On Applications of Discrete Isoperimetric and Hypercontractive Inequalities

by

Yumou Fei

S.B. Peking University (2024)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2026

Authored by: Yumou Fei
Department of Electrical Engineering and Computer Science
November 24, 2025

Certified by: Dor Minzer
Associate Professor of Mathematics
Thesis Supervisor

Certified by: Ronitt Rubinfeld
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

On Applications of Discrete Isoperimetric and Hypercontractive Inequalities

by

Yumou Fei

Submitted to the Department of Electrical Engineering and Computer Science
on November 24, 2025 in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

ABSTRACT

Isoperimetric and hypercontractive inequalities are two fundamental tools in analysis. Their discrete counterparts have been widely applied in theoretical computer science, with significant impact on areas such as property testing, computational complexity, and Markov chain analysis. In this thesis, we develop new discrete isoperimetric and hypercontractive inequalities and demonstrate their applications to Markov chain mixing times and lower bounds for streaming algorithms.

The emphasis of the thesis will be placed on presenting the additional nontrivial ingredients that enable these inequalities to be applied to the problems of interest, rather than on proving the inequalities themselves. We will also provide a high-level explanation of what makes these inequalities particularly effective in the corresponding problems.

The thesis is based on the author's joint works with Renato Ferreira Pinto Jr., Dor Minzer and Shuo Wang.

Thesis supervisor: Dor Minzer
Title: Associate Professor of Mathematics

Thesis supervisor: Ronitt Rubinfeld
Title: Professor of Electrical Engineering and Computer Science

Contents

Chapter 1

Introduction

Sobolev-type inequalities are among the central tools of modern analysis, with far-reaching applications in areas such as partial differential equations, the calculus of variations, and harmonic analysis. Broadly speaking, these inequalities show how upper bounds on suitable norms of a function's derivatives yield bounds on the function itself. Since derivatives quantify the local variation of a function, the validity of Sobolev-type inequalities typically relies on a "local-to-global" property of the underlying space — namely, that local control can be efficiently propagated to global control. For instance, the classical Sobolev inequalities exploit precisely such a local-to-global property of the Euclidean space $\mathbb{R}^n$.

In discrete mathematics and theoretical computer science, one often seeks discrete spaces that exhibit strong local-to-global properties. For instance, the ability to propagate norm bounds for functions on a space is closely tied to the propagation of probability distributions by random walks on that space. As a consequence, establishing Sobolev-type inequalities on discrete spaces has become a powerful tool for analyzing the mixing times of discrete random walks (see, e.g., [24, 26]). Such mixing-time bounds, in turn, often yield efficient sampling algorithms (see, e.g., [1, 18]). Local-to-global properties also play a central role in other algorithmic areas such as property testing (see, e.g., [13, 14]), and in computational complexity, where combinatorial structures with strong local-to-global behavior are frequently used both in reductions and in unconditional lower bounds (see, e.g., [17]). Given these broad applications, Sobolev-type inequalities have been increasingly studied and applied in discrete settings.

In this thesis, we provide an exposition of two new discrete Sobolev-type inequalities proved in [7, 8], along with their applications.

1.1 Expansion of Monotone Sets (Chapter 2)

The first result discussed in this thesis concerns the mixing time of discrete random walks. For a random walk on a discrete space, the mixing time is the number of steps after which the distribution of the particle's location — starting from any initial state — becomes close to the stationary distribution (typically uniform over all states). A classical example is the simple random walk on the hypercube $\{0,1\}^n$, where at each step the particle moves to a uniformly chosen neighbor differing in exactly one coordinate. By the well-known "coupon

collector" argument from elementary probability, the mixing time of this walk is $\Theta(n \log n)$.

A natural question is what happens when some states of the hypercube $\{0,1\}^n$ are marked as "forbidden," so that any attempted move into a forbidden state leaves the particle at its current position. If $A \subseteq \{0,1\}^n$ denotes the set of allowed states, one might hope that after only a small number of steps this "censored" random walk would become nearly uniform over A.

Unfortunately, this need not hold even when only an $o(1)$ fraction of states is forbidden. For example, if all states $x \in \{0,1\}^n$ with Hamming weight $\lfloor n/2 \rfloor$ are forbidden — just an $O(n^{-1/2})$ fraction of all states — then a walk starting from 1^n can never reach 0^n. Even when the subgraph induced by A is connected, the mixing time can be extremely slow: if all but one of the states of Hamming weight $\lfloor n/2 \rfloor$ are forbidden, then a walk starting from 1^n is very unlikely to reach any state of Hamming weight less than $\lfloor n/2 \rfloor$ without taking an exponential number of steps.

In the seminal work of [5], it is shown that if A is a monotone set,[1] then the mixing time of the random walk censored to A has a good mixing time, as long as A contains a constant fraction of states in the hypercube.

Theorem 1.1 ([5]). *Suppose $A \subseteq \{0,1\}^n$ is a monotone set with density $\mu(A) = |A|/2^n$. Then the mixing time of the random walk censored to A is at most $O(n^3/\mu(A)^2)$.*

Since a monotone subset of the hypercube no longer inherits a product-space structure, the naïve coupon-collector argument breaks down. As mentioned at the start of this chapter, a standard approach to proving mixing-time bounds for random walks is to establish a Sobolev-type inequality, which formalizes the notion of "good expansion" of the underlying space. Indeed, the main technical step in the proof of Theorem 1.1 of [5] is precisely the derivation of such an inequality for $\{0,1\}$-valued functions on the monotone set A. This inequality, in turn, is obtained via a "directed isoperimetric inequality" originating in the property-testing literature [14] (recall from the start of this chapter that the fields of Markov-chain analysis and property testing share an interest in Sobolev-type inequalities).

Using a completely different technique, [4] improves the mixing time bound of Theorem 1.1 to $O(n^3/\mu(A))$, eliminating a $1/\mu(A)$ factor. However, as conjectured in [5], one expects that the censored random walk on constant-density (or at least $(1 - o(1))$-density) monotone subsets should mix in time $O(n \log n)$, matching the uncensored walk on the full hypercube $\{0,1\}^n$.

Conjecture 1.2 ([5]). *Fix a constant $\varepsilon > 0$, and let $A \subseteq \{0,1\}^n$ be a monotone set of density $\mu(A) \geqslant \varepsilon$. Then the mixing time of the random walk censored to A is at most $O_\varepsilon(n \log n)$.*

In [7], we make progress toward this conjecture by establishing the following:

Theorem 1.3 ([7]). *Suppose $A \subseteq \{0,1\}^n$ is a monotone set with density $\mu(A)$. Then the mixing time of the random walk censored to A is at most $O(n^2/\mu(A))$.*

Our approach is conceptually closer to the original work of [5] than to the improved bound of [4]. Specifically, we establish a new directed L^2-Poincaré inequality on the hypercube,

[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 $x_i \leqslant y_i$ for every $i \in [n]$.

drawing again on ideas from the property testing literature [9]. To apply this new directed Poincaré inequality, we introduce a novel "approximate FKG inequality," which generalizes the classical FKG inequality on the hypercube [11].

Our work [7] also yields an optimal bound on the spectral expansion of general monotone sets $A \subseteq \{0,1\}^n$, improving upon the bounds of [4, 5] by a factor of n. Formally, we establish the following inequality for functions on A.

Theorem 1.4 ([7]). *Let $A \subseteq \{0,1\}^n$ be a non-empty monotone set. For all $f : A \to \mathbb{R}$, we have*

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

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

Here, $\mathrm{Var}_A[f]$ represents the norm of f itself, while $\mathcal{E}_A(f)$ (formally defined in Definition 2.1) can be viewed as the norm of a "derivative" of f. Theorem 1.4 is therefore a Poincaré-type inequality. In fact, the special case $\mu(A) = 1$ of Theorem 1.4 recovers exactly the classical Poincaré inequality on the hypercube (see [25, Section 2.3]).

1.2 Lower Bound for Streaming Max-Cut (Chapter 3)

One of the principal ways Sobolev-type inequalities are applied in discrete settings is through their extension to *log-Sobolev inequalities*. In many applications there is an underlying semigroup of operators (for example, a random walk defines a Markov semigroup), and log-Sobolev inequalities with respect to the generator of the semigroup are typically equivalent to a *hypercontractivity* property: the semigroup "mixes" so rapidly that any function it acts on converges quickly to a constant. Concretely, for an operator T in the semigroup, one can often show that the L^q-norm of Tf is bounded above by the L^p-norm of f, for some $q > p$.

Such results have proved extremely useful in discrete mathematics. For instance, suppose A is a small subset of a large discrete space Ω, and one wishes to show that an operator T (e.g. a random-walk operator) mixes especially quickly when acting on the indicator function $1_A : \Omega \to \{0,1\}$. This fits naturally into the framework of hypercontractivity: mixing is usually measured in the L^2-norm, but the smallness of A means that among functions with the same L^2-norm, 1_A has a particularly small L^1-norm. Therefore, a hypercontractive inequality controlling the L^2-norm of Tf by the L^1-norm of f does exactly what is needed. More generally, whenever discrete conditions such as "smallness" correspond to different norms, hypercontractivity becomes especially powerful.

Interestingly, hypercontractivity has also found applications in the area of *communication complexity*. In a communication game, several players are each given separate inputs and wish to compute a function depending jointly on all of their inputs while communicating as little as possible. From the lower-bound perspective, the goal is to show that with only a small total communication budget, the players cannot achieve a certain task regardless of the protocol used.

It turns out that the analysis of such communication protocols can sometimes be carried out in an L^1 space of functions over a discrete domain. A short message sent by a player corresponds to a function f in this L^1-space with a particularly small L^2-norm; when other

players extract useful information from the message, an "extractor" operator T acts on this function. To show that they obtain only limited information toward their goal, one needs to bound the L^q-norm of Tf for some $q > 2$. Voilà, hypercontractivity comes in handy again.

A specific problem where this type of analysis has been particularly successful is *Max-Cut*, which is the focus of the second main result of this thesis. In the associated communication game, the edge set of a graph G is partitioned into several subsets, and each player receives one subset as input. Their collective goal is to approximately compute the Max-Cut value of the graph.

When communication proceeds in a strictly *one-way* fashion — meaning there is a fixed order of players and each may speak only once in this order — the model closely corresponds to the *single-pass streaming* model (whose formal definition appears in Chapter 3). Leveraging hypercontractivity and the connection between one-way communication and single-pass streaming, the seminal work [19] established the following result:

Theorem 1.5 ([19])**.** *For any fixed $\varepsilon > 0$, any single-pass streaming algorithm that achieves a $(1/2 + \varepsilon)$-approximation of the Max-Cut value of a graph on n vertices requires space $\Omega_\varepsilon(\sqrt{n})$.*

Achieving a 1/2-approximation, on the other hand, is trivial for single-pass streaming algorithms using only $O(\log n)$ space, so Theorem 1.5 is optimal with respect to the approximation ratio. The space lower bound of $\Omega_\varepsilon(\sqrt{n})$ was subsequently strengthened by [20] to $\Omega_\varepsilon(n)$.

In the same communication setting — where each player receives a subset of the edge set — if players are allowed to communicate in arbitrary order and multiple times, communication lower bounds for such games can be translated into *multi-pass* streaming lower bounds. Exploiting this fact, but using a technique very different from [19, 20], the works [2, 3] showed that achieving a $(1 - \varepsilon)$-approximation requires space $n^{1-O(p\varepsilon)}$ for p-pass streaming algorithms. This still left open the possibility of obtaining a nontrivial $(> 1/2)$-approximation using space-efficient multi-pass algorithms. In [8], we answer this question in the negative by proving the following:

Theorem 1.6 ([8])**.** *For any fixed $\varepsilon > 0$, any p-pass streaming algorithm that achieves a $(1/2 + \varepsilon)$-approximation of the Max-Cut value of a graph on n vertices requires space $\Omega_\varepsilon\big(n^{1/3}/p\big)$.*

Our approach is much closer in spirit to the original work of [19] than to the methods of [2, 3]. In particular, we also employ hypercontractive inequalities, albeit in a different way from [19]. We establish a hypercontractive inequality that applies not to all functions on a space but to a special class of "global" functions. Such inequalities, known as *global hypercontractivity*, were first developed by [23].

1.3 Organization

In Chapter 2, we will provide an exposition of Theorems 1.3 and 1.4 from [7], while Theorem 1.6 (from [8]) is covered in Chapter 3. In Chapter 4, We conclude the thesis and states some open problems.

Chapter 2

Mixing on Monotone Sets

In this chapter, we provide an exposition of the proofs of Theorems 1.3 and 1.4 from [7]. Recall from Section 1.1 that we have a nonempty monotone set $A \subseteq \{0,1\}^n$, and our goal is to show that A has a good "expansion" property. We first formally define the energy functional $\mathcal{E}_A(\cdot)$ appearing in the statement of Theorem 1.4.

Definition 2.1. Fix a monotone set $A \subseteq \{0,1\}^n$. For all $f : A \to \mathbb{R}$, we define

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

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

Note that in the case $A = \{0,1\}^n$, the preceding definition is exactly the "total influence" [25, Definition 2.27] of the function $f : \{0,1\}^n \to \mathbb{R}$.

2.1 Markov Chain Preliminaries

In this section, we review some basic concepts of Markov chains. In particular, we present the (now standard) spectral argument in Markov chain theory, which reduces Theorem 1.3 to Theorem 1.4.

A Markov chain is a stochastic process that, at each step, depends only on its current state and not on the path taken to reach it. The specific Markov chain of interest here is the following process, which is also known as the *Glauber dynamics* on the set A.

Definition 2.2 (Censored random walk, [5])**.** 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).

A standard way to mathematically characterize Markov chains is through a matrix of transition probabilities, called a *Markov kernel*.

Definition 2.3. We let $\mathbf{P}_A : A \times A \to [0, \infty)$ be a matrix where for each pair of states $x, y \in A$, the entry $\mathbf{P}_A(x,y)$ is the probability of moving to state y in one step from state x.

2.1.1 The Spectral Viewpoint

We use $L^2(A)$ to denote the vector space of all real valued functions $f : A \to \mathbb{R}$, equipped with the inner product

$$\langle f, g\rangle := \mathop{\mathbb{E}}_{x\in A}\left[f(x)g(x)\right].$$

The Markov kernel $\mathbf{P}_A$ can also be viewed as a linear operator acting on the space $L^2(A)$.

Definition 2.4. For any function $f \in L^2(A)$, we define the image function $\boldsymbol{P}_A[f] \in L^2(A)$ by

$$\boldsymbol{P}_A[f](x) := \sum_{y\in A} \mathbf{P}_A(x, y)f(y).$$

Note that we denote this pull-back operator by the italic bold symbol $\boldsymbol{P}_A[\cdot]$, distinguishing it from the matrix expression $\mathbf{P}_A(\cdot, \cdot)$ to reflect that, while formally distinct, the two represent the same underlying Markov transition on the state space A.

The advantage of adopting the operator viewpoint is that we can now perform spectral analysis on $\boldsymbol{P}_A$. The following statement is standard and holds for all "lazy" Markov chains, which in particular includes our Definition 2.2.

Proposition 2.5. The operator $\boldsymbol{P}_A$ is self-adjoint with respect to the inner product on $L^2(A)$, and its eigenvalues lie in the range $[0, 1]$.

Proof. The self-adjointness follows from the fact that $\boldsymbol{P}_A(\cdot, \cdot)$ is a symmetric matrix. Suppose $f \in L^2(A)$ is an eigenvector of $\boldsymbol{P}_A$ with eigenvalue λ. Take an element $x \in A$ such that $|f(x)|$ attains its maximum value, and without loss of generality assume $f(x) \geqslant 0$. Then it is easy to see that

$$\boldsymbol{P}_A[f](x) = \sum_{y\in A} \mathbf{P}_A(x, y)f(y) \leqslant f(x) \sum_{y\in A} \mathbf{P}_A(x, y) \leqslant f(x).$$

Since $\mathbf{P}_A(x, x) \geqslant 1/2$ (this property is referred to as laziness of the chain), we also have

$$\boldsymbol{P}_A[f](x) = \mathbf{P}_A(x, x)f(x) + \sum_{y\in A\setminus\{x\}} \mathbf{P}_A(x, y)f(y) \geqslant f(x) \cdot \left(\mathbf{P}_A(x, x) - \sum_{y\in A\setminus\{x\}} \mathbf{P}_A(x, y)\right) \geqslant 0.$$

Therefore, due to the assumption $\boldsymbol{P}_A[f](x) = \lambda \cdot f(x)$, it follows that $0 \leqslant \lambda \leqslant 1$. □

The energy functional defined in Definition 2.1 is related to the Markov operator $\boldsymbol{P}_A$ by the following.

Proposition 2.6. For any $f \in L^2(A)$, we have $\langle f, f - \boldsymbol{P}_A[f]\rangle = \frac{1}{n} \cdot \mathcal{E}_A(f)$.

Proof. By Definition 2.2, we have

$$\begin{aligned}\langle f, f - \boldsymbol{P}_A[f]\rangle &= \mathop{\mathbb{E}}_{x\in A}\left[f(x)\left(\frac{1}{2} \cdot \mathop{\mathbb{E}}_{i\in[n]}\left[\left(f(x) - f(x^{\oplus i})\right) \cdot \mathbb{1}\left\{x^{\oplus i} \in A\right\}\right]\right)\right] \\ &= \frac{1}{4} \cdot \mathop{\mathbb{E}}_{\substack{x\in A\\ i\in[n]}}\left[f(x)\left(f(x) - f(x^{\oplus i})\right) \cdot \mathbb{1}\left\{x^{\oplus i} \in A\right\}\right] +\end{aligned}$$

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

In the second transition above, we used a switch of variables between x and $x^{\oplus i}$, the legitimacy of which relies on the fact that the uniform distribution on A is stationary under the Markov chain $\boldsymbol{P}_A$. □

2.1.2 Mixing Time

In this subsection, we introduce the concept of mixing time and explain how to deduce Theorem 1.3 from Theorem 1.4 using the perspective of Section 2.1.1. For convenience, we define an additional operator on $L^2(A)$, which we call the *averaging operator*.

Definition 2.7. The *averaging operator* $\boldsymbol{E}_A \colon L^2(A) \to L^2(A)$ is defined as follows. For any function $f \in L^2(A)$ and any $x \in A$, we let $\boldsymbol{E}_A[f](x) = \mathop{\mathbb{E}}_{y \in A}[f(y)]$.

Note that $\boldsymbol{E}_A \circ \boldsymbol{P}_A = \boldsymbol{E}_A = \boldsymbol{E}_A \circ \boldsymbol{P}_A$, which implies the following.

Proposition 2.8. For any integer $k \geqslant 1$, we have $(\boldsymbol{P}_A - \boldsymbol{E}_A)^k = \boldsymbol{P}_A^k - \boldsymbol{E}_A$.

We now define the mixing time of our Markov chain.

Definition 2.9. The *mixing time* of the Markov chain $\boldsymbol{P}_A$ is the smallest integer $t \geqslant 0$ such that, for every state $x \in A$ and every subset $S \subseteq A$, the probability that a walk started at x is in S after t steps lies in the interval

$$\left[\frac{|S|}{|A|} - \frac{1}{4}, \frac{|S|}{|A|} + \frac{1}{4} \right].$$

Equivalently, it is the smallest $t \geqslant 0$ such that

$$\left\| \boldsymbol{P}_A^t[1_S] - \boldsymbol{E}_A[1_S] \right\|_\infty \leqslant \frac{1}{4} \quad \text{for all } S \subseteq A,$$

where $1_S \colon A \to \{0, 1\}$ denotes the indicator function of S.

We are now ready to show that Theorem 1.4 implies Theorem 1.3.

