Generalizations and Applications of Hypercontractivity and Small-Set Expansion

Yu Zhao

CMU-CS-21-137

August 26, 2021

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Ryan O'Donnell (Chair)
Anupam Gupta
Venkatesan Guruswami
Rocco Servedio (Columbia University)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This research was sponsored by National Science Foundation awards: CCF1909310; CCF1717606; CCF1618679; and CCF1319743. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: Hypercontractivity, Expander Graphs, Fourier Analysis, Boolean Functions, Hardness of Approximation, Communication Complexity, Information Theory, Decoupling, Query Complexity, Property Testing, Distribution Testing

To my parents.

Abstract

Hypercontractive inequalities and small-set expansion are two fundamental topics closely related to each other and play important roles in many fields, including hardness of approximation, probability theory, social choice theory, information theory, and cryptography. This thesis studies generalizations and applications of hypercontractivity and small-set expansion in the following areas:

- The recent breakthrough proof of the 2-to-2 games conjecture was completed by showing a pseudorandom-set expansion result on Grassmann graphs [KMS18]. A similar property has also been shown on Johnson graphs [KMMS18]. These results can be seen as an improved version of small-set expansion on pseudorandom sets. We prove the pseudorandom-set expansion result on general product probability spaces, with a very clean and short proof. A key step in the proof involves a new hypercontractive-style inequality.
- The communication-assisted agreement distillation problem is about two parties with noisy private randomness trying to extract a common random string via communication. We give the optimal upper bound on the amount of communication necessary for achieving constant success probability for this problem. In addition, we calculate the optimal communication for the reverse binary erasure channel case by studying properties of extreme points in its hypercontractivity region. The proof technique is highly related to the equivalence of hypercontractivity and small-set expansion.
- "Decoupling" refers to the idea of analyzing a complicated random sum involving dependent random variables by comparing it to a simpler random sum where some independence is introduced between the variables. We present a new kind of "one-block decoupling" with better parameters than the classical results. We use decoupling and hypercontractivity to show tight tail bounds of low-degree Boolean functions and tight versions of several theorems from [DFKO07].
- A probability distribution over $\{-1, 1\}^n$ is k-wise uniform if its marginal distribution on every subset of k coordinates is the uniform distribution. These k-wise uniform distributions have the property that all low-degree Fourier coefficients of their density functions are equal to zero. Motivated by this, we use hypercontractive inequalities to study the properties of low-degree Fourier weights of Boolean function. In particular, we show better bounds for the Closeness and Testing problems of k-wise uniformity.

Acknowledgments

I'm thankful for my advisor Ryan O'Donnell. I still remember the first day I met Ryan in his class of Analysis of Boolean functions. From that day on Ryan showed me a splendid wonderland and took me on an exciting adventure. Ryan taught me so much about how to think, speak, teach and write. His brilliant mind, good taste for elegance and simplicity, humility and kindness always inspire me. I appreciate his consistent patience and support to me.

I'd like to thank the other members of my thesis committee, Anupam Gupta, Venkatesan Guruswami and Rocco Servedio, for their time and suggestions during the dissertation process. In addition I'd like to thank my other coauthors, Xiaorui Sun, Li-Yang Tan and John Wright.

Carnegie Mellon University has an incredible group of administrative staff. I'd like to thank Deb Cavlovich, who helped me with every administrative task and tolerated all my mistakes, Charlotte Yano, who guided me when I was volunteering to organize Open House, and Tracy Farbacher, who was helpful when I was in the master program. I also thank Angela Lusk for her counseling and suggestions during the thesis process.

I want to thank my friends in CSD who make my PhD life enjoyable: John Wright, who shared lots of common hobbies with me and is always my best buddy, David Witmer, who helped me polishing my first conference talk, Vijay Bhattiprolu, who spent many working and game-playing nights with me. I'll also miss those baskeball nights with Nicolas Resch and climbing with Colin White. Matthew Mukerjee and David Naylor have been my great officemates and I enjoyed those fun conversations with them. Shen Chen Xu has been my longest housemate and I appreciate his friendship and support. I'd like to thank my countrymen: Haoxian Chen, Yan Gu, Yihan Sun and Yuchen Wu.

I'm also grateful for my friends outside CMU. I'd like to thank Tengyu Ma, who answered my naive questions and gave me advice at the inchoate stage of my academic career, and Huacheng Yu, who hosted me in Stanford Open House. I want to thank Yao Ding, Tiancheng Liu, Wei Lu, Yuanzhi Ma, Jie Shuai, Yifan Wang and Wei Zhang for hanging out with me in my free time. Special thanks for Kun Wang and Cheng Zhang for their hosting during my last period of the thesis process, Ge Fang and Yuzhu Zhang for their accompany and support for my thesis writing.

Finally I would like to thank my parents for their undying support and patience.

Contents

Chapter 1

Introduction

Hypercontractivity and small-set expansion are two fundamental and related topics which play important roles in a variety fields, including several recent breakthroughs in theoretical computer science. In this thesis, we show more generalizations and applications of hypercontractivity and small-set expansion in several domains. Before further discussion, we first give a brief introduction for these two topics.

Hypercontractivity is a property of a finite probability space. Let $p, q \in [1, \infty]$. We say a finite probability space (Ω, π) is (p, q)-*hypercontractive* on some function operator T, if for every function $f : \Omega \to \mathbb{R}$, the following inequality holds:

$$\|\mathrm{T}f\|_q \leq \|f\|_p,$$

where $\|f\|_p = \mathbf{E}[|f(\boldsymbol{x})|^p]^{1/p}$.

Here "-contractivity" means that the function operator T is a contraction of the norms, while "hyper-" means that the inequality is established on two different norms, the L^p- and L^q-norms.

There are two motivations for studying hypercontractive inequalities from interpreting the form literally. First of all, hypercontractivity is a tool which allows one to transfer between different norms, and is especially useful for converting between general L^q-norms and the more standard L^2- and L^1-norms, which are often easier to deal with. Another motivation is that there might exist some useful operator T which is very complicated to understand. In this case, a hypercontractive inequality allows us to study the L^p-norm of the original function instead of the L^q-norm of the function with T acting on it.

As we will show in later chapters, for some specific kinds of function operators, there is a two-function version of the hypercontractive inequality which can be shown equivalent to our above definition via Hölder's inequality. More precisely, we say that a finite joint probability space $(\Omega_x \times \Omega_y, \mu)$ is (p, q)-hypercontractive for some $p, q \in [1, \infty]$, if for any real-valued functions $f : \Omega_x \to \mathbb{R}$, $g : \Omega_y \to \mathbb{R}$, the following inequality holds:

$$\mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[f(\boldsymbol{x})g(\boldsymbol{y})] \leq \|f\|_p \|g\|_{q'},$$

where q' is the Hölder conjugate of q (i.e. the number q' such that $1/q + 1/q' = 1$), and here the L^p- and $L^{q'}$-norms are defined for the marginal distributions on Ω_x and Ω_y. The Hölder conjugate

q' here looks asymmetric, but stating it in this manner allows it to be consistent with the one-function version of hypercontractivity. This two-function hypercontractive inequality can in fact be seen as a generalization of Hölder's inequality.

If we look at functions f and g with a binary output of $\{0,1\}$, there is an interesting combinatorial interpretation of two-function hypercontractivity. Let $A \subseteq \Omega_x$ be the set where $f(x) = 1$ and $B \subseteq \Omega_y$ be the set where $g(y) = 1$. Then we can rewrite the two-function hypercontractive inequality into the following form:

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq |A|^{1/p}|B|^{1/q'},$$

where $|A|$ and $|B|$ are the volumes of the set A and B, respectively, based on the marginal distributions. With proper parameters p and q, this inequality says that the joint distribution μ is a good "small-set expander", meaning that when $\boldsymbol{x}$ is drawn from set A, then with high probability $\boldsymbol{y}$ will end up outside of set B, and vice versa. Therefore, a statement of this form is called a *small-set expansion theorem*. Small-set expansion is related to expander graphs and therefore has several applications in hardness of approximation and coding theory.

The remainder of this chapter expands upon this introduction. We survey the classic results of hypercontractivity in Section 1.1 and small-set expansion in Section 1.2. We explain our contributions in Section 1.3. We give the outline of the rest of the thesis in Section 1.4.

1.1 Hypercontractivity

1.1.1 Classical hypercontractivity results on uniform ± 1 bits

The most important and useful hypercontractivity results deal with functions on the Boolean hypercube with the uniform distribution. Before stating the hypercontractivity results, we first start by introducing the definitions of norms and noise operators in this setting.

Norm and Noise operator

The domain of a Boolean function

$$f : \{-1,1\}^n \to \mathbb{R},$$

is the Boolean hypercube $\{-1,1\}^n$. Here f maps each Boolean string of length n into a real number.

For $p \in [1,\infty)$, we define the usual norm

$$\|f\|_p = (\mathbf{E}[|f(\boldsymbol{x})|^p])^{1/p}.$$

We consider the expectation with respect to a uniformly random $\boldsymbol{x} \sim \{-1,1\}^n$. This is the case for most parts of this thesis when we use the probability notation $\mathbf{Pr}$ and the expectation notation $\mathbf{E}$, unless otherwise specified. As p approaches ∞, the p-norm approaches the infinity norm $\|f\|_\infty = \max_{x\in\{-1,1\}^n}\{|f(x)|\}$.

One of the most important operators in the analysis of Boolean functions is the noise operator T_ρ. The noise operator T_ρ, when applied to a function f, is defined as the expectation of the output of f while adding to each bit of the input an independent amount of noise according to the parameter ρ. Here is the formal definition of the noise operator T_ρ:

Definition 1.1.1. Let $\rho \in [0,1]$. For fixed $x \in \{-1,1\}^n$, we write $\boldsymbol{y} \sim N_\rho(x)$ to denote the random variable $\boldsymbol{y} \in \{-1,1\}^n$ drawn as follows: for each $i \in [n]$,

$$\boldsymbol{y}_i = \begin{cases} x_i & \text{with probability } \rho, \\ \text{uniformly random} & \text{with probability } 1-\rho, \end{cases}$$

independently. We define the *noise operator* T_ρ as the following linear operator:

$$\mathrm{T}_\rho f(x) = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim N_\rho(x)}[f(\boldsymbol{y})].$$

The noise operator was introduced to the analysis of Boolean functions by Bonami [Bon70]. Beckner used the notation T_ρ in [Bec75] and the notation was standardized in [KKL88].

The noise operator is strongly related to other fundamental concepts of Boolean functions, such as degree and influence. In coding theory and information theory, the random variable $\boldsymbol{y} \sim N_\rho(x)$ can be seen as the random string received when passing x through a binary symmetric channel with "flipping probability" of $\frac{1-\rho}{2}$. As a result, the noise operator can be seen as the expectation of the output of a function where the input is received through a binary symmetric channel. Finally, the noise operator also appears in social choice theory; see [Kal02, FKN02] for two examples.

Example 1.1.2. Let $A \subseteq \{-1,1\}^n$ be a subset of the Boolean hypercube with volume α; i.e., $\mathbf{Pr}[\boldsymbol{x} \in A] = \alpha$. We write $1_A : \{-1,1\}^n \to \{0,1\}$ for the indicator function of A; i.e.,

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have $\|1_A\|_p = \alpha^{1/p}$ for any $1 \le p < \infty$ and

$$\mathrm{T}_\rho 1_A(x) = \mathop{\mathbf{Pr}}_{\boldsymbol{y} \sim N_\rho(x)}[\boldsymbol{y} \in A].$$

Example 1.1.3. Consider the single bit function $f : \{-1,1\} \to \mathbb{R}$ in which $f(x) = 1 + \epsilon x$, with parameter $0 \le \epsilon \le 1$. Then we have $\|f\|_p = \left(\frac{1}{2}(1+\epsilon)^p + \frac{1}{2}(1-\epsilon)^p\right)^{1/p}$ for any $1 \le p < \infty$ and $\mathrm{T}_\rho f(x) = 1 + \rho\epsilon x$.

The Hypercontractivity Theorem

Now we are ready to present the Hypercontractivity Theorem. In 1970, Bonami proved the Hypercontractivity Theorem for uniform ± 1 bits in [Bon70]:

Theorem 1.1.4 (The (p,q)-Hypercontractivity Theorem)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$, and let $1 \le p \le q \le \infty$. Then*

$$\|\mathrm{T}_\rho f\|_q \le \|f\|_p$$

for $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$.

The term "hypercontractivity" was introduced in [SHK72]. The suffix "contractivity" describes the fact that T_ρ is a "contraction" or a "smoothing" operator while the prefix "hyper-" indicates that it can be even viewed as a contractive operator from $L^p(\{-1,1\}^n)$ to $L^q(\{-1,1\}^n)$.

Earlier works [Pal32, Bon68, Kie69, Sch69] focus on the hypercontrative properties of homogeneous functions. In 1970 Bonami published her Ph.D. Thesis [Bon70], which contains the full Hypercontractivity Theorem. She first gave a proof for the case $n = 1$ and then extended it to general n by induction. Nelson [Nel73] gave the full Hypercontractivity Theorem in the Gaussian setting independently. Gross [Gro75] derived Nelson's result from the Log-Sobolev Inequalities. See more history of hypercontractivity in [O'D14].

A special case of Theorem 1.1.4 which is of particular interest is when $p = 2$ and $q = 4$, because of the importance of the 2nd and 4th moments. The $(2, 4)$-Hypercontractivity Theorem is also called *the Bonami Lemma* and is strong enough for many well-known applications, as we will see in the next part.

The value $\rho = \sqrt{\frac{p-1}{q-1}}$ in Theorem 1.1.4 is optimal. There are several sharp cases, including Example 1.1.2 when A is a Hamming ball with volume $\alpha \to 0$ and dimension $n \to \infty$, and even the single-bit case from Example 1.1.3 when $\epsilon \to 0$.

1.1.2 Applications of classical hypercontractivity results

Theorem 1.1.4 has applications in many fields of theoretical computer science, such as expander graphs [HLW06], probability theory [BLM13], circuit complexity [LMN89], coding theory [CCH10], and hardness of approximation [KKMO07, DS05].

The Hypercontractivity Theorem for Boolean functions with binary output is also called the Small-Set Expansion Theorem. We will have a detailed introduction to the Small-Set Expansion Theorem and its consequence, the Kahn-Kalai-Linial Theorem, in the next section. Here we start with some other applications of Boolean functions with general real-valued outputs.

Fourier expansion and low-degree functions

One basic fact is that a Boolean function can be represented as a unique multilinear polynomial

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

where parity function $x^S = \prod_{i \in S} x_i$. This is called *Fourier expansion* of the function f, and $\widehat{f}(S)$ is the *Fourier coefficient* of f on set S. The degree of f is defined to be the degree of its Fourier expansion, i.e. the size of the largest S for which $\widehat{f}(S)$ is nonzero.

The parity functions x^S in the Fourier expansion are orthogonal with respect to the uniform measure on $\{-1,1\}^n$. Therefore we can express the p-norm of a Boolean function using its Fourier coefficients whenever p is an integer (though the function needs to be nonnegative when p is odd). In particular, the square of the 2-norm of a Boolean function f can be calculated by the sum of squares of f's Fourier coefficients; i.e., for any $f : \{-1,1\}^n \to \mathbb{R}$,

$$\|f\|_2^2 = \sum_{S \subseteq n} \widehat{f}(S)^2.$$

This fact is called *Parseval's Theorem*. Another useful fact is that $\|f\|_1 = \widehat{f}(\emptyset)$ for any nonnegative Boolean function f.

From the Fourier expansion perspective, the noise operator T_ρ can be interpreted as follows:

$$\mathrm{T}_\rho f(x) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) x^S.$$

The Fourier coefficient of each term shrinks (resp., expands) exponentially in its degree when $0 \leq \rho < 1$ (resp., $\rho > 1$). This is an illustration of why the noise operator tends to be contractive when $0 \leq \rho < 1$.

Combining the Fourier expansion, the $(q, 2)$-Hypercontractivity Theorem and Parseval's Theorem, it is easy to show that a low-degree polynomial of uniform ± 1 bits has a "reasonable" behavior, in the sense that its general q-norm is not too large compared to its 2-norm:

Theorem 1.1.5. *For any Boolean function $f : \{-1, 1\}^n \to \mathbb{R}$ with degree at most k,*

$$\|f\|_q \leq \sqrt{q-1}^k \|f\|_2,$$

for any $q \geq 2$.

Bonami stated Theorem 1.1.5 in [Bon68]. In particular, Theorem 1.1.5 with case $q = 4$ is called *the Bonami Lemma*.

Not only does Theorem 1.1.5 upper-bound the higher norms of a low-degree polynomial, it can also be used to *lower-bound* the p-norm for $1 \leq p \leq 2$ using Hölder's Inequality (as shown in [Jan97]). This is stated as follows.

Theorem 1.1.6. *For any Boolean function $f : \{-1, 1\}^n \to \mathbb{R}$ with degree at most k,*

$$\|f\|_2 \leq (e^{\frac{2}{p}-1})^k \|f\|_p,$$

for any $1 \leq p \leq 2$.

Theorem 1.1.6 with $p = 1$ is useful when dealing with nonnegative Boolean functions with low or fixed expectation, for example the probability mass function of a distribution on Boolean strings of fixed length. We will present an application of this in Chapter 5.

Theorems 1.1.5 and 1.1.6 can be used to get a strong tail bound for low-degree Boolean functions showing that a low-degree function cannot exceed its standard deviation with high probability, but they can also be used to show that it *does* in fact exceed its mean with noticeable probability, at least in a "one-sided" sense (from [PZ78, Bor79, Jan97]).

Theorem 1.1.7. *For any Boolean function $f : \{-1, 1\}^n \to \mathbb{R}$ with degree at most k,*

$$\Pr_{\boldsymbol{x} \sim \{-1,1\}^n}[|f(\boldsymbol{x})| \geq t\|f\|_2] \leq \exp\left(-\frac{k}{2e} t^{2/k}\right),$$

for any $t \geq (2e)^{k/2}$. On the other hand, if f is not a constant function,

$$\Pr_{\boldsymbol{x} \sim \{-1,1\}^n}[f(\boldsymbol{x}) > \mathbf{E}[f]] \geq \frac{1}{4} e^{-2k}.$$

We will utilize Theorem 1.1.7 and present another tail bound for low-degree functions in Chapter 4.

These small and crucial facts of low-degree Boolean functions, especially The Bonami Lemma, have become essential tools appearing in numerous places throughout theoretical computer science. We will now show an influential application related to Chapter 4: the Invariance Principle. Then we will finish this section with a brief summary of hypercontractivity using in other areas.

The Invariance Principle

A multilinear polynomial function whose inputs are independent standard Gaussian random variables can be seen as a special case of a Boolean function, because one can use the sum of many Boolean random bits to "simulate" Gaussian random variables. It is easy to check that the hypercontractivity results for Boolean functions also hold for the low-degree polynomials with independent standard Gaussian random variables. On the other hand, Gaussian random variables also have many good behaviors and are sometimes easier to work with compared to Boolean bits. Mossel et al. [MOO10] showed that in some situations it is also possible to use Gaussian random variables to "simulate" Boolean bits:

Theorem 1.1.8 (The Invariance Principle (informal))**.** *Let the Boolean function f be a low-degree multilinear polynomial with small influences on all coordinates. Replacing the input bits by standard Gaussian independent random variables for f does not change the distribution of its output much.*

The Invariance Principle is one of the most important applications of hypercontractivity to theoretical computer science. The Bonami Lemma for uniform Boolean bits and standard Gaussian random variables plays a key role in its proof.

The main purpose of studying the Invariance Principle in [MOO10] came from the field of hardness of approximation. They deduced that the Majority function is "stablest" on the Boolean hypercube from the Invariance Principle and proved a tight upper-bound on the approximability of the Max-Cut problem assuming the Unique Games Conjecture holds.

1.1.3 Generalizations of hypercontractivity

Product spaces

The Boolean hypercube $\{-1,1\}^n$ is an example of a product domain. In general, we can consider the case of functions $f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ where the domain has the product probability distribution $\pi_1 \otimes \cdots \otimes \pi_n$. We begin by extending ρ-correlation to these general domains.

Definition 1.1.9. Consider a product probability space (Ω, π), where $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and $\pi = \pi_1 \otimes \cdots \otimes \pi_n$. We say $\boldsymbol{y}$ is *ρ-correlated* to $x \in \Omega$ to denote that the random string $\boldsymbol{y}$ is drawn as follows: for each $i \in [n]$ independently,

$$\boldsymbol{y}_i = \begin{cases} x_i & \text{with probability } \rho, \\ \text{under distribution } \pi_i & \text{with probability } 1-\rho; \end{cases}$$

We say that the pair $(\boldsymbol{x}, \boldsymbol{y})$ is *ρ-correlated under* π if $\boldsymbol{x}$ is drawn under distribution π, and then $\boldsymbol{y}$ is ρ-correlated to $\boldsymbol{x}$.

The definition of the noise operator remains the same as in the case of uniform ± 1 bits.

In this situation, the following general hypercontractivity theorem was shown in the earlier works [BKK$^+$92, Tal94, FK96, Fri98]:

Theorem 1.1.10 (The General Hypercontractivity Theorem)**.** *Let $f \in L^2(\Omega, \pi)$, where $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and $\pi = \pi_1 \otimes \cdots \otimes \pi_n$. Let λ be the minimum probability of all outcomes among the distributions $\pi_1, \ldots, \pi_n$. Let $2 \leq q \leq \infty$. Then*

$$\|\mathrm{T}_\rho f\|_q \leq \|f\|_2$$

for $0 \leq \rho \leq \frac{1}{\sqrt{q-1}} \lambda^{1/2-1/q}$.

