \right\rangle$. The main conceptual ingredient of the proof shows that this correlation is nonnegative, which is established by an argument reminiscent of the analysis of the "edge tester" for monotonicity of functions on the hypercube [25][6]. We prove this fact below.

▶ **Lemma 50.** *For every $i, j \in [n]$ with $i \neq j$, we have $\left\langle \mathcal{L}^{(i)} f, \mathcal{L}^{(j)} f \right\rangle \geq 0$.*

Proof. Suppose without loss of generality that $i = 1$ and $j = 2$. We first observe that it suffices to consider each "square" obtained by fixing all but the first two coordinates, since the inner product decomposes along these squares. Concretely, for each $y \in \{0,1\}^{n-2}$, let $g_y : \{0,1\}^2 \to \mathbb{R}$ be given by $g_y(x) := f(x, y)$ for each $x \in \{0,1\}^2$, where we write $f(x,y)$ for the value of f at the input obtained by concatenating x and y. Then

[6] [25] define a *switch operator* S_i which fixes all violations of monotonicity of a Boolean function f along direction i, by switching the values of f along violating edges. A key lemma in that paper shows that the application of S_i can only make the number of violations of monotonicity along a different direction j smaller – informally, the work along direction i "helps" toward the work along direction j. In our L^2 setting, this is captured by the positive correlation $\left\langle \mathcal{L}^{(i)} f, \mathcal{L}^{(j)} f \right\rangle \geq 0$ between the contributions of directions i and j to the action of $\mathcal{L}$.

$$\begin{aligned}
\left\langle \mathcal{L}^{(1)} f, \mathcal{L}^{(2)} f \right\rangle &= \frac{1}{2^n} \sum_{z \in \{0,1\}^n} \Big((\mathcal{L}^{(1)} f)(z) \Big) \Big((\mathcal{L}^{(2)} f)(z) \Big) \\
&= \frac{1}{2^n} \sum_{x \in \{0,1\}^2} \sum_{y \in \{0,1\}^{n-2}} \Bigg[\frac{1}{2} \left(f(x^{\oplus 1}, y) - f(x,y) \right) \mathbb{1}\left[f(x^{1\to 0}, y) > f(x^{1 \to 1}, y) \right] \\
&\qquad \cdot \frac{1}{2} \left(f(x^{\oplus 2}, y) - f(x,y) \right) \mathbb{1}\left[f(x^{2\to 0}, y) > f(x^{2 \to 1}, y) \right] \Bigg] \\
&= \frac{1}{2^n} \sum_{y \in \{0,1\}^{n-2}} \sum_{x \in \{0,1\}^2} \mathcal{L}^{(1)} g_y(x) \cdot \mathcal{L}^{(2)} g_y(x) \,.
\end{aligned}$$

We will show that for any $g : \{0,1\}^2 \to \mathbb{R}$, the sum $\sum_{x\in\{0,1\}^2} \mathcal{L}^{(1)} g(x) \cdot \mathcal{L}^{(2)} g(x)$ is nonnegative, which will complete the proof. Define

$$c := g(0,1)\,, \qquad d := g(1,1)\,,$$
$$a := g(0,0)\,, \qquad b := g(1,0)\,.$$

Then we have

$$\begin{aligned}
&\sum_{x \in \{0,1\}^2} \mathcal{L}^{(1)} g(x) \cdot \mathcal{L}^{(2)} g(x) \\
&= \sum_{x \in \{0,1\}^2} \Big[\left(g(x^{\oplus 1}) - g(x) \right) \mathbb{1}\left[g(x^{1\to 0}) > g(x^{1\to 1}) \right] \\
&\qquad \cdot \left(g(x^{\oplus 2}) - g(x) \right) \mathbb{1}\left[g(x^{2\to 0}) > g(x^{2\to 1}) \right] \Big] \\
&= (b-a)\mathbb{1}[a > b]\,(c-a)\mathbb{1}[a>c] + (a-b)\mathbb{1}[a>b]\,(d-b)\mathbb{1}[b>d] \\
&\qquad + (d-c)\mathbb{1}[c>d]\,(a-c)\mathbb{1}[a>c] + (c-d)\mathbb{1}[c>d]\,(b-d)\mathbb{1}[b>d] \\
&= (a-b)^+(a-c)^+ - (a-b)^+(b-d)^+ - (c-d)^+(a-c)^+ + (c-d)^+(b-d)^+ \\
&= \left[(a-b)^+ - (c-d)^+\right]\left[(a-c)^+ - (b-d)^+\right] \geq 0\,,
\end{aligned}$$

where the last inequality is proved in Lemma 51 below. ◀

▶ **Lemma 51.** *For any $a, b, c, d \in \mathbb{R}$, we have $[(a-b)^+ - (c-d)^+]\,[(a-c)^+ - (b-d)^+] \geq 0$.*

Proof. Let $X := (a-b)^+ - (c-d)^+$ and $Y := (a-c)^+ - (b-d)^+$, so that our goal is to show that $XY \geq 0$, or equivalently that $X > 0 \implies Y \geq 0$ and $Y > 0 \implies X \geq 0$. By switching the roles of b and c, it suffices to prove the first implication, i.e. that $X > 0 \implies Y \geq 0$.