Proof of Theorem 1.3 assuming Theorem 1.4. The operator $\boldsymbol{E}_A$ is precisely the kernel of the Markov chain on A that, from any state $x \in A$, transitions to a uniformly random state $y \in A$. In particular, Proposition 2.5 applies to $\boldsymbol{E}_A$. It is easy to see that $\boldsymbol{E}_A$ has only two eigenvalues, 0 and 1, with constant functions being the sole eigenvectors of eigenvalue 1. Since constant functions are also eigenvectors of $\boldsymbol{P}_A$ with eigenvalue 1, it follows that $\boldsymbol{P}_A - \boldsymbol{E}_A$ is positive semidefinite. Moreover, for any $f \in L^2(A)$,

$$\langle f, (\boldsymbol{P}_A - \boldsymbol{E}_A)[f] \rangle = \langle f, f \rangle - \frac{1}{n} \cdot \mathcal{E}_A[f] - \langle f, \boldsymbol{E}_A f \rangle = \mathrm{Var}_A[f] - \frac{1}{n} \cdot \mathcal{E}_A[f].$$

Applying Theorem 1.4 yields

$$\langle f, (\boldsymbol{P}_A - \boldsymbol{E}_A)[f] \rangle \leqslant \left(1 - \frac{1 - \sqrt{1 - \mu(A)}}{n}\right) \cdot \mathrm{Var}_A[f] \leqslant \left(1 - \frac{1 - \sqrt{1 - \mu(A)}}{n}\right) \cdot \|f\|_2^2 .$$

Thus every eigenvalue of the self-adjoint operator $\boldsymbol{P}_A - \boldsymbol{E}_A$ is at most

$$\lambda^* := 1 - \frac{1 - \sqrt{1 - \mu(A)}}{n}.$$

For some $t = O(n^2/\mu(A))$ we have $(\lambda^*)^t \leqslant 2^{-n/2-2}$. Hence, for this t and any $f \in L^2(A)$,

$$\|\boldsymbol{P}_A[f] - \boldsymbol{E}_A[f]\|_\infty \leqslant \sqrt{|A|} \cdot \|\boldsymbol{P}_A[f] - \boldsymbol{E}_A[f]\|_2 \leqslant \sqrt{|A|} \cdot (\lambda^*)^t \|f\|_2 \leqslant \frac{1}{4}\|f\|_2.$$

In particular, for every indicator function $1_S \colon A \to \{0,1\}$,

$$\left\|\boldsymbol{P}_A^t[1_S] - \boldsymbol{E}_A[1_S]\right\|_\infty \leqslant \frac{1}{4}.$$

Therefore the mixing time of the Markov chain $\boldsymbol{P}_A$ is at most $t = O\big(n^2/\mu(A)\big)$. □

2.2 Ideas from Prior Work

In this section, we discuss the main high-level ideas that led to the conception of our proof of Theorem 1.4. Along the way, we provide an overview of several prior works from which our ideas are drawn.

2.2.1 Boolean Isoperimetric Inequalities

The classical Poincaré inequality for the hypercube (see [25, Section 2.3]) states that, for any function $f : \{0,1\}^n \to \mathbb{R}$, we have

$$\mathrm{Var}_A[f] \leqslant \mathcal{E}_A(f), \quad \text{for } A = \{0,1\}^n. \tag{2.1}$$

When $A = \{0,1\}^n$, we will omit the subscript A in the functionals $\mathrm{Var}_A[\cdot]$ and $\mathcal{E}_A(\cdot)$.

If we specialize to the case where f is $\{0,1\}$-valued, the Poincaré inequality (2.1) can be interpreted as an "isoperimetric inequality" as follows. Suppose f is the indicator function of a subset $S \subseteq \{0,1\}^n$. We let $E\big(S, \{0,1\}^n \setminus S\big)$ denote the set of edges $\{x, y\}$ of the hypercube graph with $x \in S$ and $y \in \{0,1\}^n \setminus S$. It is easy to see that

$$\mathrm{Var}[1_S] = 2^{-2n} \cdot |S| \cdot |\{0,1\}^n \setminus S| \quad \text{and} \quad \mathcal{E}(1_S) = 2^{-n-1} \cdot \big|E\big(S, \{0,1\}^n \setminus S\big)\big| .$$

Therefore, (2.1) implies that

$$\big|E\big(S, \{0,1\}^n \setminus S\big)\big| \geqslant \frac{|S| \cdot |\{0,1\}^n \setminus S|}{2^{n-1}} \geqslant \min\big\{|S|, |\{0,1\}^n \setminus S|\big\}. \tag{2.2}$$

The inequality (2.2) asserts that, whenever S contains at most half of the vertices of the hypercube, the number of edges leaving S (i.e., connecting a vertex in S to one outside S) is bounded below by $|S|$. This type of statement is reminiscent of the classical isoperimetric inequality in Euclidean space, which lower bounds the surface area of a compact set in terms of its volume. Accordingly, (2.2) may be regarded as a "Boolean isoperimetric inequality."

2.2.2 A Directed Analogue

Continuing from Section 2.2.1, consider a set $S \subseteq \{0,1\}^n$. Define

$$E^-\big(S, \{0,1\}^n \setminus S\big) \subseteq E\big(S, \{0,1\}^n \setminus S\big)$$

to be the set of edges $\{x, y\}$ of the hypercube graph with $x \in S$, $y \in \{0,1\}^n \setminus S$, and $x \preceq y$. In other words, $E^-(S, \{0,1\}^n \setminus S)$ consists of the edges going "upward" from S to its complement — what we may call the *upward boundary edges* of S.

Analogously to (2.2), the size of this upward boundary can also be bounded below, though now not by the volume of S but by its "distance" from being monotone, as formalized by the following result of [14].

Theorem 2.10 ([14])**.** *For any subset $S \subseteq \{0,1\}^n$,*

$$\big|E^-\big(S, \{0,1\}^n \setminus S\big)\big| \geqslant \mathrm{dist}^{\mathsf{mono}}(S), \tag{2.3}$$

where $\mathrm{dist}^{\mathsf{mono}}(S)$ denotes the smallest symmetric difference between S and any monotone set $A \subseteq \{0,1\}^n$.

Observe that the right-hand side of (2.2), $\min\big\{|S|, |\{0,1\}^n \setminus S|\big\}$, is precisely the smallest symmetric difference between S and any "trivial" set — if we view the collection of trivial sets as consisting of the empty set and the whole hypercube $\{0,1\}^n$. Thus Theorem 2.10 indeed provides a directed analogue of the Boolean isoperimetric inequality (2.2).

Remark 2.11. The main motivation behind the directed isoperimetric inequality of [14] comes from the algorithmic task of *monotonicity testing*. A monotonicity tester has query access to a Boolean function $f : \{0,1\}^n \to \{0,1\}$, and its goal is to accept all monotone functions[1] and reject all functions f that are far from monotone; that is, for some fixed $\varepsilon > 0$,

$$\mathrm{dist}_1^{\mathsf{mono}}(f) := \min\Big\{\|f - g\|_1 \;\Big|\; g : \{0,1\}^n \to \{0,1\} \text{ is monotone}\Big\} \geqslant \varepsilon.$$

The directed isoperimetric inequality (2.3) implies that any such function f must have many *violating edges*, i.e., edges $\{x, y\}$ of the hypercube graph with $x \preceq y$ but $f(x) > f(y)$. The abundance of violating edges enables a tester to reject these functions simply by sampling random edges and checking for such violations of monotonicity.

2.2.3 From Directed to Undirected Isoperimetry: the Boolean Case

It is not hard to see that Theorem 2.10, the directed isoperimetric inequality, implies the undirected version (2.2), up to a constant factor. Indeed, for any given subset $S \subseteq \{0,1\}^n$, we may let $A \subseteq \{0,1\}^n$ be the monotone set that minimizes the size of the symmetric difference $|S \setminus A| + |A \setminus S|$. Similarly, let $B \subseteq \{0,1\}^n$ be the monotone set closest to $S^c := \{0,1\}^n \setminus S$ in terms of the symmetric difference. Then we have

$$|E(S, S^c)| = |E^-(S, S^c)| + |E^-(S^c, S)| \geqslant \mathrm{dist}^{\mathsf{mono}}(S) + \mathrm{dist}^{\mathsf{mono}}(S^c)$$

[1] A function $f : \{0,1\}^n \to \mathbb{R}$ is said to be *monotone* if $f(x) \leqslant f(y)$ for all $x, y \in \{0,1\}^n$ such that $x \preceq y$.

$$= |S \setminus A| + |A \setminus S| + |S^c \setminus B| + |B \setminus S^c|. \tag{2.4}$$

Using the classical FKG inequality [11] for the hypercube, we have

$$|A \cap B| = \frac{|A| \cdot |B|}{2^n} \geqslant \frac{(|S| - |S \setminus A|)(|S^c| - |S^c \setminus B|)}{2^n} \geqslant \frac{1}{2} \min\big\{|S|, |S^c|\big\} - |S \setminus A| - |S^c \setminus B|. \tag{2.5}$$

Finally, it is clear that

$$|A \cap B| \leqslant |S \cap S^c| + |A \setminus S| + |B \setminus S^c|. \tag{2.6}$$

Combining (2.4), (2.5) and (2.6) yields the undirected isoperimetric inequality

$$|E(S, S^c)| \geqslant \frac{1}{2} \min\big\{|S|, |S^c|\big\}.$$

To the best of the author's knowledge, the above argument was independently observed in [5] and [22, Section 9.2], each in a different generalized form. Khot, Minzer, and Safra [22] extended the argument to a broader class of (directed and undirected) Boolean isoperimetric inequalities, with a particular focus on the applications of the directed versions to monotonicity testing. Ding and Mossel [5], on the other hand, generalized the argument in terms of the *ambient space*: they showed that even when the ambient space $\{0,1\}^n$ is replaced by a *large monotone subset* of $\{0,1\}^n$, the implication from directed to undirected isoperimetry continues to hold. This observation enabled Ding and Mossel to prove the following theorem:

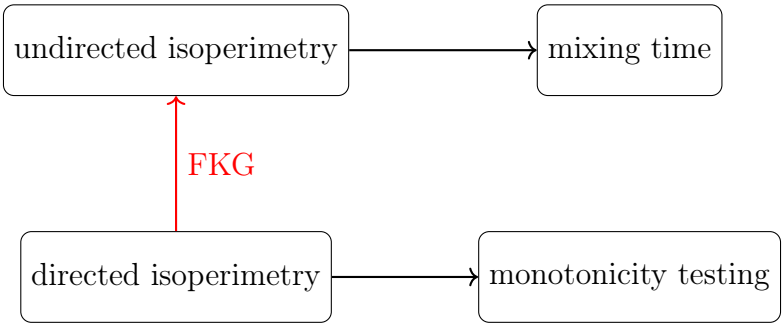
Theorem 2.12. *Suppose $A \subseteq \{0,1\}^n$ is a monotone set with density $\mu(A) = |A|/2^n$. Then for any subset $S \subseteq A$, we have*

$$|E(S, A \setminus S)| \geqslant \frac{\mu(A)}{16} \min\big\{|S|, |A \setminus S|\big\}.$$

Indeed, Theorem 2.12 is the main ingredient in the proof of Theorem 1.1 by [5]: the latter theorem follows by combining Theorem 2.12 with standard arguments in Markov chain theory (namely, the spectral viewpoint covered in Section 2.1 and Cheeger's inequality).

We note that although there are several possible approaches to proving (2.2), the route via the directed analogue (Theorem 2.10) has the distinctive advantage of remaining naturally robust under restrictions of the ambient space to monotone subsets. This robustness is precisely why the idea discussed in this subsection plays a central role in establishing mixing-time bounds for censored random walks, both in [5] and in our work [7], as will be elaborated in the next section.

The following is a conceptual diagram of the interrelation between undirected and directed isoperimetry, and their respective main applications.



2.3 From Directed to Undirected Poincaré Inequalities

Recall from Section 2.2.1 that the Boolean isoperimetric inequality (2.2) is a special case of the L^2-Poincaré inequality (2.1). In a sense, the Poincaré inequality (2.1) can be interpreted as an isoperimetric inequality for *functions* on the hypercube, rather than for subsets of it. As discussed in Section 2.2.3, for many forms of isoperimetric inequalities, a directed version can imply its undirected counterpart via an FKG-type argument, as shown in [22, Section 9.2]. This naturally raises the question of whether there exists a directed analogue of the Poincaré inequality (2.1) that implies the undirected version. Such a result could potentially allow one to extend the inequality from the entire hypercube $\{0,1\}^n$ to large monotone subsets, as in [5], thereby establishing the desired Poincaré inequality (Theorem 1.4) for these restricted domains.

A key conceptual contribution of our work [7] is that this is indeed the case: the argument of [5] used to prove Theorem 2.12 can be adapted to the L^2 setting to establish Theorem 1.4. Most of the components discussed in Section 2.2.3 admit natural analogues in the L^2 setting, as summarized in Table 2.1.

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 $-f$
$\lvert E(S, A \setminus S)\rvert$	$\mathcal{E}_A(f)$
$\min\{\lvert S\rvert, \lvert A \setminus S\rvert\}$	$\mathrm{Var}_A[f]$
directed isoperimetric inequality [14]	directed L^2-Poincaré inequality (Theorem 2.15)
classical FKG inequality [11]	approximate FKG inequality (Theorem 2.17)

Table 2.1: The analogies between the discrete and L^2 settings

To find an L^2-analogue for the directed isoperimetric inequality Theorem 2.10, 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 2.13 (Distance to monotonicity)**.** For a function $f : \{0,1\}^n \to \mathbb{R}$, we define

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

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

Definition 2.14 (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 $(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)$.

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

Theorem 2.15 (Directed Poincaré inequality)**.** *For all functions $f : \{0,1\}^n \to \mathbb{R}$, we have*

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

In contrast to the directed isoperimetric inequality, the other key ingredient listed in Table 2.1 — the FKG inequality — was originally proved by [11] already in the L^2 setting. However, for technical reasons, we cannot directly apply the classical FKG inequality as [5] did. The result of [11] 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 $f(x)$ and $g(x)$ are nonnegatively correlated. In our proof, however, we crucially need a lower bound on the correlation ratio of any two increasing functions defined only on the subset $A \subseteq \{0,1\}^n$, rather than on the entire hypercube.

Unfortunately, the nonnegative correlation in the classical FKG inequality does not necessarily hold for the restricted domain A. We therefore seek an *approximate* version of the FKG inequality, one that guarantees the correlation ratio is bounded away from -1, even if not necessarily nonnegative.

Definition 2.16 (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 2.17 (Approximate FKG inequality)**.** *For any monotone set $A \subseteq \{0,1\}^n$ with at least 2 elements, we have $\delta(A) \geqslant -\sqrt{1-\mu(A)}$.*

In the rest of this section, we develop a proof of Theorem 1.4 assuming Theorems 2.15 and 2.17, by adapting the argument of [5]. The proofs of Theorem 2.15 and Theorem 2.17 will be discussed in Sections 2.4 and 2.5, respectively. In fact, the goal of this section is to prove the following more general result assuming Theorem 2.15.

Theorem 2.18. *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*

$$\mathcal{E}_A(f) \geqslant (1 + \delta(A)) \cdot \mathrm{Var}_A[f].$$

It is clear that Theorems 2.17 and 2.18 together imply Theorem 1.4 (see Figure 2.1).

2.3.1 Domain Extension

Since the target result, Theorem 2.18, focuses solely on the subset A of the hypercube, while Theorem 2.15 applies only to functions defined on the entire hypercube, we first introduce a simple method for extending function domains to the whole hypercube.

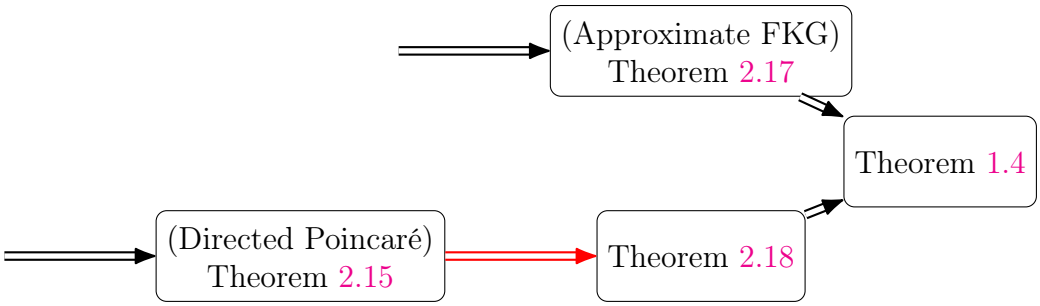


Figure 2.1: The structure of the proof of Theorem 1.4. The step indicated by the red arrow is the focus of this thesis.

Definition 2.19. 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 2.15.

Proposition 2.20. 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. Since $T[f]$ is constant on $\{0,1\}^n \setminus A$, and since $T[f](x) \leqslant T[f](y)$ for all $x \in \{0,1\}^n \setminus A$ and $y \in A$, we know that for $x \in \{0,1\}^n$ and $i \in [n]$,

$$T[f](x^{i \to 1}) < T[f](x^{i \to 0}) \quad \text{can hold only if} \quad x, x^{\oplus i} \in A.$$

So we have

$$\begin{aligned} \mathcal{E}^-(T[f]) &= \frac{1}{4} \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\sum_{i=1}^n \min \left\{ 0, T[f](x^{i \to 1}) - T[f](x^{i \to 0}) \right\}^2 \right] \\ &= \frac{1}{4} \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\sum_{i=1}^n \min \left\{ 0, T[f](x^{i \to 1}) - T[f](x^{i \to 0}) \right\}^2 \cdot \mathbb{1} \left\{ x, x^{\oplus i} \in A \right\} \right] \\ &= \frac{\mu(A)}{4} \cdot \mathop{\mathbb{E}}_{x \in A} \left[\sum_{i=1}^n \min \left\{ 0, f(x^{i \to 1}) - f(x^{i \to 0}) \right\}^2 \cdot \mathbb{1} \left\{ x^{\oplus i} \in A \right\} \right] \end{aligned}$$

Applying the above argument to $T[-f]$ instead of $T[f]$, we obtain

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

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

Adding the above two equations together yields

$$\begin{aligned}\mathcal{E}^-(T[f]) + \mathcal{E}^-(T[-f]) &= \frac{\mu(A)}{4} \cdot \mathop{\mathbb{E}}_{x \in A} \left[\sum_{i=1}^{n} \left(f(x^{i\to 1}) - f(x^{i\to 0})\right)^2 \cdot \mathbb{1}\left\{x^{\oplus i} \in A\right\} \right] \\ &= \mu(A) \cdot \mathcal{E}_A(f). \end{aligned}$$

□

Proposition 2.21. For every function $f : A \to \mathbb{R}$, there exists a monotone increasing function $g : A \to \mathbb{R}$ such that

$$\|f - g\|_2 \leqslant \mu(A)^{-1/2} \cdot \mathrm{dist}_2^{\mathsf{mono}}(T[f]),$$

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 $\|T[f] - \widetilde{g}\|_2 = \mathrm{dist}_2^{\mathsf{mono}}(T[f])$, 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,

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

□

2.3.2 Correlation Analysis

In this subsection, we lay some groundwork about correlation of functions (or equivalently, random variables) that will help prove Theorem 2.18. We begin with the following natural definition of correlation ratios.

Definition 2.22. 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 2.18. 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 2.23. Consider three non-constant functions $f, g, h : A \to \mathbb{R}$. We have

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

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 $\max\{0,\rho(f,h)\}^2 \leqslant 1$. Similarly if $\rho(f,h) < 0$, the conclusion is also trivial. In the following, we assume that $\rho(f,g) \geqslant 0$ and $\rho(f,h) \geqslant 0$.

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 $\boldsymbol{\lambda}^T B \boldsymbol{\lambda} = \mathrm{Var}_A[\lambda_1 f + \lambda_2 g + \lambda_3 h] \geqslant 0$. So B is a positive semi-definite matrix. This means $\det B \geqslant 0$, and we can expand it into

$$1 + 2\rho(f,g)\rho(f,h)\rho(g,h) \geqslant \rho(f,g)^2 + \rho(f,h)^2 + \rho(g,h)^2. \tag{2.7}$$

If $\rho(g,h) < 0$, then (2.7) implies $1 \geqslant \rho(f,g)^2 + \rho(f,h)^2$ and we arrive at the conclusion. In the following we assume $\rho(g,h) \geqslant 0$.