This bound has the correct asymptotic dependence on λ. One sharp case is a subcube indicator of a Boolean hypercube with a very biased distribution, in which $\lambda \to 0$. However, when $\lambda \to 0$, this inequality becomes very weak and hard to use, partially due to the fact that the range of allowed ρ's is very limited. Researchers have considered several techniques to get rid of this situation. Earlier works [FB99] utilized randomization/symmetrization techniques to reduce biased random variables to symmetric random variables (i.e. uniform ± 1 bits). Keevash et al. [KLLM21] studied global/pseudorandom functions which excludes those bad cases. We study this as well in our thesis; see Chapter 2 for more discussion.

The optimal value of ρ was calculated in [LO94, Wol07] for the case in which each $\boldsymbol{x}_i$ is a mean-zero random variable. The correct asymptotic bound for general (p, q)-hypercontractivity is still unknown to our knowledge.

Similar to the uniform ± 1 case, researchers have found useful applications of the General Hypercontractivity Theorem to the study of low-degree polynomials, as in Section 1.1.2. As mentioned above, these inequalities are useful when λ is large, and become trivial when $\lambda \to 0$.

Non-product spaces

Hypercontractivity has also been studied on Johnson graphs and Grassmann graphs, which was motivated by proving the 2-to-2 Games Conjecture in [KMMS18, KMS18]. These two graphs are tightly related to each other, as well as to Boolean hypercube: the Grassmann graph is known as a "2-analog" of the Johnson graph, and the Johnson graph can be seen as a slice of the Boolean hypercube. See Chapter 2 for more discussion.

Other studies include hypercontractivity results for the multislice [FOW18], the symmetric group [FKLM20], and the Poisson semigroup on the hypersphere [FI21].

1.1.4 Two-function hypercontractivity

There is an equivalent two-function version of hypercontractivity on uniform ± 1 bits:

Theorem 1.1.11 (Two-Function Hypercontractivity Theorem for the uniform distribution)**.** *Let $f, g : \{-1, 1\}^n \to \mathbb{R}$, and let $1 \leq p \leq q \leq \infty$. Then*

$$\mathop{\mathbf{E}}_{\substack{(\boldsymbol{x},\boldsymbol{y}) \\ \rho\text{-correlated}}} [f(\boldsymbol{x})g(\boldsymbol{y})] \leq \|f\|_p \|g\|_{q'}$$

for $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$, where q' is the Hölder conjugate of q: i.e., $\frac{1}{q} + \frac{1}{q'} = 1$.

The Two-Function Hypercontractivty Theorem and its equivalence to the Hypercontractivity Theorem are from [Nev76]. The proof of the equivalence uses Hölder's inequality in both directions. O'Donnell [O'D14] uses two-function hypercontractivity to facilitate induction and simplify the proof of the Hypercontractivity Theorem.

In the previous discussion we focused on ρ-correlated random variables $(\boldsymbol{x}, \boldsymbol{y})$ on finite domains. From the form of the two-function version it is natural to generalize the definition of hypercontractivity to any joint distribution μ of random variables $(\boldsymbol{x}, \boldsymbol{y})$ on the finite domain $\Omega_x \times \Omega_y$. We use $\mu_{\boldsymbol{x}}$ and $\mu_{\boldsymbol{y}}$ to denote the marginal distributions of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. We also use $\mu_{\boldsymbol{x}|y}$ to denote the conditional distribution of $\boldsymbol{x}$ given $\boldsymbol{y} = y$.

In this setting we define the operator T as $\mathrm{T}f(y) = \mathbf{E}_{\mu_{\boldsymbol{x}|y}}[f(\boldsymbol{x})]$. Notice that if the function f is in $L^2(\Omega_x, \mu_x)$ then $\mathrm{T}f$ is in $L^2(\Omega_y, \mu_y)$. The equivalence of the one-function and two-function versions of hypercontractivity still holds in this setting:

Theorem 1.1.12. *In the above setting,*

$$\|\mathrm{T}f\|_q \leq \|f\|_p$$

holds for all $f \in L^2(\Omega_x, \mu_x)$ if and only if

$$\mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[f(\boldsymbol{x})g(\boldsymbol{y})] \leq \|f\|_p \|g\|_{q'}$$

holds for all $f \in L^2(\Omega_x, \mu_x)$ and $g \in L^2(\Omega_y, \mu_y)$, where q' is the Hölder conjugate of q.

The proof is the same as the proof of the ρ-correlated case in [Nev76] with only notational changes. In this setting, we say that random variables $(\boldsymbol{x}, \boldsymbol{y}) \sim \mu$ is (p, q)-*hypercontractive*. It is slightly awkward that in the two-function version p and q are not symmetric (again, p and q' are), but we want to keep the definition consistent with our one-function hypercontractivity results.

The Two-Function Hypercontractivity Theorem is convenient in the field of communication complexity and information theory, where two parties (or the transmitter and receiver) may have different domains and compute different functions.

1.2 Small-set expansion

An expander graph is a graph that has strong connectivity properties. Expander graphs have found extensive applications in computer science, especially in complexity theory, computer networks, coding theory and cryptography. See [HLW06] for a summary.

Consider a regular graph $G = (V, E)$ and a non-empty set of vertices $A \subseteq V$. We write $\boldsymbol{u} \sim \boldsymbol{v}$ to denote that we draw random vertex the $\boldsymbol{v}$ from the uniform distribution on V, and then draw the random vertex $\boldsymbol{u}$ uniformly from $\boldsymbol{v}$'s neighbors. The *edge expansion* of A is defined as:

$$\Phi(A) = \mathop{\mathbf{Pr}}_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{u} \notin A | \boldsymbol{v} \in A].$$

A regular graph G is a good expander if $\Phi(A)$ is high for all $|A| \leq \frac{1}{2}|V|$.

In recent years there has been considerable interest in *small-set expanders*, which are graphs where only sets containing a small fraction of the nodes are required to expand. In this section, we will introduce classical results of small-set expansion on noisy Boolean hypercubes, some generalizations to other graphs and applications, and the relationship between small-set expansion and hypercontractvity.

1.2.1 Small-set expansion on noisy Boolean hypercube

The $(p, 2)$-Hypercontractivity Theorem on uniform ± 1 bits says for any function $f : \{-1, 1\}^n \to \mathbb{R}$ and $1 \leq p \leq 2$,

$$\|\mathrm{T}_{\sqrt{p-1}} f\|_2 \leq \|f\|_p.$$

This theorem does not have a good combinatorial meaning by itself. However, the noise operator T can be interpreted in terms of an important concept known as noise stability:

$$\mathbf{Stab}_\rho[f] = \langle f, \mathrm{T}_\rho f \rangle = \|\mathrm{T}_{\sqrt{\rho}} f\|_2^2 \leq \|f\|_{1+\rho}^2.$$

By focusing on a binary-output function $f : \{-1, 1\}^n \to \{0, 1\}$, as in Example 1.1.2, [KKL88] gave the following interesting way of interpreting the Hypercontractivity Theorem:

Theorem 1.2.1 (Small-Set Expansion Theorem)**.** *Let $A \subseteq \{-1, 1\}^n$ have volume α; i.e., let $1_A : \{-1, 1\}^n \to \{0, 1\}$ satisfy $\mathbf{E}[1_A] = \alpha$. Then for any $0 \leq \rho \leq 1$,*

$$\mathbf{Stab}_\rho[1_A] = \mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x} \sim \{-1,1\}^n \\ \boldsymbol{y} \sim N_\rho(\boldsymbol{x})}} [\boldsymbol{x} \in A, \boldsymbol{y} \in A] \leq \alpha^{\frac{2}{1+\rho}}.$$

That is to say,

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x} \sim \{-1,1\}^n \\ \boldsymbol{y} \sim N_\rho(\boldsymbol{x})}} [\boldsymbol{y} \in A | \boldsymbol{x} \in A] \leq \alpha^{\frac{1-\rho}{1+\rho}}.$$

Consider the edge-weighted hypercube graph $G = (V, E)$ with vertices $V = \{-1, 1\}^n$ and edges $E = V \times V$, where the weight of edge (x, y) is equal to $\mathbf{Pr}[(\boldsymbol{x}, \boldsymbol{y}) = (x, y)]$ when $\boldsymbol{x}, \boldsymbol{y}$ are ρ-correlated ($\rho < 1$). Then Theorem 1.2.1 suggests that for any subset A with small volume α, choosing a random vertex $\boldsymbol{x} \in A$ and a random edge out of $\boldsymbol{x}$ with probability proportional to its weight, we will go outside of A with high probability, at least $1 - \alpha^{\frac{1-\rho}{1+\rho}}$. Therefore this hypercube graph is a good small-set expander.

This graph can also be thought of as the discrete-time Markov chain on state space $\{-1, 1\}^n$ in which a step from state $x \in \{-1, 1\}^n$ consists of moving to state $\boldsymbol{y} \sim N_\rho(x)$. This is a reversible chain with the uniform stationary distribution. Each discrete step is equivalent to running the usual continuous-time Markov chain on the hypercube for time $t = \ln(1/\rho)$ (assuming $0 \leq \rho \leq 1$).

Theorem 1.2.1 is essentially sharp when A is a Hamming ball, as mentioned in the discussion of the Hypercontractivity Theorem.

We can also deduce a general small-set expansion result from Theorem 1.1.10:

Theorem 1.2.2 (General Small-Set Expansion Theorem)**.** *In the same setting as Theorem 1.1.10, let $A \subseteq \Omega$ have volume α; i.e., let $1_A : \Omega \to \{0, 1\}$ satisfy $\mathbf{E}[1_A] = \alpha$. Let $q > 2$. Then for any $0 \leq \rho \leq \frac{1}{q-1} \lambda^{1-2/q}$,*

$$\mathbf{Stab}_\rho[1_A] = \mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x} \sim (\Omega, \pi) \\ \boldsymbol{y} \sim N_\rho(\boldsymbol{x})}} [\boldsymbol{x} \in A, \boldsymbol{y} \in A] \leq \alpha^{2-2/q}.$$

This bound is useful when the minimum outcome parameter λ is large and becomes trivial when $\lambda \to 0$.

1.2.2 Generalizations and applications of small-set expansion

The Kahn-Kalai-Linial Theorem and its generalization

Kahn, Kalai and Linial first mentioned small set expansion of the uniform distribution of the Boolean hypercube in [KKL88], and their main application is the following:

Theorem 1.2.3 (The Kahn-Kalai-Linial Theorem (informal))**.** *For $f : \{-1,1\}^n \to \{-1,1\}$, there exists some coordinate $i \in [n]$, such that the probability that flipping x_i affects the outcome of f is at least $\mathbf{Var}[f] \cdot \Omega\big(\frac{\log n}{n}\big)$.*

The Kahn-Kalai-Linial Theorem is famous for two reasons: first, it pioneers multiple concepts in theoretical computer science, including Fourier expansion, noise stability, influence, and small-set expansion. Second, the KKL Theorem has plenty of applications, including to social choice theory [FKN02] and cryptography [BGM16]. The Kahn-Kalai-Linial Theorem is tight and the best bound is achieved by the tribes functions.

One might think that we can similarly deduce a KKL-style theorem on general product spaces from Theorem 1.2.2. But unfortunately the general small-set expansion result in Theorem 1.2.2 is too weak to prove a good KKL result when λ is small. Bourgain [Bou02] showed that those functions which cannot get a good KKL result must satisfy some specific constraints:

Theorem 1.2.4 (Bourgain's Sharp Threshold Theorem (informal))**.** *For $f \in L^2(\Omega^n, \pi^{\otimes n})$ be $\{0,1\}$ valued. Assume the probability that flipping x_i affects the outcome is small for all coordinates $i \in [n]$ (i.e., the total influence of f is small), and $\mathbf{Var}[f] \geq .01$. Then for a typical input string x, there is a large chance that it contains a constant-sized substring such that the restriction of f to this substring can change the expectation of f by a large amount.*

The constraint looks slightly complicated, but it is very similar to the definition of non-pseudorandomness mentioned in the next part. One key trick in the proof is the randomization/symmetrization technique which reduces biased distributed random variables into uniform ± 1 bits.

Pseudorandom-set expansion

The recent breakthrough proof of the 2-to-2 Games Conjecture is related to the pseudomrandom-set expansion in Grassmann graphs [KMS18]. In Grassmann graphs, not every small set has a nearly perfect expansion. However, [KMS18] showed that most of the small sets expand perfectly except for those sets which are not pseudorandom. Roughly speaking, a pseudorandom set is one in which any constant-size restriction cannot change the density of the set by a lot. This definition is very similar to the constraint in the Bourgain's Sharp Threshold Theorem. Pseudorandomness is also studied on Johnson graphs in [KMMS18]. We will have a detailed discussion and study pseudorandomness on the biased Boolean hypercube in Chapter 2.

1.2.3 Two-set version and equivalence to hypercontractivity

By focusing on $f, g : \{-1,1\}^n \to \{0,1\}$, the Two-Function Hypercontractivity Theorem (Theorem 1.1.11) can also be interpreted as a two-set generalization of the Small-Set Expansion Theorem due to [MOR$^+$06]:

Theorem 1.2.5. *In the setting of Theorem 1.1.12. Let $A \subseteq \Omega_x^n$, $B \subseteq \Omega_y^n$ and random variables $(\boldsymbol{x}, \boldsymbol{y}) \sim ((\Omega_x \times \Omega_y)^n, \mu^{\otimes n})$. Then*

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq |A|^{\frac{1}{p}}|B|^{\frac{1}{q'}},$$

if domain $(\Omega_x \times \Omega_y, \mu)$ is (p, q)-hypercontractive.

For sets A, B with fixed volume, one can take the infimum of $|A|^{1/p}|B|^{1/q'}$ among all p and q such that $(\Omega_x \times \Omega_y, \mu)$ is (p, q)-hypercontractive. Here is an example for ρ-correlated random variables on Boolean hypercube from [O'D14].

Example 1.2.6. Suppose $(\boldsymbol{x}, \boldsymbol{y})$ are ρ-correlated unbiased random variables on $\{-1, 1\}^n$ with $0 < \rho \leq 1$. Let $A, B \subseteq \{-1, 1\}^n$ have volume $\exp(-\frac{a^2}{2})$ and $\exp(-\frac{b^2}{2})$, and let $0 \leq \rho a \leq b \leq \frac{a}{\rho}$. Then

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq \exp\left(-\frac{1}{2}\frac{a^2 - 2\rho ab + b^2}{1 - \rho^2}\right).$$

This bound is sharp in the case when A and B are concentric Hamming balls.

Until recently, the Small-Set Expansion Theorem was seen as a binary-output special case of the Hypercontractivity Theorem. However, [Nai14] showed that these two theorems are in fact equivalent.

Theorem 1.2.7. *In the setting of Theorem 1.1.12, $(\Omega_x \times \Omega_y, \mu)$ is (p, q)-hypercontractive if and only if*

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq |A|^{\frac{1}{p}}|B|^{\frac{1}{q'}},$$

holds for any n, $A \subseteq \Omega_x^n$, and $B \subseteq \Omega_y^n$.

We remark that for hypercontractivity the cases of $n = 1$ and general n are equivalent by induction, but the small set expansion statement requires n to be general, and in fact the sharp case may happen when $n \to \infty$.

Nair studied this equivalence of hypercontractivity and small-set expansion using information measures, and this result might be overlooked by other fields. We will extend the discussion, restate Nair's proof, and give an application in communication complexity in Chapter 3.

1.3 Problem studied

In this thesis, we will calculate the parameter domain of hypercontractivity and small-set expansion for some special cases. We will also give applications of hypercontractivity and small-set expansion for some problems in coding theory and complexity theory.

Pseudorandom-set expansion. A recent breakthrough of proving the 2-to-2 games conjecture is completed by showing the pseudorandom-set expansion on Grassmann graphs [KMS18]. Roughly speaking, if any subset of vertices on Grassmann graph is "pseudorandom" enough, it will have almost full expansion on the graph. A similar property is also shown on Johnson graphs [KMMS18]. These pseudorandom-set expansion results can be seen as an improvement of small-set expansion for special cases. We prove the pseudorandom-set expansion on biased Boolean

cube as an analog of that on Johnson graphs, with a very short and comprehensive proof. Our goal is to give an analog of Grassmann graph expansion and hope to inspire further directions for the unique games conjecture.

Communication distillation. The communication distillation problem is about two parties with noisy private randomness trying to extract a common random string via communication. We show that the upper and lower bounds of this problem are both related to the small-set expansion based on the work of [AC98, GR11]. We also show that communication distillation with high probability is related to some properties of extreme points in the hypercontractivity domain.

Decoupling. The decoupling method refers to the idea of analyzing a complicated random sum involving dependent random variables by comparing it to a simpler random sum where some independence is introduced between the variables. Decoupling applies in multiple areas, including randomly stopped processes and unbiased estimation. Roughly speaking, the decoupling method transforms a polynomial of random variables into its "multilinear version". This multilinear property is convenient. For example we can apply well-studied hypercontractivity results.

Let $f(x) = f(x_1, \ldots, x_n) = \sum_{|S| \le k} a_S \prod_{i \in S} x_i$ be an n-variate real multilinear polynomial of degree at most k, where $S \subseteq [n] = \{1, 2, \ldots, n\}$. For its *one-block decoupled* version,

$$\breve{f}(y, z) = \sum_{|S| \le k} a_S \sum_{i \in S} y_i \prod_{j \in S \setminus \{i\}} z_j,$$

we show tail-bound comparisons of the form

$$\mathbf{Pr}\Big[\Big|\breve{f}(\boldsymbol{y}, \boldsymbol{z})\Big| > C_k t\Big] \le D_k \, \mathbf{Pr}\Big[|f(\boldsymbol{x})| > t\Big].$$

Our constants C_k, D_k are significantly better than those known for "full decoupling". For example, when $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are independent Gaussians we obtain $C_k = D_k = O(k)$; when $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are ± 1 random variables we obtain $C_k = O(k^2)$, $D_k = k^{O(k)}$. By contrast, for full decoupling only $C_k = D_k = k^{O(k)}$ is known in these settings.

We describe consequences of these results for query complexity (related to conjectures of Aaronson and Ambainis) and for analysis of Boolean functions (including an optimal sharpening of the DFKO Inequality).

Property testing on k-wise uniformity. The density function of a k-wise uniform distribution has zero coefficients for any monomial of degree at most k, except for the constant term (which is always equal to 1). In other word, the Fourier expansion of the function has only high-degree monomials, which is the opposite case of the low-degree function. However we can transform the optimization of k-wise uniform distribution into analyzing low-degree polynomials via duality. Then we can apply the low-degree inequalities, as an application of the hypercontractivity.

A probability distribution over $\{-1, 1\}^n$ is (ϵ, k)-*wise uniform* if, roughly, it is ϵ-close to the uniform distribution when restricted to any k coordinates. We consider the problem of how far an (ϵ, k)-wise uniform distribution can be from any globally k-wise uniform distribution. We show that every (ϵ, k)-wise uniform distribution is $O(n^{k/2}\epsilon)$-close to a k-wise uniform distribution in

total variation distance. In addition, we show that this bound is optimal for all even k: we find an (ϵ, k)-wise uniform distribution that is $\Omega(n^{k/2}\epsilon)$-far from any k-wise uniform distribution in total variation distance. For $k = 1$, we get a better upper bound of $O(\epsilon)$, which is also optimal.

One application of our closeness result is to the sample complexity of testing whether a distribution is k-wise uniform or δ-far from k-wise uniform. We give an upper bound of $O(n^k/\delta^2)$ (or $O(\log n/\delta^2)$ when $k = 1$) on the required samples. We show an improved upper bound of $\tilde{O}(n^{k/2}/\delta^2)$ for the special case of testing fully uniform vs. δ-far from k-wise uniform. Finally, we complement this with a matching lower bound of $\Omega(n/\delta^2)$ when $k = 2$.

Our results improve upon the best known bounds from [AAK$^+$07], and have simpler proofs.

1.4 Outline

The thesis is organized as follows: Chapter 2 covers pseudorandom-set expansion on product probability spaces; Chapter 3 covers the equivalence of hypercontractivity and small-set expansion and its application to communication agreement distillation; Chapter 4 covers decoupling results; Chapter 5 covers property testing of k-wise uniformity.

Chapter 4 is based on work from [OZ16]. Chapter 5 is based on work from [OZ18]. Chapter 2 is based on unpublished joint work with Ryan O'Donnell; a similar result was also shown in [KLLM21] independently from our work. Most of Chapter 3 is based on unpublished joint work with Venkat Guruswami and Ryan O'Donnell, except for Section 3.6, which is from [NW16].

Chapter 2

Pseudorandom-Set Expansion on Product Probability Spaces

2.1 Introduction

In this chapter, we focus on expansion properties of product probability spaces, especially on the noisy Boolean hypercube with biased distribution. In particular, we are interested in the class of sets on this graph satisfying a property called pseudorandomness.

2.1.1 Pseudorandom-set expansion

Pseudorandom-set expansion appears in the milestone breakthrough proof of the 2-to-2 Games Conjecture. A line of works [KMS17, DKK$^+$18] reduces an NP-hard problem to the problem of 2-to-2 games on Grassmann graphs, and in the process these works end up showing an expansion property of small sets in Grassmann graphs [DKK$^+$21]. Unfortunately, not every small set has nearly perfect expansion, as one might hope, but luckily to prove the 2-to-2 Games Conjecture it suffices to work with only *pseudorandom* sets, in the sense that under any small co-dimensional subspace restriction, the density of the set will not increase too much (this turns out to be very similar to the idea of closeness to k-wise uniformity mentioned in Chapter 5). Pseudorandom-set expansion was finally proved in [KMS18], which finished the proof of the 2-to-2 Game Conjecture (at least with imperfect completeness).