Suppose $X > 0$. Then $a - b > 0$, since otherwise we would have $(a-b)^+ = 0$ and hence $X \leq 0$. Therefore $a - b = (a-b)^+ > (c-d)^+ \geq c - d$. Moreover, if $(b-d)^+ = 0$ then $Y \geq 0$ and we are done, so we may assume that $b - d = (b-d)^+ > 0$. We conclude that

$$Y = (a-c)^+ - (b-d)^+ \geq (a-c) - (b-d) = (a-b) - (c-d) \geq 0\,,$$

where the first inequality holds since $(a-c)^+ \geq a - c$ while $(b-d)^+ = b - d$, and the second inequality holds since $a - b \geq c - d$. ◀

Combining Theorems 48 and 49, we conclude

▶ **Corollary 52** (Directed Poincaré inequality for the hypercube; refinement of Theorem 12). *For all $f \in L^2(V)$, it holds that $\mathrm{dist}_2^{\mathrm{mono}}(f)^2 \leq \|f - P_\infty f\|_2^2 \leq \frac{1}{\lambda^-(H_n)}\, \mathcal{E}^-(f) = \mathcal{E}^-(f)$.*

References

1 Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1198–1211, 2020. doi:10.1145/3357713.3384317.

2 Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1–12, 2019. doi:10.1145/3313276.3316385.

3 Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer Cham, 2014.

4 Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. L_p-testing. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 164–173, 2014. doi:10.1145/2591796.2591887.

5 Hadley Black, Deeparnab Chakrabarty, and C Seshadhri. A $d^{1/2+o(1)}$ monotonicity tester for boolean functions on d-dimensional hypergrids. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1796–1821, 2023. doi:10.1109/FOCS57990.2023.00110.

6 Hadley Black, Deeparnab Chakrabarty, and C Seshadhri. Directed isoperimetric theorems for boolean functions on the hypergrid and an $\widetilde{O}(n\sqrt{d})$ monotonicity tester. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 233–241, 2023.

7 Hadley Black, Deeparnab Chakrabarty, and Comandur Seshadhri. A $o(d) \cdot \mathrm{poly}\log n$ monotonicity tester for boolean functions over the hypergrid $[n]^d$. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2133–2151, 2018.

8 Hadley Black, Iden Kalemaj, and Sofya Raskhodnikova. Isoperimetric inequalities for real-valued functions with applications to monotonicity testing. *Random Structures & Algorithms*, 2024.

9 Vladimir Igorevich Bogachev. *Measure theory*, volume 1. Springer, 2007.

10 Mark Braverman, Subhash Khot, Guy Kindler, and Dor Minzer. Improved monotonicity testers via hypercube embeddings. In *Proceedings of the Innovations in Theoretical Computer Science Conference (ITCS)*, 2023.

11 D Chakrabarty and C Seshadhri. An $o(n)$ monotonicity tester for boolean functions over the hypercube. *SIAM Journal on Computing*, 45(2):461–472, 2016. doi:10.1137/13092770X.

12 Deeparnab Chakrabarty, Xi Chen, Simeon Ristic, C Seshadhri, and Erik Waingarten. Monotonicity testing of high-dimensional distributions with subcube conditioning, 2025. doi:10.48550/arXiv.2502.16355.

13 Fan Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9:1–19, 2005.

14 Emma Cohen. Problems in catalan mixing and matchings in regular hypergraphs. *Georgia Tech PhD thesis*, 2016.

15 Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-sobolev inequalities for strongly log-concave distributions. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1358–1370, 2019. doi:10.1109/FOCS.2019.00083.

16 Jian Ding and Elchanan Mossel. Mixing under monotone censoring. *Electronic Communications in Probability*, 19:1–6, 2014.

17 Martin Dyer. Approximate counting by dynamic programming. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 693–699, 2003. doi:10.1145/780542.780643.

18 Weiming Feng and Ce Jin. Approximately counting knapsack solutions in subquadratic time. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1094–1135, 2025. doi:10.1137/1.9781611978322.32.