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

$$2 + 2\rho(g,h) \geqslant (\rho(f,g) + \rho(f,h))^2 \geqslant 4\rho(f,g)\rho(f,h). \tag{2.8}$$

Multiplying both sides of (2.8) by $\rho(g,h)/2$ and then adding it to (2.7), we get the desired conclusion

$$1 + \rho(g,h) \geqslant \rho(f,g)^2 + \rho(f,h)^2. \qquad \square$$

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

Definition 2.24. For functions $f, g : A \to \mathbb{R}$, we define

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

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

Proposition 2.25. 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

$$\begin{aligned} \tau(f,g)^2 &= \min_{a \in \mathbb{R}_{\geqslant 0}, b \in \mathbb{R}} \|f - (ag+b)\|_2^2 = \min_{a \in \mathbb{R}_{\geqslant 0}} \mathrm{Var}_A[f - ag] \\ &= \min_{a \in \mathbb{R}_{\geqslant 0}} \left(a^2 \cdot \mathrm{Var}_A[g] - 2a \cdot \mathrm{Cov}_A[f,g] + \mathrm{Var}_A[f]\right). \end{aligned} \tag{2.9}$$

If $\rho(f,g) < 0$, then $\mathrm{Cov}_A[f,g] < 0$, and the quadratic polynomial in the right hand side of (2.9) is minimized at $a = 0$. Therefore $\tau(f,g)^2 = \mathrm{Var}_A[f]$, as desired.

If $\rho(f,g) \geqslant 0$, then $\mathrm{Cov}_A[f,g] \geqslant 0$, and the quadratic polynomial in the right hand side of (2.9) is minimized at $a = \mathrm{Cov}_A[f,g] / \mathrm{Var}_A[g]$. Therefore (2.9) 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. $\square$

2.3.3 Proof of Theorem 2.18

We are now ready to prove Theorem 2.18 assuming Theorem 2.15.

Proof of Theorem 2.18 assuming Theorem 2.15. We have

$$\begin{aligned}
\mathcal{E}_A(f) &= \mu(A)^{-1} \cdot \mathcal{E}^-(T[f]) + \mu(A)^{-1} \cdot \mathcal{E}^-(T[-f]) && \text{(by Proposition 2.20)} \\
&\geqslant \mu(A)^{-1} \cdot \mathrm{dist}_2^{\mathsf{mono}}(T[f])^2 + \mu(A)^{-1} \cdot \mathrm{dist}_2^{\mathsf{mono}}(T[-f])^2 && \text{(by Theorem 2.15)} \\
&\geqslant \|f - g_0\|_2^2 + \|-f - h_0\|_2^2 && \text{(by Proposition 2.21),}
\end{aligned} \tag{2.10}$$

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

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

Similarly we pick a non-constant increasing function $h : A \to \mathbb{R}$ such that $\|-f - h_0\|_2^2 \geqslant \tau(-f, h)^2$. We can then continue from (2.10) and have

$$\begin{aligned}
\mathcal{E}_A(f) &\geqslant \tau(f, g)^2 + \tau(-f, h)^2 \\
&= \Big(1 - \max\{0, \rho(f, g)\}^2\Big) \cdot \mathrm{Var}_A[f] + \\
&\qquad \Big(1 - \max\{0, \rho(-f, h)\}^2\Big) \cdot \mathrm{Var}_A[f] && \text{(by Proposition 2.25)} \\
&= \Big(2 - \max\{0, \rho(f, g)\}^2 - \max\{0, \rho(f, -h)\}^2\Big) \cdot \mathrm{Var}_A[f] \\
&\geqslant \Big(1 - \max\{0, \rho(g, -h)\}\Big) \cdot \mathrm{Var}_A[f] && \text{(by Proposition 2.23)} \\
&= \Big(1 + \min\{0, \rho(g, h)\}\Big) \cdot \mathrm{Var}_A[f] \geqslant (1 + \delta(A)) \cdot \mathrm{Var}_A[f] && \text{(by Definition 2.16).}
\end{aligned}$$

□

2.4 The Directed Poincaré Inequality

In this section, we outline the proof of Theorem 2.15. Readers interested in full details are referred to the paper [7] or the thesis [10] by Ferreira Pinto Jr.

The proof of Theorem 2.15 is based on analyzing the energy functional defined in Definition 2.14 via a *continuous-time* dynamical system that we call the *directed heat process*. The (undirected) heat process is a classical concept originating from mathematical physics and has long been known to be deeply connected to Poincaré-type inequalities, dating back to Poincaré's original work in the 19th century. In particular, the classical Poincaré inequality on the hypercube can be interpreted as a lower bound on the convergence rate of the heat process for functions on the hypercube (see, e.g., [16, Chapter 2]). In [7], an analogous connection is established between the directed heat process and the directed Poincaré inequality. The proof of Theorem 2.15 thus proceeds by analyzing the convergence behavior of this directed heat process.

Remark 2.26. The idea of using the directed heat process to prove directed isoperimetric inequalities originates from Ferreira Pinto Jr.'s work [9]. The work [9] develops a directed Poincaré inequality for functions over the solid hypercube $[0,1]^n$, motivated by questions in monotonicity testing.

2.4.1 Directed Laplacian and Energy Functional

Recall from Definition 2.14 that $\mathcal{E}^-(\cdot)$ is a real-valued functional defined on the space $L^2(\{0,1\}^n)$. The directed heat process can be naturally interpreted as the gradient descent dynamics associated with this functional: a "particle" in $L^2(\{0,1\}^n)$ evolves by moving in the direction of the negative gradient of $\mathcal{E}^-(\cdot)$ at its current position. Note that the functional $\mathcal{E}^-(\cdot)$ is minimized exactly at the collection of monotone functions in $L^2(\{0,1\}^n)$. Therefore, the gradient descent dynamics could potentially converge any function $f \in L^2(\{0,1\}^n)$ to a limit monotone function, thus providing a candidate monotone function which we may argue that f is close to, as required in Theorem 2.15.

The following operator plays the role of the (negative) gradient of $\mathcal{E}^-(\cdot)$.

Definition 2.27. For any function $f : \{0,1\}^n \to \mathbb{R}$, we define the function $\boldsymbol{L}^-[f] : \{0,1\}^n \to \mathbb{R}$ by letting

$$\boldsymbol{L}^-[f](x) := \frac{1}{2}\sum_{i=1}^{n}\Big(\big(f(x^{\oplus i}) - f(x)\big)^+ \cdot \mathbb{1}\{x_i = 1\} - \big(f(x) - f(x^{\oplus i})\big)^+ \cdot \mathbb{1}\{x_i = 0\}\Big).$$

In this thesis, we use the notation $x^+ := \max\{x, 0\}$, for $x \in \mathbb{R}$.

Proposition 2.28. The functional $\mathcal{E}^- : L^2(\{0,1\}^n) \to \mathbb{R}$ is continuously differentiable and convex. Furthermore, its gradient operator is $-2^{1-n}\boldsymbol{L}^-[\cdot]$.

Note that $L^2(\{0,1\}^n)$ is a finite-dimensional Euclidean space, so we may use the notion of gradient from standard multi-variable calculus.

In fact, the directed energy functional $\mathcal{E}^-$ can be expressed in terms of the directed Laplacian in the following way.

Proposition 2.29. For any $f \in L^2(\{0,1\}^n)$, we have $\mathcal{E}^-(f) = -\langle f, \boldsymbol{L}^-[f]\rangle$.

Remark 2.30. If $\boldsymbol{L}^-$ were linear, self-adjoint and negative semi-definite, then Proposition 2.28 would directly follow from 2.29. Although the directed Laplacian operator $\boldsymbol{L}^-$ we consider here is not a linear operator, it is *piecewise* linear, self-adjoint and negative semi-definite, so Proposition 2.29 still implies Proposition 2.28.

2.4.2 The Directed Heat Process

Due to the Lipschitz continuity of the operator $\boldsymbol{L}^-$, standard theory of ordinary differential equations implies the following claim.

Proposition 2.31. There exists a unique family of continuous operators $P_t : L^2(\{0,1\}^n) \to L^2(\{0,1\}^n)$, for $t \in [0, +\infty)$, such that the following hold:

(1) For any $f \in L^2(\{0,1\}^n)$, we have $P_0 f = f$.

(2) For any $f \in L^2(\{0,1\}^n)$, the map $t \mapsto P_t f$ is a differentiable map from $[0,\infty)$ to $L^2(\{0,1\}^n)$, and its derivative at each $s \in [0,\infty)$ equals $\boldsymbol{L}^-[P_s f]$.

The dynamical process $(P_t f)_{t \geqslant 0}$ specified by Proposition 2.31 is called the directed heat process starting from f. To analyze the convergence of this process, we prove the following key lemma.

Lemma 2.32. *For any $f \in L^2(\{0,1\}^n)$ we have*

$$\left\|\boldsymbol{L}^-[f]\right\|_2^2 \geqslant \mathcal{E}^-(f).$$

Note that due to Proposition 2.28, the directed heat process always moves a function $f \in L^2(\{0,1\}^n)$ along the direction that (greedily) minimizes the directed energy value $\mathcal{E}^-(f)$, and hence it may be called a "gradient descent process." Lemma 2.32 then serves to *quantify* the speed of this directed energy decay: it gives a lower bound on the magnitude of velocity vectors of the directed heat process. Furthermore, the bound in Lemma 2.32 is good enough to show that the directed energy decays at an exponential rate. Analyzing the dynamical process from this perspective, we are able to deduce the following conclusion.

Corollary 2.33. *For each $f \in L^2(\{0,1\}^n)$, there exists a monotone function $P_\infty f : \{0,1\}^n \to \mathbb{R}$ such that $P_\infty f = \lim_{t \to +\infty} P_t f$. Furthermore, we have*

$$\|f - P_\infty f\|_2 \leqslant \int_0^{+\infty} \left\|\boldsymbol{L}^-[P_t f]\right\|_2 \mathrm{d}t = \int_0^{+\infty} \sqrt{-\frac{1}{2} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{E}^-(P_t f)}\, \mathrm{d}t \leqslant \sqrt{\mathcal{E}^-(f)}. \tag{2.11}$$

Note that Theorem 2.15 follows directly from Corollary 2.33. We also note that the first two transitions in (2.11) are direct consequences of Propositions 2.28 and 2.31, while the third transition is due to differential inequality

$$-\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{E}^-(P_t f) \geqslant 2\mathcal{E}^-(P_t f),$$

which in turn is a consequence of Lemma 2.32.

2.5 The Approximate FKG Inequality

In this subsection, we outline the proof of the approximate FKG inequality (Theorem 2.17). Readers interested in full details are referred to [7, Section 2].

We first note that the case where the functions take values in $\{0,1\}$ is easy. Indeed, we have the following simple lemma.

Lemma 2.34. *Assume that $f, g : A \to \{0,1\}$ are monotone increasing functions. Then we have $\mathop{\mathbb{E}}_{x \in A}[f(x)g(x)] \geqslant \mu(A) \cdot \mathop{\mathbb{E}}_{x \in A}[f(x)] \cdot \mathop{\mathbb{E}}_{x \in A}[g(x)]$.*

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 [11] we know that $\mu(B \cap C) \geqslant \mu(B) \cdot \mu(C)$. Therefore,

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

It is straightforward to deduce from Lemma 2.34 that for monotone increasing functions $f, g : A \to \{0,1\}$, the desired approximate FKG inequality

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

holds.

The main challenge in Theorem 2.17 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 2.35. *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 $a, b \in \mathbb{R}$,*

$$\mathbb{P}[X \geqslant a, Y \geqslant b] \geqslant c \cdot \mathbb{P}[X \geqslant a] \cdot \mathbb{P}[Y \geqslant b], \tag{2.12}$$

then we must have

$$\mathrm{Cov}[X,Y] \geqslant -\sqrt{(1-c) \cdot \mathrm{Var}[X] \cdot \mathrm{Var}[Y]}. \tag{2.13}$$

Proof of Theorem 2.17 assuming Theorem 2.35. 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

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

since they are clearly monotone increasing 0/1-valued functions, we can apply Lemma 2.34 to f_a and g_b to deduce that

$$\mathbb{P}[X \geqslant a, Y \geqslant b] \geqslant \mu(A) \cdot \mathbb{P}[X \geqslant a] \cdot \mathbb{P}[Y \geqslant b].$$

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 2.35 to the random variable pair (X,Y) with constant $c = \mu(A)$, which yields exactly the desired conclusion. $\square$

The proof of Theorem 2.35 turns out to be surprisingly nontrivial. To illustrate the complexity behind this inequality, we note that equality in (2.13) holds for a wide range of joint distributions of (X,Y) beyond the case captured by Lemma 2.34, i.e. where X and Y take only two possible values.

Example 2.36. Let (X, Y) follow a discrete distribution supported on the grid $\{0, 2, 3\}^2$. Specifically, let

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

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

It is easy to check that (2.12) holds for $c = 45/49$ and all $a, b \in \mathbb{R}$. On the other hand, we have $\operatorname{Cov}[X, Y] = -8/45$ and $\operatorname{Var}[X] = \operatorname{Var}[Y] = 28/45$, so equality in (2.13) holds for $c = 45/49$ as well.

A key challenge in Theorem 2.35 lies in its lack of "centrosymmetry" with respect to (X, Y). While the assumption (2.12) does not remain invariant under the substitution $X \mapsto -X$ and $Y \mapsto -Y$, the conclusion is unaffected by such substitutions. The first step in the proof of Theorem 2.35 is to extract the "symmetric information" inherent in the condition (2.12). In the second step, we use the Cauchy-Schwarz inequality in a careful way that takes in to account all equality cases similar to Example 2.36. The main techniques in the second step are "constructive" applications of the Fubini-Tonelli theorem (i.e. exchanging the order of integration), inspired by the author's previous work [6].

Chapter 3

Max-Cut in Multi-Pass Streaming

In this chapter, we provide an exposition of the proof of Theorem 1.6 from [8]. We first define the "Max-Cut value" of graphs.

Definition 3.1. Given a finite vertex set V and a multi-set E of non-self-loop edges on V, the *Max-Cut value* of the graph $G = (V, E)$ is defined to be

$$\mathrm{MaxCut}(G) := \min_{x \in \mathbb{F}_2^V} \frac{1}{|E|} \sum_{\{u,v\} \in E} \mathbb{1}\left\{x_u + x_v = 1\right\}.$$

In words, the Max-Cut value is the maximum fraction of edges cut by any bipartition of the vertex set. We consider the complexity of the MaxCut problem under the model of streaming algorithms, formally defined as follows.

Definition 3.2. A *deterministic space-S streaming algorithm* for graphs over the vertex set V is specified by:

- a transition function $\mathcal{T} : \{0,1\}^S \times \binom{V}{2} \to \{0,1\}^S$, and
- an output function $\mathcal{O} : \{0,1\}^S \to [0,1]$.

When the algorithm reads an edge $\{u, v\} \in \binom{V}{2}$ while in memory state $z \in \{0,1\}^S$, it updates its memory to $\mathcal{T}(z, \{u, v\})$. After processing all constraints in the input stream and reaching a final memory state $z \in \{0,1\}^S$, it outputs the value $\mathcal{O}(z)$.

Definition 3.3. A *randomized space-S streaming algorithm* is a probability distribution over deterministic space-S streaming algorithms.

In this thesis, streaming algorithms are always assumed to be randomized, if not stated otherwise.

Definition 3.4. Let p be a positive integer. Given an input graph (V, E) whose edges are presented as an ordered list $(e_1, e_2, \ldots, e_m)$, a *p-pass streaming algorithm* sequentially scans the edges in the given order p times. Each scan over the entire edge list is referred to as a *pass*.

We are now ready to state the formal version of Theorem 1.6.

Theorem 3.5 ([8]). *For any fixed constant $\varepsilon > 0$, there is a constant integer $K > 0$ such that the following holds. Let n, p, S be positive integers. Suppose $\mathcal{A}$ is a p-pass space-S streaming algorithm such that given any input graph $G = ([n], E)$ with $|E| \leqslant Kn$ and any ordering of the edge set E, it achieves the following:*

(1) If $\mathrm{MaxCut}(G) = 1$, then the algorithm accepts with probability at least $2/3$.

(2) If $\mathrm{MaxCut}(G) \leqslant \frac{1}{2} + \varepsilon$, then the algorithm rejects with probability at least $2/3$.

Then we must have $pS \geqslant \Omega_\varepsilon(n^{1/3})$.

3.1 The Communication Complexity Approach

The standard approach to proving space lower bounds for streaming algorithms is through the framework of communication complexity. In this section, we introduce the communication game used in the proof of Theorem 3.5. The game itself is only a minor adaptation from prior works such as [19, 20], but we will present it in a way that best suits the subsequent analysis in our paper [8].

3.1.1 Labeled Matchings

Labeled matchings are combinatorial objects that play a central role in previous lower bounds for streaming Max-Cut [19, 20]. Compared to these prior works, our work [8] places a more significant emphasis on the *space* of labeled matchings and the Fourier analytic properties of this space.

Definition 3.6. Fix a ground set U and an integer $m \leqslant |U|/2$. An element $y \in \{-1, 0, 1\}^{\binom{U}{2}}$ is said to be a *labeled matching* over U if the edge set $\mathrm{supp}(y) := \left\{\{u, v\} \in \binom{U}{2} : y_{\{u,v\}} \neq 0\right\}$ consists of vertex disjoint edges. In that case, we think of the support of y as a graph, and of the label of an edge $\{u, v\}$ as $y_{\{u,v\}}$.

Definition 3.7. The space of labeled matchings, denoted by $\Omega^{U,m} \subseteq \{-1, 0, 1\}^{\binom{U}{2}}$, is defined as

$$\Omega^{U,m} := \left\{ y \in \{-1, 0, 1\}^{\binom{U}{2}} : \mathrm{supp}(y) \text{ is a matching with } m \text{ edges} \right\}.$$

Throughout this chapter we will mostly consider the case the ground set U is $[n]$, and the matching size m is αn where $\alpha > 0$ is a small absolute constant. When both the ground set and the matching size are clear from context, we often abbreviate notations and write Ω instead of $\Omega^{[n],\alpha n}$. In particular, we denote by Ω^K the Cartesian product of K copies of Ω.

3.1.2 The Communication Game (DIHP)

The communication game we will define is called the *Distributional Implicit Hidden Partition (DIHP)* problem. In this game there are K players, each receiving a labeled matching from $\Omega^{[n],\alpha n}$ as an input. Their goal is to be able to tell if the labels of their matchings are consistent, in the sense that there is a bipartition of the vertex set U so that edges that cross

this bipartition are labeled by -1, and edges that stay within one side of the bipartition are labeled by 1. We stress that each player only gets to see the edges they received, so while that player can find bipartitions of the vertices consistent with their edges, the challenge here is that the bipartition should be consistent with the edges of other players as well.

To prove that this communication problem is hard, [20] introduce two distributions over the inputs of the players. In the YES distribution, all of the labels follow from one common bipartition of the vertices, whereas in the NO distribution each matching is labeled independently. Formally, we define these distributions as follows:

Definition 3.8. Let $K \in \mathbb{N}$ be a constant. Define two distributions $\mathcal{D}_{\text{yes}}$ and $\mathcal{D}_{\text{no}}$ over Ω^K:

1. The *no distribution*: define $\mathcal{D}_{\text{no}}$ to be the uniform distribution over Ω^K.
2. The *yes distribution*: sample a uniformly random vector $x \in \mathbb{F}_2^n$, then independently and uniformly sample K matchings $M^{(1)}, M^{(2)}, \ldots, M^{(K)}$ of size αn. For each $i \in [K]$, we let $y^{(i)} \in \Omega$ have support $\text{supp}(y^{(i)}) = M^{(i)}$ and be defined as $y^{(i)}_{uv} = (-1)^{x_u + x_v}$ for $\{u, v\} \in M^{(i)}$. We define $\mathcal{D}_{\text{yes}}$ to be the joint distribution of $(y^{(1)}, \ldots, y^{(K)})$ obtained by this procedure.

The work [20] shows an $\Omega(n)$ communication lower bound for this problem for protocols in which player 1 sends a message to player 2, which sends a message to player 3 and so on until its player K turn to speak. In that step, player K should decide whether to accept or reject. These type of restricted protocols suffice for establishing single-pass lower bounds, and are typically easier to prove.