In this chapter we will prove the near-perfect expansion property of pseudorandom sets on product probability spaces. This is an analog of the results in [KMS18, KMMS18] for product probability spaces. For example, the Boolean hypercube with biased distribution is a special case of this generalized domain in which $\Omega = \{-1, 1\}^n$ and $\pi = \pi_\lambda^{\otimes n}$ where $\pi_\lambda(-1) = \lambda$, $\pi_\lambda(1) = 1 - \lambda$. Here is the general definition of pseudorandomness on product probability spaces.

Definition 2.1.1. Let (Ω, π) be a finite probability space, where the finite set $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and the distribution $\pi = \pi_1 \otimes \cdots \otimes \pi_n$ are n-dimentional. A subset $A \subseteq \Omega$ is called (k, ϵ)-*pseudorandom* if for any restriction $J|z$ with size $|J| \leq k$, we have

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim (\Omega,\pi)}[\boldsymbol{x} \in A | \boldsymbol{x} \in (z \times \Omega_{\bar{J}})] \leq \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim (\Omega,\pi)}[\boldsymbol{x} \in A] + \epsilon.$$

This definition is an analog of pseudorandom sets on Johnson graphs in [KMMS18] and Grassmann graphs in [KMS18]. The restriction part is more similar to the definition on Grassmann graphs in [KMS18], in the sense that having different values for each coordinate of z is an analog of the "zoom-in" and "zoom-out" subspace restrictions in [KMS18]. The definition of pseudorandomness on Johnson graphs in [KMMS18] is comparable to only considering z being the all-ones vector, and a similar definition also appears in [KLLM21].

To describe the expansion property of sets on general product probability spaces, we first extend our classical definitions of the noise operator to general product probability spaces.

Definition 2.1.2. We say $\boldsymbol{y}$ is *ρ-correlated* to $x \in \Omega$ to denote that the random string $\boldsymbol{y}$ is drawn as follows: for each $i \in [n]$ independently,

$$\boldsymbol{y}_i = \begin{cases} x_i & \text{with probability } \rho, \\ \text{under distribution } \pi_i & \text{with probability } 1-\rho; \end{cases}$$

We say pair $(\boldsymbol{x}, \boldsymbol{y})$ is *ρ-correlated* if $\boldsymbol{x}$ is drawn under distribution π, and then $\boldsymbol{y}$ is ρ-correlated to $\boldsymbol{x}$.

We want to emphasize the relationship between the noisy Boolean hypercube with biased distribution and the Johnson graph. The Johnson graph $J(n,l,t)$ is the graph whose nodes are sets of size l in a universe of size n. Two sets have an edge if and only if the cardinality of their intersection is equal to t. Let $(\boldsymbol{x}, \boldsymbol{y})$ be ρ-correlated variables on the Boolean hypercube $\{-1,1\}^n$ with λ-biased distribution. It is also easy to check that in expectation there are λn bits of $\boldsymbol{x}$ and $\boldsymbol{y}$ with value -1. It is easy to check that when fixing $\boldsymbol{x}$ to have λn bits of -1's, the random variable $\boldsymbol{y}$ will still have λn bits of -1's in expectation and will share $(\rho + (1-\rho)\lambda)\lambda n$ bits of -1's with $\boldsymbol{x}$ in expectation. Therefore, the Johnson graph $J(n,l,t)$ can be seen as an "expectation version" of the ρ-correlated weighted graph on the Boolean hypercube with a λ-biased distribution where $l = \lambda n, t = (\rho + (1-\rho)\lambda)\lambda n$.

Our main result is the following nearly perfect expansion property for pseudorandom sets:

Theorem 2.1.3 (Pseudorandom-set expansion)**.** *In the setting of Definition 2.1.1 and 2.1.2, for every $0 \le \rho < 1$ and $0 < \eta \le 1$, when $k \in \mathbb{N}$ and $\epsilon > 0$ satisfy $(1152)^{k/3}\epsilon \le (\eta - \rho^k)^3$, the following holds. If subset $A \subseteq \Omega$ is (k,ϵ)-pseudorandom, then*

$$\Pr_{\substack{(\boldsymbol{x},\boldsymbol{y}) \\ \rho\text{-correlated}}} [\boldsymbol{y} \in A | \boldsymbol{x} \in A] \le \eta.$$

This result is most related to the pseudorandom set expansion on Johnson Graphs studied in [KMS18]. The Johnsonn Graph can be treated as one slice of the Boolean hypercube, and therefore it has a lot of similar properties as the Boolean hypercube under a biased distribution. The proof in [KMMS18] served as the inspiration for the paper [KMS18], which gave a proof of pseudorandom set expansion on Grassmann graphs, which can be seen as q-analog of Johnson graphs.

The advantage of our result is that the proof is very simple and clear and has a short length. The product probability spaces is easier to study than the Johnson graphs because there are more tools available. The key part of our proof is a hypercontractive inequality, which we prove in Lemma 2.1.4 below. We hope this result can lead to a simplification and improvement of the proof of the 2-to-2 Games Conjecture.

2.1.2 Hypercontractivity on biased Boolean hypercube

The classical hypercontractivity theorem for uniform ± 1 bits is tight and has a bunch of fundamental applications in the analysis of Boolean functions, such as the Kahn-Kalai-Linial Theorem [KKL88] and the Invariance Principle [MOO10]. However the hypercontractivity theorem seems to be not that powerful for studying extremely biased distributions. For example, in the useful $(2, 4)$-Hypercontractivity Theorem,

$$\|\mathrm{T}_\rho f\|_4 \leq \|f\|_2,$$

we can set $\rho = 1/\sqrt{3}$ to ensure the inequality holds for any $f \in L^2(\{-1, 1\}^n, \pi_\lambda^{\otimes n})$ with $\lambda = 1/2$, or in other words the norms are calculated for uniform ± 1 bits. However when we consider the general biased distribution, in particular when the lowest probability $\lambda = o(1)$, then the correlation parameter ρ can only taken to be $\rho \leq O(\lambda^{1/4})$ due to [LO94]. This correlation parameter ρ is so small that the General Hypercontractivity Theorem becomes ineffective for analogs of the KKL Theorem and other applications on extremely biased ± 1 bits. One tight case might be the dictatorship function, e.g. $f(x) = x_1$. One can easy to check that in this case $\|T_\rho f\|_4 \geq \rho\lambda^{1/4}$ while $\|f\|_2 = \lambda^{1/2}$. Keevash et al. [KLLM21] observed this phenomenon, and noted that the key property of these "bad" examples is *locality*, in the sense that a small number of coordinates can significantly influence the output of the function. They suggested that by excluding these local functions, a stronger hypercontractivity theorem may hold for *global* functions, and they showed several strengthened biased distributed analogs of the KKL Theorem and other applications when focusing on these global functions.

Our result of hypercontractivity inequality is following:

Lemma 2.1.4. *Let $f \in L^2(\Omega_1 \times \cdots \times \Omega_n, \pi_1 \otimes \cdots \otimes \pi_n)$ and $k \leq n$. There exists a restriction $J \subseteq [n]$ and $z \in \Omega_J$ with $|J| \leq k$ such that*

$$\|f^{\leq k}\|_4^4 \leq (1152)^k \|f^{\leq k}\|_2^2 \|f_{\bar{J}|z}\|_2^2.$$

Notice that this inequality does not depend on any parameter of the distribution π at all. No matter how biased the distribution is, this inequality is strong whenever f has small $\|f_{\bar{J}|z}\|_2$ for any restriction $J \subseteq [n]$ and $z \in \Omega_J$. Keevash et al. [KLLM21] obtained a similar result with slightly different conditions. They only consider restrictions with an all-ones vector z, but they need the function f to be monotone or small density to make the hypercontractive inequality hold. Another difference is that our result applies not only on the Boolean hypercube, but also on any arbitrary product probability spaces.

The proof is inspired by randomization/symmetrization tricks used in the proof of Bourgain's Sharp Threshold Theorem [FB99].

2.1.3 Related work

Hypercontractivity on general finite spaces. Earlier works [BKK$^+$92, Tal94, FK96, Fri98] had established forms of the General Hypercontractivity Theorem for λ-biased bits, giving as applications KKL-type theorems in this setting with the correct asymptotic dependence on λ. The optimal $(2, q)$-hypercontractivity parameter for λ-biased Boolean bits is obtained in [LO94].

The case of general discrete random variables is a reduction to the two-valued case due to Wolff [Wol07]. Keevash et al. [KLLM21] studied the hypercontractivity of global functions for λ-biased Boolean bits independently from our work. The definitions of pseudorandomness/globalness are slightly different between their work and ours. As discussed above, their condition of globalness/pseudorandomness is slightly less restrictive but they need extra small density or monotonicity properties to ensure the pseudorandom-set expansion property. The noise stability of monotone functions for biased Boolean bits is also considered in [LM19].

Pseudorandom sets expansion & the 2-to-2 Games Conjecture. The definition of pseudorandom sets on Grassmann graphs first pops up in [DKK$^+$21], where they suggest a better understanding of these expansion properties of pseudorandom sets is required for the 2-to-2 Games problem. Previous works [KMS17, DKK$^+$18] connected Grassmann graphs to the 2-to-2 Games Conjecture by showing a reduction from an NP-hard problem to an instance of the 2-to-2 games problem on a Grassmann graph. Pseudorandom set expansion on Grassmann graphs was proved in [KMS18], which completed the proof of the 2-to-2 Games Conjecture (with imperfect completeness). The proof of pseudorandom set expansion on Grassmann graphs is inspired by the proof of an easier analog, pseudorandom set expansion on Johnson graphs in [KMMS18].

Biased Boolean hypercubes vs. Johnson graphs. The slice of the Boolean hypercube has been previously studied in algebraic combinatorics, where it is referred to as the "Johnson association scheme", and in spectral graph theory in relation to the Johnson graphs. An earlier work [LY98] proved a hypercontractivity property of Johnson graphs. Yuval Filmus [Fil16] presents an orthogonal basis for functions over a slice of the Boolean hypercube, which generalized lots of results from the Boolean hypercube to Johnson schemes/graphs, e.g., Friedgut's Theorem [Wim14], linearity testing [DDG$^+$17], the Invariance Principle [FM19, FKMW18], and low-degree spectral concentration [FI19]. The pseudorandom set expansion in the Johnson graph was proved in [KMMS18]. They mentioned the analog of their result in the noisy biased Boolean hypercube as an open problem, which is what we prove in this chapter.

Johnson graphs vs. Grassmann graphs and q-analogs. Many of the parameters of Grassmann graphs are q-analogs of the parameters of Johnson graphs (the $q = 2$ case of Grassmann graphs is the case relevant to 2-to-2 Games Conjecture), and Grassmann graphs have several of the same graph properties as Johnson graphs. The Johnson graph analysis is a special but crucial case of the Grassmann graph analysis in the proof of pseudorandom set expansion [KMMS18, KMS18]. Boolean analysis of q-eposet (q-simplicial complex) is studied in [DDFH18], with an application of pseudorandom set expansion on q-simplicial complexes in [HKL20]. One future direction of our work might be extending the result to the q-analog of the Boolean hypercube, and doing so might simplify the proof of pseudorandom set expansion on Grassmann graphs.

Randomization/symmetrization. Kahane [Kah93] has been credited with the early development of the randomization/symmetrization technique for random variables. The comparison of norms of a function and its randomization/symmetrization is due to [Bou79]. Our

proof is inspired by the proof of Bourgain's Theorem in [FB99, Bal13] utilizing the randomization/symmetrization tricks.

2.2 Preliminaries

Let $\pi = \pi_1 \times \cdots \times \pi_n$ be a product probability distribution on finite set $\Omega = \Omega_1 \times \cdots \times \Omega_n$.

2.2.1 Orthogonal decomposition on generalized domains

Theorem 2.2.1. *Every function $f \in L^2(\Omega, \pi)$ has a unique decomposition as follows:*

$$f = \sum_{S \subseteq [n]} f^{=S},$$

where $f^{=S}$ depends only on the coordinates in S, and for any function $g \in L^2(\Omega, \pi)$ which only depends on coordinates $T \subsetneq S$, $\langle f^{=S}, g \rangle = 0$.

This decomposition is orthogonal so it is similar to the Fourier expansion in many ways.

Proposition 2.2.2. *For any $f, g \in L^2(\Omega, \pi)$,*

1. *$\langle f^{=S}, g^{=T} \rangle = 0$ if $S \neq T$;*
2. *$\langle f, g \rangle = \sum_{S \subseteq [n]} \langle f^{=S}, g^{=S} \rangle$;*
3. *$\mathrm{T}_\rho f(x) = \mathbf{E}_{\boldsymbol{y} \sim N_\rho(x)}[f(\boldsymbol{y})] = \sum_{S \subseteq [n]} \rho^{|S|} f^{=S}(x)$.*

The definition of pseudo-random sets involves fixing some coordinates to constant values. We introduce the following notation:

Definition 2.2.3. Let $f \in L^2(\Omega, \pi)$. For set $J \subseteq [n]$ we denote $\Omega_J = \otimes_{i \in J} \Omega_i$ and $\pi_J = \otimes_{i \in J} \pi_i$. Let $(J, \bar{J})$ be a partition of [n]. Let $z \in \Omega_J$. We write $f_{\bar{J}|z} \in L^2(\Omega_{\bar{J}}, \pi_{\bar{J}})$ for the subfunction of f by fixing the coordinates in J to z.

Note that for indicator function 1_A of subset $A \subseteq \Omega$, its restricted function $(1_A)_{\bar{J}|z}$ indicates the subset $A \cap (z \times \Omega_{\bar{J}})$. Hence

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim (\Omega, \pi)}[\boldsymbol{x} \in A | \boldsymbol{x} \in z \times \Omega_{\bar{J}}] = \mathop{\mathbf{E}}_{\boldsymbol{x}_{\bar{J}} \sim (\Omega_{\bar{J}}, \pi_{\bar{J}})}[(1_A)_{\bar{J}|z}] = \|(1_A)_{\bar{J}|z}\|_q^q$$

for any $q > 0$.

One easily checks the following formula between f and $f_{\bar{J}|z}$:

Proposition 2.2.4. *For any $S \subseteq \bar{J}$,*

$$(f_{\bar{J}|z})^{=S}(x_S) = \sum_{I \subseteq J} f^{=I \cup S}(z_I, x_S).$$

In other words,

$$f^{=J \cup S}(z, x_S) = \sum_{I \subseteq J} (-1)^{|J|-|I|} (f_{\bar{I}|z_I})^{=S}(x_S).$$

See Chapter 8 in [O'D14] for more details and the origins of these discussions.

2.2.2 Randomization/symmetrization technique

One key part of the proof of Bourgain's Sharp Threshold Theorem is the randomization/symmetrization technique, which introduces independent uniformly random bits and reduces the analysis to uniformly random ± 1 bits. Here is the formal definition:

Definition 2.2.5. Let $f \in L^2(\Omega, \pi)$. The randomization/symmtrization of f is the function $\widetilde{f} \in L^2(\Omega \times \{-1,1\}^n, \pi \times \pi_{1/2}^{\otimes n})$ defined by

$$\widetilde{f}(\boldsymbol{r}, \boldsymbol{x}) = \sum_{S \subseteq [n]} \boldsymbol{r}^S f^{=S}(\boldsymbol{x}).$$

The essential feature of the randomization/symmetrization is that the q-norms do not change that much, as shown in [Bou79].

Proposition 2.2.6. *Let $f \in L^2(\Omega, \pi)$,*

1. $\|\widetilde{f}\|_2 = \|f\|_2$*;*
2. $\|f\|_q \le \|\widetilde{T_2 f}\|_q$ *for $q \ge 1$.*

2.3 Proofs

2.3.1 Proof of Theorem 2.1.3

We prove the following function version of Theorem 2.1.3.

Theorem 2.3.1. *Let $f \in L^2(\Omega, \pi)$ and $0 \le \rho \le 1$ and $k \in \mathbb{N}$. There exists a restriction $J \subseteq [n]$ and $z \in \Omega_J$ with $|J| \le k$ such that*

$$\langle f, \mathrm{T}_\rho f \rangle \le (1152)^{k/3} \|f\|_{4/3}^{4/3} \|f_{\bar{J}|z}\|_2^{2/3} + \rho^k \|f\|_2^2.$$

Theorem 2.1.3 can be seen as a special case of Theorem 2.3.1 on the indicator function 1_A.

Theorem 2.3.1 can be easily deduced from the following two steps combining with 2.1.4.

Claim 2.3.2. *Let $f \in L^2(\Omega, \pi)$ and $0 \le \rho \le 1$ and $k \in \mathbb{N}$. Then*

$$\langle f, \mathrm{T}_\rho f \rangle \le \|f^{\le k}\|_2^2 + \rho^k \|f\|_2^2.$$

Proof. Writing f in terms of its orthogonal decomposition, we have:

$$\langle f, \mathrm{T}_\rho f \rangle = \sum_{S \subseteq [n]} \rho^{|S|} \langle f^{=S}, f^{=S} \rangle = \sum_{S \subseteq [n]} \rho^{|S|} \|f^{=S}\|_2^2.$$

Then we conclude

$$\begin{aligned} \sum_{S \subseteq [n]} \rho^{|S|} \|f^{=S}\|_2^2 &= \sum_{|S| \le k} \rho^{|S|} \|f^{=S}\|_2^2 + \sum_{|S| > k} \rho^{|S|} \|f^{=S}\|_2^2 \\ &\le \sum_{|S| \le k} \|f^{=S}\|_2^2 + \rho^k \sum_{|S| > k} \|f^{=S}\|_2^2 \\ &\le \|f^{\le k}\|_2^2 + \rho^k \|f\|_2^2. \end{aligned}$$

□

Claim 2.3.3. *Let $f \in L^2(\Omega, \pi)$ and $k \in \mathbb{N}$. Then*

$$\|f^{\leq k}\|_2^2 \leq \|f\|_{4/3}\|f^{\leq k}\|_4.$$

Proof. Apply orthogonality of the decomposition and then Hölder's inequality:

$$\|f^{\leq k}\|_2^2 = \langle f, f^{\leq k}\rangle \leq \|f\|_{4/3}\|f^{\leq k}\|_4. \qquad \square$$

2.3.2 Proof of Lemma 2.1.4

Proof of Lemma 2.1.4. We start with the symmetrization technique. Let $g = \mathrm{T}_2 f^{\leq k}$. Let $\boldsymbol{r} \sim \{-1,1\}^n$ be independent uniformly random bits. Then we have

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}\mathop{\mathbf{E}}_{\boldsymbol{r}}\left[\widetilde{g}(\boldsymbol{x},\boldsymbol{r})^4\right] \leq \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[9^k \mathop{\mathbf{E}}_{\boldsymbol{r}}\left[\widetilde{g}(\boldsymbol{x},\boldsymbol{r})^2\right]^2\right] = 9^k \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\left(\sum_{|S|\leq k} g^{=S}(\boldsymbol{x})^2\right)^2\right].$$

The inequality is by Bonami's Lemma and the equality follows from Parseval's Theorem on the random variable $\boldsymbol{r}$. Recalling the definition of g we conclude

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[f^{\leq k}(\boldsymbol{x})^4] \leq \mathop{\mathbf{E}}_{\boldsymbol{x}}\mathop{\mathbf{E}}_{\boldsymbol{r}}\left[\widetilde{g}(\boldsymbol{x},\boldsymbol{r})^4\right] \leq 9^k \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\left(\sum_{|S|\leq k}(2^k f^{=S}(\boldsymbol{x}))^2\right)^2\right] = (144)^k \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\left(\sum_{|S|\leq k}(f^{=S}(\boldsymbol{x}))^2\right)^2\right].$$

Next, we open up the square inside the expectation, and break up the resulting double-sum over S, T, according to $I := S \cup T$:

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\left(\sum_{|S|\leq k}(f^{=S}(\boldsymbol{x}))^2\right)^2\right] = \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\sum_{|I|\leq k}\sum_{\substack{|S|\leq k\\ S\supseteq I}} f^{=S}(\boldsymbol{x})^2 \left(\sum_{\substack{|T|\leq k\\ S\cap T = I}} f^{=T}(\boldsymbol{x})^2\right)\right]. \tag{2.1}$$

We carefully switch the order of product and expectations based on $S \cap T = I$. Then we include the additional nonnegative terms by dropping the condition that $S \cap T = I$. Thus

$$(2.1) \leq \sum_{|I|\leq k}\mathop{\mathbf{E}}_{\boldsymbol{x}_I}\left[\left(\sum_{\substack{|S|\leq k\\ S\supseteq I}}\mathop{\mathbf{E}}_{\boldsymbol{x}_{S\setminus I}}[f^{=S}(\boldsymbol{x}_S)^2]\right)\left(\sum_{\substack{|T|\leq k\\ T\supseteq I}}\mathop{\mathbf{E}}_{\boldsymbol{x}_{T\setminus I}}[f^{=T}(\boldsymbol{x}_T)^2]\right)\right].$$

We upper-bound the second factor by taking a maximum over the set I and its assignment y_I:

$$(2.1) \leq \left(\sum_{|I|\leq k}\sum_{\substack{|S|\leq k\\ S\supseteq I}}\mathop{\mathbf{E}}_{\boldsymbol{x}_S}[f^{=S}(\boldsymbol{x}_S)^2]\right)\max_{\substack{|I|\leq k\\ y_I\in\Omega_I}}\sum_{\substack{|T|\leq k\\ T\supseteq I}}\mathop{\mathbf{E}}_{\boldsymbol{x}_{T\setminus I}}[f^{=T}(y_I,\boldsymbol{x}_{T\setminus I})^2]. \tag{2.2}$$

In the first factor, for each set S, $f^{=S}$ is counted $2^{|S|}$ times, and therefore

$$\sum_{|I|\le k}\sum_{\substack{|S|\le k\\ S\supseteq I}}\mathop{\mathbf{E}}_{\boldsymbol{x}_S}[f^{=S}(\boldsymbol{x}_S)^2]=\sum_{|S|\le k}2^{|S|}\mathop{\mathbf{E}}_{\boldsymbol{x}_S}[f^{=S}(\boldsymbol{x}_S)^2]\le 2^k\mathop{\mathbf{E}}_{\boldsymbol{x}}[\sum_{|S|\le k}f^{=S}(\boldsymbol{x})^2]=2^k\|f^{\le k}\|_2^2. \quad (2.3)$$

For the second factor, one easily checks that

$$f^{=I\cup S}(y_I,x_S)=\sum_{J\subseteq I}(-1)^{|I|-|J|}(f_{J|y_J})^{=S}(x_S).$$

Hence by dropping the condition that $|T|\le k$ and including the additional nonnegative terms we get

$$\sum_{T\supseteq I}\mathop{\mathbf{E}}_{\boldsymbol{x}_{T\setminus I}}[f^{=T}(y_I,\boldsymbol{x}_{T\setminus I})^2]\le 2^{|I|}\sum_{J\subseteq I}\sum_{S\subseteq\bar{I}}\mathop{\mathbf{E}}_{\boldsymbol{x}_S}[(f_{J|y_J})^{=S}(\boldsymbol{x}_S)^2]\le 4^k\max_{J\subseteq I}\|f_{J|y_J}\|_2^2. \quad (2.4)$$

The first ineqauality is Cauchy-Schwarz and the second inequality is from adding nonnegative terms by changing $S\subseteq\bar{I}$ to $S\subseteq\bar{J}$. We thus conclude the proof by plugging (2.3) and (2.4) into (2.2). □

Chapter 3

Communication Assisted Agreement Distillation and Hypercontractivity Region

3.1 Introduction

3.1.1 Communication assisted agreement distillation

Consider the following communication problem: let μ be an arbitrary joint distribution on the finite domain $\Omega_x \times \Omega_y$. Consider a pair of random strings $(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n}$ drawn from the product probability on $(\Omega_x \times \Omega_y)^n$. Alice receives string $\boldsymbol{x}$ and Bob receives string $\boldsymbol{y}$. The goal is for Alice and Bob to agree on a uniformly random string in $\{0,1\}^k$, using as little communication between each other as possible.