19 Renato Ferreira Pinto Jr. Directed poincaré inequalities and L^1 monotonicity testing of lipschitz functions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, 2023. doi:LIPIcs.APPROX/RANDOM.2023.61.

20 Renato Ferreira Pinto Jr. Directed isoperimetry and monotonicity testing: A dynamical approach. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2295–2305. IEEE, 2024. doi:10.1109/FOCS61266.2024.00134.

21 James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *The annals of applied probability*, pages 62–87, 1991.

22 Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22:89–103, 1971.

23 Kaito Fujii, Tasuku Soma, and Yuichi Yoshida. Polynomial-time algorithms for submodular laplacian systems. *Theoretical Computer Science*, 892:170–186, 2021. doi:10.1016/J.TCS.2021.09.019.

24 Pawel Gawrychowski, Liran Markin, and Oren Weimann. A faster FPTAS for #knapsack. In *Proceedings of the International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 64–1, 2018. doi:10.4230/LIPIcs.ICALP.2018.64.

25 Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 3(20):301–337, 2000. doi:10.1007/S004930070011.

26 Parikshit Gopalan, Adam Klivans, Raghu Meka, Daniel Štefankovic, Santosh Vempala, and Eric Vigoda. An fptas for #knapsack and related counting problems. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 817–826, 2011. doi:10.1109/FOCS.2011.32.

27 Venkatesan Guruswami. Rapidly mixing markov chains: A comparison of techniques (a survey), 2016. arXiv:1603.01512.

28 Masahiro Ikeda, Atsushi Miyauchi, Yuuki Takai, and Yuichi Yoshida. Finding cheeger cuts in hypergraphs via heat equation. *Theoretical Computer Science*, 930:1–23, 2022. doi:10.1016/J.TCS.2022.07.006.

29 Mark Jerrum. Mathematical foundations of the markov chain monte carlo method. In *Probabilistic methods for algorithmic discrete mathematics*, pages 116–165. Springer, 1998.

30 Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity.* Springer Science & Business Media, 2003.

31 Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric-type theorems. *SIAM Journal on Computing*, 47(6):2238–2276, 2018. doi:10.1137/16M1065872.

32 Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, and Akshay Ramachandran. The paulsen problem, continuous operator scaling, and smoothed analysis. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 182–189, 2018. doi:10.1145/3188745.3188794.

33 Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Cheeger inequalities for directed graphs and hypergraphs using reweighted eigenvalues. In *Proceedings of the ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1834–1847, 2023. doi:10.1145/3564246.3585139.

34 Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Fast algorithms for directed graph partitioning using flows and reweighted eigenvalues. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 591–624, 2024. doi:10.1137/1.9781611977912.22.

35 David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

36 Pierre Mathieu. Log-sobolev and spectral gap inequalities for the knapsack markov chain. *Markov Process. Relat. Fields*, 8:595–610, 2002.

37 Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006. doi:10.1561/0400000003.

38 Ben Morris and Alistair Sinclair. Random walks on truncated cubes and sampling 0-1 knapsack solutions. *SIAM journal on computing*, 34(1):195–226, 2004. doi:10.1137/S0097539702411915.

39 Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

40 Ramesh Krishnan S Pallavoor, Sofya Raskhodnikova, and Erik Waingarten. Approximating the distance to monotonicity of boolean functions. *Random Structures & Algorithms*, 60(2):233–260, 2022. doi:10.1002/RSA.21029.

41 Romeo Rizzi and Alexandru I Tomescu. Faster fptases for counting and random generation of knapsack solutions. *Information and Computation*, 267:135–144, 2019. doi:10.1016/J.IC.2019.04.001.

42 Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. A deterministic polynomial-time approximation scheme for counting knapsack solutions. *SIAM Journal on Computing*, 41(2):356–366, 2012. doi:10.1137/11083976X.

43 Michel Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric & Functional Analysis*, 3(3):295–314, 1993.

44 Gerald Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Soc., 2012.

45 Kam Chuen Tung. *Reweighted Eigenvalues: A New Approach to Spectral Theory beyond Undirected Graphs*. PhD thesis, University of Waterloo, 2025.

46 Ramon van Handel. Probability in high dimension (lecture notes), 2014.

47 Yuichi Yoshida. Nonlinear laplacian for digraphs and its applications to network analysis. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 483–492, 2016. doi:10.1145/2835776.2835785.

48 Yuichi Yoshida. Cheeger inequalities for submodular transformations. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2582–2601, 2019. doi:10.1137/1.9781611975482.160.