A subtle difference between the communication problem we consider here and the one used in [20], is that in our case each player only knows their matching. In contrast, in the setting of [20], player i knows all of the matchings $M^{(1)}, \ldots, M^{(i)}$, and only the labels are private. This does not matter for [20], as their setting is specifically designed to facilitate their single-pass analysis. This distinction is crucial in the multi-pass setting, though, as without it the problem is no longer hard. The multi-pass setting corresponds to protocols similar to the above, except that in the end, player K sends a message to player 1, and then the protocol continues in the same way. Note that with high probability the matchings $M^{(1)}, \ldots, M^{(K)}$ will not be edge-disjoint. Hence, if the matchings were public, the last player to act could announce a common edge to two of the matchings, and in the second pass the two corresponding players could broadcast their labels of that edge and compare it. In the no distribution, with probability $1/2$ the labels would not match each other.

In light of this, we will consider the version of this problem wherein both the matchings and the labels are private.

Definition 3.9. We define $\mathsf{DIHP}(n, \alpha, K)$ to be the K-player communication game in the number-in-hand (NIH) communication model where the i-th player gets as private input $y^{(i)} \in \Omega$, and their goal is to decide whether $(y^{(1)}, \ldots, y^{(K)})$ comes from $\mathcal{D}_{\text{yes}}$ or $\mathcal{D}_{\text{no}}$. For a protocol Π, we define its *advantage* to be

$$\text{adv}(\Pi) := \left| \mathop{\mathbb{P}}_{y^{(1)}, \ldots, y^{(K)} \sim \mathcal{D}_{\text{yes}}} \left[\Pi\left(y^{(1)}, \ldots, y^{(K)}\right) = 1 \right] - \mathop{\mathbb{P}}_{y^{(1)}, \ldots, y^{(K)} \sim \mathcal{D}_{\text{no}}} \left[\Pi\left(y^{(1)}, \ldots, y^{(K)}\right) = 1 \right] \right|.$$

Our main result regarding the $\mathsf{DIHP}(n, \alpha, K)$ is the following communication complexity lower bound:

Theorem 3.10 ([8]). *Let $\alpha \in (0, 10^{-7}]$ be a constant. Any communication protocol Π for $\mathsf{DIHP}(n, \alpha, K)$ with advantage at least 0.1 requires $\Omega\left(n^{1/3}K^{-2}\right)$ bits of communication*[1].

3.1.3 Streaming Lower Bound from Communication Complexity

We next show that Theorem 3.10 implies Theorem 3.5. In the analysis, we will need the following martingale concentration inequality.

Proposition 3.11 ([20, Lemma 2.5]). Let $p \in (0, 1)$ and let $X_1, \ldots, X_n$ be Bernoulli random variables such that for every $i \in [n]$, $\mathbb{E}\left[X_i \mid X_1, \ldots, X_{i-1}\right] \leqslant p$. For any $\Delta > 0$, we have

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geqslant \varepsilon + \sum_{i=1}^{n} p_i\right] \leqslant \exp\left(-\frac{\varepsilon^2}{2\varepsilon + 2\sum_{i=1}^{n} p_i}\right).$$

We are now ready to deduce Theorem 3.5 from Theorem 3.10.

Proof of Theorem 3.5 assuming Theorem 3.10. Assume without loss of generality that $\varepsilon < 1/10$. Let $\alpha \in (0, 10^{-7}]$ and $K \in \mathbb{N}$ be constants depending on ε that are to be determined later. We divide the proof into the following steps.

Step 1: the reduction map. Recall that in the $\mathsf{DIHP}(n, \alpha, K)$ game, the i-th player has a labeled matching $y^{(i)}$ in hand. In this step, we show how to map a joint input $Y = (y^{(1)}, \ldots, y^{(K)}) \in \Omega^K$ in the communication game to a graph $G_Y = ([n], E_Y)$. The construction of G_Y is as follows.

(1) For each $i \in [K]$, let M_i be the sub-matching of $\mathrm{supp}(y^{(i)})$ consisting of all edges $\{u, v\} \in \mathrm{supp}(y^{(i)})$ such that $y^{(i)}_{uv} = -1$. We let E_Y be the multi-set union of the M_i's, for $i \in [K]$.

(2) Note that as G_Y is meant to be fed to a hypothetical streaming algorithm, we also need to specify the order in which the edges in E_Y appear in the stream. This is straightforward: we fix an arbitrary ordering of edges in each sub-matching M_i, and concatenate the edge sequences M_i with respect to natural order on $[K]$.

Note that we always have $|E_Y| \leqslant Kn$, so G_Y is a valid input graph in the context of Theorem 3.5.

[1] The communication complexity is the total number of bits broadcast by all players in all rounds during the communication game.

Step 2: reduction justification. It is easy to see that a multi-pass streaming algorithm taking input G_Y can be translated back to a communication protocol for $\mathsf{DIHP}(n, \alpha, K)$: in any pass whenever the streaming algorithm finishes processing a sub-matching M_i, the i-th player in the communication game correspondingly broadcasts the memory state. In this way, any p-pass streaming algorithm $\mathcal{A}$ that achieves

$$\left| \mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{yes}}} [\mathcal{A}(G_Y) = 1] - \mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{no}}} [\mathcal{A}(G_Y) = 1] \right| \geqslant 0.1 \tag{3.1}$$

using S bits of memory implies a communication protocol Π for $\mathsf{DIHP}(n, \alpha, K)$ with $\mathrm{adv}(\Pi) \geqslant 0.1$ using pKS total bits of communication. We thus conclude from Theorem 3.10 that any p-pass streaming algorithm that achieves (3.1) must use at least $(pK)^{-1} \cdot \Omega_\alpha(n^{1/3}/K^2) = \Omega(n^{1/3}/p)$ bits of memory.

The next step is to show the completeness and soundness of the reduction. For the completeness, it is clear that when Y is sampled from the support of $\mathcal{D}_{\text{yes}}$, the graph G_Y is bipartite and thus has a Max-Cut value of 1. It remains to show the soundness guarantee

$$\mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{no}}} \left[\mathrm{MaxCut}(G_Y) \leqslant \frac{1}{2} + \varepsilon \right] \geqslant 1 - o(1), \tag{3.2}$$

where $o(1)$ denotes a term that converges to 0 as $n \to +\infty$. Given (3.2), it would follow that any p-pass streaming algorithm $\mathcal{A}$ that satisfies the description in Theorem 3.5 must achieve (3.1), and thus have memory size at least $\Omega(n^{1/3}/p)$, as desired.

Step 3: soundness. In order to upper bound $\mathrm{MaxCut}(G_Y)$ with high probability, we upper bound the value

$$\mathrm{Cut}(x, Y) := \frac{1}{|E_Y|} \sum_{\{u,v\} \in E_Y} \mathbb{1}\{x_u + x_v = 1\}$$

for any fixed bipartition $x \in \mathbb{F}_2^n$ with high probability over the random joint input $Y \sim \mathcal{D}_{\text{no}}$.

It is clear that the random variable $|E_Y|$ is the sum of $K \cdot \alpha n$ independent uniform Bernoulli random variables, so we can apply Proposition 3.11 and get

$$\mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{no}}} \left[|E_Y| \leqslant \left(\frac{1}{2} - \frac{\varepsilon}{4} \right) K\alpha n \right] \leqslant \exp\left(-\frac{K\alpha n}{64} \cdot \varepsilon^2 \right). \tag{3.3}$$

We can think of each random matching $\mathrm{supp}(y^{(i)})$ as the result of a sequential random selection of edges in $\binom{[n]}{2}$, without replacement of vertices. In the t-th selection step of the sequential random selection process, the number of available vertices at each selection step is $u_t = n - 2(t-1) \geqslant n/2$. Therefore, conditioned on the previously selected $(t-1)$ edges, the probability that the t-th selected edge $\{u, v\}$ satisfies $x_u + x_v = 1$ is at most

$$\frac{u_t^2/4}{\binom{u_t}{2}} = \frac{u_t}{2(u_t - 1)} \leqslant \frac{1}{2} + \frac{2}{n}.$$

We consider the combined process of selecting the K random matchings $\mathrm{supp}(y^{(1)}), \ldots, \mathrm{supp}(y^{(K)})$. This is a process that has $K \cdot \alpha n$ selection steps. Applying Proposition 3.11

to this process, we deduce that with probability at most $\exp(-K\alpha n \cdot \varepsilon^2/64)$ over the joint input $Y = (y^{(1)}, \ldots, y^{(K)}) \sim \mathcal{D}_{\text{no}}$ we have

$$\sum_{i=1}^{K} \sum_{\{u,v\} \in \text{supp}(y^{(i)})} \mathbb{1}\left\{y_{uv} = -1 \text{ and } x_u + x_v = 1\right\} \geqslant \left(\frac{1}{4} + \frac{1}{n} + \frac{\varepsilon}{4}\right) K\alpha n.$$

In other words, we have

$$\mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{no}}} \left[\sum_{\{u,v\} \in E_Y} \mathbb{1}\left\{x_u + x_v = 1\right\} \geqslant \left(\frac{1}{4} + \frac{1}{n} + \frac{\varepsilon}{4}\right) K\alpha n \right] \leqslant \exp\left(-\frac{K\alpha n}{64} \cdot \varepsilon^2\right). \tag{3.4}$$

For sufficiently large n, combining (3.3) and (3.4) yields

$$\mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{no}}} \left[\text{Cut}(x, Y) \geqslant \frac{1}{2} + \varepsilon \right] \leqslant 2 \exp\left(-\frac{K\alpha n}{64} \cdot \varepsilon^2\right).$$

Taking a union bound over all $x \in \mathbb{F}_2^n$, we conclude that

$$\mathop{\mathbb{P}}_{Y \sim \mathcal{D}_{\text{no}}} \left[\text{MaxCut}(G_Y) \geqslant \frac{1}{2} + \varepsilon \right] \leqslant 2 \cdot 2^n \exp\left(-\frac{K\alpha n}{64} \cdot \varepsilon^2\right) = \exp(-\Omega(n)),$$

as long as $K \geqslant 100\alpha^{-1}\varepsilon^{-2}$. This proves (3.2) and hence the whole theorem. □

3.2 Main Ideas

In this section, we provide an exposition of the main ideas behind the proof of Theorem 3.10. Sections 3.2.1 to 3.2.3 discuss the main ideas in the prior work [19], but from our perspective. We then explain how these ideas are extended by our work [8] to prove Theorem 3.10, in Sections 3.2.4 and 3.2.5.

3.2.1 The Markov Kernel

Recall from Definition 3.8 that in the YES distribution $\mathcal{D}_{\text{yes}}$, each labeled matching $y^{(i)} \in \Omega^{[n],\alpha n}$ is generated based on an underlying bipartition $x \in \mathbb{F}_2^n$. We formalize this process of generating $y^{(i)}$ from x as a Markov transition from the space $\mathbb{F}_2^n$ to the space $\Omega^{[n],\alpha n}$.

Definition 3.12. Fix a ground set U and an integer $m \leqslant |U|/2$. We define a right stochastic matrix $\mathbf{P}^{U,m} : \mathbb{F}_2^U \times \Omega^{U,m} \to [0, \infty)$ as follows. For each $x \in \mathbb{F}_2^U$ and $y \in \Omega^{U,m}$, we let

$$\mathbf{P}^{U,m}(x, y) := \frac{1}{|\mathcal{M}_{U,m}|} \cdot \mathbb{1}\left\{y_{uv} = (-1)^{x_u + x_v} \text{ for all } \{u, v\} \in \text{supp}(y)\right\},$$

where $\mathcal{M}_{U,m}$ denotes the collection of all matchings over U with exactly m edges.

Notice the similarity of the above definition with Definition 2.3. We may now define a pull-back operator from $L^2(\Omega^{U,m})$ to $L^2(\mathbb{F}_2^U)$, analogous to what we did in Section 2.1.1.

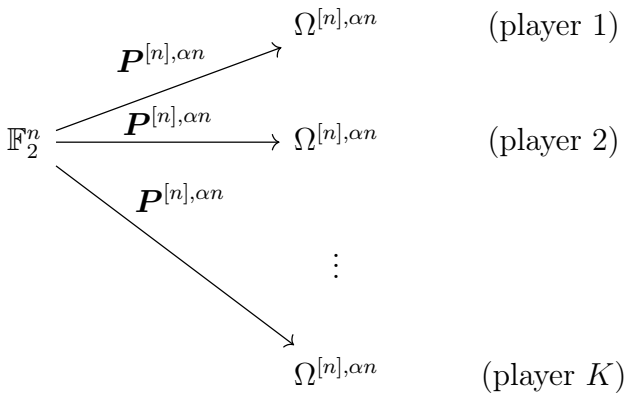


Figure 3.1: The Markov transitions generating $\mathcal{D}_{\text{yes}}$ in $\mathsf{DIHP}(n, \alpha, K)$

Definition 3.13. For any function $f \in L^2(\Omega^{U,m})$, we define the pull-back image $\boldsymbol{P}^{U,m}[f] \in L^2(\mathbb{F}_2^U)$ by

$$\boldsymbol{P}^{U,m}[f](x) := \sum_{y \in \Omega^{U,m}} \mathbf{P}^{U,m}(x,y) f(y).$$

The benefit of switching from the matrix representation $\mathbf{P}^{U,m}$ to the operator formulation $\boldsymbol{P}^{U,m}$ is that the latter makes it easier to conduct spectral analysis, as is the case in Section 2.1.1. A key difference between our operator $\boldsymbol{P}^{U,m}$ with the operator in Definition 2.4 is that the latter is a linear "endomorphism" on an inner product space, while $\boldsymbol{P}^{U,m}$ is a linear map between different inner product spaces. Therefore, whereas we analyzed the eigenvalues of the operator in Section 2.1.1, here the correct analogy is to analyze the *singular value decomposition* of the $\boldsymbol{P}^{U,m}$.

The Markov-kernel view of the YES distribution in the $\mathsf{DIHP}(n, \alpha, K)$ game is illustrated in Figure 3.1.

Note that the matrix $\mathbf{P}^{U,m}$ can be viewed as the (normalized) adjacency matrix of a biregular bipartite graph between the left vertex set $\mathbb{F}_2^U$ and the right vertex set $\Omega^{U,m}$. We denote this bipartite graph by $G^{U,m}$. Due to the biregularity of $G^{U,m}$, there is a canonical Markov transition from $\Omega^{U,m}$ back to $\mathbb{F}_2^U$: on each state $y \in \Omega^{U,m}$, one picks a random neighbor $x \in \mathbb{F}_2^U$ of y in $G^{U,m}$ and moves to x. This Markov transition gives rise to a pull-back operator

$$\left(\boldsymbol{P}^{U,m}\right)^\dagger : L^2(\mathbb{F}_2^U) \longrightarrow L^2(\Omega^{U,m}),$$

in the same way as Definition 3.12 gives rise to Definition 3.13. It is not hard to see that the operator $\left(\boldsymbol{P}^{U,m}\right)^\dagger$ is exactly the *adjoint* of the operator $\boldsymbol{P}^{U,m}$, in the sense that

$$\left\langle \boldsymbol{P}^{U,m}[f], g \right\rangle = \left\langle f, \left(\boldsymbol{P}^{U,m}\right)^\dagger [g] \right\rangle, \quad \text{for all } f \in L^2(\Omega^{U,m}) \text{ and } g \in L^2(\mathbb{F}_2^U).$$

3.2.2 Hypercontractivity

As introduced in Section 1.2, *hypercontractivity* refers to the phenomenon where an operator T not only *contracts* a given L^p-norm, in the sense that

$$\|Tf\|_p \lesssim \|f\|_p,$$

but in fact *improves* the norm, contracting from an L^p-norm to a stronger L^q-norm, i.e.

$$\|Tf\|_q \lesssim \|f\|_p$$

for some $q > p$.

In many applications in discrete mathematics, hypercontractivity admits a combinatorial interpretation as a *small-set expansion* property. For instance, a key property of the bipartite graph $G^{[n],\alpha n}$ between $\mathbb{F}_2^n$ and $\Omega^{[n],\alpha n}$ is that, for any sufficiently small subset of vertices $B \subseteq \mathbb{F}_2^n$, the edges emanating from B to $\Omega^{[n],\alpha n}$ have few "collisions" among their endpoints. In other words, the vertex set B *expands well* in this bipartite graph — hence the term *small-set expansion*.

To formalize this perspective, for each subset $B \subseteq \mathbb{F}_2^n$ we define its *density function* $\phi_B : \mathbb{F}_2^n \to [0, \infty)$ by

$$\phi_B(x) = \frac{2^n}{|B|} \cdot \mathbb{1}\{x \in B\}.$$

The rate of "pairwise collisions" among the edges emanating from B to $\Omega^{[n],\alpha n}$ is then captured, up to normalization, by the ratio

$$\frac{\left\| \left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger [\phi_B] \right\|_2^2}{\|\phi_B\|_2^2}.$$

Note that the case $B = \mathbb{F}_2^n$ corresponds to the worst possible collision rate, for which the above ratio equals 1; hence, a value $\ll 1$ indicates *good expansion*. Accordingly, for the bipartite graph between $\mathbb{F}_2^n$ and $\Omega^{[n],\alpha n}$ to qualify as a small-set expander (from left to right), we would like a statement of the form:

Statement 3.14 (Small-set expander)**.** For any constant $\varepsilon > 0$, there exists a constant $\delta > 0$ such that for all $n \geqslant 1$ and all subsets $B \subseteq \mathbb{F}_2^n$ of size $|B| \leqslant \delta \cdot 2^n$, we have

$$\left\| \left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger [\phi_B] \right\|_2^2 \leqslant \varepsilon \|\phi_B\|_2^2.$$

Such statements can usually be deduced from some hypercontractive inequalities, such as the following.

Statement 3.15 (Hypercontractivity)**.** For any nonempty subset $B \subseteq \mathbb{F}_2^n$, we have

$$\left\| \left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger [\phi_B] \right\|_2 \leqslant \|\phi_B\|_{5/4}.$$

Note that Statement 3.14 follows from Statement 3.15 because $\|\phi_B\|_{5/4} = (2^n/|B|)^{1/5} = \|\phi_B\|_2^{2/5}$.

Remark 3.16. Although we have used the bipartite graph between $\mathbb{F}_2^n$ and $\Omega^{[n],\alpha n}$ as an illustrative example, the preceding discussion on hypercontractivity and small-set expansion applies to any biregular bipartite graph. For the specific graph $G^{[n],\alpha n}$, the paper [12] implicitly established that Statement 3.15 holds, and consequently, so does Statement 3.14.

3.2.3 Expander vs. Extractor

A combinatorial notion closely related to expander graphs is that of *extractors*. We again use the biregular bipartite graph $G^{[n],\alpha n}$ as an example. For this graph to qualify as an extractor (from left to right), we would like a statement of the following form:

Statement 3.17 (Extractor)**.** There exists a constant $\delta > 0$ such that for all $n \geqslant 1$ and all subsets $B \subseteq \mathbb{F}_2^n$ of size $|B| \geqslant 2^{n-\delta}$, we have

$$\left\| \left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger [\phi_B] \right\|_2 \leqslant 1 + o(1),$$

where $o(1)$ denotes a quantity that tends to 0 as $n \to \infty$.

Such a statement expresses that the uniform distribution on a large subset $B \subseteq \mathbb{F}_2^n$ (whose density function is ϕ_B) is *pushed forward* by the Markov transition $\mathbf{P}^{[n],\alpha n}$ to a distribution on $\Omega^{[n],\alpha n}$ that is asymptotically close to uniform.

Unlike the small-set expander statement (Statement 3.14), this extractor statement cannot be deduced from the hypercontractivity inequality in Statement 3.15. Nevertheless, for the specific bipartite graph $G^{[n],\alpha n}$, the following stronger hypercontractivity inequality was (implicitly) proved in [12].

Lemma 3.18 ([12])**.** *Suppose $\alpha \in (0, 10^{-2})$ is a fixed constant. For any function $f \in L^2(\mathbb{F}_2^n)$, we have*

$$\left\| \left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger [f] \right\|_2 \leqslant \|f\|_{1+n^{-1/2}} + 2^{-\sqrt{n}} \|f\|_2.$$

By applying Lemma 3.18 to the function ϕ_B, we get the following corollary which clearly implies Statement 3.17.

Corollary 3.19. *Suppose $\alpha \in (0, 10^{-2})$ is a fixed constant. For any constant $\varepsilon > 0$, there exists a constant $\delta > 0$ such that for all $n \geqslant 1$ and all subsets $B \subseteq \mathbb{F}_2^n$ of size $|B| \geqslant 2^{n-\delta\sqrt{n}}$, we have*

$$\left\| \left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger [\phi_B] \right\|_2 \leqslant 1 + \varepsilon.$$

The extractor property of the operator $\left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger$ plays a crucial role in proving streaming lower bounds for approximating Max-Cut, as we next explain.