The agreement distillation problem is very natural in the context of imperfect sharing of randomness. The pair of random strings $(\boldsymbol{x}, \boldsymbol{y})$ from the joint distribution μ represents "imperfectly shared randomness", something in the middle of "private randomness" and "perfectly shared randomness". An efficient scheme of agreement distillation will convert any communication protocol with public (perfectly shared) randomness into one that relies only on imperfect shared randomness. The work [BGI14] studied simulating any communication protocol with public randomness by a protocol with imperfect shared randomness in the simultaneous message passing setting. See [CGMS17] for more discussion of communication with imperfectly shared randomness. The agreement distillation is also related to the communication problem of testing independence of the joint distribution μ, studied in [ST18, ST21], and estimating correlation parameters of the joint distribution μ, studied in [HLPS19]. The communication-free version of agreement distillation relates to the problem of extracting a unique identification string from process variations. See [BM11] for further discussion of this motivation.

The agreement distillation problem is one typical category in the larger area of Common Randomness Generation and Secret Key Generation problems. The study of these problems dates back to the seminal work of Shannon on secrecy systems [Sha49] in information theory. See [STW19] for an up-to-date survey on such problems.

Earlier works for the agreement distillation problem focused on the zero-communication version. Witsenhausen [Wit75] studied the probability of extracting one bit without communication. The work [BM11] showed the probability Alice and Bob can agree on a k-bit string is

exponentially small in k without communication when the joint distribution is correlated unbiased Boolean case (equivalent to the binary symmetric channel). O'Donnell and Wright [OW12] proved the optimal probability of extracting one bit with zero-communication over the binary erasure channel. The work [GR16] established the precise trade-off between communication and probability of agreement for BSC and BEC distribution cases. See [GR16] for a more detailed history before their work. Extracting common random bits from correlated distribution on large alphabets is considered in [CMN14]. For the interactive multi-round communication version of this problem, round complexity and communication-round trade-offs are discussed in [GS20, SGGB19].

These studies mentioned above, as well as our work, focus on communication complexity and the success probability of the agreement distillation problem. We do not try to optimize n, the number of correlated samples we use. We allow $n \to \infty$ for any value of k, the number of common random bits to be generated. Another sequence of earlier works [AC98, ZC11] also studied the ratio of the number of source samples to the number of common random bits generated. The work [GJ18] gave a resource-efficient communication protocol matching the sharp trade-off between communication and probability of agreement in [GR16] for correlated Boolean and Gaussian cases, with only $\mathrm{poly}(k)$ samples used.

Let us be more precise on the setting. Suppose we only allow one-way communication from Alice to Bob. Ideally Alice decides her uniformly random string $h_A(\boldsymbol{x})$, where the "generator function" $h_A : (\Omega_x)^n \to \{0,1\}^k$ satisfies $\mathbf{Pr}[h_A(\boldsymbol{x}) = z] = 2^{-k}$ for any $z \in \{0,1\}^k$. Alice also decides the message $\boldsymbol{\pi}$ she sends to Bob. (The transcript $\boldsymbol{\pi}$ can also be seen as a function of random string $\boldsymbol{x}$.) Bob will try to guess the string Alice generated based on his random string $\boldsymbol{y}$ and the communication transcript $\boldsymbol{\pi}$. We define Bob's guess as function $h_B(\boldsymbol{y}, \boldsymbol{\pi})$. The success probability of this agreement distillation protocol is $\mathbf{Pr}[h_A(\boldsymbol{x}) = h_B(\boldsymbol{y}, \boldsymbol{\pi})]$.

However, for a general distribution μ it might be impossible to construct a function h_A such that the output is exactly uniform on $\{0,1\}^k$. Here we slightly revise the constraint of h_A to be $h_A : (\Omega_x)^n \to \{\{0,1\}^k, \text{"FAIL"}\}$ satisfying $\mathbf{Pr}[h_A(\boldsymbol{x}) = z | h_A(\boldsymbol{x}) \neq \text{"FAIL"}] = 2^{-k}$ for any $z \in \{0,1\}^k$. In other words, we want Alice to generate a uniformly random string with length k most of the time, but also allow Alice to give up and output "FAIL" by some small probability.

The work of [GR16] studied the special cases of this problem in which distribution μ is given by sending $\boldsymbol{x}$ through a binary symmetric channel (BSC) or a binary erasure channel (BEC) and letting $\boldsymbol{y}$ be the result. They pinpoint the exact trade-off between the communication and success probability required in order for Alice and Bob to agree on k bits of common randomness, when an unlimited number of correlated samples are available.

Though [GR16] only focused on BSC and BEC cases, their lower bound on communication complexity can be generalized to arbitrary joint distributions μ:

Theorem 3.1.1 (General lower bound from [GR16])**.** *Suppose there is a protocol with $\mathbf{Pr}[h_A(\boldsymbol{x}) = z] \leq 2^{-k}$ for any $z \in \{0,1\}^k$, and Alice sends ck bits to Bob after which Bob is able to guess $h_A(\boldsymbol{x})$ with probability at least $2^{-\gamma k}$. Then for any (p,q)-hypercontractive distribution μ,*

$$c \geq 1 - \frac{q'}{p'} - q'\gamma,$$

where p', q' are Hölder conjugates of p, q.

In particular, to achieve constant agreement probability, the communication lower bound would be

$$c \geq 1 - s^*(\boldsymbol{x}; \boldsymbol{y}) - o(1),$$

where

$$\begin{aligned} s^*(\boldsymbol{x}; \boldsymbol{y}) &= \lim_{q' \to \infty} \inf_{(p,q)\text{-hypercontractive}} \frac{q'}{p'} \\ &= \lim_{p \to 1} \inf_{(p,q)\text{-hypercontractive}} \frac{p-1}{q-1}. \end{aligned}$$

The function $s^*(\boldsymbol{x}; \boldsymbol{y})$ is the chordal slope of the boundary of the hypercontractivity region at the infinity, and is determined by the distribution μ. We will study it in further detail in Section 3.1.3.

The work of [GR16] showed that this lower bound is tight for BSC and some, though not all, of the BEC cases. One major open question in [GR16] determining the situation for more general joint distributions beyond the BSC and BEC cases.

The main result of this chapter is to construct a communication assisted agreement distillation protocol for general distribution μ with constant success probability, communicating $(1 - s^*(\boldsymbol{x}; \boldsymbol{y}) + o(1))k$ bits, which is optimal.

Theorem 3.1.2 (General upper bound for constant success probability)**.** *There is a protocol in which $h_A(\boldsymbol{x})$ is uniformly distributed in $\{0,1\}^k$, conditioned on not outputting "FAIL", such that Alice sends ck bits to Bob, who then succeeds in guessing $h_A(\boldsymbol{x})$ with probability $\Theta(1)$, in which the communication rate is given by*

$$c \leq 1 - s^*(\boldsymbol{x}; \boldsymbol{y}) + o(1).$$

Our agreement distillation scheme is explicit, including in the cases of the BSC and BEC channels. This is another advantage compared to the protocol given in [GR16], which is analyzed using the probabilistic method and only gives an existential result. In our scheme, the number of samples n will be exponential in k.

For any arbitrary imperfect shared randomness, one can study the hypercontractivity region on finite probability space $(\Omega_x \times \Omega_y, \mu)$ and construct an agreement distillation protocol with constant success probability. We further study the chordal slope of the boundary of the hypercontractivity region for the binary erasure channel and can show the following application:

Corollary 3.1.3 (Upper bound for reverse BEC)**.** *Consider the joint distribution μ in which $\boldsymbol{x}$ is the output of the random string $\boldsymbol{y}$ going through the binary erasure channel with erasure rate ϵ. There is a protocol such that $h_A(x)$ is uniformly distributed in $\{0,1\}^k$, conditioned on not outputting "FAIL", in which Alice sends ck bits to Bob, who then succeeds in guessing $h_A(x)$ with probability $\Theta(1)$. Furthermore, it has communication rate*

$$c = \begin{cases} \epsilon & \text{if } 0 \leq \epsilon \leq \frac{1}{2}; \\ 1 - \left(-\log \frac{1-\epsilon}{2}\right)^{-1} & \text{if } \frac{1}{2} < \epsilon < 1. \end{cases}$$

This communication rate matches the tight lower bound.

As we can see in the statement of Theorem 3.1.1 and 3.1.2, there is a tight relationship between this problem and hypercontractive inequalities. Indeed, the optimal bound in Theorem 3.1.2 shows how constructing the agreement distillation protocol and analyzing its communication rate reduced to calculating hypercontractivity parameters for the joint distribution μ. Our proof relies on hypercontractive inequalities, small-set expansion inequalities, inequalities involving the Kullback-Leibler-divergence, and the equivalences among them. The original proof of Theorem 3.1.1 in [GR16] relies only on hypercontractive inequalities. In this chapter we will rewrite the proof in a more comprehensive way with the help of the Small-Set Expansion Theorem. Our construction of the communication protocols in Theorem 3.1.2 is an extension of the proof of the equivalence between inequalities involving the KL divergence and small-set expansion inequalities in [Nai14].

3.1.2 Equivalence of general hypercontractivity and small-set expansion

We start by extending the definition of hypercontractivity to general pairs of random variable $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$, where μ is a joint distribution on the finite domain $\Omega_x \times \Omega_y$. Let μ_x and μ_y be the marginal distribution of $\boldsymbol{x}$ and $\boldsymbol{y}$, and let $\mu_{x|y}$ be the marginal distribution of $\boldsymbol{x}$ with $\boldsymbol{y} = y$.

Definition 3.1.4. We say a pair of random variables $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$ is (p, q)-*hypercontractive* for $1 \leq p \leq q \leq \infty$ if

$$\|\mathrm{T}f(\boldsymbol{y})]\|_q \leq \|f(\boldsymbol{x})\|_p \qquad \forall f : \Omega_x \to \mathbb{R},$$

where $\mathrm{T}f(y) = \mathbf{E}_{\boldsymbol{x} \sim \mu_{x|\boldsymbol{y}=y}}[f(\boldsymbol{x})]$, the left-hand side is the q-norm on the finite probability space (Ω_y, μ_y), and the right-hand side is the p-norm on the finite probability space (Ω_x, μ_x).

The earliest roots of the operator $\mathrm{T}f(y) = \mathbf{E}_{\boldsymbol{x} \sim \mu_{x|y}}[f(\boldsymbol{x})]$ are in the famous work of Markov [Mar06] which discusses stochastic matrices, the transition matrices in a Markov Chain. The operator T was later known as a Markov operator. The hypercontracting property of the Matrix operator was independently considered in theoretical physics as mentioned in [AG76]. Ahlswede and Gács studied the infimum of p/q on the hypercontractivity region in [AG76]. Hypercontractivity for general distributions is well-studied in the case of noisy channels in multi-user information theory. The special case of Gaussian and Boolean random variables and their history is mentioned in previous chapters. Hypercontractivity on the binary erasure channel is studied in [NW16, NW17].

Many classical hypercontractivity properties still hold for hypercontractivity on general finite probability spaces. For example, we also have an equivalent two-function version of the hypercontractivity statement.

Proposition 3.1.5. *The random variable pair $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$ is (p, q)-hypercontractive if and only if*

$$\mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y}) \sim \mu}[f(\boldsymbol{x})g(\boldsymbol{y})] \leq \|f(\boldsymbol{x})\|_p \|g(\boldsymbol{y})\|_{q'} \qquad \forall f : \Omega_x \to \mathbb{R}, g : \Omega_y \to \mathbb{R},$$

where q' is the Hölder conjugate of q, i.e. $\frac{1}{q} + \frac{1}{q'} = 1$.

The original hypercontractivity definition only focuses on functions of a single random variable $\boldsymbol{x}$. The two-function version hypercontractivity is more natural and symmetric on $\boldsymbol{x}$ and $\boldsymbol{y}$.

We will mainly use the two-function version of hypercontractivity in the rest of this chapter, while we keep the notation of p and q' to be aligned with classical definitions.

The fact that the $n = 1$ case of hypercontractivity can be extended to the general n case by induction also works for general joint probability distributions. Hence hypercontractive inequalities can be extended to general functions $f : (\Omega_x)^n \to \mathbb{R}$ and $g : (\Omega_y)^n \to \mathbb{R}$. For $(\boldsymbol{x}, \boldsymbol{y})$ which are (p, q)-hypercontractive, if we look at indicator functions f and g, we can get a small-set expansion inequality:

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq |A|^{\frac{1}{p}}|B|^{\frac{1}{q'}} \tag{3.1}$$

for any integer n, $A \subseteq (\Omega_x)^n$, and $B \subseteq (\Omega_y)^n$.

The small-set expansion phenomena and its relationship to hypercontractivity were previously mentioned in a series of studies [CF63, AGK76, AG76]. For the correlated unbiased Boolean case, the one-set small-set expansion inequalities were first shown in [KKL88] in order to study the distribution of Hamming distances in a vector set. The two-set generalization of the small-set expansion on the correlated unbiased Boolean case was first mentioned in [O'D14], inspired by a reverse version appeared in [MOR$^+$06], with an application to random walks.

The two-set generalization of small-set expansion arises naturally in communication with imperfect shared randomness. Specifically, in our agreement distillation problem, for any Boolean string output z of length k, let

$$A = \{x \mid h_A(x) = z\}, \qquad B = \{y \mid \exists \pi, h_B(\pi, y) = z\}.$$

That is to say, A is the set of all strings x letting Alice output z, and B is the set of all strings y such that it is possible for Bob to output z with some communication transcript. Then $\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B]$ naturally becomes an upper bound on the probability of Alice and Bob agreeing on the output z. We use small-set expansion to simplify the proof of Theorem 3.1.1 in [GR16]. See Section 3.3 for the summary.

For given sets A and B, one can take the infimum among all hypercontractive pairs (p, q) on the right-hand side of (3.1) and get the best upper bound. In [O'D14], it was shown that this upper bound is essentially sharp for ρ-correlated unbiased random Boolean strings in the case that A and B are concentric Hamming Balls. However, this constructive proof cannot be extended to small-set expansion on general finite domains. In general, the small-set expansion inequality looks weaker than the hypercontractive inequality since it only focuses on the subset of indicator functions $f : (\Omega_x)^n \to \{0, 1\}$, $g : (\Omega_y)^n \to \{0, 1\}$. Surprisingly, Chandra Nair showed in [Nai14] that the hypercontractivity inequality and the small-set expansion inequality are identical on general finite probability spaces, and are also identical to an inequality involving the Kullback-Leibler divergence:

Theorem 3.1.6 (From [Nai14])**.** *Let μ be a joint distribution on the finite domain $\Omega_x \times \Omega_y$. The following statements are equivalent:*

1) The random variable $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$ is (p, q)-hypercontractive;

2) $D(\nu\|\mu) \geq \frac{1}{p}D(\nu_x\|\mu_x) + \frac{1}{q'}D(\nu_y\|\mu_y)$ for any distribution $\nu \ll \mu$ on finite domain $\Omega_x \times \Omega_y$;

3) $\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq |A|^{\frac{1}{p}}|B|^{\frac{1}{q'}}$ for any integer n, $A \subseteq (\Omega_x)^n$, $B \subseteq (\Omega_y)^n$.

Here we denote the measure ν is absolutely continuous with respect to the measure μ as $\nu \ll \mu$ and let $D(\nu\|\mu)$ be the relative entropy between ν and μ when $\nu \ll \mu$.

Chandra Nair [Nai14] was focusing on equivalent formulations of hypercontractivity using information measures. The equivalence of the small-set expansion was not mentioned explicitly in the his theorem statement but appeared in the proof. We summarize and simplify the proof of Theorem 3.1.6 in Section 3.6.

There are some follow-up works which study Brascamp-Lieb inequalities [LCV15, LCCV16], which are similar to multi-party versions of hypercontractivity inequalities.

The deduction from the small-set expansion inequality to the inequality involving the KL divergence also gives us a way to construct sharp cases of sets A and B. We extend this deduction with a more precise Stirling's approximation to construct our communication protocol for the agreement distillation protocol. See the proof of our upper bound for the agreement distillation problem in Section 3.4.

The proofs of these results on the communication-assisted agreement distillation problem demonstrate the advantage of all three equivalent formulations in Theorem 3.1.6. First, we show a comprehensive proof of the lower bounds on communication complexity with the help of small-set expansion. Next, to show the tightness of these lower bounds, we use the KL-divergence form to construct sharp cases of sets and communication protocol. Finally, for a specific distribution μ, calculating the exact lower bound relies on the properties and constraints of hypercontractivity parameters p, q which are mostly shown and proven in the original hypercontractivity form. We believe these techniques can be applied to more problems. The equivalence of hypercontractivity and small-set expansion seems very promising and we hope it will prove useful in the future.

3.1.3 Boundary of hypercontractivity region

In the statement of Theorem 3.1.1 and 3.1.2, the ratio q'/p' of hypercontractivity parameters appears. If we focus on the communication protocols achieving constant agreement probability, it would be related to the ratio q'/p' where $p', q' \to \infty$. Therefore to calculate the exact lower bound on a specific distribution μ, we need to calculate the infimum of q'/p' constrained to $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$ being (p, q)-hypercontractive.

Because the L^p norm is monotone increasing, if $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$ is (p_0, q)-hypercontractive, then $(\boldsymbol{x}, \boldsymbol{y})$ is (p, q)-hypercontractive for any $p \geq p_0$. For a given $q > 1$, we define

$$p^*(q) = \inf\{p \mid (p, q)\text{-hypercontractive}\}.$$

Then $p^*(q)$ is the curve function of the boundary of the hypercontractivity region.

The early study of slope $p^*(q)/q$ appears in [AG76] by Ahlswede and Gács. They showed that $p^*(q)/q$ is monotonically decreasing in q. In [KA16] they studied the chordal slope $(p^*(q) - 1)/(q - 1)$, as well as

$$s^*(\boldsymbol{x}; \boldsymbol{y}) = \lim_{q \to 1} \frac{p^*(q) - 1}{q - 1}, \qquad s^*(\boldsymbol{y}; \boldsymbol{x}) = \lim_{q \to \infty} \frac{p^*(q) - 1}{q - 1}.$$

They showed that these quantities are connected to the maximal correlation from information theory. One can check [AGKN14] for a summary and some further discussion. $s^*(\boldsymbol{x}; \boldsymbol{y})$ also has an equivalent KL-divergence definition shown in [KA16]. We will use this property in proving the upper bound. See Section 3.2.2 for details.

One can easily check that if $(\boldsymbol{x}, \boldsymbol{y})$ is (p, q)-hypercontractive, then $(\boldsymbol{y}, \boldsymbol{x})$ is (q', p')-hypercontractive based on the two-function version of the hypercontractivity. Therefore we can use prior results to calculate the infimum of q'/p' in our agreement distillation problem.