As discussed in Section 3.1.2, to establish lower bounds against single-pass streaming algorithms for Max-Cut, it suffices to prove a communication lower bound for *one-way* protocols for $\mathsf{DIHP}(n, \alpha, K)$. To this end, we aim to show that in any efficient one-way protocol, no player can reliably distinguish whether the joint input $(y^{(1)}, \ldots, y^{(K)})$ is drawn from $\mathcal{D}_{\text{yes}}$ or $\mathcal{D}_{\text{no}}$, based solely on her received input $y^{(i)}$ and the messages sent by previous players. Formally, for each player index $i \in [K]$, we would like to show that, conditioned on the messages of the first $(i-1)$ players, the distributions $\mathcal{D}_{\text{yes}}$ and $\mathcal{D}_{\text{no}}$ have almost identical marginals on the i-th player. Since the marginal of $\mathcal{D}_{\text{no}}$ (even after arbitrary conditioning on the inputs of the first $(i-1)$ players) is uniform over $\Omega^{[n],\alpha n}$, it therefore suffices to show that under $(y^{(1)}, \ldots, y^{(K)}) \sim \mathcal{D}_{\text{yes}}$, the distribution of $y^{(i)}$ conditioned on the first $(i-1)$ messages is close to uniform. Due to the construction of the distribution $\mathcal{D}_{\text{yes}}$ (see Figure 3.1), this amounts to asking for some extractor property of the operator $\left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger$.

Using Corollary 3.19 and the argument above, it was proved in [19] that any one-way protocol for $\mathsf{DIHP}(n, \alpha, K)$ achieving an advantage of at least 0.1 must have total message length at least $\Omega(\sqrt{n})$. This in turns imply the single-pass streaming lower bound of Theorem 1.5.

3.2.4 Reverse Extractor

The extractor property of $\left(\boldsymbol{P}^{[n],\alpha n}\right)^{\dagger}$ can be interpreted as a form of *information loss* when the Markov kernel $\mathbf{P}^{[n],\alpha n}$ pushes a probability distribution on $\mathbb{F}_2^n$ forward to a distribution on $\Omega^{[n],\alpha n}$ (see Figure 3.1 for an illustration). From this perspective, Corollary 3.19 informally captures the idea that although a subset $B \subseteq \mathbb{F}_2^n$ of size $|B| \geqslant 2^{n-\delta\sqrt{n}}$ may encode up to $\delta\sqrt{n}$ bits of "information," this information is largely lost under the Markov transition to $\Omega^{[n],\alpha n}$, as the resulting distribution is nearly uniform.

The single-pass analysis of [19] relies primarily on this forward-direction loss of information from $\mathbb{F}_2^n$ to $\Omega^{[n],\alpha n}$. However, such an argument breaks down in the multi-pass setting, due to the more intricate "information flow" that arises in non–one-way communication protocols.

A key conceptual contribution of our work [8] is to show that the *backward* transition from $\Omega^{[n],\alpha n}$ to $\mathbb{F}_2^n$ also exhibits a loss-of-information property, and that this property can be leveraged to prove communication lower bounds for general (possibly multi-pass) protocols.

Ideally, we would like to establish a "backward" analogue of Corollary 3.19, of the following form.

Statement 3.20 (False)**.** For any constant $\varepsilon > 0$, there exists a constant $\delta > 0$ such that for all $n \geqslant 1$ and all subsets $A \subseteq \Omega^{[n],\alpha n}$ of size $|A| \geqslant 2^{-\delta\sqrt{n}} \cdot \left|\Omega^{[n],\alpha n}\right|$, we have

$$\left\|\boldsymbol{P}^{[n],\alpha n}[\phi_A]\right\|_2 \leqslant 1 + \varepsilon,$$

where $\phi_A = \left(\left|\Omega^{[n],\alpha n}\right| / |A|\right) \cdot 1_A$ is the density function of A.

Unfortunately, the above statement is false, due to the following counterexample.

Example 3.21. Let $A \subseteq \Omega^{[n],\alpha n}$ be defined by

$$A := \left\{ y \in \Omega^{[n],\alpha n} \,\middle|\, y_{\{1,2\}} = 1 \right\}.$$

Then for all $x \in \mathbb{F}_2^n$ we have

$$\boldsymbol{P}^{[n],\alpha n}[\phi_A](x) = \begin{cases} 2, & \text{if } x_1 + x_2 = 0, \\ 0, & \text{if } x_1 + x_2 = 1. \end{cases}$$

We thus have $\left\|\boldsymbol{P}^{[n],\alpha n}[\phi_A]\right\|_2 = \sqrt{2}$, while $|A| \geqslant n^{-O(1)} \cdot \left|\Omega^{[n],\alpha n}\right|$.

3.2.5 Global Hypercontractivity

At this point, the technique of *global hypercontractivity* comes to the rescue. This approach originates from the seminal work of Khot, Minzer, and Safra [23], which addressed the

challenge that a certain graph of interest — known as the Grassmann graph — fails to be a small-set expander. They discovered the crucial phenomenon that what drives expansion in the Grassmann graph is not the *size* of a set, but rather its *pseudorandomness*. To show that any pseudorandom set exhibits good expansion, they established a hypercontractive inequality of the form

$$\|Tf\|_q \lesssim \|f\|_p \quad \text{for all pseudorandom functions } f,$$

where T denotes the Markov operator corresponding to the adjacency matrix of the Grassmann graph. In words, Khot, Minzer and Safra showed that although the operator T is not hypercontractive, it does exhibit hypercontractivity when acted on certain pseudorandom functions.

Since both small-set expansion and extractor properties often comes from hypercontractivity of operators, it is natural to ask whether one can fix Statement 3.20 by adding certain pseudorandomness condition to the set $A \subseteq \Omega^{[n],\alpha n}$, in analogy with how [23] fixes the failure of small-set expansion in the Grassmann graph.

To define the notion of pseudorandomness we shall be interested in restrictions of $\Omega^{[n],\alpha n}$, by which we mean the resulting space of labeled matchings once we fix a few of the coordinates.

Definition 3.22 (Restrictions)**.** For each string $z \in \{-1,0,1\}^{\binom{[n]}{2}}$, we let $\Omega_z^{[n],\alpha n} \subseteq \Omega^{[n],\alpha n}$ be the restricted domain defined by

$$\Omega_z^{[n],\alpha n} := \left\{ y \in \Omega^{[n],\alpha n} : y_{uv} = z_{uv} \text{ for all } \{u,v\} \in \mathrm{supp}(z) \right\}.$$

We call z a "restriction" if $\Omega_z^{[n],\alpha n} \neq \emptyset$. In particular, for z to qualify as a restriction, the edge set $\mathrm{supp}(z)$ must be a matching of size no more than αn.

Note that in the space $\Omega^{[n],\alpha n}$, if we fix the entry corresponding to edge $\{i,j\} \in \binom{[n]}{2}$ to a non-zero value, then the entry corresponding to any other edge $\{i,j'\}$ and $\{i',j\}$ must be given the value 0. Indeed, this is because each element in $\Omega^{[n],\alpha n}$ is a labeled matching. This motivates the following definition:

Definition 3.23. For a string $z \in \{-1,0,1\}^{\binom{[n]}{2}}$, we denote by $N(z) \subseteq [n]$ the set of vertices incident to some pair in $\mathrm{supp}(z)$.

Note that the restricted space $\Omega_z^{[n],\alpha n}$ is "isomorphic" to $\Omega^{[n]\setminus N(z),\, \alpha n - |\mathrm{supp}(z)|}$.

We will often want to further restrict an already restricted space. Towards this end, for a restriction z' we wish to consider the restrictions z that extend/subsume it:

Definition 3.24. For two strings $z, z' \in \{-1,0,1\}^{\binom{[n]}{2}}$, we say z subsumes z' if $\mathrm{supp}(z') \subseteq \mathrm{supp}(z)$ and for all $\{u,v\} \in \mathrm{supp}(z')$ we have that $z_{uv} = z'_{uv}$.

Armed with the notion of restrictions, we can now formally define the notion of global (i.e. pseudorandom) sets.

Definition 3.25. A subset $A \subseteq \Omega^{[n],\alpha n}$ is said to be z'-global if $A \subseteq \Omega^{[n],\alpha n}_{z'}$, and for all restrictions z that subsume z' we have

$$\frac{\left|A \cap \Omega^{[n],\alpha n}_z\right|}{\left|\Omega^{[n],\alpha n}_z\right|} \leqslant 2^{|\mathrm{supp}(z)|-|\mathrm{supp}(z')|} \cdot \frac{\left|A \cap \Omega^{[n],\alpha n}_{z'}\right|}{\left|\Omega^{[n],\alpha n}_{z'}\right|}.$$

When $z' = \vec{0}$ is the trivial restriction, we simply say that A is global (omitting the z').

In words, for a set A and a restriction z', we say that A is z'-global if any further restrictions z that subsumes z' increases the relative density of A by factor at most $2^{|\mathrm{supp}(z)|-|\mathrm{supp}(z')|}$.

It is easy to see that the set A defined in Example 3.21 is not a global set. In fact, we are now ready to state the following "global" version of Statement 3.20 proved in [8], which rules out any counterexample that is a global set.

Theorem 3.26 ([8])**.** *Let $\alpha \in (0, 10^{-7})$ be a fixed constant. For any constant $\varepsilon > 0$, there exists a constant $\delta > 0$ such that for all $n \geqslant 1$ and all global subset $A \subseteq \Omega^{[n],\alpha n}$ of size $|A| \geqslant 2^{-\delta n} \cdot \left|\Omega^{[n],\alpha n}\right|$, we have*

$$\left\|\boldsymbol{P}^{[n],\alpha n}[\phi_A]\right\|_2 \leqslant 1 + \varepsilon.$$

As clean as Theorem 3.26 is, we are not able to directly apply this result in the proof of the desired communication lower bound, due to technical reasons. In the next section, we will introduce a more complicated variant of Theorem 3.26 — Lemma 3.33 — and explain how to use it to prove the communication lower bound in Theorem 3.10.

3.3 The Discrepancy Method

As is standard in communication complexity, a subset $R \subseteq \Omega^K$ of the space of joint inputs is called a *rectangle* if it is the Cartesian product of sets $A^{(i)} \subseteq \Omega$, one for each $i \in [K]$; that is, $R = \prod_{i=1}^K A^{(i)}$. One of the most classical techniques in communication complexity is the *discrepancy method*, which usually takes the following two steps:

1. Decomposition step: show that the output of any efficient communication protocol must be constant on each of a collection of (large) rectangles that partition the joint input space.

2. Discrepancy step: show for each such rectangle R that $\mathcal{D}_{\mathrm{yes}}(R)$ and $\mathcal{D}_{\mathrm{no}}(R)$ are close, where $\mathcal{D}_{\mathrm{yes}}$ and $\mathcal{D}_{\mathrm{no}}$ are the YES and NO distributions, respectively.

Our proof of Theorem 3.10 adopts the general outline of the discrepancy method. The additional nuance in our setting is that the rectangles in the partition of Ω^K must all satisfy some globalness properties, as specified by the following definitions.

Definition 3.27 (Structured rectangles)**.** Let $R = A^{(1)} \times \cdots \times A^{(K)} \subseteq \Omega^K$ be a rectangle, and let $\zeta = (z^{(1)}, \ldots, z^{(K)})$ be a sequence of restrictions. We say that R is ζ-global if for all $i \in [K]$, the set $A^{(i)}$ is $z^{(i)}$-global. When a rectangle R is ζ-global, we also say that the pair (ζ, R) is a structured rectangle.

Definition 3.28 (Good rectangles)**.** Let W be a positive real number. We say a structured rectangle (ζ, R), where $R = \prod_{i=1}^{k} A^{(i)}$ and $\zeta = \left(z^{(i)}\right)_{i\in[K]}$, is W-good if the following conditions hold:

(1) The edge sets $\left(\mathrm{supp}(z^{(i)})\right)_{i\in[K]}$ are pairwise disjoint, and their union does not contain any cycle.

(2) $\sum_{i=1}^{K} \left|\mathrm{supp}(z^{(i)}))\right| \leqslant W$.

(3) $\left|A^{(i)}\right| / \left|\Omega_{z^{(i)}}\right| \geqslant 2^{-W}$ for all $i \in [K]$.

The goals of our decomposition step and discrepancy step are to prove the following two lemmas:

Lemma 3.29 (Decomposition lemma)**.** *Fix an integer $K > 0$ and a parameter $\alpha > 0$. There exists a constant $\beta > 0$ such that given any communication protocol Π for $\mathsf{DIHP}(n, \alpha, K)$ with $|\Pi| \leqslant \beta n^{1/3}$,*[2] *there exists a collection $\mathcal{R}$ of pairwise-disjoint structured rectangles (ζ, R) in the space Ω^K such that the following conditions hold:*

(1) $\mathcal{D}_{\mathrm{no}}\left(\bigcup_{(\zeta,R)\in\mathcal{R}} R\right) \geqslant 0.99$.

(2) Each $(\zeta, R) \in \mathcal{R}$ is $(10^5 \cdot |\Pi|)$-good.

(3) For each $(\zeta, R) \in \mathcal{R}$, there exists $a_R \in \{0, 1\}$ such that $\Pi(Y) = a_R$ for every $Y \in R$.

Lemma 3.30 (Discrepancy lemma)**.** *Fix an integer $K > 0$ and a parameter $\alpha \in (0, 10^{-7})$. There exists a constant $\gamma > 0$ such that for any $(\gamma n^{1/3})$-good structured rectangle (ζ, R), we have*

$$|\mathcal{D}_{\mathrm{yes}}(R) - \mathcal{D}_{\mathrm{no}}(R)| \leqslant 10^{-3} \cdot \mathcal{D}_{\mathrm{no}}(R).$$

Once we have Lemmas 3.29 and 3.30, Theorem 3.10 will easily follow:

Proof of Theorem 3.10 assuming Lemmas 3.29 and 3.30. Let β and γ be the constants obtained from Lemmas 3.29 and 3.30, respectively. We fix a communication protocol Π with $|\Pi| \leqslant \min\{\beta, 10^{-5}\gamma\} \cdot n^{1/3}$, and we proceed to show that $\mathrm{adv}(\Pi) \leqslant 0.1$.

We first apply Lemma 3.29 to obtain a collection $\mathcal{R}$ of structured rectangles. We know that each pair $(\zeta, R) \in \mathcal{R}$ is $(10^5|\Pi|)$-good, which is also $(\gamma n^{1/3})$-good because $10^5|\Pi| \leqslant \gamma n^{1/3}$. By Lemma 3.30 we have

$$|\mathcal{D}_{\mathrm{yes}}(R) - \mathcal{D}_{\mathrm{no}}(R)| \leqslant 10^{-3} \cdot \mathcal{D}_{\mathrm{no}}(R)$$

for each $(\zeta, R) \in \mathcal{R}$.

Note that for each $(\zeta, R) \in \mathcal{R}$, the output of Π is constant on R. By Definition 3.9 we have

$$\mathrm{adv}(\Pi) = \left| \mathop{\mathbb{P}}_{Y\sim\mathcal{D}_{\mathrm{yes}}} [\Pi(Y) = 1] - \mathop{\mathbb{P}}_{Y\sim\mathcal{D}_{\mathrm{no}}} [\Pi(Y) = 1] \right|$$

[2] We let $|\Pi|$ denote the maximum total number of bits broadcast by the players, over any joint input.

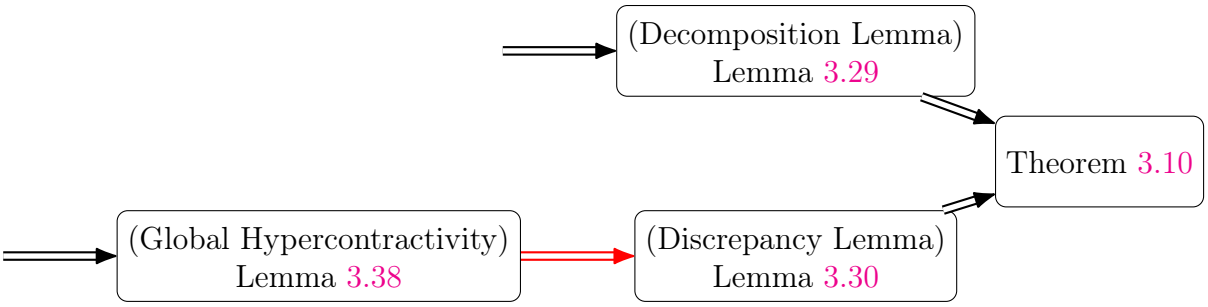


Figure 3.2: The structure of the proof of Theorem 3.10. The step indicated by the red arrow is the focus of this thesis.

$$\begin{aligned}
&\leqslant \mathop{\mathbb{P}}_{Y\sim\mathcal{D}_{\mathrm{yes}}}\left[Y \notin \bigcup_{(\zeta,R)\in\mathcal{R}} R\right] + \mathop{\mathbb{P}}_{Y\sim\mathcal{D}_{\mathrm{no}}}\left[Y \notin \bigcup_{(\zeta,R)\in\mathcal{R}} R\right] + \sum_{(\zeta,R)\in\mathcal{R}} |\mathcal{D}_{\mathrm{yes}}(R) - \mathcal{D}_{\mathrm{no}}(R)| \\
&\leqslant 2\cdot \mathop{\mathbb{P}}_{Y\sim\mathcal{D}_{\mathrm{no}}}\left[Y \notin \bigcup_{(\zeta,R)\in\mathcal{R}} R\right] + 2\cdot \sum_{(\zeta,R)\in\mathcal{R}} |\mathcal{D}_{\mathrm{yes}}(R) - \mathcal{D}_{\mathrm{no}}(R)| \\
&\leqslant 2(1-0.99) + 2\sum_{(\zeta,R)\in\mathcal{R}} 0.001\cdot\mathcal{D}_{\mathrm{no}}(R) < 0.1,
\end{aligned}$$

as desired. □

In the rest of this section, we develop a proof of the discrepancy lemma assuming Lemma 3.38, which is a variant of the "reverse extractor" statement in Theorem 3.26. The proofs of Lemma 3.38 and the decomposition lemma will be discussed in Sections 3.4 and 3.5, respectively. The overall structure of the proof of Theorem 3.10 is illustrated in Figure 3.2.

3.3.1 Discrepancy Calculation

The main reason a hypercontractive inequality for the pull-back operator $\boldsymbol{P}^{[n],\alpha n}$ may be useful for proving the discrepancy lemma is that the discrepancy $|\mathcal{D}_{\mathrm{yes}}(R) - \mathcal{D}_{\mathrm{no}}(R)|$ of a rectangle R can be expressed in terms of the operator $\boldsymbol{P}^{[n],\alpha n}$, as shown in the following lemma.

Lemma 3.31. *For $\Omega = \Omega^{[n],\alpha n}$ and for any rectangle $R = A^{(1)}\times\cdots\times A^{(K)} \subseteq \Omega^K$, we have*

$$\mathcal{D}_{\mathrm{yes}}(R) = \mathcal{D}_{\mathrm{no}}(R)\cdot \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}\left[\prod_{i=1}^{K} \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x)\right]. \tag{3.5}$$

Proof. Recall that in the sampling process of $\mathcal{D}_{\mathrm{yes}}$, we first uniformly sample a vector $x\in\mathbb{F}_2^n$, and for each $i\in[K]$, sample $y^{(i)}$ according to the probability density function $\mathbf{P}^{[n],\alpha n}(x,\cdot)$. We can thus calculate

$$\mathcal{D}_{\mathrm{yes}}(R) = \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}\left[\prod_{i=1}^{K}\sum_{y\in A^{(i)}} \mathbf{P}^{[n],\alpha n}(x,y)\right]$$

$$= \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}\left[\prod_{i=1}^{K}\sum_{y\in\Omega}\mathbf{P}^{[n],\alpha n}(x,y)\phi_{A^{(i)}}(y)\right]\cdot\prod_{i=1}^{K}\frac{|A^{(i)}|}{|\Omega|}$$

$$= \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}\left[\prod_{i=1}^{K}\boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x)\right]\cdot\mathcal{D}_{\mathrm{no}}(R). \qquad \square$$

3.3.2 Bounding Discrepancy via K-norms

We would like to bound the right-hand side of (3.5) using Hölder's inequality, which requires an upper bound on the K-norm of the pull-back function $\boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right]$. Compared with the statement of Theorem 3.26, which provides a bound on the 2-norm of the pull-back function, this may seem like a natural strengthening — after all, the hypercontractivity of an operator typically allows one to control stronger norms of its image from bounds on weaker norms of the original function.

Another reason Theorem 3.26 does not apply well for proving the discrepancy bound in Lemma 3.30 is that while Theorem 3.26 works for sets A that are $\vec{0}$-global (see Definition 3.25), the sets arising from the structured rectangles in Lemma 3.30 are only z-global for some restrictions z. In particular, we cannot hope that the distribution $\boldsymbol{P}^{[n],\alpha n}[\phi_A]$ is close to uniform on $\mathbb{F}_2^n$ if A is only assumed to be z-global, for some nonzero restriction z. To address this issue, we define for each restriction z a linear subspace $L(z)\subseteq\mathbb{F}_2^n$ such that any large z-global set of Ω "extracts" to a near-uniform distribution on $L(z)$, under the operator $\boldsymbol{P}^{[n],\alpha n}$.

Definition 3.32. For every string $z\in\{-1,0,1\}^{\binom{U}{2}}$, we define the affine subspace $L(z)\subseteq\mathbb{F}_2^U$ by

$$L(z) := \bigcap_{\{u,v\}:\, z_{uv}=1}\left\{x\in\mathbb{F}_2^U : x_u+x_v=0\right\}\cap\bigcap_{\{u,v\}:\, z_{uv}=-1}\left\{x\in\mathbb{F}_2^U : x_u+x_v=1\right\}.$$

When used in this context, we call the string z a "constraint."

We can now state our main goal in the respect of extracting near-uniform distributions from z-global sets.

Lemma 3.33. *Let $A\subseteq\Omega_{z'}$ be a z'-global set, let z be a constraint that subsumes z' and define the function $h: L(z)\to\mathbb{R}$ by $h(x)=2^{-|\mathrm{supp}(z')|}\cdot\boldsymbol{P}^{[n],\alpha n}[\phi_A](x)-1$. Suppose that the following conditions hold for constants $\gamma,\eta\in(0,\frac{1}{10})$:*

1. *$\mathrm{supp}(z)$ does not contain any cycles;*
2. *$|\mathrm{supp}(z)|\leqslant\gamma n^{1/3}$;*
3. *$|A|/|\Omega_{z'}|\geqslant 2^{-\eta n^{1/3}}$.*

Then we have that $\|h\|_K\leqslant\sqrt{4\eta\gamma^2K^2+4\eta+2\gamma}+o(1)$. Here, the K-norm of h is with respect to the uniform distribution on the subspace $L(z)$, i.e., $\|h\|_K=\left(\mathbb{E}_{x\in L(z)}\left[|h(x)|^K\right]\right)^{1/K}$.

The proof of Lemma 3.33 is deferred to the following subsections. We conclude this subsection by showing how the discrepancy lemma follows from Lemma 3.33.

Proof of Lemma 3.30 assuming Lemma 3.33. We pick a constant $\gamma > 0$ such that

$$\exp\left(K\sqrt{4\gamma^3 K^2 + 6\gamma}\right) - 1 \leqslant 10^{-4}.$$

Let (ζ, R) be any $(\gamma n^{-1/3})$-good structured rectangle. Let $\zeta = (z^{(1)}, \dots, z^{(k)})$ and $R = \prod_{i=1}^{K} A^{(i)}$. By Lemma 3.31, it suffices to show

$$\left| \mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n} \left[\prod_{i=1}^{K} \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x) \right] - 1 \right| \leqslant 10^{-3}.$$

Define $z \in \{-1, 0, 1\}^{\binom{[n]}{2}}$ by

$$z_{uv} = \begin{cases} z^{(i)}_{uv}, & \text{if } \{u, v\} \in \operatorname{supp}(z^{(i)}) \text{ for some } i \in [K], \\ 0, & \text{if } \{u, v\} \notin \operatorname{supp}(z^{(i)}) \text{ for all } i \in [K]. \end{cases}$$

We note that z is well defined since the supports in the sequence $\left(\operatorname{supp}(z^{(i)})\right)_{i \in [K]}$ are disjoint. Also, using this fact and the fact their union does not contain any cycles, we have

$$\dim(L(z)) = n - |\operatorname{supp}(z)| = n - \sum_{i=1}^{K} \left|\operatorname{supp}(z^{(i)})\right|.$$

Since $L(z) = \bigcap_{i=1}^{K} L(z^{(i)})$ and each density function $\boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right]$ is supported on $L(z^{(i)})$, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n} \left[\prod_{i=1}^{K} \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x) \right] &= 2^{|\operatorname{supp}(z)|} \cdot \mathop{\mathbb{E}}_{x \in L(z)} \left[\prod_{i=1}^{K} \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x) \right] \\ &= \mathop{\mathbb{E}}_{x \in L(z)} \left[\prod_{i=1}^{K} \left(2^{|\operatorname{supp}(z^{(i)})|} \cdot \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x) \right) \right]. \end{aligned}$$

Defining $h_i : L(z) \to \mathbb{R}$ by $h_i(x) := 2^{|\operatorname{supp}(z^{(i)})|} \cdot \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x) - 1$, we get

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n} \left[\prod_{i=1}^{K} \boldsymbol{P}^{[n],\alpha n}\left[\phi_{A^{(i)}}\right](x) \right] - 1 \right| &= \left| \mathop{\mathbb{E}}_{x \in L(z)} \left[\prod_{i=1}^{K} (1 + h_i(x)) \right] - 1 \right| \\ &= \left| \sum_{T \subseteq [K],\, T \neq \emptyset} \mathop{\mathbb{E}}_{x \in L(z)} \left[\prod_{i \in T} h_i(x) \right] \right| \\ &\leqslant \sum_{T \subseteq [K],\, T \neq \emptyset} \mathop{\mathbb{E}}_{x \in L(z)} \left[\left| \prod_{i \in T} h_i(x) \right| \right], \end{aligned}$$

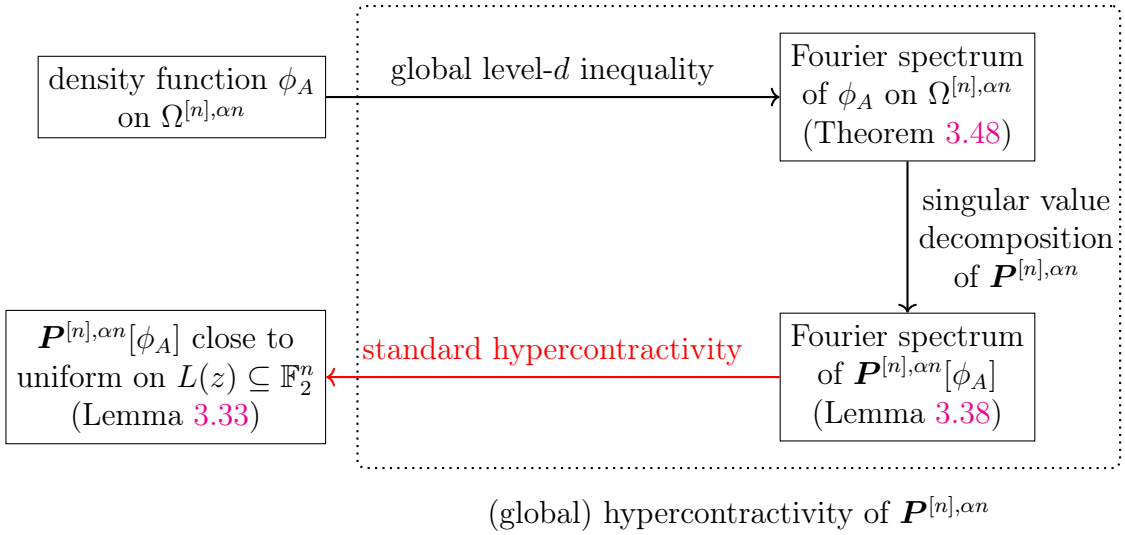


Figure 3.3: The structure of the proof of Lemma 3.33. The step indicated by the red arrow is the focus of this section.

where the last transition is by the triangle inequality. Using Hölder's inequality, the last expression is at most

$$\sum_{\substack{T\subseteq[K]\\T\neq\emptyset}}\prod_{i\in T}\|h_i\|_{|T|} \leqslant \sum_{\substack{T\subseteq[K]\\T\neq\emptyset}}\prod_{i\in T}\|h_i\|_K = \prod_{i=1}^{K}(1+\|h_i\|_K)-1 \leqslant \exp\left(K\sqrt{4\gamma^3K^2+6\gamma}+o(1)\right)-1,$$

where the last transition is by the inequality $1+s \leqslant \exp(s)$ and by Lemma 3.33. □

3.3.3 Hypercontractivity via Fourier Analysis

As discussed in Section 3.2, extractor properties such as Lemma 3.33 are usually due to the hypercontractivity of the underlying operator, which is $\boldsymbol{P}^{[n],\alpha n}$ in this case. However, the author is not aware of any direct way of proving such hypercontractivity. What we *can* do with the operator $\boldsymbol{P}^{[n],\alpha n}$ is to analyze its singular value decomposition (as briefly mentioned in Section 3.2.1). Fortunately, the singular vectors produced by the decomposition of $\boldsymbol{P}^{[n],\alpha n}$ turns out to be nice-behaving *Fourier characters*,[3] and Fourier analysis has provided successful approaches to proving hypercontractivity of many operators of interest (see, e.g., [25, Chapter 9]). In light of this, we will take the Fourier analytic approach to prove Lemma 3.33; the road map of the proof is illustrated in Figure 3.3.

Remark 3.34. The above discussion also applies to the hypercontractivity of the "forward operator" $\left(\boldsymbol{P}^{[n],\alpha n}\right)^\dagger$ (Theorem 3.18) proved by [12]. The high-level structure of the proof Lemma 3.33 is basically the same as that of [12], except for a reversing of the roles of the spaces $\mathbb{F}_2^n$ and $\Omega^{[n],\alpha n}$. On a more technical level, however, there are two important caveats.

[3] In fact, this is the main reason behind the choice of $\boldsymbol{P}^{[n],\alpha n}$ in the construction of the YES distribution $\mathcal{D}_{\text{yes}}$ in Definition 3.8.

First, there are no restrictions or pseudorandomness notions in [12], while we crucially need those in our setting due to the failure of hypercontractivity otherwise (see Statement 3.20). Second, while the proof in [12] could directly use the classical level-d inequality for the hypercube in the first step (see Figure 3.3), we have to prove a new *global* level-d inequality for the non-product space $\Omega^{[n],\alpha n}$.

We assume basic familiarity with discrete Fourier analysis on the hypercube $\mathbb{F}_2^n$ (see, e.g., [25]). For each subset $S \subseteq [n]$, let $\chi_S : \mathbb{F}_2^n \to \mathbb{R}$ denote the associated character function. Every function $f : \mathbb{F}_2^n \to \mathbb{R}$ admits a unique Fourier expansion

$$f = \sum_{S \subseteq [n]} \widehat{f}(S)\, \chi_S,$$

where the scalars $\widehat{f}(S)$ are the *Fourier coefficients* of f.

This expansion induces the usual *degree decomposition* of f, namely

$$f = \sum_{d=0}^{n} f^{=d}, \qquad \text{where} \qquad f^{=d} = \sum_{\substack{S \subseteq [n] \\ |S|=d}} \widehat{f}(S)\, \chi_S.$$

Informally, characters indexed by larger sets S "oscillate" more rapidly, so the decomposition $f = \sum_d f^{=d}$ separates the "low-oscillation" components of f (small $|S|$) from the "high-oscillation" ones. If the degree-d mass

$$\left\| f^{=d} \right\|_2 = \left(\sum_{S \subseteq [n],\ |S|=d} \widehat{f}(S)^2 \right)^{1/2}$$

decays quickly as d increases, then f behaves "smoothly" in the sense that its high-frequency content is small. Fourier-degree decay is therefore a standard proxy for smoothness.

In our setting, being "close to uniform" is precisely a kind of smoothness condition on a distribution's density function. Thus, a natural approach to proving Theorem 3.33 is to establish an appropriate Fourier-decay bound for the density functions on $\mathbb{F}_2^n$ that are pull-backs of the operator $\boldsymbol{P}^{[n],\alpha n}$.

The following definition of decay of Fourier coefficients is specifically designed for our context.

Definition 3.35. Let w be a real number in the range $(0, n)$, and let c be a positive real number. We say a function $f : \mathbb{F}_2^n \to \mathbb{R}$ is (w, δ, c)-decaying if

1. $\widehat{f}(\emptyset)^2 \leqslant \delta$,
2. for every $d \geqslant 1$, $f^{=2d-1} = 0$, and
3. for every $1 \leqslant d \leqslant n/2$, $\left\| f^{=2d} \right\|_2^2 \leqslant c^{-d} F(n, d, w)$, where $F(n, d, w)$ is defined by

$$F(n, d, w) = \begin{cases} \left(\frac{w}{n}\right)^d, & \text{if } 0 \leqslant d \leqslant w, \\ \left(\frac{d}{4n}\right)^d \cdot 2^{2w}, & \text{if } d > w. \end{cases}$$

We remark that for $d = 0$ the value $F(n, d, w)$ is defined to be 1. Although this case is irrelevant for the definition of decaying functions, adopting this convention will be useful for subsequent analysis.

We will need the following proposition ensuring that the piecewise definition of the Fourier weight bound $F(n, d, w)$ is well-behaved and convenient to work with.

Proposition 3.36. The function $F(n, d, w)$ has the following properties:

(1) For fixed n and d, the bound $F(n, d, w)$ is increasing in w, and for all $t \geqslant 1$

$$\frac{F(n, d, tw)}{F(n, d, w)} \geqslant t^{\min\{d,w\}}.$$

(2) For fixed n and $w < n$, the bound $F(n, d, w)$ is decreasing in d, and

$$\frac{F(n, d, w)}{F(n, d-1, w)} \leqslant \frac{\max\{d, w\}}{n}.$$

Proof. We begin with the first item, and we fix n and d. The function $F(n, d, w)$ is continuous and piecewise differentiable in w, in the range $w \in (0, n)$. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}w} \ln F(n, d, w) &= \begin{cases} \ln 4, & \text{if } 0 < w < d \\ d/w, & \text{if } w > d \end{cases} \\ &\geqslant \min\{d, w\}/w.\end{aligned}$$

It follows that

$$\int_w^{tw} \frac{\mathrm{d}}{\mathrm{d}r} \ln F(n, d, r) dr \geqslant \int_w^{tw} \frac{\min\{d, r\}}{r} dr \geqslant \int_w^{tw} \frac{\min\{d, w\}}{r} dr = \min\{d, w\} \cdot \ln t,$$

which translates to $F(n, d, tw)/F(n, d, w) \geqslant t^{\min\{d,w\}}$, giving the first item.

For the second item, we have

$$\frac{F(n, d, w)}{F(n, d-1, w)} \leqslant \max\left\{\frac{w}{n}, \frac{d^d}{4(d-1)^{d-1} \cdot n}\right\} \leqslant \frac{\max\{d, w\}}{n},$$

as desired. □

The following results shows that fast decay of Fourier coefficients coupled with mild bounds on the L^∞-norm implies good bounds for the K-norm of a function.

Lemma 3.37. *Let $K \geqslant 2$ be an integer and let $\gamma \in (0, \frac{1}{4})$ be a constant. Suppose $h : \mathbb{F}_2^n \to \mathbb{R}$ is $(\gamma n, \delta, 1)$-decaying, and that $\|h\|_\infty \leqslant 2^{o(n)}$. Then $\|h\|_K \leqslant \sqrt{\delta + 2\gamma K^2} + o_n(1)$, where $o_n(1)$ hides a term that tends to 0 as $n \to +\infty$.*

Proof. By classical hypercontractivity (see [25], proof of Theorem 9.21), we have

$$\left\|h^{\leqslant 2\gamma n}\right\|_K \leqslant \left(\sum_{d=0}^{\lfloor \gamma n\rfloor}(K-1)^{2d}\left\|h^{=2d}\right\|_2^2\right)^{1/2} \leqslant \left(\delta + \sum_{d=1}^{\lfloor \gamma n\rfloor} K^{2d}\gamma^d\right)^{1/2} \leqslant \sqrt{\delta + 2\gamma K^2}.$$

Using Parseval and Proposition 3.36 we also get

$$\left\|h^{>2\gamma n}\right\|_2 \leqslant \left(\sum_{d=\lceil \gamma n\rceil}^{n/2} F(n,d,\gamma n)\right)^{1/2} \leqslant \sqrt{2F(n,\lceil \gamma n\rceil,\gamma n)} \leqslant \sqrt{2\gamma^{\gamma n}}.$$

Therefore, using Hölder's inequality and Cauchy-Schwarz we get that

$$\begin{aligned}\|h\|_K^K = \left\langle h^{K-1}, h\right\rangle &= \left\langle h^{K-1}, h^{\leqslant 2\gamma n}\right\rangle + \left\langle h^{K-1}, h^{>2\gamma n}\right\rangle \\ &\leqslant \left\|h^{K-1}\right\|_{K/(K-1)}\left\|h^{\leqslant 2\gamma n}\right\|_K + \left\|h^{K-1}\right\|_2\left\|h^{>2\gamma n}\right\|_2 \\ &\leqslant \sqrt{\delta + 2\gamma K^2}\,\|h\|_K^{K-1} + \sqrt{2\gamma^{\gamma n}}\|h\|_\infty^{K-1} \\ &\leqslant \sqrt{\delta + 2\gamma K^2}\,\|h\|_K^{K-1} + o_n(1),\end{aligned}$$

and the conclusion follows by rearranging. □

We now ready to state the main decay lemma, asserting that if A is global, then applying $\boldsymbol{P}^{[n],\alpha n}$ to ϕ_A produces a decaying function:

Lemma 3.38. *Suppose that U has $|U| \geqslant 10^7 m$ where $m \geqslant 10(w+1)$, and let $A \subseteq \Omega^{U,m}$ be a global set with $|A| = 2^{-w}\cdot\left|\Omega^{U,m}\right|$. Then the function $f : \mathbb{F}_2^U \to \mathbb{R}$ defined by $f(x) := \boldsymbol{P}^{U,m}[\phi_A](x) - 1$ is $(w/2, 0, 2)$-decaying.*

The proof of Lemma 3.38 will be the subject of the Section 3.4. In the next few subsections, we explain how to deduce Lemma 3.33 from Lemma 3.38.

3.3.4 Fourier Analytic Setup

Recall the goal of Lemma 3.33: we want to show that the operator $\boldsymbol{P}^{[n],\alpha n}$ maps any sufficiently large global set in $\Omega_{z'}$ to a distribution that is close to uniform on $L(z)$. Both the space $\Omega_{z'}$ and the space $L(z)$ require some Fourier-analytic setup before we can delve into the main part of the analysis. In this section, we only focus on the space $L(z)$; analysis on the more complicated space $\Omega_{z'}$ will the subject of Section 3.4.

As a linear subspace of $\mathbb{F}_2^n$, the space $L(z)$ still possesses the familiar hypercube-type Fourier structure. To present a meaningful identification of the subspace $L(z)$ with a Boolean cube, we give a different way of presenting subspaces of the form of $L(z)$.

Definition 3.39. Let $B = (B_1, \ldots, B_k)$ be a partition of $[n]$, and let $b = (b_1, \ldots, b_n) \in \mathbb{F}_2^n$. We define the affine subspace $V^{B,b} \subseteq \mathbb{F}_2^n$ by

$$V^{B,b} := \bigcap_{\ell=1}^{k}\left\{x \in \mathbb{F}_2^n : x_i + b_i = x_j + b_j,\ \forall i,j \in B_\ell\right\}.$$

We note that not every affine subspaces of $\mathbb{F}_2^n$ can be represented by a partition and a string as in Definition 3.39. However, it is easy to see that for any constraint $z \in \{-1,0,1\}^{\binom{[n]}{2}}$, the subspace $L(z)$ can indeed be expressed in this form.[4] The main benefit of working with Definition 3.41 is that it naturally gives rise to a canonical identification between $V^{B,b}$ and a Boolean cube, and thus with a canonical Fourier basis for functions over $V^{B,b}$.