The classical Hypercontractivity Theorem on ρ-correlated unbiased Boolean random variables says that $\frac{p^*(q)-1}{q-1} = \rho^2$ for any $q > 1$ in this case. The case of the binary erasure channel with erasure rate ϵ, i.e. $\boldsymbol{y} \sim \mathrm{BEC}_\epsilon(\boldsymbol{x})$, is studied in [NW16, NW17]. In [NW16] they showed that $\frac{p^*(q)-1}{q-1} = 1-\epsilon$ when $\epsilon - \frac{1}{2} \le \frac{3}{2}(q'-1)$. In this chapter we calculate the chordal slope $\frac{p^*(q)-1}{q-1}$ when $\epsilon > \frac{1}{2}$ and $q \to \infty$, which implies Corollary 3.1.3 as an application.

3.2 Preliminaries

3.2.1 Properties of the slope of hypercontractivity boundary

In this chapter we focus on a general finite domain $\Omega_x \times \Omega_y$ with joint distribution μ. We write $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$ to denote that the random variable pair $(\boldsymbol{x}, \boldsymbol{y})$ is chosen from $\Omega_x \times \Omega_y$ with distribution μ. We also write μ_x, μ_y, and $\mu_{x|\boldsymbol{y}=y}$ to denote the marginal distribution of $\boldsymbol{x}$, $\boldsymbol{y}$ and the conditional distribution $\boldsymbol{x}$ given $\boldsymbol{y} = y$. The Markov operator T on functions $f : \Omega_x \to \mathbb{R}$ is defined by $\mathrm{T}f(y) = \mathbf{E}_{\boldsymbol{x}\sim\mu_{x|\boldsymbol{y}=y}}[f(\boldsymbol{x})]$. Finally we use the notation $\|f(\boldsymbol{x})\|_p = \mathbf{E}_{\boldsymbol{x}\sim\mu_x}[|f(\boldsymbol{x})|^p]^{1/p}$.

Getting back to the hypercontractivity region, for a real number $q > 1$, we define

$$p^*(q) = \inf\Big\{p \mid \|\mathrm{T}f(\boldsymbol{y})]\|_q \le \|f(\boldsymbol{x})\|_p \quad \forall f : \Omega_x \to \mathbb{R}\Big\}.$$

In other words, $p^*(q)$ is the infimum of p such that $(\boldsymbol{x}, \boldsymbol{y})$ is (p, q)-hypercontractive, and so $p^*(q)$ is the curve function of the hypercontractivity boundary. One important property of the slope $p^*(q)/q$ is its monotonicity, as shown in [AG76].

Proposition 3.2.1. *$\frac{p^*(q)}{q}$ is monotonically decreasing as q increases.*

Another important quantity is the chordal slope $(p^*(q)-1)/(q-1)$. For correlated Boolean random variables/the binary symmetric channel and the binary erasure channel with erasure probability $\epsilon \le \frac{1}{2}$, the chordal slope is a constant for any q. Here we write $\boldsymbol{y} \sim \mathrm{BSC}_{\frac{1-\rho}{2}}(\boldsymbol{x})$ to denote that $\boldsymbol{y}$ is the result of transmitting the uniform Boolean string $\boldsymbol{x}$ through a BSC with crossover probability $\frac{1-\rho}{2}$, i.e. $(\boldsymbol{x}, \boldsymbol{y})$ is ρ-correlated. Similarly, we write $\boldsymbol{y} \sim \mathrm{BEC}_\epsilon(\boldsymbol{x})$ to denote that $\boldsymbol{y}$ is the result of transmitting the uniform Boolean variable $\boldsymbol{x}$ through a BEC with erasure probability ϵ.

Proposition 3.2.2 (From [O'D14, NW16])**.** *For the BSC and BEC cases, we have*

- $\frac{p^*(q)-1}{q-1} = \rho^2$ *for any $q > 1$ when $\boldsymbol{y} \sim \mathrm{BSC}_{\frac{1-\rho}{2}}(\boldsymbol{x})$;*
- $\frac{p^*(q)-1}{q-1} = 1-\epsilon$ *when $\boldsymbol{y} \sim \mathrm{BEC}_\epsilon(\boldsymbol{x})$ and $(\epsilon - \frac{1}{2})(q-1) \le \frac{3}{2}$.*

We are specifically interested in the chordal slopes of the hypercontractivity boundary at $q \to 1$ and $q \to \infty$. We define

$$s^*(\boldsymbol{x}; \boldsymbol{y}) = \lim_{q\to 1} \frac{p^*(q)-1}{q-1}.$$

We will present an equivalent definition in terms of the KL divergence in Section 3.2.2.

Let p', q' be the conjugate Hölder indices of p and q, e.g. $\frac{1}{p} + \frac{1}{p'} = 1$. It is easy to check that if $(\boldsymbol{x}, \boldsymbol{y})$ is (p, q)-hypercontractive, then $(\boldsymbol{y}, \boldsymbol{x})$ is (q', p')-hypercontractive from the two-function version of hypercontractivity in Proposition 3.1.5. Similarly we define

$$q'^*(p') = \inf\Big\{q' \mid \|\mathrm{T}g(\boldsymbol{x})]\|_{p'} \leq \|g(\boldsymbol{y})\|_{q'} \quad \forall g : \Omega_y \to \mathbb{R}\Big\},$$

where $\mathrm{T}g(x) = \mathbf{E}_{\boldsymbol{y}\sim\mu_{y|\boldsymbol{x}=x}}[g(\boldsymbol{y})]$.

Proposition 3.2.3. *The slopes and the choral slopes have the following properties:*

- $\frac{q'^*(p')}{p'}$ *is monotonically decreasing as* p' *increases;*
- $s^*(\boldsymbol{x}; \boldsymbol{y}) = \lim_{q\to 1} \frac{p^*(q)-1}{q-1} = \lim_{p'\to\infty} \frac{q'^*(p')}{p'}$;
- $s^*(\boldsymbol{y}; \boldsymbol{x}) = \lim_{q\to\infty} \frac{p^*(q)}{q} = \lim_{p'\to 1} \frac{q'^*(p')-1}{p'-1}$.

3.2.2 Kullback-Leibler divergence

For probability distributions ν and μ defined on finite domain Ω, we denote the measure ν is absolutely continuous with respect to the measure μ as $\nu \ll \mu$.The Kullback-Leibler divergence from μ to ν when $\nu \ll \mu$ is defined to be

$$D(\nu\|\mu) = \sum_{x\in\Omega} \nu(x) \log \frac{\nu(x)}{\mu(x)}.$$

We assume $\nu \ll \mu$ for the rest discussion of this chapter.

The Kullback-Leibler divergence is an important ingredient in proving the equivalence of hypercontractivity and small set expansion. On the one hand, the distribution ν is a function over the domain $\Omega_x \times \Omega_y$ so it is naturally related to hypercontractivity. On the other hand, the Kullback-Leibler divergence has the following statistical interpretation:

Claim 3.2.4. *Let* $\boldsymbol{x} \sim (\Omega^n, \mu^{\otimes n})$ *be* n *i.i.d. random variables under distribution a* μ*. Let* c *be the histogram with counts* $c_a = \nu(a)n$ *for all* $a \in \Omega$*, where the measure* $\nu \ll \mu$*. Let* A *be the set of all strings following the histogram* c*. I.e.,* $x \in A$ *if and only if the number of coordinates with value* a *in* x *is equal to* c_a *for all* $a \in \Omega$*. Then*

$$-\log \mathbf{Pr}[\boldsymbol{x} \in A] = D(\nu\|\mu)n + \frac{1}{2}\left(\log n - \sum_{a\in\Omega} \log c_a\right) + \Theta(1).$$

We remark that in this claim we assume $c_a = \nu(a)n$ is an integer, which might not be true for most of the cases. But it is easy to prove that rounding $\nu(a)n$ will only introduce constant difference, and so we omit this part for the simplicity. The proof follows Section 12.1 in [Cov99], with a more precise Stirling's approximation.

The slopes and the chordal slopes of the hypercontractivity boundary at $q \to 1$ and $q \to \infty$ have an equivalent KL-divergence form proven in [KA12].

Theorem 3.2.5. *Consider a pair of random variables* $(\boldsymbol{x}, \boldsymbol{y}) \sim (\Omega_x \times \Omega_y, \mu)$*. Then*

$$s^*(\boldsymbol{x}; \boldsymbol{y}) = \sup_{\substack{\nu(x,y)=\nu_x(x)\mu_{y|\boldsymbol{x}=x}(y) \\ \nu_x\neq\mu_x}} \frac{D(\nu_y\|\mu_y)}{D(\nu_x\|\mu_x)},$$

where the supremum is over all distributions ν satisfying $\nu_x(x,y) = \nu(x)\mu_{y|\boldsymbol{x}=x}(y)$ such that the marginal distribution ν_x is not equal to μ_x.

3.3 Lower bound

In this section, we simplify the proof of the lower bound in Theorem 3.1.1 shown in [GR16] in the language of small set expansion. Small set expansion naturally arises when we look at the sets of Alice's $\boldsymbol{x}$ and Bob's $\boldsymbol{y}$ in which they would agree on a specific common random string.

Proof of Theorem 3.1.1. On Alice's side,we define

$$A_z = \{x \in (\Omega_x)^n \mid h_A(x) = z\}$$

for all $z \in [2^k]$. On Bob's side, we also define

$$B_z = \{y \in (\Omega_y)^n \mid \exists \pi \text{ s.t. } h_B(y,\pi) = z\}.$$

Because there are 2^{ck} different possible transcripts, Bob can only output 2^{ck} different values given a fixed y. Therefore $\sum_z |B_z| \le 2^{ck}$. Alice and Bob can agree with the same string only if $x \in A_z$ and $y \in B_z$ for some z. Therefore

$$2^{-\gamma k} \le \mathbf{Pr}[\text{Protocol success}] \le \sum_z \mathbf{Pr}[\boldsymbol{x} \in A_z, \boldsymbol{y} \in B_z] \le \sum_z |A_z|^{\frac{1}{p}} |B_z|^{\frac{1}{q'}}.$$

for any (p,q)-hypercontractive. The last inequality holds by Theorem 3.1.6.

Then we conclude

$$2^{-\gamma k} \le \sum_z |A_z|^{\frac{1}{p}} |B_z|^{\frac{1}{q'}} \le 2^{-\frac{k}{p}} \sum_z |B_z|^{\frac{1}{q'}} \le 2^{-\left(\frac{1}{p} + \frac{1-c}{q'} - 1\right)k},$$

where the last inequality follows from Jensen's inequality. We conclude the proof by rearranging the inequality to

$$c \ge 1 - \frac{q'}{p'} - q'\gamma.$$

For achieving constant agreement probability, we plug in $\gamma = \Theta(1/k)$, $q' = \sqrt{k}$ and choose the maximum p such that μ is (p,q)-hypercontractive. □

3.4 Upper bound

In this section we will prove the upper bound in Theorem 3.1.2. We first present a construction of a general one-way communication protocol with any arbitrary sets A and B. This protocol construction is inspired by the idea in [AC98].

Theorem 3.4.1. *For any integer m, any nonempty sets $A \subsetneq (\Omega_x)^m$, $B \subsetneq (\Omega_y)^m$. Let $k = \lfloor \log \frac{1}{|A|} \rfloor$, $k' = \lfloor \log \frac{1}{|B|} \rfloor$. There is a protocol in which $h_A(\boldsymbol{x})$ is uniformly distributed in $\{0,1\}^k$, conditioned on not outputting "FAIL", such that Alice sends $(k - k')$ bits to Bob, who then succeeds in guessing $h_A(x)$ with probability at least $\mathbf{Pr}[\boldsymbol{y} \in B | \boldsymbol{x} \in A] \cdot \frac{1}{2e^2}$.*

Proof. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be $n = 2^k m$ i.i.d. random variables under joint distribution μ. We write $\boldsymbol{x} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(2^k)})$ and $\boldsymbol{y} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(2^k)})$, where each $(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)})$ is a block of m i.i.d. random variables under joint distribution μ. We then create new random variables which indicate whether $\boldsymbol{x}^{(i)}$ is in A, and $\boldsymbol{y}^{(i)}$ is in B:

$$\boldsymbol{z}_i = \begin{cases} 1 & \text{if } \boldsymbol{x}^{(i)} \in A, \\ 0 & \text{otherwise;} \end{cases} \qquad \boldsymbol{w}_i = \begin{cases} 1 & \text{if } \boldsymbol{y}^{(i)} \in B, \\ 0 & \text{otherwise.} \end{cases}$$

In our setting, Alice knows the random variable $\boldsymbol{x}$ and therefore $\boldsymbol{z}$; Bob knows the random variable $\boldsymbol{y}$ and therefore $\boldsymbol{w}$. Alice decides her uniform random string $h_A(x)$ as follows:

$$h_A(x) = \begin{cases} i & \text{if } z_i = 1 \text{ and } z_j = 0 \text{ for all } j \neq i; \\ \text{“FAIL”} & \text{otherwise.} \end{cases}$$

From the symmetricity of the i.i.d. random variable $\boldsymbol{x}$ we know that $h_A(x)$ will be uniformly random in $\{0,1\}^k$ conditioned on not outputting “FAIL”.

Alice further divides $\boldsymbol{z}$ into $2^{k-k'}$ blocks with size $2^{k'}$ each. We denote

$$\Omega_\pi = \{[1, 2^{k'}], [2^{k'}+1, 2^{k'}+2^{k'}], \ldots, [2^k - 2^{k'} + 1, 2^k]\}$$

for the index sets of all these $2^{k-k'}$ blocks. If Alice does not output “FAIL”, there exists only one index i such that $\boldsymbol{z}_i = 1$. Alice chooses the transcript $\pi \in \Omega_\pi$ such that $\pi \ni i$ and send π to Bob. Because $|\Omega_\pi| = 2^{k-k'}$, Alice will send $(k - k')$ bits to Bob.

Bob has his random variables $\boldsymbol{y}$ and therefore $\boldsymbol{w}$. He also receives a block π from Alice. Bob only needs to check if there exists exactly one $z_{i'}$ with value 1 in block π and outputs index i':

$$h_B(y, \pi) = \begin{cases} i' & \text{if } w_{i'} = 1 \text{ for some } i' \in \pi \text{ and } w_j = 0 \text{ for all other } j \in \pi, j \neq i'; \\ \text{“FAIL”} & \text{otherwise.} \end{cases}$$

The rest of the proof shows a lower bound on the success probability of Alice and Bob agreeing with each other, i.e. $\mathbf{Pr}[h_A(\boldsymbol{x}) = h_B(\boldsymbol{y}, \boldsymbol{\pi})]$. Consider the case of $h_A(x) = h_B(y, \pi) = 1$. On Alice's side Alice should have $z_1 = 1$ and $z_j = 0$ for all $j \neq 1$. Alice then sends $\pi = [1, 2^{k'}]$ to Bob. On Bob's side he must have $w_1 = 1$ and $w_j = 0$ for all $j \in [1, 2^{k'}]$, $j \neq 1$. Therefore

$$\begin{aligned}
&\mathbf{Pr}[h_A(\boldsymbol{x}) = h_B(\boldsymbol{y}, \boldsymbol{\pi}) = 1] \\
&= \mathbf{Pr}[\boldsymbol{z}_1 = 1, \boldsymbol{w}_1 = 1] \left(\prod_{j=2}^{2^{k'}} \mathbf{Pr}[\boldsymbol{z}_j = 0, \boldsymbol{w}_j = 0] \right) \left(\prod_{j=2^{k'}+1}^{2^k} \mathbf{Pr}[\boldsymbol{z}_j = 0] \right) \\
&\geq \mathbf{Pr}[\boldsymbol{z}_1 = 1, \boldsymbol{w}_1 = 1] \left(\prod_{j=2}^{2^{k'}} \mathbf{Pr}[\boldsymbol{w}_j = 0] \right) \left(\prod_{j=2^{k'}+1}^{2^k} \mathbf{Pr}[\boldsymbol{z}_j = 0] \right) \\
&\geq \mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \cdot (1 - |B|)^{2^{k'}} \cdot (1 - |A|)^{2^k} \\
&\geq \mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \cdot e^{-2}.
\end{aligned}$$

The last inequality holds since $k = \lfloor \log \frac{1}{|A|} \rfloor$, $k' = \lfloor \log \frac{1}{|B|} \rfloor$.

In total, $h_A(x)$ and $h_B(y, \pi)$ have at least $2^k > \frac{1}{2|A|}$ different values. Therefore the total success probability will be at least $\mathbf{Pr}[\boldsymbol{y} \in B | \boldsymbol{x} \in A] \cdot \frac{1}{2e^2}$. □

The following lemma shows that we can always find some sets A and B with proper density which lie on the hypercontractivity boundary.

Lemma 3.4.2. *For any $\epsilon > 0$, there exists some large $q_0' \geq 2$, such that for any $q' \geq q_0'$, there exist integer m, nonempty sets $A \subsetneq \mathcal{X}^m$, $B \subsetneq \mathcal{Y}^m$, satisfying:*

$$\mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] > |A|^{1/p} |B|^{1/q'}, \tag{3.2}$$

where $\frac{q'}{p'} = s^*(X; Y) - \epsilon$, and

$$\log \frac{1}{|A|} = Cq' \tag{3.3}$$

for some constant C.

We can achieve constant agreement probability by plugging in sets given from Lemma 3.4.2 into the protocol constructed in Theorem 3.4.1.

Proof of Theorem 3.1.2. First of all, we choose q satisfying $Cq' = k$, so $k = \lfloor \log \frac{1}{|A|} \rfloor$ for set A in Lemma 3.4.2.

Secondly, we need to bound the communication complexity. By (3.2), we know that

$$|A| > \mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] > |A|^{1/p} |B|^{1/q'},$$

which means $|B| < |A|^{q'/p'}$ and the communication complexity in the protocol in Theorem 3.4.1 will be $k - \lfloor \log \frac{1}{|B|} \rfloor < (1 - s^*(Y; X) + \epsilon)k$.

The last part is showing that $\mathbf{Pr}[\boldsymbol{y} \in B | \boldsymbol{x} \in A]$ is a constant. Using (3.2) again we have

$$|B| > \mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] > |A|^{1/p} |B|^{1/q'},$$

which means $|B| > |A|^{q/p}$. Therefore

$$\mathbf{Pr}[\boldsymbol{y} \in B | \boldsymbol{x} \in A] > |A|^{\frac{1}{p}-1} |B|^{\frac{1}{q'}} > |A|^{\frac{1}{p}+\frac{q}{pq'}} = |A|^{\frac{1}{q'-1}\left(1-\frac{q'}{p'}\right)} > |A|^{\frac{2}{q'}},$$

where the last inequality holds by assuming $q' \geq 2$. Combining with (3.3) we get $\mathbf{Pr}[\boldsymbol{y} \in B | \boldsymbol{x} \in A] > \Omega(1)$. □

We will prove Lemma 3.4.2 in the rest of this section. We start the construction of the sharp sets from finding the proper joint distribution ν in the inequality of KL-divergence form.

Claim 3.4.3. *For any constant $\epsilon > 0$, there exists a joint distribution ν on $\Omega_x \times \Omega_y$, satisfying:*

$$D(\nu \| \mu) - \frac{1}{p} D(\nu_x \| \mu_x) - \frac{1}{q'} D(\nu_y \| \mu_y) < -\Omega\left(\frac{1}{q'}\right),$$

for any $\frac{q'}{p'} = s^*(X; Y) - \epsilon$.

Proof. From Theorem 3.2.5, we have

$$s^*(\boldsymbol{x};\boldsymbol{y}) = \sup_{\nu_x \neq \mu_x} \frac{D(\nu_y\|\mu_y)}{D(\nu_x\|\mu_x)}$$

where ν_y denotes the y-marginal distribution of $\nu(x,y) = \nu(x)\frac{\mu(x,y)}{\mu_x(x)}$. Therefore there exists some distribution ν satisfying

$$\frac{D(\nu_y\|\mu_y)}{D(\nu_x\|\mu_x)} = s^*(\boldsymbol{x};\boldsymbol{y}) - \frac{\epsilon}{2}$$

and $\nu(x,y) = \nu(x)\frac{\mu(x,y)}{\mu_x(x)}$, where $\nu_x \neq \mu_x$. Then clearly $D(\nu\|\mu) = D(\nu_x\|\mu_x) > 0$. Therefore

$$\begin{aligned} D(\nu\|\mu) - \frac{1}{p}D(\nu_x\|\mu_x) - \frac{1}{q'}D(\nu_y\|\mu_y) &= \frac{1}{p'}D(\nu_x\|\mu_x) - \frac{1}{q'}D(\nu_y\|\mu_y) \\ &= \frac{D(\nu_x\|\mu_x)}{q'}\left(\frac{q'}{p'} - \frac{D(\nu_y\|\mu_y)}{D(\nu_x\|\mu_x)}\right) \\ &= -\frac{D(\nu_x\|\mu_x)\epsilon}{2q'}. \end{aligned}$$

Notice that the construction of ν is only depended on μ and ϵ, not q, so we conclude the proof. □

Now we can construct sets A and B based on distribution ν. The proof is a more precise version of the reduction from the Small Set Expansion Inequality to the KL-divergence Inequality in Theorem 3.1.6.

Proof of Lemma 3.4.2. Let m be any integer. Define the histogram c on $\Omega_x \times \Omega_y$ where $c(x,y) = \nu(x,y)m$ with distribution ν in Claim 3.4.3. (We omit the rounding process of $\nu(x,y)m$ for the simplicity of the proof.) Let set A denote all strings in Ω_x^m following the histogram c_x and B denote all strings in Ω_y^m following the histogram c_y. Then using Claim 3.2.4 on sets $A \times B$, A, B, we get

$$\begin{aligned} -\log \mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] &= D(\nu\|\mu)m + \frac{1}{2}\left(\log m - \sum_{\mu(x,y)>0} \log c(x,y)\right) + \Theta(1); \\ -\log|A| &= D(\nu_x\|\mu_x)m + \frac{1}{2}\left(\log m - \sum_{x\in\Omega_x} \log c(x)\right) + \Theta(1); \\ -\log|B| &= D(\nu_y\|\mu_y)m + \frac{1}{2}\left(\log m - \sum_{y\in\Omega_y} \log c(y)\right) + \Theta(1). \end{aligned}$$

Notice that

$$\sum_{\mu(x,y)>0} \log c(x,y) - \frac{1}{p}\sum_{x\in\Omega_x} \log c(x) - \frac{1}{q'}\sum_{y\in\Omega_y} \log c(y) \geq -\frac{1}{q'}\sum_{y\in\Omega_y} \log c(y) = -\frac{|\Omega_y|}{q'}\log m + \Theta(1)$$

when $c(x,y) \geq 2$ for any x, y. Therefore to make (3.2) hold, we need m to satisfy

$$\left(D(\nu\|\mu) - \frac{1}{p}D(\nu_x\|\mu_x) - \frac{1}{q'}D(\nu_y\|\mu_y)\right)m + \frac{|\Omega_y|}{q'}\log m = -\Omega(1).$$

Then by Claim 3.4.3 we know that we can choose $m = O(q')$ to make (3.2) hold. Adjusting m such that $-\log|A| = Cq'$ for some large constant C will conclude the proof. Here C is related to distribution μ and parameter ϵ. □

3.5 Examples and hypercontractivity boundary for BEC

3.5.1 Examples

We will construct protocols of the binary symmetric channel (BSC) and the binary erasure channel (BEC) cases studied in [GR16]. Our protocols are explicit rather than the probabilistic methods in [GR16]. We start from the example when Bob receives $\boldsymbol{y}$ by transmitting Alice's $\boldsymbol{x}$ through a BSC.

Examples 3.5.1. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be ρ-correlated unbiased ± 1 random strings, which is equivalent to say that random string $\boldsymbol{y}$ is uniformly random string $\boldsymbol{x}$ going through a binary symmetric channel (BSC) with crossover probability $\frac{1-\rho}{2}$. Consider the agreement distillation where Alice has $\boldsymbol{x}$ and Bob has $\boldsymbol{y}$. The following hold.

- Upper bound: Let $A, B \subseteq \{-1,1\}^n$ be concentric Hamming balls with volumes $|A| = \alpha$, $|B| = \alpha^{(1-o(1))\rho^2}$. Then $\mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] > .99\,\mathbf{Pr}[\boldsymbol{x} \in A]$, when ρ is fixed and α is small enough. The protocol in Theorem 3.4.1 with such sets A, B satisfies that Alice generates $h_A(\boldsymbol{x})$ uniformly distributed in $\{-1,1\}^k$ (conditioned on not outputting "FAIL"), Alice sends $(1-\rho^2+o(1))k$ bits to Bob, and then Bob succeeds in guessing $h_A(\boldsymbol{x})$ with constant probability;
- Matching Lower bound: According to Theorem 3.1.1 and Proposition 3.2.2, suppose there is a protocol with $\mathbf{Pr}[h_A(\boldsymbol{x}) = z] \leq 2^{-k}$ for any $z \in \{0,1\}^k$, and Alice sends ck bits to Bob after which Bob is able to guess $h_A(\boldsymbol{x})$ with probability at least $2^{-\gamma k}$. Then

  $$c \geq (1-\rho^2)(1-\gamma) - 2\rho\sqrt{(1-\rho^2)\gamma}.$$

  In particular, the optimal communication approaches $(1-\rho^2)k$ to achieve constant agreement probability, for large k.

The optimal upper and lower bounds are the same as the results in [GR16]. Notice that in the upper bound part, we used the well-studied sets of concentric Hamming balls from [Jan97, O'D14] to construct the protocol in Theorem 3.4.1 rather then the sets generated in Lemma 3.4.2. In fact we remark that by adjusting the densities of concentric Hamming balls of A and B properly, we can get a protocol matching the lower bound for any agreement probability. The underlying reason is that the pair of concentric Hamming balls is always the optimal case of the small-set expansion.

Another example is Bob receiving $\boldsymbol{y}$ by transmitting Alice's $\boldsymbol{x}$ through a BEC, which is also studied in [GR16].

Examples 3.5.2. Let random string $\boldsymbol{x}$ be uniform in $\{-1,1\}^N$ and random string $\boldsymbol{y}$ in $\{-1,0,1\}^N$ be $\boldsymbol{x}$ going through a binary erasure channel with erasure probability ϵ. Consider the agreement distillation where Alice has $\boldsymbol{x}$ and Bob has $\boldsymbol{y}$. The following hold.

- Upper bound: Let $A = \{\sum x_i \geq t_A n\} \subseteq \{-1,1\}^n$ and $B = \{\sum y_i \geq t_B n\} \subseteq \{-1,0,1\}^n$, where parameters t_A, t_B are chosen to achieve the volumes $|A| = \alpha$, $|B| = \alpha^{(1-o(1))(1-\epsilon)}$. Then $\mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] > .99\,\mathbf{Pr}[\boldsymbol{x} \in A]$, when ρ is fixed and α is small enough. The protocol in Theorem 3.4.1 with such sets A, B satisfies that Alice generates $h_A(\boldsymbol{x})$ uniformly distributed in $\{-1,1\}^k$ (conditioned on not outputting "FAIL"), Alice sends $(\epsilon + o(1))k$ bits to Bob, and then Bob succeeds in guessing $h_A(\boldsymbol{x})$ with constant probability;
- Matching Lower bound: According to Theorem 3.1.1 and Proposition 3.2.2, suppose there is a protocol with $\mathbf{Pr}[h_A(\boldsymbol{x}) = z] \leq 2^{-k}$ for any $z \in \{0,1\}^k$, and Alice sends ck bits to Bob after which Bob is able to guess $h_A(\boldsymbol{x})$ with probability at least $2^{-\gamma k}$. Then

$$c \geq \epsilon(1-\gamma) - 2\sqrt{\epsilon(1-\epsilon)\gamma}.$$

  In particular, the optimal communication approaches ϵk to achieve constant agreement probability, for large k.

The construction of sets A and B in the upper bound follows the similar idea of the concentric Hamming balls in the BSC case. The pair of A and B is the essential sharp case for the BEC when $|B| \geq |A|$, according to Proposition 3.2.2. Therefore by adjusting the parameters of t_A, t_B, we can also get an optimal protocol for any agreement probability.

The setup for BEC is not symmetric between Alice and Bob. What can be done if Alice and Bob switch roles? Alice receives $\boldsymbol{x}$ by transmitting Bob's $\boldsymbol{y}$ through a BEC channel. This setup is mentioned as an open problem in [GR16]. We obtain tight communication complexity upper and lower bounds for the constant agreement probability as in Corollary 3.1.3:

Examples 3.5.3. Let random string $\boldsymbol{y}$ be uniform in $\{-1,1\}^N$ and random string $\boldsymbol{x}$ in $\{-1,0,1\}^N$ be $\boldsymbol{x}$ going through a binary erasure channel with erasure probability ϵ. Consider the agreement distillation where Alice has $\boldsymbol{x}$ and Bob has $\boldsymbol{y}$. The following hold.

- Upper bound: If $\epsilon < \frac{1}{2}$, we use the protocol in Example 3.5.2 and communicate $(\epsilon + o(1))k$ bits to achieve constant agreement probability. Otherwise let A, B be the sets with only the all-one string of length $l = \left(-\log \frac{1-\epsilon}{2}\right)^{-1} k$. In this case $\mathbf{Pr}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] = \mathbf{Pr}[\boldsymbol{x} \in A]$ and the protocol in Theorem 3.1.2 with such sets A and B will communicate $\left(1 - \left(-\log \frac{1-\epsilon}{2}\right)^{-1}\right)k$ bits to achieve constant agreement probability;
- Matching Lower bound: Combining Theorem 3.1.1, Proposition 3.2.2 and Theorem 3.5.4, suppose there is a protocol with $\mathbf{Pr}[h_A(\boldsymbol{x}) = z] \leq 2^{-k}$ for any $z \in \{0,1\}^k$, and Alice sends ck bits to Bob after which Bob is able to guess $h_A(\boldsymbol{x})$ with constant probability. Then optimal communication rate

$$c = \begin{cases} \epsilon & \text{if } 0 \leq \epsilon \leq \frac{1}{2}; \\ 1 - \left(-\log \frac{1-\epsilon}{2}\right)^{-1} & \text{if } \frac{1}{2} < \epsilon < 1. \end{cases}.$$

The construction of sets A and B in the case $\frac{1}{2} \leq \epsilon < 1$ follows the proofs of Lemma 3.4.2 and Theorem 3.5.4. The exact lower bound is presented in Theorem 3.5.4 for the case $\frac{1}{2} \leq \epsilon < 1$ and in Proposition 3.2.2 for the case $\epsilon < \frac{1}{2}$.

3.5.2 Limit of gradient at infinity for BEC hypercontractivity boundary

In this section we will calculate $s^*(\boldsymbol{x};\boldsymbol{y})$ of the BEC when erasure probability $\frac{1}{2} \le \epsilon < 1$. The parameter for the case $\epsilon < \frac{1}{2}$ is mentioned in Proposition 3.2.2, which is proved in [NW16]. Plugging Theorem 3.5.4 and Proposition 3.2.2 into Theorem 3.1.2 and 3.1.1, we get Corollary 3.1.3, the tight communication complexity upper and lower bounds achieving constant agreement probability in the setting of the reverse BEC.

Theorem 3.5.4. *Let $\boldsymbol{y}$ be uniformly random in $\{-1,1\}$ and let $\boldsymbol{x}$ be the result of sending $\boldsymbol{y}$ through a binary erasure channel with erasure probability ϵ. Then*

$$s^*(\boldsymbol{x};\boldsymbol{y}) = \lim_{q\to 1} \frac{p^*(q)-1}{q-1} = \left(-\log \frac{1-\epsilon}{2}\right)^{-1},$$

when $\frac{1}{2} \le \epsilon < 1$.

Proof. We first show that $s^*(\boldsymbol{x};\boldsymbol{y}) \ge \left(-\log \frac{1-\epsilon}{2}\right)^{-1}$. It is easy to check that $(\boldsymbol{x},\boldsymbol{y})$ being (p,q)-hypercontractive is equivalent to $(\boldsymbol{y},\boldsymbol{x})$ being (q',p')-hypercontractive. From Proposition 3.2.3, $s^*(\boldsymbol{x};\boldsymbol{y}) = \lim_{p'\to\infty} \frac{q'^*(p')}{p'}$. Let $s = q'/p'$. Then $(\boldsymbol{y},\boldsymbol{x})$ being (q',p')-hypercontractive can be interpreted as

$$\|\mathrm{T}'g(\boldsymbol{x})\|_{p'} \le \|g(\boldsymbol{y})\|_{p's},$$

where $\mathrm{T}'g(x) = \mathbf{E}_{\boldsymbol{y}\sim\mu_{y|\boldsymbol{x}=x}}[f(\boldsymbol{y})]$. Henceforth, we are going to calculate the infimum of s at $p' \to \infty$ such that this inequality holds for all $g : \{-1,1\} \to \mathbb{R}$.

We consider the case that $g(1) = 1+\delta$, $g(-1) = 1-\delta$ with the constraint $|\delta| \le 1$. In this case the inequality becomes

$$\left(\frac{1-\epsilon}{2}(1-\delta)^{p'} + \frac{1-\epsilon}{2}(1+\delta)^{p'} + \epsilon\right)^{\frac{1}{p'}} \le \left(\frac{1}{2}(1-\delta)^{p's} + \frac{1}{2}(1+\delta)^{p's}\right)^{\frac{1}{p's}}.$$

Consider the case $\delta = 1$:

$$\left((1-\epsilon)2^{p'-1} + \epsilon\right)^{\frac{1}{p'}} \le 2^{1-\frac{1}{p's}}.$$

Hence,

$$s \ge \frac{1}{p' - \log((1-\epsilon)2^{p'-1} + \epsilon)}.$$

Then we take the limit at $p' \to \infty$ and conclude $s^*(\boldsymbol{x};\boldsymbol{y}) \ge \left(-\log \frac{1-\epsilon}{2}\right)^{-1}$.

We use the KL-divergence form to show that $s^*(\boldsymbol{x};\boldsymbol{y}) \le \left(-\log \frac{1-\epsilon}{2}\right)^{-1}$. From Theorem 3.2.5, we have

$$s^*(\boldsymbol{x};\boldsymbol{y}) = \sup_{\substack{\nu(x,y)=\nu_x(x)\mu_{y|\boldsymbol{x}=x}(y)\\ \nu_x\neq\mu_x}} \frac{D(\nu_y\|\mu_y)}{D(\nu_x\|\mu_x)}.$$

We wish show that

$$D(\nu_y\|\mu_y) - sD(\nu_x\|\mu_x) \le 0$$

where $s = \left(-\log \frac{1-\epsilon}{2}\right)^{-1}$ for any distribution ν satisfies $\nu(x,y) = \nu_x(x)\mu_{y|\boldsymbol{x}=x}(y)$.

We write $\nu_x(-1) = \nu_{-1}, \nu_x(1) = \nu_1, \nu_x(0) = \nu_0$. Then $\nu_y(-1) = \nu_{-1} + \frac{1}{2}\nu_0$ and $\nu_y(1) = \nu_1 + \frac{1}{2}\nu_0$. It is easy to check at the boundaries that the worst case is $\nu_1 = 1, \nu_{-1} = \nu_0 = 0$ and $D(\nu_y \| \mu_y) - sD(\nu_x \| \mu_x) = 0$.

The rest of the proof is to show that $D(\nu_y \| \mu_y) - sD(\nu_x \| \mu_x) \leq 0$ on all interior stationary points. That is:

$$\begin{aligned}\left(\nu_{-1} + \frac{1}{2}\nu_0\right)\ln(2\nu_{-1} + \nu_0) + \left(\nu_1 + \frac{1}{2}\nu_0\right)\ln(2\nu_1 + \nu_0) \\ - s\left(\nu_{-1}\ln\frac{2\nu_{-1}}{1-\epsilon} + \nu_1\ln\frac{2\nu_1}{1-\epsilon} + \nu_0\ln\frac{\nu_0}{\epsilon}\right) \leq 0\end{aligned} \tag{3.4}$$

Here we use base e for logarithm of KL-divergence to make differentiation easier. For any strictly interior stationary points, the Lagrange conditions yield:

$$k = \ln(2\nu_{-1} + \nu_e) + 1 - s\left(\ln\frac{2\nu_{-1}}{1-\epsilon} + 1\right), \tag{3.5}$$

$$k = \ln(2\nu_1 + \nu_e) + 1 - s\left(\ln\frac{2\nu_1}{1-\epsilon} + 1\right), \tag{3.6}$$

$$k = \frac{1}{2}\ln(2\nu_{-1} + \nu_0) + \frac{1}{2}\ln(2\nu_1 + \nu_0) + 1 - s\left(\ln\frac{\nu_0}{\epsilon} + 1\right). \tag{3.7}$$

Equating $\frac{1}{2}((3.5) + (3.6))$ and (3.7) yields:

$$\sqrt{\nu_{-1}\nu_1} = \frac{1-\epsilon}{2\epsilon}\nu_0.$$

Let $\nu_{-1} = \frac{1-\epsilon}{2\epsilon}\nu_0 t$ and $\nu_1 = \frac{1-\epsilon}{2\epsilon}\nu_0 t^{-1}$ where $t \geq 0$. Plug them into the equation of (3.5) and (3.6):

$$(1-\epsilon)t + \epsilon = (1-\epsilon)t^{-(1-2s)} + \epsilon t^{2s}. \tag{3.8}$$

Define

$$h(t) = (1-\epsilon)t + \epsilon - (1-\epsilon)t^{-(1-2s)} - \epsilon t^{2s}$$

and $h''(t) = 2t^{-(3-2s)}(1-2s)(\epsilon + s - 1 + \epsilon s(t-1))$. Because $s = \left(-\log\frac{1-\epsilon}{2}\right)^{-1}$ and $\frac{1}{2} < \epsilon < 1$, we have $0 < s < \frac{1}{2}$ and $\epsilon + s - 1 > 0$, so $h(t)$ is convex when $t \geq 1$. Because $h(1) = 0$, this means there will be at most two more roots for (3.8), $t = t_0$ and $t = t_0^{-1}$.

We only need to show that $t = 1$ is a strict local maximum for the left-hand side of (3.4). Then those two more stationary points could only be local minimum. Plug $\nu_{-1} = \frac{1-\epsilon}{2\epsilon}\nu_0 t$, $\nu_1 = \frac{1-\epsilon}{2\epsilon}\nu_0 t^{-1}$, $\nu_{-1} + \nu_1 + \nu_0 = 1$ into the left-hand side of (3.4). Now it is a function of t and we conclude the proof by taking the second derivative at $t = 1$ which is $-(1-\epsilon)(\epsilon + s - 1) < 0$. □

3.6 Hypercontractivity and Small Set Expansion are equivalent

In this section we summarize the proof of Theorem 3.1.6 shown in [Nai14].

Proof of 3.1.6, 1) → 3). From a trivial induction of hypercontractivity, for any positive integer n, $(\boldsymbol{x}, \boldsymbol{y}) \sim ((\Omega_x)^n \times (\Omega_y)^n, \mu^{\otimes n})$, we have

$$\mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[f(\boldsymbol{x})g(\boldsymbol{y})] \leq \|f(\boldsymbol{x})\|_p \|g(\boldsymbol{y})\|_{q'} \qquad \forall f : (\Omega_x)^n \to \mathbb{R}, g : (\Omega_y)^n \to \mathbb{R}.$$

Then we can get 3) by setting $f = 1_A$ and $g = 1_B$, where 1_A and 1_B are indicator functions of sets A and B. □

Proof of 3.1.6, 3) → 2). Let

$$\begin{aligned}
A &:= \{x \in (\Omega_x)^n : \forall a \in \Omega_x, |\{i : x_i = a\}| = c_x(a)\}, \\
B &:= \{y \in (\Omega_y)^n : \forall b \in \Omega_y, |\{i : y_i = b\}| = c_y(b)\}, \\
C &:= \{(x, y) \in (\Omega_x)^n \times (\Omega_y)^n : \forall a \in \Omega_x, b \in \Omega_y, |\{i : (x_i, y_i) = (a, b)\}| = c_{x,y}(a, b)\},
\end{aligned}$$

for some histogram c where $n = \sum_{a,b} c_{x,y}(a, b)$, $c_x(a) = \sum_b c_{x,y}(a, b)$, $c_y(b) = \sum_a c_{x,y}(a, b)$. If $(x, y) \in C$, we have $x \in A$ and $y \in B$. Therefore,

$$|C| \leq \mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x} \in A, \boldsymbol{y} \in B] \leq |A|^{\frac{1}{p}} |B|^{\frac{1}{q'}}. \tag{3.9}$$

If we set $\frac{c_{x,y}}{n} \to \nu(x, y)$ as $n \to \infty$, then Claim 3.2.4 shows that

$$\frac{1}{n} \log |A| \to -D(\nu_x \| \mu_x), \quad \frac{1}{n} \log |B| \to -D(\nu_y \| \mu_y), \quad \frac{1}{n} \log |C| \to -D(\nu \| \mu).$$

Hence 2) holds by taking the logarithm on both sides of (3.9). □

Proof of 3.1.6, 2) → 1). Without loss of generalization we can assume $\mathbf{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu}[f(\boldsymbol{x})g(\boldsymbol{y})] = 1$, so we only need to prove $\|f(\boldsymbol{x})\|_p \|g(\boldsymbol{y})\|_{q'} \geq 1$. We set

$$\nu(x, y) = \mu(x, y) f(x) g(y),$$

Therefore

$$\begin{aligned}
D(\nu \| \mu) &= \sum_{x,y} \nu(x, y) \log \frac{\nu(x, y)}{\mu(x, y)} \\
&= \sum_{x,y} \nu(x, y) \log(f(x)g(y)) \\
&= \sum_x \nu(x) \log f(x) + \sum_y \nu(y) \log g(y).
\end{aligned}$$

Then we have

$$
\begin{aligned}
0 &\le D(\nu\|\mu) - \frac{1}{p}D(\nu_X\|\mu_X) - \frac{1}{q'}D(\nu_Y\|\mu_Y) \\
&= \sum_x \nu(x)\log f(x) + \sum_y \nu(y)\log g(y) - \frac{1}{p}\sum_x \nu(x)\log\frac{\nu(x)}{\mu(x)} - \frac{1}{q'}\sum_y \nu(y)\log\frac{\nu(y)}{\mu(y)} \\
&= \frac{1}{p}\sum_x \nu(x)\log\frac{f(x)^p\mu(x)}{\nu(x)} + \frac{1}{q'}\sum_y \nu(y)\log\frac{g(y)^{q'}\mu(y)}{\nu(y)} \\
&\le \frac{1}{p}\log\left(\sum_x \mu(x)f(x)^p\right) + \frac{1}{q'}\log\left(\sum_y \mu(y)g(y)^{q'}\right).
\end{aligned}
$$

The last inequality holds by Jenson's inequality. Then we can conclude by taking the exponential on both sides of the inequality. □

Chapter 4

A New Homogeneous Tail Bound for Boolean Functions via One-block Decoupling

4.1 Introduction

Broadly speaking, *decoupling* refers to the idea of analyzing a complicated random sum involving dependent random variables by comparing it to a simpler random sum where some independence is introduced between the variables. For perhaps the simplest example, if $(a_{ij})_{i,j=1}^n \in \mathbb{R}$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ are independent uniform ± 1 random variables, we might ask how the moments of

$$\sum_{i,j=1}^n a_{ij}\boldsymbol{x}_i\boldsymbol{x}_j, \text{ and its "decoupled version" } \sum_{i,j=1}^n a_{ij}\boldsymbol{x}_i\boldsymbol{y}_j$$

compare. The theory of decoupling inequalities developed originally in the study of Banach spaces, stochastic processes, and U-statistics, mainly between the mid-'80s and mid-'90s; see [dlPG99] for a book-length treatment.