Definition 3.40. Define the canonical identification map $\mathfrak{id} : V^{B,b} \to \mathbb{F}_2^k$ that maps $x \in V^{B,b}$ to $z \in \mathbb{F}_2^k$ defined by $x_i + b_i = z_\ell$, where $\ell \in [k]$ is such that $i \in B_\ell$.

We note that $\mathfrak{id}$ is well defined, as by definition the value of $x_i + b_i$ is the same for all $i \in B_\ell$. We also note that $\mathfrak{id}$ is a 1-to-1 map, as x can be recovered from z (when b and B are thought of as fixed). Finally, we note that as the sets $B_1, \ldots, B_k$ are disjoint, sampling $x \in V^{B,b}$ uniformly, the distribution of $\mathfrak{id}(x)$ is uniform over $\mathbb{F}_2^k$. Hence it makes sense to define the Fourier basis of the space $V^{B,b}$ using the Fourier basis over $\mathbb{F}_2^k$.

Definition 3.41. For a subset $S \subseteq [k]$, we define the character function $\chi_S : V^{B,b} \to \{-1,1\}$ by

$$\chi_S(x) := \prod_{\ell \in S} (-1)^{(\mathfrak{id}(x))_\ell}.$$

Definition 3.42. For a function $f : V^{B,b} \to \mathbb{R}$ and a subset $S \subseteq [k]$, we define the corresponding Fourier coefficient of f by

$$\widehat{f}(S) := \frac{1}{2^k} \sum_{x \in V^{B,b}} f(x)\chi_S(x).$$

Thus, the Fourier expansion of a function $f : V^{B,b} \to \mathbb{R}$ is given as $f(x) = \sum_{S \subseteq [k]} \widehat{f}(S)\chi_S(x)$. Fourier analysis on $L(z)$ is in essence identical to the analysis on Boolean cubes $\mathbb{F}_2^k$. We will therefore adopt many of the notations therein, and specifically define the degree d part of a function f as $f^{=d}(x) = \sum_{|S|=d} \widehat{f}(S)\chi_S(x)$.

3.3.5 Unrefinements and Fourier Decay

Recall the goal of Lemma 3.33: we want to show that the operator $\boldsymbol{P}^{[n],\alpha n}$ maps any sufficiently large global set in $\Omega_{z'}$ to a distribution that is close to uniform on $L(z)$. For a set $A \subseteq \Omega_{z'}$, the pull-back function $\boldsymbol{P}^{[n],\alpha n}[\phi_A]$ is a density function supported on $L(z') \subseteq \mathbb{F}_2^n$, and it is straightforward to combine this with Lemma 3.38 to conclude that $\boldsymbol{P}^{[n],\alpha n}[\phi_A]$ is close to uniform *on* $L(z')$.

However, this does *not* automatically imply that the restriction of $\boldsymbol{P}^{[n],\alpha n}[\phi_A]$ to the smaller subspace $L(z) \subseteq L(z')$ remains close to uniform on $L(z)$. Closeness to uniformity is a delicate property of probability distributions, and in general it is not preserved under restricting the domain.

[4] In fact, it is easily seen that the family of subspaces that can be expressed as in Definition 3.32 coincides with the family of subspaces that can be expressed as in Definition 3.39.

The key observation is that the Fourier decay established in Lemma 3.38, being a stronger structural condition than closeness to uniformity, does exhibit some robustness under restriction. In this subsection we formalize this idea and show how the decay properties on $L(z')$ can be leveraged to control the behavior of the restricted density on $L(z)$.

Recall that from Section 3.3.4 that Fourier analysis on $L(z')$ proceeds by thinking of $L(z')$ as a space $V^{B',b}$, and then considering the cube given by the image of the map $\mathfrak{id}$. Each block B'_i then corresponds to a coordinate, and the Fourier expansion is defined correspondingly. Thinking of $\mathrm{supp}(z')$ as a graph, the blocks $\{B'_i\}$ are its connected components. To get $L(z)$ into the picture we need to explain how its representation relates to $V^{B',b}$. The simplest example for z that subsumes z' is z that is identical to z', except that z_{uv} is non-zero for some $\{u,v\} \notin \mathrm{supp}(z')$. In that case, the graph of z is the graph of z' with one additional edge, which may lead to a merge of two of the connected component of z'. More generally, the graph of z that subsumes z' is the graph of z with additional edges, which may lead to merges between connected components of z's. In other words, the connected components of z are unions of B'_i's, and this is what we refer to as "unrefinement". The unrefinement operation gives a representation of $L(z)$ as the space $V^{B,b}$ where each B_i is a union of possibly several B'_j's. In the other direction, B' may be viewed as a refinement of B.

The following result asserts that for a function $g : L(z') \to \mathbb{R}$ that has Fourier decay, applying a mild unrefinement operation one gets a function that also has a decent Fourier decay.

Lemma 3.43. *Suppose B and B' are partitions of the set $[n]$, and $b \in \mathbb{F}_2^n$ is a string of bits. Suppose B' is a refinement of B, that is, every component of B is a union of components of B'. Let $\gamma, \eta \in (0, \frac{1}{10})$, and suppose that*

$$\sum_{\ell} |B_\ell| \cdot \mathbb{1}\{|B_\ell| \geqslant 2\} \leqslant \gamma n^{1/3}.$$

If $g : V^{B',b} \to \mathbb{R}$ is $\big(\eta n^{1/3}, 0, 1\big)$-decaying, then $h = g_{|V^{B,b}}$ is $(4\eta\gamma^2|B|, 8\eta + 2\gamma, 1)$-decaying.

Proof. Suppose $|B| = k$ and $|B'| = k'$. Note that the conditions imply $k' \geqslant k \geqslant n/2$. For any $S \subseteq [k]$, we define $\mathcal{T}(S)$ to be the collection of subsets $T \subseteq [k']$ such that:

1. for every $\ell \in S$, the number of elements $j \in T$ such that $B'_j \subseteq B_\ell$ is odd;
2. for every $\ell \in [k] \setminus S$, the number of elements $j \in T$ such that $B'_j \subseteq B_\ell$ is even.

We will first show that the Fourier coefficient $\widehat{h}(S)$ is equal to the sum of the coefficients $\widehat{g}(T)$ for $T \in \mathcal{T}(S)$. Towards this end, note that if $\chi'_T : V^{B',b} \to \mathbb{R}$ is character on $V^{B',b}$ for $T \subseteq [k']$, and $\chi_S : V^{B,b} \to \mathbb{R}$ is the character on $V^{B,b}$ for $S \subseteq [k]$, then

$$\sum_{x \in V^{B,b}} \chi'_T(x)\chi_S(x) = \begin{cases} 2^k, & \text{if } T \in \mathcal{T}(S), \\ 0, & \text{if } T \notin \mathcal{T}(S). \end{cases}$$

Therefore,

$$\widehat{h}(S) = \frac{1}{2^k} \sum_{x \in V^{B,b}} h(x)\chi_S(x) = \frac{1}{2^k} \sum_{x \in V^{B,b}} g(x)\chi_S(x) = \frac{1}{2^k} \sum_{x \in V^{B,b}} \sum_{T \subseteq [k']} \widehat{g}(T)\chi'_T(x)\chi_S(x)$$

$$= \frac{1}{2^k} \sum_{T \subseteq [k']} \widehat{g}(T) \sum_{x \in V^{B,b}} \chi'_T(x)\chi_S(x)$$
$$= \sum_{T \in \mathcal{T}(S)} \widehat{g}(T). \tag{3.6}$$

Next, to relate the squares of coefficients we want to have an upper bound on $|\mathcal{T}(S)|$. Denote $m = \sum_{\ell=1}^{k} |B_\ell| \cdot \mathbb{1}\{|B_\ell| \geqslant 2\}$, so that by assumption $m \leqslant \gamma n^{1/3}$. Note that the collections $\{\mathcal{T}(S)\}_{S \subseteq [n]}$ have the following property:

1. For $S_1, S_2 \subseteq [k]$ such that $S_1 \neq S_2$, we have $\mathcal{T}(S_1) \cap \mathcal{T}(S_2) = \emptyset$.

Next, let $Q \subseteq [k']$ be the set of all indices $j \in [k']$ such that B'_j is contained in the set $\bigcup_{\ell \in [k]:\, |B_\ell| \geqslant 2} B_\ell$. Since the latter set has size m, we know that Q has size at most m. By definition of $\mathcal{T}(S)$, it is not hard to see that a member $T \in \mathcal{T}(S)$ is uniquely determined by the set $Q \cap T$. Also, $|Q \cap T| \leqslant |T| - |S| + s$ where s is the number of elements $\ell \in S$ such that $|B_\ell| \geqslant 2$. We thus get the following additional properties of $\mathcal{T}(S)$:

2. Fix $T \in \mathcal{T}(S)$. As each $\ell \in S$ contributes an odd number of elements to T and each $\ell \notin S$ contributes an even number of elements to T, we get that $|T| - |S|$ is an even integer. By the above, we get that it is in the range $\big[0, |Q|\big] \subseteq [0, m]$.

3. For integer $j \in [0, m/2]$, the number of sets T in $\mathcal{T}(S)$ with $|T| - |S| = 2j$ is at most $\binom{|Q|}{2j+s}$, which we upper bound as $\binom{|Q|}{2j+s} \leqslant \binom{m}{2j} \cdot m^{|S|}$. Indeed, this is an upper bound on the number of choices for $Q \cap T$, and each such choice determines a unique element in $\mathcal{T}(S)$.

From this point onward in the proof, we will refer to the three properties mentioned listed above as the first, second and third properties of $\mathcal{T}(S)$.

By the second property of $\mathcal{T}(S)$, we know that if $|S|$ is odd then $|T|$ is odd for every $T \in \mathcal{T}(S)$, which implies $\widehat{g}(T) = 0$ by the assumption on g, and thus $\widehat{h}(S) = 0$ by (3.6). We are now ready to show that h satisfies the desired Fourier weight bound on even degrees, and we split into cases.

Case 1: the low degree case. Suppose that $|S| = 2d$ where $0 \leqslant d \leqslant \eta n^{1/3}$. Using (3.6), Cauchy-Schwarz and the third property of $\{\mathcal{T}(S)\}$ above we have:

$$\begin{aligned}
\widehat{h}(S)^2 &\leqslant \left(\sum_{T \in \mathcal{T}(S)} n^{-(|T|-|S|)/3} \right) \left(\sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2 \cdot n^{(|T|-|S|)/3} \right) \\
&\leqslant \left(\sum_{j=0}^{\lfloor m/2 \rfloor} \binom{m}{2j} m^{|S|} n^{-2j/3} \right) \left(\sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2 \cdot n^{(|T|-|S|)/3} \right) \\
&\leqslant m^{|S|}(1 + n^{-1/3})^m \sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2 \cdot n^{(|T|-|S|)/3} \\
&\leqslant e^{\gamma} m^{|S|} \sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2 \cdot n^{(|T|-|S|)/3},
\end{aligned}$$

where we used $m \leqslant \gamma n^{1/3}$ in the last transition. Using the first property of $\{\mathcal{T}(S)\}$, the decaying assumption of g, and then Proposition 3.36(2), we get

$$\begin{aligned}
\sum_{|S|=2d} \widehat{h}(S)^2 &\leqslant e^{\gamma} m^{2d} \sum_{j=0}^{\lfloor m/2 \rfloor} n^{2j/3} \cdot \left(\sum_{|T|=2d+2j} \widehat{g}(T)^2 \right) \\
&\leqslant e^{\gamma} m^{2d} \left\| g^{=2d} \right\|_2^2 + e^{\gamma} m^{2d} \sum_{j=1}^{\lfloor m/2 \rfloor} n^{2j/3} \cdot F\left(k', d+j, \eta n^{1/3}\right). \\
&\leqslant e^{\gamma} m^{2d} \left\| g^{=2d} \right\|_2^2 + e^{\gamma} m^{2d} F\left(k', d, \eta n^{1/3}\right) \cdot \sum_{j=1}^{\lfloor m/2 \rfloor} n^{2j/3} \left(\frac{2\eta n^{1/3}}{k'} \right)^j \\
&\leqslant e^{\gamma} m^{2d} \left\| g^{=2d} \right\|_2^2 + e^{\gamma} m^{2d} \frac{4\eta}{1-4\eta} \cdot F\left(k', d, \eta n^{1/3}\right),
\end{aligned}$$

where we used $k' \geqslant k \geqslant n/2$ in the last inequality. For the special case $d = 0$ (recall that we defined $F(\cdot, 0, \cdot) = 1$) we have

$$\left\| h^{=0} \right\|_2^2 \leqslant e^{\gamma} \cdot \frac{4\eta}{1-4\eta} \leqslant 8\eta + 2\gamma,$$

using $\gamma, \eta \in (0, \frac{1}{10})$. For $d > 0$, using $\left\| g^{=2d} \right\|_2^2 \leqslant F\left(k', d, \eta n^{1/3}\right)$ we get

$$\begin{aligned}
\left\| h^{=2d} \right\|_2^2 \leqslant \frac{e^{\gamma} m^{2d}}{1-4\eta} F\left(k', d, \eta n^{1/3}\right) &\leqslant 2(\gamma n^{1/3})^{2d} \cdot F\left(k', d, \eta n^{1/3}\right) \\
&\leqslant F\left(k', d, 2\eta\gamma^2 n\right) && \text{(by Proposition 3.36(1))} \\
&\leqslant F\left(k', d, 4\eta\gamma^2 k\right) && \text{(using } k' \geqslant k \geqslant n/2\text{)} \\
&\leqslant F(k, d, 4\eta\gamma^2 k),
\end{aligned}$$

where the last inequality is because $k' \geqslant k$.

Case 2: the high degree case. Suppose that $|S| = 2d$ where $d > \eta n^{1/3}$. Then by (3.6), Cauchy-Schwarz and the third property of $\{\mathcal{T}(S)\}$ above we have:

$$\begin{aligned}
\widehat{h}(S)^2 &\leqslant \left(\sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2 \right) \left(\sum_{T \in \mathcal{T}(S)} 1 \right) \\
&\leqslant \left(\sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2 \right) \left(\sum_{j=1}^{\lfloor m/2 \rfloor} \binom{m}{2j+s} \right) \leqslant 2^{m-1} \sum_{T \in \mathcal{T}(S)} \widehat{g}(T)^2.
\end{aligned}$$

So using the first property of $\{\mathcal{T}(S)\}$, we get:

$$\sum_{|S|=2d} \widehat{h}(S)^2 \leqslant 2^{m-1} \sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{|T|=2d+2j} \widehat{g}(T)^2$$

$$\leqslant 2^{m-1} \sum_{j=0}^{\lfloor m/2 \rfloor} F\left(k', d+j, \gamma n^{1/3}\right) \qquad \text{(by assumption on } g\text{)}$$

$$\leqslant 2^{m-1} F\left(k', d, \gamma n^{1/3}\right) \sum_{j=0}^{\lfloor m/2 \rfloor} 2^{-j} \qquad \text{(by Proposition 3.36(2))}$$

$$\leqslant 2^{m} F\left(k', d, \gamma n^{1/3}\right) \leqslant F\left(k', d, 2^{\gamma/\min\{\gamma,\eta\}} \cdot \gamma n^{1/3}\right) \qquad \text{(by Proposition 3.36(1))}$$

$$\leqslant F\left(k, d, 2^{\gamma/\min\{\gamma,\eta\}} \cdot \gamma n^{1/3}\right) \qquad \text{(using } k' \geqslant k\text{)}.$$

Since γ, η are constants, $k \geqslant n/2$ and n is assumed to be sufficiently large, the bound obtained is at most $F(k, d, 4\eta\gamma^2 k)$. □

3.3.6 Finishing the Proof

We are now ready to finish the proof of Lemma 3.33.

Proof of Lemma 3.33 assuming Lemma 3.38. The proof proceeds in the following steps.

Step 1: bringing the subspaces $L(z), L(z')$ to form. let $G = ([n], \mathrm{supp}(z))$ be the graph formed by the constraint z, and let $B_1, B_2, \ldots, B_k$ be the connected components of G. For each $\ell \in [k]$, pick an arbitrary vertex $v_\ell \in B_\ell$ and set $b_{v_\ell} = 0$. For any other vertex $u \in B_\ell$, pick the unique simple path $(u_0, u_1, \ldots, u_j)$ in G such that $u_0 = u$, $u_j = v_\ell$, and define

$$b_u = \sum_{i=1}^{j} z_{u_{i-1}u_i}.$$

By inspection, the affine subspace $V^{B,b}$ associated with the partition B and string $b = (b_1, \ldots, b_n)$ is exactly the affine subspace $L(z)$.

Similarly, letting $G' = ([n], \mathrm{supp}(z'))$ be the graph formed by the constraint z', and letting $B'_1, \ldots, B'_{k'}$ be the connected components of G', the graph G' is a subgraph of G and so B' is a refinement of B. Using a similar reasoning to before we see that $V^{B',b}$ coincides with $L(z')$.

Since z' is a restriction in the sense of Definition 3.22, we have that each one of $B'_1, \ldots, B'_{k'}$ is a set of cardinality 1 or 2. Denoting $t = |\mathrm{supp}(z')|$, we assume without loss of generality that $B'_1, \ldots, B'_{n-2t}$ are the singletons among them, and $B'_{n-2t+1}, \ldots, B'_{n-t}$ have size 2. We also assume without loss of generality that $B'_j = \{j\}$ for every $j \in [n-2t]$.

Noting that $V^{B,b} \subseteq V^{B',b}$, we let $\iota \colon V^{B,b} \to V^{B',b}$ be the inclusion map. We recall the map $\mathfrak{id} \colon V^{B',b} \to \mathbb{F}_2^{k'}$ as in Definition 3.40. Finally, consider the projection map $\pi : \mathbb{F}_2^{k'} \to \mathbb{F}_2^{n-2t}$ defined by $(x_1, \ldots, x_{k'}) \mapsto (x_1 + b_1, \ldots, x_{n-2t} + b_{n-2t})$. These maps give rise to the commutative diagram in Figure 3.4.