The powerful tool of decoupling seems to be relatively under-used in theoretical computer science. ([BM13] proves a variant of Hanson-Wright Inequality using decoupling inequalities with degree two; a recent work of Makarychev and Sviridenko [MS14] provides another exception, though they use a much different kind of decoupling than the one studied in this chapter.) In this work we will observe several places where decoupling can be used in a "black-box" fashion to solve or simplify problems quite easily.

The main topic of this chapter, however, is to study a partial form decoupling that we call "one-block decoupling". The advantage of one-block decoupling is that for degree-k polynomials we can achieve bounds with only *polynomial* dependence on k, as opposed to the exponential dependence on k that arises for the standard full decoupling. Although one-block decoupling does not introduce as much independence as full decoupling does, we show several applications where one-block decoupling is sufficient.

The applications we describe in this chapter are the following:

- (Theorem 4.2.8.) Aaronson and Ambainis's conjecture concerning the generality of their [AA18, Theorem 4] holds. I.e., there is a sublinear-query algorithm for estimating any bounded, constant-degree Boolean function.
- (Theorem 4.2.13.) The Aaronson–Ambainis Conjecture [Aar08, AA14] holds if and only if it holds for one-block decoupled functions. We also show how the best known result towards the conjecture can be proven extremely easily (4.1) in the case of one-block decoupled functions.
- (Corollary 4.3.6.) An optimal form of the DFKO Fourier Tail Bound [DFKO07]: any bounded Boolean function f that is far from being a junta satisfies $\sum_{|S|>k} \widehat{f}(S)^2 \geq \exp(-O(k^2))$. Relatedly (Corollary 4.3.5), any degree-k real-valued Boolean function with $\Omega(1)$ variance and small influences must exceed 1 in absolute value with probability at least $\exp(-O(k^2))$; this can be further improved to $\exp(-O(k))$ if f is homogeneous.

4.1.1 Definitions

Throughout this section, let f denote a multilinear polynomial of degree at most k in n variables $x = (x_1, \ldots, x_n)$, with coefficients a_S from a Banach space:

$$f(x) = \sum_{\substack{S \subseteq [n] \\ |S| \leq k}} a_S x_S,$$

where we write $x_S = \prod_{i \in S} x_i$ for brevity. (The coefficients a_S will be real in all of our applications; however we allow them to be from a Banach space since the proofs are no more complicated.)

We begin by defining our notion of partial decoupling:

Definition 4.1.1. The *one-block decoupled* version of f, denoted $\breve{f}$, is the multilinear polynomial over $2n$ variables $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$ defined by

$$\breve{f}(y, z) = \sum_{\substack{S \subseteq [n] \\ 1 \leq |S| \leq k}} a_S \sum_{i \in S} y_i z_{S \setminus i}.$$

In other words, each monomial term like $x_1 x_3 x_7$ is replaced with $y_1 z_3 z_7 + z_1 y_3 z_7 + z_1 z_3 y_7$. In case f is homogeneous we have the relation $\breve{f}(x, x) = k f(x)$.

Let us also recall the traditional notion of decoupling:

Definition 4.1.2. The *(fully) decoupled* version of f, which we denote by $\widetilde{f}$, is a multilinear polynomial over k *blocks* $x^{(1)}, \ldots, x^{(k)}$ of n variables; each $x^{(i)}$ is $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$. It is formed as follows: for each monomial x_S in f, we replace it with the average over all ways of assigning its variables to different blocks. More formally,

$$\widetilde{f}(x^{(1)}, \ldots, x^{(k)}) = a_\emptyset + \sum_{\substack{S \subseteq [n] \\ 1 \leq |S| \leq k}} \frac{(k - |S|)!}{k!} \cdot a_S \sum_{\substack{\text{injective} \\ b : S \to [k]}} \prod_{i \in S} x_i^{(b(i))}.$$

The definition is again simpler if f is homogeneous. For example, if f is homogeneous of degree 3, then each monomial in f like $x_1x_3x_7$ is replaced in $\widetilde{f}$ with

$$\frac{1}{6}(w_1y_2z_3 + w_1z_2y_3 + y_1w_2z_3 + y_1z_2w_3 + z_1w_2y_3 + z_1y_2w_3).$$

(Here we wrote w, y, z instead of $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, for simplicity.) Note that $\widetilde{f}(x, x, \dots, x) = f(x)$ always holds, even if f is not homogeneous.

We conclude by comparing the two kinds of decoupling. Assume for simplicity that f is homogeneous of degree k. The fully decoupled version $\widetilde{f}(x^{(1)}, \dots, x^{(k)})$ is in "block-multilinear form"; i.e., each monomial contains exactly one variable from each of the k "blocks". This kind of structure has often been recognized as useful in theoretical computer science; see, e.g., [KN08, Lov10, KM13, AA18]. By contrast, the one-block decoupling $\breve{f}(y, z)$ does not have such a simple structure; we only have that each monomial contains exactly one y-variable. Nevertheless we will see several examples in this chapter where having one-block decoupled form is just as useful as having fully decoupled form. And as mentioned, we will show that it is possible to achieve one-block decoupling with only $\mathrm{poly}(k)$ parameter losses, whereas full decoupling in general suffers exponential losses in k.

Remark 4.1.3. We have also chosen different "scalings" for the two kinds of decoupling. For example, in the homogeneous case, we have $\widetilde{f}(y, z, z, \dots, z) = \frac{1}{k} \cdot \breve{f}(y, z)$ and also $\mathbf{Var}[\widetilde{f}] = \frac{1}{k \cdot k!} \mathbf{Var}[\breve{f}]$ for $f : \{\pm 1\}^n \to \mathbb{R}$.

4.1.2 A useful inequality

Several times we will use the following basic inequality from analysis of Boolean functions, which relies on hypercontractivity; see [O'D14, Theorems 9.24, 10.23].

Theorem 4.1.4. *Let $f(x) = \sum_{|S| \le k} a_S x_S$ be a nonconstant n-variate multilinear polynomial of degree at most k, where the coefficients a_S are real. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be independent uniform ± 1 random variables. Then*

$$\mathbf{Pr}\big[f(\boldsymbol{x}) > \mathbf{E}[f]\big] \ge \tfrac{1}{4}e^{-2k}.$$

This also holds if some of the $\boldsymbol{x}_i$'s are standard Gaussians.[1] *Finally, if the $\boldsymbol{x}_i$'s are not uniform ± 1 random variables, but they take on each value ± 1 with probability at least λ, then we may replace $\frac{1}{4}e^{-2k}$ by $\frac{1}{4}(e^2/2\lambda)^{-k}$.*

4.2 Decoupling theorems, and query complexity applications

4.2.1 Classical decoupling inequalities, and an application in query complexity

Traditional decoupling inequalities compare the probabilistic behavior of f and $\widetilde{f}$ under independent random variables (usually symmetric ones; e.g., standard Gaussians). The easier forms

[1] Although it is not stated in [O'D14], an identical proof works since Gaussians have the same hypercontractivity properties as uniform ± 1 random variables.

of the inequalities compare expectations under a convex test function; e.g., they can be used to compare p-norms. The following was essentially proved in [dlP92]; see [dlPG99, Theorem 3.1.1,(3.4.23)–(3.4.27)]:

Theorem 4.2.1. *Let $\Phi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ be convex and nondecreasing. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ consist of independent real random variables with all moments finite, and let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}$ denote independent copies of $\boldsymbol{x}$. Then*

$$\mathbf{E}\Big[\Phi\Big(\Big\|\widetilde{f}\big(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}\big)\Big\|\Big)\Big] \leq \mathbf{E}\Big[\Phi\Big(C_k \|f(\boldsymbol{x})\|\Big)\Big],$$

where $C_k = k^{O(k)}$ is a constant depending only on k.

Remark 4.2.2. A reverse inequality also holds, with worse constant $C_k = k^{-O(k^2)}$.

Another line of research gave comparisons between tail bounds for f and $\widetilde{f}$. This culminated in the following theorem from [dlPMS95, Gin98]; see also [dlPG99, Theorem 3.4.6]:

Theorem 4.2.3. *In the setting of Theorem 4.2.1, for all $t > 0$,*

$$\mathbf{Pr}\Big[\Big\|\widetilde{f}\big(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}\big)\Big\| > C_k t\Big] \leq D_k \, \mathbf{Pr}\Big[\|f(\boldsymbol{x})\| > t\Big],$$

where $C_k = D_k = k^{O(k)}$. The analogous reverse bound also holds.

Remark 4.2.4. Kwapień [Kwa87] showed that when the $\boldsymbol{x}_i$'s are α-stable random variables, the constant C_k in Theorem 4.2.1, can be improved to $k^{k/\alpha}/k!$; this is $k^{k/2}/k!$ for standard Gaussians. Furthermore, for uniform ± 1 random variables Kwapień's proof goes through as if they were 1-stable; thus in this case one may take $C_k = k^k/k! \leq e^k$. In the Gaussian setting with homogeneous f, Kwapień obtains $C_k = k^{k/2}/k!$ and $D_k = 2^k$ for Theorem 4.2.3.

For function $f(\boldsymbol{x}) = \sum_{|S| \leq k} a_S \boldsymbol{x}_S$ where coefficients a_S are real, we denote its p-norm $\|f\|_p = \mathbf{E}[f(\boldsymbol{x})^p]^{1/p}$. Furthermore if f is a bounded function with input $\boldsymbol{x}$, we denote the infinity norm

$$\|f\|_\infty = \lim_{p \to \infty} \|f\|_p = \sup_{\boldsymbol{x}} |f(\boldsymbol{x})|.$$

Corollary 4.2.5. *In the setting of Theorem 4.2.1, it holds that $\|\widetilde{f}\|_\infty \leq k^{O(k)} \|f\|_\infty$. Further, if $f : \{\pm 1\}^n \to \mathbb{R}$ then $\|\widetilde{f}\|_\infty \leq (2e)^k \|f\|_\infty$.*

Proof. The first statement is an immediate corollary of either Theorem 4.2.1 (taking $\Phi(u) = u^p$ and $p \to \infty$) or Theorem 4.2.3 (taking $t = \|f\|_\infty$). The second statement is immediate from Remark 4.2.4, with the better constant $k^k/k!$ in case f is homogeneous. In the general case, we use the fact that if $f^{=j}$ denotes the degree-j part of f, then $\|f^{=j}\|_\infty \leq 2^j \|f\|_\infty$; this is also proved by Kwapień [Kwa87, Lemma 2]. Then

$$\Big\|\widetilde{f}\Big\|_\infty = \Big\|\sum_{j=0}^k \widetilde{f^{=j}}\Big\|_\infty \leq \sum_{j=0}^k \Big\|\widetilde{f^{=j}}\Big\|_\infty \leq \sum_{j=0}^k (j^j/j!) \big\|f^{=j}\big\|_\infty \leq \sum_{j=0}^k (j^j/j!) 2^j \|f\|_\infty \leq (2e)^k \|f\|_\infty.$$

□

Remark 4.2.6. Classical decoupling theory has not been too concerned with the dependence of constants on k, and most statements like Theorem 4.2.3 in the literature simply write $D_k = C_k$

to conserve symbols. However there are good reasons to retain the distinction, since making C_k small is usually much more important than making D_k small. For example, we can deduce Corollary 4.2.5 from Theorem 4.2.3 regardless of D_k's value.

Let us give an example application of these fundamental decoupling results. In a recent work comparing quantum query complexity to classical randomized query complexity, Aaronson and Ambainis [AA18] presented[2] [3] the following:

Conjecture 4.2.7. *Let f be an N-variate degree-k homogeneous block-multilinear polynomial with real coefficients. Assume that under uniformly random ± 1 inputs we have $\|f\|_\infty \leq 1$. Then there is a randomized query algorithm making $2^{O(k)}(N/\epsilon^2)^{1-1/k}$ nonadaptive queries to the coordinates of $x \in \{\pm 1\}^N$ that outputs an approximation to $f(x)$ that is accurate to within $\pm\epsilon$ (with high probability).*

The authors "strongly conjecture[d]" that the assumption of block-multilinearity could be removed, and gave a somewhat lengthy proof of this conjecture in the case of $k = 2$, using [DFKO07] . We note that the full conjecture follows almost immediately from full decoupling:

Theorem 4.2.8. *If Aaronson and Ambainis's Conjecture 4.2.7 holds, then it holds without the assumption of block-multilinearity or homogeneity.*

Proof. Given a non-block-multilinear f on N variables ranging in $\{\pm 1\}$, consider its full decoupling $\widetilde{f}$ on kN variables. By Corollary 4.2.5 we have $\|\widetilde{f}\|_\infty \leq (2e)^k$. Let $g = (2e)^{-k}\widetilde{f}$, so that $g : \{\pm 1\}^{kN} \to [-1, +1]$ is a degree-k block-multilinear polynomial with $f(x) = (2e)^k g(x, x, \ldots, x)$. Now given query access to $x \in \{\pm 1\}^N$ and an error tolerance ϵ, we apply Conjecture 4.2.7 to $g(x, x, \ldots, x)$ with error tolerance $\epsilon_1 = (2e)^{-k}\epsilon$; note that we can simulate queries to $(x, x, \ldots, x)$ using queries to x. This gives the desired query algorithm, and it makes $2^{O(k)}(kN/\epsilon_1^2)^{1-1/k} = 2^{O(k)}(N/\epsilon^2)^{1-1/k}$ queries as claimed. There is one more minor point: Conjecture 4.2.7 requires its function to be homogeneous in addition to block-multilinear. However this assumption is easily removed by introducing k dummy variables treated as $+1$, and padding the monomials with them. □

4.2.2 Our one-block decoupling theorems, and the AA Conjecture

We now state our new versions of Theorems 4.2.1, 4.2.3 which apply only to one-block decoupling, but that have *polynomial* dependence of C_k on k. Proofs are deferred to Section 4.4.

As before, let $f(x) = \sum_{|S|\leq k} a_S x_S$ be an n-variate multivariate polynomial of degree at most k with coefficients a_S in a Banach space; let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ consist of independent real random variables with all moments finite, and let $\boldsymbol{y}$, $\boldsymbol{z}$ be independent copies. We consider three

[2] Actually, there is a small gap in their proof. In the line reading "By the concavity of the square root function…", they claim that $\|\boldsymbol{X}\|_1 \geq \|\boldsymbol{X}\|_2$ when $\boldsymbol{X}$ is a degree-k polynomial of uniformly random ± 1 bits. In fact the inequality goes the other way in general. But the desired inequality does hold up to a factor of e^k by [O'D14, Theorem 9.22], and this is sufficient for their proof.

[3] Conjecture 4.2.7 is claimed as a theorem in [AA18]. However Aaronson et al.[AAB+21] found a flaw but prove the correctness in the case $k = 1$.

slightly different hypotheses:

H1: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \sim \mathrm{N}(0,1)$ are standard Gaussians.
H2: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are uniformly random ± 1 values.
H3: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are uniformly random ± 1 values and f is homogeneous.

Theorem 4.2.9. *If $\Phi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ is convex and nondecreasing, then*

$$\mathbf{E}\Big[\Phi\Big(\Big\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big\|\Big)\Big] \leq \mathbf{E}\Big[\Phi\Big(C_k\|f(\boldsymbol{x})\|\Big)\Big].$$

Also, if $t > 0$ (and we assume f's coefficients a_S are real under **H2**, **H3***), then*

$$\mathbf{Pr}\Big[\Big\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big\| > C_k t\Big] \leq D_k \,\mathbf{Pr}\Big[\|f(\boldsymbol{x})\| > t\Big].$$

Here

$$C_k = \begin{cases} O(k) & \text{under } \mathbf{H1}, \\ O(k^2) & \text{under } \mathbf{H2}, \\ O(k^{3/2}) & \text{under } \mathbf{H3}, \end{cases} \qquad D_k = \begin{cases} O(k) & \text{under } \mathbf{H1}, \\ k^{O(k)} & \text{under } \mathbf{H2, H3}. \end{cases}$$

Remark 4.2.10. It may seem that for the Φ-inequality in the Gaussian case, Kwapień's result mentioned in Remark 4.2.4 is better than ours, since he achieves full decoupling with a better constant than we get for one-block decoupling. But actually they are incomparable; the reason is the different scaling mentioned in Remark 4.1.3.

Remark 4.2.11. As we will explain later in Remark 4.3.4, the bound $C_k = O(k)$ under **H1** is best possible (assuming that $D_k \leq \exp(O(k^2))$).

An immediate consequence of the above theorem, as in Corollary 4.2.5, is the following:

Corollary 4.2.12. *If $f : \{\pm 1\}^n \to \mathbb{R}$ then $\|\breve{f}\|_\infty \leq O(k^2)\|f\|_\infty$.*

Let us now give an example of how one-block decoupling can be as useful as full decoupling, and why it is important to obtain $C_k = \mathrm{poly}(k)$. A very notable open problem in analysis of Boolean functions is the *Aaronson–Ambainis (AA) Conjecture*, originally proposed in 2008 [Aar08, AA14]:

AA Conjecture. *Let $f : \{\pm 1\}^n \to [-1,+1]$ be computable by a multilinear polynomial of degree at most k, $f(x) = \sum_{|S| \leq k} a_S x_S$. Then $\mathbf{MaxInf}[f] \geq \mathrm{poly}(\mathbf{Var}[f]/k)$.*

Here we use the standard notations for influences and variance:

$$\mathbf{MaxInf}[f] = \max_{i \in [n]}\{\mathbf{Inf}_i[f]\}, \qquad \mathbf{Inf}_i[f] = \sum_{S \ni i} a_S^2, \qquad \mathbf{Var}[f] = \sum_{S \neq \emptyset} a_S^2, \qquad \|f\|_2^2 = \sum_S a_S^2.$$

The AA Conjecture is known to imply (and was directly motivated by) the following folklore conjecture concerning the limitations of quantum computation, dated to 1999 or before [AA14]:

Quantum Conjecture. *Any quantum query algorithm solving a Boolean decision problem using T queries can be correctly simulated on a $1-\epsilon$ fraction of all inputs by a classical query algorithm using $\mathrm{poly}(T/\epsilon)$ queries.*

Because of their importance for quantum computation, Aaronson has twice listed these conjectures as "semi-grand challenges for quantum computing theory" [Aar05, Aar10].

The best known result in the direction of the AA Conjecture [AA14] obtains an influence lower bound of $\mathrm{poly}(\mathbf{Var}[f])/\exp(O(k))$, using the DFKO Inequality [DFKO07]. Here we observe that there is a "one-line" deduction of this bound under the assumption that f is one-block decoupled.[4] To see this, suppose that f is indeed one-block decoupled, so it can be written as $f(y,z)=\sum_{i=1}^{n} y_i g_i(z)$, where $g_i(z)=\sum_{S\ni i} a_S z_{S\setminus i}$ is the ith "derivative" of f. Observe that $\|g_i\|_2^2=\mathbf{Inf}_i[f]$ and hence $\sum_{i=1}^{n}\|g_i\|_2^2\geq\mathbf{Var}[f]$. Also note that for any $z\in\{\pm1\}^n$ we must have $\sum_{i=1}^{n}|g_i(z)|\leq 1$, as otherwise we could achieve $|f(y,z)|>1$ by choosing $y\in\{\pm1\}^n$ appropriately. Taking expectations we get $\sum_{i=1}^{n}\|g_i\|_1\leq 1$, and hence

$$e^{k-1}\geq e^{k-1}\sum_{i=1}^{n}\|g_i\|_1\geq\sum_{i=1}^{n}\|g_i\|_2\geq\frac{\sum_{i=1}^{n}\|g_i\|_2^2}{\max_{i=1}^{n}\|g_i\|_2}\geq\frac{\mathbf{Var}[f]}{\max_{i=1}^{n}\sqrt{\mathbf{Inf}_i[f]}}$$
$$\Rightarrow\qquad \mathbf{MaxInf}[f]\geq e^{2-2k}\,\mathbf{Var}[f]^2, \tag{4.1}$$

where the second inequality used the basic fact in analysis of Boolean functions [O'D14, Theorem 9.22] that $\|g\|_2\leq e^{k-1}\|g\|_1$ for $g:\{\pm1\}^n\to\mathbb{R}$ of degree at most $k-1$.

The above gives a good illustration of how even one-block decoupling can already greatly simplify arguments in analysis of Boolean functions. We feel that (4.1) throws into sharp relief the challenge of improving $\exp(-O(k))$ to $1/\mathrm{poly}(k)$ for the AA Conjecture. We now use our results to show that the assumption that f is one-block decoupled is completely without loss of generality.