Step 2: working in $\Omega_{z'}$. Recall that we may identify $\Omega_{z'}$ with the space $\Omega^{[n-2t],\alpha n-t}$. The density function $\phi_A : \Omega_{z'} \to \mathbb{R}$ is thus identified with a density function $\phi_{\widetilde{A}} : \Omega^{[n-2t],\alpha n-t} \to \mathbb{R}$. Applying Lemma 3.38 to the function $\phi_{\widetilde{A}}$, we see that the function $f : \mathbb{F}_2^{n-2t} \to \mathbb{R}$ defined by

$$f(x) := \boldsymbol{P}^{[n-2t],\alpha n-t}\left[\phi_{\widetilde{A}}\right](x) - 1$$

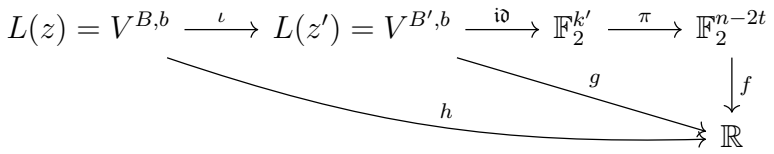


Figure 3.4: A commutative diagram of maps between sets

is $(\eta n^{1/3}/2, 0, 2)$-decaying. Also note that since the operator $\boldsymbol{P}^{[n-2t],\,\alpha n-t}$ does not increase L^∞-norm (see Propostion 2.5), we have

$$\|f\|_\infty \leqslant \left\|\phi_{\widetilde{A}}\right\|_\infty = \frac{\left|\Omega^{[n-2t],\,\alpha n-t}\right|}{\left|\widetilde{A}\right|} = \frac{|\Omega_{z'}|}{|A|} \leqslant 2^{\eta n^{1/3}}. \tag{3.7}$$

Step 3: returning to the space $L(z')$. We want to convert the information we have on f to information about the function h in the statement of the lemma. Towards that end, we first note that $A \subseteq \Omega_{z'}$ directly induces a distribution $\boldsymbol{P}^{[n],\alpha n}[\phi_A]$ supported on $L(z') = V^{B',b}$, and we define $g : L(z') \to \mathbb{R}$ by

$$g(x) := 2^{-t}\boldsymbol{P}^{[n],\alpha n}[\phi_A](x) - 1.$$

For any $x \in L(z')$, any $y \in \Omega_{z'}$ and its corresponding element $\widetilde{y} \in \Omega^{[n-2t],\,\alpha n-t}$, we have by definitions

$$\mathbf{P}^{[n],\alpha n}(x, y) = \frac{2^t\,|\Omega_{z'}|}{|\Omega|}\mathbf{P}^{[n-2t],\,\alpha n}\big(\pi(\mathfrak{id}(x)), \widetilde{y}\big).$$

Therefore, for all $x \in L(z')$ we have

$$\begin{aligned}
\boldsymbol{P}^{[n],\alpha n}[\phi_A](x) &= \frac{|\Omega|}{|A|}\sum_{y\in A}\mathbf{P}^{[n],\alpha n}(x, y) \\
&= \frac{2^t\,|\Omega_{z'}|}{|A|}\sum_{\widetilde{y}\in\widetilde{A}}\mathbf{P}^{[n-2t],\,\alpha n-t}\big(\pi(\mathfrak{id}(x)), \widetilde{y}\big) \\
&= 2^t\boldsymbol{P}^{[n-2t],\,\alpha n-t}\big[\phi_{\widetilde{A}}\big](x),
\end{aligned}$$

and hence $g(x) = f(\pi(\mathfrak{id}(x)))$ (as also indicated by Figure 3.4). Considering the Fourier expansion, we get that for all $S \subseteq [k']$

$$\begin{aligned}
\widehat{g}(S) = 2^{-n+t}\sum_{x\in V^{B',b}} g(x)\chi_S(\mathfrak{id}(x)) &= 2^{-n+t}\sum_{\xi\in\mathbb{F}_2^{n-2t}} f(\xi)\left(\sum_{x\in\mathbb{F}_2^{n-t}}\chi_S(x)\cdot\mathbb{1}\{\pi(x) = \xi\}\right) \\
&= 2^{-n+t}\sum_{\xi\in\mathbb{F}_2^{n-2t}} f(\xi)\left(\sum_{x\in\pi^{-1}(\xi)}\chi_S(x)\right) \\
&= \begin{cases} 2^{-n+t}\sum_{\xi\in\mathbb{F}_2^{n-2t}} f(\xi)\cdot 2^t\chi_S(\xi + b) & \text{if } S \subseteq [n-2t] \\ 0 & \text{if } S \not\subseteq [n-2t] \end{cases}
\end{aligned}$$

$$= \begin{cases} \chi_S(b)\widehat{f}(S), & \text{if } S \subseteq [n-2t] \\ 0, & \text{if } S \not\subseteq [n-2t]. \end{cases}$$

Therefore, for every $d \geqslant 0$, we have $\left\|g^{=d}\right\|_2 = \left\|f^{=d}\right\|_2$. Since f is $(\eta n^{1/3}/2, 0, 2)$-decaying, we know that $f^{=d} \equiv 0$ for d which is either 0 or odd, yielding that $g^{=d} \equiv 0$ for such d's. For even d's, we get

$$\left\|g^{=2d}\right\|_2^2 = \left\|f^{=2d}\right\|_2^2 \leqslant 2^{-d} F\Big(n-2t, d, \eta n^{1/3}/2\Big) \leqslant F\Big(n-t, d, \eta n^{1/3}/2\Big),$$

where the last inequality follows by the definition of F. This means g is $(\eta n^{1/3}/2, 0, 1)$-decaying.

To finish the proof, we note that $h := g_{|V^{B,b}}$, so by Lemma 3.43 we get that g is $(2\eta\gamma^2 k, 4\eta + 2\gamma, 1)$-decaying. By (3.7) we get that $\|h\|_\infty = \|f\|_\infty \leqslant 2^{\eta n^{1/3}} = 2^{o(k)}$, so applying Lemma 3.37 we conclude that

$$\|h\|_K \leqslant \sqrt{4\eta\gamma^2 K^2 + 4\eta + 2\gamma} + o(1). \qquad \square$$

3.4 Global Hypercontractivity in Ω

In this section, we outline the proof of Lemma 3.38. Readers interested in full details are referred to [8, Section 4].

3.4.1 The Level-d Inequality

The structure of the proof of Lemma 3.38 is illustrated in Figure 3.3. Broadly speaking, it consists of two steps. First, we establish a *global level-d inequality* that controls the Fourier spectrum of the density function ϕ_A on $\Omega^{U,m}$. We then transfer this control to the Fourier spectrum of $\boldsymbol{P}^{U,m}[\phi_A]$ by analyzing the singular value decomposition of $\boldsymbol{P}^{U,m}$.

Both steps require a common preliminary framework for Fourier analysis on the non-product space $\Omega^{U,m}$. Our first task is therefore to introduce a family of character functions on $\Omega^{U,m}$. These characters are indexed by "partial matchings" on the ground set U, namely matchings of size at most m. To facilitate the definition, we introduce the following notation.

Definition 3.44. For a ground set U and an integer $d \geqslant 0$, we let $\mathcal{M}_{U,d}$ denote the collection of all matchings over U of size exactly d, and let $\mathcal{M}_{U,\leqslant d} := \bigcup_{s=0}^{d} \mathcal{M}_{U,s}$.

Definition 3.45. For integers n, m such that $n \geqslant 2m \geqslant 0$, we define $\Psi(n, m, 0) := 1$, and for $1 \leqslant d \leqslant m$ we define inductively $\Psi(n, m, d) := m\binom{n}{2}^{-1} \cdot \Psi(n-2, m-1, d-1)$.

It is easy to see that $\Psi(n, m, d)$ is equal to the probability that a fixed matching of size d over a ground set of size n is contained in a uniformly random matching of size m. We now define an associated collection of character:

Definition 3.46. For a matching $M \in \mathcal{M}_{U,\leqslant m}$ and an element $y \in \Omega^{U,m}$, we define the character function $\psi_M : \Omega^{U,m} \to \mathbb{R}$ by

$$\psi_M(y) := \Psi(|U|, m, |M|)^{-1/2} \cdot \prod_{\{u,v\} \in M} y_{uv}.$$

Note that the dimension of the inner product space $L^2(\Omega^{U,m})$ is much larger than the number of partial matchings on U. In particular, the character functions we defined do not form a basis for $L^2(\Omega^{U,m})$. Nevertheless, they will be sufficient for us, as Fourier coefficients coming from h in the context of Leamm 3.38 will only be related to correlations with these character functions.

The next proposition shows the functions from Definition 3.46 form an orthonormal set.

Proposition 3.47. For matchings $M, M' \in \mathcal{M}_{U,\leqslant m}$, we have $\langle \psi_M, \psi_{M'} \rangle = \mathbb{1}\{M = M'\}$.

We are now ready to state our level-d inequality.

Theorem 3.48 (Projected level-d inequality)**.** *Suppose $|U| \geqslant 10m$ and $m \geqslant 10(d+1)$. Let $A \subseteq \Omega^{U,m}$ be a global set of size $|A| = 2^{-w} \cdot \left|\Omega^{U,m}\right|$, where $w \geqslant 2d$. Then*

$$\sum_{M \in \mathcal{M}_{U,d}} \langle \phi_A, \psi_M \rangle^2 \leqslant \left(\frac{2 \cdot 10^5 \log\left(\left|\Omega^{U,m}\right| / |A|\right)}{d} \right)^d.$$

Theorem 3.48 is obtained by first establishing a *derivative-based* global hypercontractive inequality, following the approach of [21].

3.4.2 Singular Value Decomposition

As noted in Section 3.3.3, it turns out that the singular value decomposition of $\boldsymbol{P}^{U,m}$ can be nicely described by Fourier characters.

Definition 3.49. For any subset $S \subseteq U$, let $\mathcal{M}(S)$ be the collection of perfect matchings of the vertices in S (if $|S|$ is odd then $\mathcal{M}(S) = \emptyset$).

It is not hard to establish the following.

Lemma 3.50 (Singular value decomposition)**.** *For any character function $\chi_S : \mathbb{F}_2^U \to \{-1, 1\}$, we have*

$$\left\langle \boldsymbol{P}^{U,m}[f], \chi_S \right\rangle_{L^2(\mathbb{F}_2^U)} = \sum_{M \in \mathcal{M}(S)} \left\langle f, \sqrt{\Psi(|U|, m, |M|)} \cdot \psi_M \right\rangle_{L^2(\Omega^{U,m})}$$

for any function $f \in L^2(\Omega^{U,m})$.

We can now prove Lemma 3.38 using Theorem 3.48 and Lemma 3.50.

Proof of Lemma 3.38. By the definition of f, we know that $\widehat{f}(\emptyset) = 0$. For any nonempty set $S \subseteq U$, we know from Lemma 3.50 that

$$\widehat{f}(S) = \left\langle \boldsymbol{P}^{U,m}[\phi_A], \chi_S \right\rangle = \begin{cases} 0, & \text{if } |S| \text{ is odd,} \\ \sqrt{\Psi(|U|, m, |S|/2)} \sum_{M \in \mathcal{M}(S)} \langle \phi_A, \psi_M \rangle, & \text{if } |S| \text{ is even.} \end{cases}$$

Therefore, it immediately follows that $f^{=2d-1} = 0$ for any integer $d \geqslant 1$. Furthermore, for $d \geqslant 1$ we have

$$\begin{aligned}\left\|f^{=2d}\right\|_2^2 &= \sum_{S\subseteq U,\,|S|=2d} \widehat{f}(S)^2 = \Psi(|U|,m,d) \sum_{S\subseteq U,\,|S|=2d} \left(\sum_{M\in\mathcal{M}(S)} \langle \phi_A, \psi_M\rangle\right)^2 \\ &\leqslant \Psi(|U|,m,d)\cdot(2d-1)!!\cdot \sum_{S\subseteq U,\,|S|=2d}\sum_{M\in\mathcal{M}(S)} \langle \phi_A,\psi_M\rangle^2 \\ &= \binom{m}{d}\binom{|U|}{2d}^{-1}\cdot \sum_{M\in\mathcal{M}_{U,d}} \langle \phi_A,\psi_M\rangle^2. \end{aligned} \tag{3.8}$$

If $d \leqslant w/2$, we can apply Theorem 3.48 to the right hand side of (3.8) and obtain

$$\left\|f^{=2d}\right\|_2^2 \leqslant \binom{m}{d}\binom{|U|}{2d}^{-1}\cdot\left(\frac{2\cdot 10^5 w}{d}\right)^d \leqslant \left(\frac{3m}{d}\right)^d\left(\frac{2d}{|U|}\right)^{2d}\left(\frac{2\cdot 10^5 w}{d}\right)^d \leqslant \left(\frac{w/2}{2|U|}\right)^d,$$

since $m \leqslant 10^{-7}|U|$. For $d > w/2$ we note that

$$\sum_{M\in\mathcal{M}_{U,d}} \langle \phi_A,\psi_M\rangle^2 \leqslant \|\phi_A\|_2^2 = 2^w,$$

which we plug into (3.8) and get

$$\left\|f^{=2d}\right\|_2^2 \leqslant \frac{\|\varphi\|_2^2}{\|\varphi\|_1^2}\cdot\binom{m}{d}\binom{|U|}{2d}^{-1} \leqslant 2^w\left(\frac{3m}{d}\right)^d\left(\frac{2d}{|U|}\right)^{2d} \leqslant 2^w\cdot\left(\frac{d}{8|U|}\right)^d,$$

since $m \leqslant 10^{-7}|U|$. We thus conclude for any $d \geqslant 1$ that $\left\|f^{=2d}\right\|_2^2 \leqslant 2^{-d}F(|U|,d,w/2)$, as desired. □

3.5 The Decomposition Lemma

In this section, we outline the proof of the decomposition lemma (Lemma 3.29). Readers interested in full details are referred to [8, Section 2].

The general strategy of the proof follows the structure vs. randomness framework of "lifting theorems" in communication complexity, such as in the work [15].

The first step is to establish a decomposition lemma for *sets* in $\Omega^{[n],\alpha n}$.

Lemma 3.51 (Set decomposition)**.** *Let z' be any restriction on $\Omega^{[n],\alpha n}$ (as defined in Definition 3.22), and let $A \subseteq \Omega_{z'}$ be any subset. Then we can decompose A into a disjoint union of subsets $A_{(1)}, A_{(2)}, \ldots, A_{(k)}$ such that:*

1. *Globalness: for each $i \in [k]$, there exists a restriction $z_{(i)}$ that subsumes z' such that $A_{(i)} \subseteq \Omega_{z_{(i)}}$ and $A_{(i)}$ is $z_{(i)}$-global.*

2. Size of the restrictions: the restrictions $z_{(i)}$ satisfy the following inequality:

$$\sum_{i=1}^{k} \frac{|A_{(i)}|}{|A|} \left(\left|\mathrm{supp}(z_{(i)})\right| + \log_2 \frac{|\Omega_{z_{(i)}}|}{|A_{(i)}|} \right) \leqslant |\mathrm{supp}(z')| + \log_2 \frac{|\Omega_{z'}|}{|A|} + 2.$$

We next show how to use Theorem 3.51 to transform any protocol into a "global" protocol. Before doing so, we must formally define what we mean by "global protocols" and how we measure their cost.

Definition 3.52 (Potential function of rectangles)**.** For restrictions $\zeta = (z^{(1)}, \ldots, z^{(K)})$ and a rectangle $R = A^{(1)} \times \cdots \times A^{(K)}$ such that $A^{(i)} \subseteq \Omega_{z^{(i)}}$, we define the potential of (ζ, R) as:

$$p(\zeta, R) := \sum_{i=1}^{K} |\mathrm{supp}(z^{(i)})| + \log_2 \left(\frac{|\Omega_{z^{(i)}}|}{|A^{(i)}|} \right).$$

We now formally define global protocols.

Definition 3.53. A communication protocol Π for $\mathsf{DIHP}(n, \alpha, K)$ is called an r-round global communication protocol if it specifies the following procedure of communications:

- the K players take turns to send messages according Π;
- there are at most r rounds of communications, there is only one player sending message in a single round;
- the length of message in each round of communications is not bounded; instead, from the perspective of rectangles, after each round of communications, a ζ-global rectangle R is further partitioned into several rectangles $R := R_{(1)} \cup \cdots \cup R_{(k)}$ such that: (1) $R_{(i)}$ is $\zeta_{(i)}$-global; (2) $\zeta_{(i)}$ subsumes ζ; (3) the following inequality holds:

$$\sum_{i=1}^{k} \frac{|R_{(i)}|}{|R|} p(\zeta_{(i)}, R_{(i)}) \leqslant p(\zeta, R) + 3.$$

There turns out to be a natural way of "decomposing" an arbitrary communication protocol into a global protocol, by applying the set decomposition lemma (Lemma 3.51) at each communication round.

Lemma 3.54. *Given a communication protocol Π for $\mathsf{DIHP}(n, \alpha, K)$ with communication complexity at most r, we can construct an r-round global protocol Π^{ref} for $\mathsf{DIHP}(n, \alpha, K)$ such that* $\mathrm{adv}(\Pi^{\mathrm{ref}}) \geqslant \mathrm{adv}(\Pi)$.

Note that a global protocol partitions the input space Ω^K into a collection of structured rectangles. It remains to show that for any global protocol, there always exists a subcollection of such rectangles that satisfies the condition in Lemma 3.29. This can be done with some elementary but nontrivial probability calculations.

Chapter 4

Conclusions and Open Problems

The main results of this thesis are:

1. An $O(n^2)$ upper bound on the mixing time of random walk on large monotone subsets of the hypercube. The proof makes use of a new directed isoperimetric inequality for the hypercube.

2. An $\Omega(n^{1/3}/p)$ lower bound on the memory of p-pass streaming algorithms that non-trivially approximates Max-Cut. The proof makes use of a new global hypercontractive inequality for a labeled-matching space.

Although there are many open directions left by these results, we highlight the following two open problems.

Problem 4.1. *Is it possible to improve the $O(n^2)$ mixing time bound in Theorem 1.3 to $O(n \log n)$, as conjectured by Ding and Mossel [5]?*

Problem 4.2. *Prove a polynomial-in-n space lower bound for multi-pass streaming algorithms that nontrivially approximates Max-Cut using* random order[1] *streams.*

The author believes that answering these two questions likely requires fundamentally new techniques beyond those presented in this thesis.

[1] In the random-order streaming model, the edges of the input graph is ordered randomly, instead of adversarially.

References

[1] N. Anari, K. Liu, S. O. Gharan, and C. Vinzant. "Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 1–12.

[2] S. Assadi, G. Kol, R. R. Saxena, and H. Yu. "Multi-pass graph streaming lower bounds for cycle counting, max-cut, matching size, and other problems". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 354–364.

[3] S. Assadi and V. N. "Graph streaming lower bounds for parameter estimation and property testing via a streaming XOR lemma". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 612–625.

[4] E. Cohen. "Problems in catalan mixing and matchings in regular hypergraphs". In: *Georgia Tech PhD thesis* (2016).

[5] J. Ding and E. Mossel. "Mixing under monotone censoring". In: *Electronic Communications in Probability* 19 (2014), pp. 1–6.

[6] Y. Fei. "Improved approximation to first-best gains-from-trade". In: *International Conference on Web and Internet Economics*. Springer. 2022, pp. 204–218.

[7] Y. Fei and R. Ferreira Pinto Jr. "On the Spectral Expansion of Monotone Subsets of the Hypercube". In: *arXiv preprint arXiv:2505.02685* (2025).

[8] Y. Fei, D. Minzer, and S. Wang. "Multi-Pass Streaming Lower Bounds for Approximating Max-Cut". In: *arXiv preprint arXiv:2503.23404* (2025).

[9] R. Ferreira Pinto Jr. "Directed Isoperimetry and Monotonicity Testing: A Dynamical Approach". In: *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2024, pp. 2295–2305.

[10] R. Ferreira Pinto Jr. "Analytic Property Testing: Directed Isoperimetry and Monotonicity". PhD thesis. University of Waterloo, 2025.

[11] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. "Correlation inequalities on some partially ordered sets". In: *Communications in Mathematical Physics* 22 (1971), pp. 89–103.

[12] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. De Wolf. "Exponential separations for one-way quantum communication complexity, with applications to cryptography". In: *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. 2007, pp. 516–525.

[13] O. Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

[14] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samorodnitsky. "Testing Monotonicity". In: *Combinatorica* 3.20 (2000), pp. 301–337.

[15] M. Göös, T. Pitassi, and T. Watson. "Query-to-communication lifting for BPP". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2017, pp. 132–143.

[16] R. van Handel. *Probability in high dimension (Lecture notes)*. 2014.

[17] S. Hoory, N. Linial, and A. Wigderson. "Expander graphs and their applications". In: *Bulletin of the American Mathematical Society* 43.4 (2006), pp. 439–561.

[18] M. Jerrum, A. Sinclair, and E. Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *Journal of the ACM (JACM)* 51.4 (2004), pp. 671–697.

[19] M. Kapralov, S. Khanna, and M. Sudan. "Streaming lower bounds for approximating MAX-CUT". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2015, pp. 1263–1282.

[20] M. Kapralov and D. Krachun. "An optimal space lower bound for approximating MAX-CUT". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 277–288.

[21] N. Keller, N. Lifshitz, and O. Marcus. "Sharp hypercontractivity for global functions". In: *arXiv preprint arXiv:2307.01356* (2023).

[22] S. Khot, D. Minzer, and M. Safra. "On monotonicity testing and boolean isoperimetric-type theorems". In: *SIAM Journal on Computing* 47.6 (2018), pp. 2238–2276.

[23] S. Khot, D. Minzer, and M. Safra. "Pseudorandom sets in Grassmann graph have near-perfect expansion". In: *Annals of Mathematics* 198.1 (2023), pp. 1–92.

[24] D. A. Levin and Y. Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[25] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[26] A. Sinclair and M. Jerrum. "Approximate counting, uniform generation and rapidly mixing Markov chains". In: *Information and Computation* 82.1 (1989), pp. 93–133.