Theorem 4.2.13. *The AA Conjecture holds if and only if it holds for one-block decoupled functions f.*

Proof. Suppose $f:\{\pm1\}^n\to[-1,+1]$ has degree at most k. By Corollary 4.2.12 we get that $\|\breve{f}\|_\infty\leq C_k=O(k^2)$. Now $g=C_k^{-1}\breve{f}$ is one-block decoupled and has range $[-1,+1]$. Assuming the AA Conjecture holds for it, we get some $i\in[2n]$ such that $\mathbf{Inf}_i[g]\geq\mathrm{poly}(\mathbf{Var}[g]/k)$. Certainly this implies $\mathbf{Inf}_i[\breve{f}]\geq\mathrm{poly}(\mathbf{Var}[\breve{f}]/k)$. It is easy to see that $\mathbf{Inf}_i[f]=\mathbf{Inf}_i[\breve{f}]$ and $\mathbf{Inf}_i[f]\geq\mathbf{Inf}_{i+n}[\breve{f}]/(k-1)$ for all $i\in[n]$. Therefore letting $i'=\max\{i,i-n\}\in[n]$, we have $\mathbf{Inf}_{i'}[f]\geq\mathbf{Inf}_i[\breve{f}]/(k-1)$, and also $\mathbf{Var}[\breve{f}]\geq\mathbf{Var}[f]$. Thus $\mathbf{Inf}_{i'}[f]\geq\mathrm{poly}(\mathbf{Var}[f]/k)$, confirming the AA Conjecture for f. □

In particular, by combining this with (4.1) we recover the known $\mathrm{poly}(\mathbf{Var}[f])/\exp(O(k))$ lower bound for the AA Conjecture as applied to general f.

[4]This observation is joint with John Wright.

4.3 Tight versions of the DFKO theorems

This section is concerned with analysis of Boolean functions $f : \{\pm 1\}^n \to \mathbb{R}$. We will use traditional Fourier notation, writing $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x_S$. A key theme in this field is the dichotomy between functions with "Gaussian-like" behavior and functions that are essentially "juntas". Recall that f is said to be an (ϵ, C)-junta if $\|f - g\|_2^2 \leq \epsilon$ for some $g : \{\pm 1\}^n \to \mathbb{R}$ depending on at most C input coordinates. Partially exemplifying this theme is a family of theorems stating that any Boolean function f which is not essentially a junta must have a large "Fourier tail" — something like $\sum_{|S|>k} \widehat{f}(S)^2 > \delta$. Examples of such results include Friedgut's Average Sensitivity Theorem [Fri98], the FKN Theorem [FKN02] (sharpened in [JOW15, O'D14]), the Kindler–Safra Theorem [KS02, Kin02], and the Bourgain Fourier Tail Theorem [Bou02]. The last of these implies that any $f : \{\pm 1\}^n \to \{\pm 1\}$ which is not a $(.01, k^{O(k)})$-junta must satisfy $\sum_{|S|>k} \widehat{f}(S)^2 > k^{-1/2+o(1)}$. This $k^{-1/2+o(1)}$ bound was made more explicit in [KN06], and the optimal bound of $\Omega(k^{-1/2})$ was obtained in [KO12]. These "Fourier tail" theorems have had application in fields such as PCPs and inapproximability [Kho02, Din07], sharp threshold theory [FK96], extremal combinatorics [EFF12], and social choice [FKN02].

All of the aforementioned theorems concern Boolean-*valued* functions; i.e., those with range $\{\pm 1\}$. By contrast, the DFKO Fourier Tail Theorem [DFKO07] is a result of this flavor for *bounded* functions; i.e., those with range $[-1, +1]$.

DFKO Fourier Tail Theorem. *Suppose $f : \{\pm 1\}^n \to [-1, +1]$ is not an $(\epsilon, 2^{O(k)}/\epsilon^2)$-junta. Then*

$$\sum_{|S|>k} \widehat{f}(S)^2 > \exp(-O(k^2 \log k)/\epsilon).$$

Most applications do not use this Fourier tail theorem directly. Rather, they use a key intermediate result, [DFKO07, Theorem 3], which we will refer to as the "DFKO Inequality". This was the case, for example, in a recent work on approximation algorithms for the Max-kXOR problem [BMO+15].

DFKO Inequality. *Suppose $f : \{\pm 1\}^n \to \mathbb{R}$ has degree at most k and $\mathbf{Var}[f] \geq 1$. Let $t \geq 1$ and suppose that $\mathbf{MaxInf}[f] \leq 2^{-O(k)}/t^2$. Then $\mathbf{Pr}[|f(\boldsymbol{x})| > t] \geq \exp(-O(t^2 k^2 \log k))$.*

Returning to the theme of "Gaussian-like behavior" versus "junta" behavior, we may add that the DFKO results straightforwardly imply (by the Central Limit Theorem) analogous, simpler-to-state results concerning functions on Gaussian space and Hermite tails. We record these generic consequences here; see, e.g., [O'D14, Sections 11.1, 11.2] for a general discussion of such implications, and the definitions of Hermite coefficients $\widehat{f}(\alpha)$.

Corollary 4.3.1. *Any $f : \mathbb{R}^n \to [-1, +1]$ satisfies the Hermite tail bound*

$$\sum_{|\alpha|>k} \widehat{f}(\alpha)^2 > \exp(-O(k^2 \log k)/\mathbf{Var}[f]).$$

Furthermore, suppose $\boldsymbol{z}$ is a standard n-dimensional Gaussian random vector and $t \geq 1$. Then any n-variate polynomial f of degree at most k with $\mathbf{Var}[f(\boldsymbol{z})] \geq 1$ satisfies $\mathbf{Pr}[|f(\boldsymbol{z})| > t] \geq \exp(-O(t^2k^2 \log k))$.

Even though the Gaussian results in Corollary 4.3.1 are formally easier than their Boolean counterparts, we are not aware of any way to prove them — even in the case $n = 1$ — except via DFKO.

Tightness of the bounds. In [DFKO07, Section 6] it is shown that the results in Corollary 4.3.1 are tight, up to the $\log k$ factor in the exponent; this implies the same statement about the DFKO Fourier Tail Theorem and the DFKO Inequality. The tight example in both cases is essentially the univariate, degree-k Chebyshev polynomial.[5] In the next section we will show how to use our one-block decoupling result to remove the $\log k$ in the exponential from both DFKO theorems. The results immediately transfer to the Gaussian setting, and we therefore obtain the tight $\exp(-\Theta(k^2))$ bound for all versions of the inequality.

Our method of proof is actually to *first* prove the results in the Gaussian setting, where the one-block decoupling makes the proofs quite easy. Then we can transfer the results to the Boolean setting by using the Invariance Principle [MOO10]. This methodology — proving the more natural Gaussian tail bound first, then transferring the result to the Boolean setting via Invariance — is quite reminiscent of how the optimal form of Bourgain's Fourier Tail Theorem was recently obtained [KO12].

There is actually an additional, perhaps unexpected, bonus of our proof methodology; we show that the bound in the DFKO Inequality can be improved from $\exp(-O(t^2k^2))$ to $\exp(-O(t^2k))$ whenever f is *homogeneous*.

4.3.1 Proofs of the tight DFKO theorems

We begin with a tail-probability lower bound for one-block decoupled polynomials of Gaussians.

Lemma 4.3.2. *Suppose $f(y, z) = \sum_{i=1}^{n} y_i g_i(z)$ is a one-block decoupled polynomial on $n + n$ variables, with real coefficients and degree at most k. Let $\boldsymbol{y}, \boldsymbol{z} \in \mathrm{N}(0,1)^n$ be independent standard n-dimensional Gaussians and write*

$$\sigma^2 = \mathbf{Var}[f(\boldsymbol{y}, \boldsymbol{z})] = \sum_{i=1}^{n} \|g_i\|_2^2. \tag{4.2}$$

Then for $u > 0$ we have $\mathbf{Pr}[|f(\boldsymbol{y}, \boldsymbol{z})| > u] \geq \exp(-O(k + u^2/\sigma^2))$.

Proof. Let $v(z) = \sum_{i=1}^{n} g_i(z)^2$, a polynomial of degree at most $2(k-1)$ in $z_1, \ldots, z_n$. By (4.2) we have $\mathbf{E}[v(\boldsymbol{z})] = \sigma^2$. We now use Theorem 4.1.4 to get

$$\mathbf{Pr}[v(\boldsymbol{z}) > \sigma^2] \geq \frac{1}{4} e^{-2(2k-1)} = \exp(-O(k)).$$

[5] Formally speaking, [DFKO07, Section 6] only argues tightness of the Boolean theorems, but their constructions are directly based on the degree-k Chebyshev polynomial applied to a single standard Gaussian.

On the other hand, for any outcome $\boldsymbol{z} = z$ we have that $f(\boldsymbol{y}, z) \sim \mathrm{N}(0, v(z))$. Thus

$$v(z) > \sigma^2 \qquad \Longrightarrow \qquad \mathbf{Pr}[|f(\boldsymbol{y}, z)| > u] \geq \Omega(e^{-u^2/2\sigma^2}).$$

Combining the previous two statements completes the proof, since $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent. □

We can now prove an optimal version of the DFKO Inequality in the Gaussian setting. It is also significantly better in the homogeneous case.

Theorem 4.3.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most k, and let $\boldsymbol{x} \sim \mathrm{N}(0,1)^n$ be a standard n-dimensional Gaussian vector. Assume $\mathbf{Var}[f(\boldsymbol{x})] \geq 1$. Then for $t \geq 1$ it holds that $\mathbf{Pr}[|f(\boldsymbol{x})| > t] \geq \exp(-O(t^2k^2))$. Furthermore, if f is multilinear and homogeneous then the lower bound may be improved to $\exp(-O(t^2k))$.*

Proof. For any n-variate polynomial of Gaussians, we can find an N-variate multilinear polynomial of Gaussians of no higher degree that is arbitrarily close in Lévy distance (see, e.g., [Kan11, Lemma 15], or use the CLT to pass to ± 1 random variables, then Invariance to pass back to Gaussians). Note, however, that this transformation does not preserve homogeneity. In any case, we can henceforth assume f is multilinear, $f(x) = \sum_{|S| \leq k} a_S x_S$.

For independent $\boldsymbol{y}, \boldsymbol{z} \sim \mathrm{N}(0,1)^n$, observe that

$$\mathbf{Var}[\breve{f}(\boldsymbol{y}, \boldsymbol{z})] = \sum_{j=1}^{k} j \sum_{|S|=j} a_S^2 \geq \sum_{S \neq \emptyset} a_S^2 = \mathbf{Var}[f(\boldsymbol{x})] \geq 1,$$

and if f is homogeneous we get the better bound $\mathbf{Var}[\breve{f}(\boldsymbol{y}, \boldsymbol{z})] \geq k$. By our Theorem 4.2.9 on one-block decoupling, we have

$$\mathbf{Pr}\Big[\Big|f(\boldsymbol{x})\Big| > t\Big] \geq D_k^{-1} \, \mathbf{Pr}\Big[\Big|\breve{f}(\boldsymbol{y}, \boldsymbol{z})\Big| > C_k t\Big],$$

where $C_k = D_k = O(k)$. The theorem is now an immediate consequence of Lemma 4.3.2. □

Remark 4.3.4. A consequence of this proof is that — assuming $D_k \leq \exp(O(k^2))$ — it is impossible to asymptotically improve on our $C_k = O(k)$ in Theorem 4.2.9 in the Gaussian setting **H1**. Otherwise, we would achieve a bound of $\exp(-o(k^2))$ in Theorem 4.3.3, contrary to the example in [DFKO07, Section 6].

We can now obtain the sharp DFKO Inequality in the Boolean setting by using the Invariance Principle.

Corollary 4.3.5. *Theorem 4.3.3 holds when $\boldsymbol{x} \sim \{\pm 1\}^n$ is uniform and we additionally assume that $\mathbf{MaxInf}[f] \leq \exp(-Ct^2k^2)$, or just $\exp(-Ct^2k)$ in the homogeneous case. Here C is a universal constant.*

Proof. This follows immediately from the Lévy distance bound in [MOO10, Theorem 3.19, Hypothesis 4]. We only need to ensure that the Lévy distance is noticeably less than the target lower bound we're aiming for. (We also remark that the Invariance Principle transformation preserves variance and homogeneity.) □

Next, we obtain the sharp DFKO Fourier Tail Theorem. Its deduction from the DFKO Inequality in [DFKO07] is unfortunately not "black-box", so we will have to give a proof.

Corollary 4.3.6. *Suppose $f : \{\pm 1\}^n \to [-1, +1]$ is not an $(\epsilon, 2^{O(k^2/\epsilon)})$-junta. Then*

$$\sum_{|S|>k} \widehat{f}(S)^2 > \exp(-O(k^2)/\epsilon). \tag{4.3}$$

Proof. We use notation and basic results from [O'D14]. Given $f : \{\pm 1\}^n \to [-1, +1]$, let $J = \{i \in [n] : \mathbf{Inf}_i^{\leq k}[f] > \exp(-Ak^2/\epsilon)\}$, where A is a large constant to be chosen later. Since $\|f\|_2^2 \leq 1$ it follows easily that $|J| \leq 2^{O(k^2/\epsilon)}$. Now define $g = f - f^{\subseteq J}$; note that g has range in $[-2, +2]$ since $f^{\subseteq J}$ has range in $[-1, +1]$, being an average of f over the coordinates outside J. If $\|g\|_2^2 < \epsilon/2$ then f is $\epsilon/2$-close to the $2^{O(k^2/\epsilon)}$-junta $f^{\subseteq J}$ and we are done. Otherwise, $\|g\|_2^2 \geq \epsilon/2$ and we let $h = g^{\leq k}$. If $\|h - g\|_2^2 > \epsilon/4$ then we immediately conclude that $\sum_{|S|>k} \widehat{f}(S)^2 > \epsilon/4$, which is more than enough to be done. Otherwise $\|h - g\|_2^2 \leq \epsilon/4$, from which we conclude $\|h\|_2^2 \geq \epsilon/4$. Now h has degree at most k and satisfies $\mathbf{Inf}_i[h] \leq \exp(-Ak^2/\epsilon)$ for all $i \notin J$. Let $\widetilde{h}$ denote the mixed Boolean/Gaussian function which has the same multilinear form as h, but where we think of the coordinates in J as being ± 1 random variables and the coordinates not in J as being standard Gaussians. We now "partially" apply the Invariance Principle [MOO10, Theorem 3.19] to h, in the sense that we only hybridize over the coordinates not in J. We conclude that the Lévy distance between h and $\widetilde{h}$ is at most $\exp(-\Omega(Ak^2/\epsilon))$. Our goal is now to show that

$$\mathbf{Pr}[|\widetilde{h}| > 3] \geq \exp(-O(k^2/\epsilon)), \tag{4.4}$$

where the constant in the $O(\cdot)$ does not depend on A. Having shown this, by taking A large enough the Lévy distance bound lets us deduce (4.4) for h as well. But then since $|g| \leq 2$ always, we may immediately deduce $\|g - h\|_2^2 \geq \exp(-O(k^2)/\epsilon)$ and hence (4.3).

It remains to verify (4.4). For each restriction x_J to the J-coordinates, the function $\widetilde{h}_{x_J}$ is a multilinear polynomial in independent Gaussians with some variance $\sigma_{x_J}^2$. From Theorem 4.3.3 we can conclude that $\mathbf{Pr}[|\widetilde{h}_{x_J}| > 3] \geq \exp(-O(k^2/\sigma_{x_J}^2))$. Thus if we can show $\sigma_{\boldsymbol{x}_J}^2 \geq \Omega(\epsilon)$ with probability at least $2^{-O(k)}$ when $\boldsymbol{x}_J \in \{\pm 1\}^J$ is uniformly random, we will have established (4.4). But this follows similarly as in Lemma 4.3.2. Note that $\sigma_{x_J}^2 = \mathbf{E}[\widetilde{h}_{x_J}^2]$, since h has no constant term. Now $\sigma_{x_J}^2$ is a degree-$2k$ polynomial in x_J, and its expectation is simply $\|h\|_2^2 \geq \epsilon/4$, so Theorem 4.1.4 indeed implies that $\mathbf{Pr}[\sigma_{\boldsymbol{x}_J}^2 \geq \epsilon/4] \geq 2^{-O(k)}$ and we are done. □

Remark 4.3.7. We comment that the dependence of $\mathbf{MaxInf}[f]$ on t in Corollary 4.3.5, and the junta size in Corollary 4.3.6, are not as good as in [DFKO07]. This seems to be a byproduct of the use of Invariance.

A similar (but easier) proof can be used to derive the following Gaussian version of Corollary 4.3.6; alternatively, one can use a generic CLT argument, noting that the only "junta" a Gaussian function can be close to is a constant function:

Corollary 4.3.8. *Any $f : \mathbb{R}^n \to [-1, +1]$ satisfies the Hermite tail bound*

$$\sum_{|\alpha|>k} \widehat{f}(\alpha)^2 > \exp(-O(k^2)/\mathbf{Var}[f]).$$

This strictly improves upon Corollary 4.3.1.

4.4 Proofs of our one-block decoupling theorems

In this section we prove Theorem 4.2.9. The key idea of the proof is to express $\breve{f}(y,z)$ as a "small" linear combination of expressions of the form $f(\alpha_i x + \beta_i y)$, where $\alpha_i^2 + \beta_i^2 = 1$ (in the Gaussian case) or $|\alpha_i| + |\beta_i| = 1$ (in the Boolean case). The following is the central lemma.

Lemma 4.4.1. *In the setting of Theorem 4.2.9, there exists $m = O(k)$ and $\alpha, \beta, c \in \mathbb{R}^m$ such that*

- $\breve{f}(y,z) = \sum_{i=1}^m c_i f(\alpha_i y + \beta_i z)$*;*
- $\sum_{i=1}^m |c_i| \le C_k$*;*
- $\alpha_i^2 + \beta_i^2 = 1$ *for all* $i \in [m]$ *under* **H1***, and* $|\alpha_i| + |\beta_i| = 1$ *for all* $i \in [m]$ *under* **H2**, **H3***;*
- $|\alpha_i|, |\beta_i| \ge 1/O(C_k)$ *for all* $i \in [m]$.

With Lemma 4.4.1 in hand, the proof of Theorem 4.2.9 is quite straightforward in the Gaussian case, and not much more difficult in the Boolean case. We show these deductions first.

Proof of Theorem 4.2.9 under Hypothesis **H1**. By Lemma 4.4.1, for any convex nondecreasing function $\Phi : \mathbb{R}^{\ge 0} \to \mathbb{R}^{\ge 0}$ we have

$$\begin{aligned}
\mathbf{E}\Big[\Phi\Big(\Big\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\Big\|\Big)\Big] &= \mathbf{E}\Big[\Phi\Big(\Big\|\sum_{i=1}^m c_i f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\Big\|\Big)\Big] \\
&\le \mathbf{E}\Big[\Phi\Big(\sum_{i=1}^m |c_i|\Big\|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\Big\|\Big)\Big] \\
&\le \sum_{i=1}^m \frac{|c_i|}{C_k}\,\mathbf{E}[\Phi(C_k\|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\|)] \\
&= \sum_{i=1}^m \frac{|c_i|}{C_k}\,\mathbf{E}[\Phi(C_k\|f(\boldsymbol{x})\|)] \\
&\le \mathbf{E}[\Phi(C_k\|f(\boldsymbol{x})\|].
\end{aligned}$$

Here the inequalities follow from the convexity and monotonicity of Φ, and the second equality holds because $\alpha_i \boldsymbol{y} + \beta \boldsymbol{z} \sim \mathrm{N}(0,1)^n$ due to $\alpha_i^2 + \beta_i^2 = 1$.

As for the tail-bound comparison, by Lemma 4.4.1, whenever y, z are such that $\|\breve{f}(y,z)\| > C_k t$, the triangle inequality implies that there must exist at least one $i \in [m]$ with $\|f(\alpha_i y + \beta_i z)\| > t$. It follows that there must exist at least one $i \in [m]$ such that

$$\mathbf{Pr}[\|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\| > t] \ge \frac{1}{m}\,\mathbf{Pr}[\|\breve{f}(\boldsymbol{y},\boldsymbol{z})\| > C_k t].$$

This completes the proof, since $\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z} \sim \mathrm{N}(0,1)^n$ and $m = O(k)$. □

Proof of Theorem 4.2.9 under Hypotheses **H2**, **H3**. We define ± 1 random variables as follows:

$$\boldsymbol{x}_j^{(i)} = \begin{cases} \mathrm{sgn}(\alpha_i)\boldsymbol{y}_j & \text{with probability } |\alpha_i|, \\ \mathrm{sgn}(\beta_i)\boldsymbol{z}_j & \text{with probability } |\beta_i|, \end{cases}$$

for all $i \in [m]$ and $j \in [n]$ independently. Notice that each $\boldsymbol{x}^{(i)}$ is distributed uniformly on $\{\pm 1\}^n$, though they are not independent. To prove the desired inequality concerning Φ, we can repeat the proof in the Gaussian case, except that we no longer have the identity

$$\mathbf{E}[\Phi(C_k\|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\|)] = \mathbf{E}[\Phi(C_k\|f(\boldsymbol{x})\|)].$$

In fact we will show that the left-hand side is at most the right-hand side. Notice that for all fixed $y, z \in \{\pm 1\}^n$, the multilinearity of f implies that

$$f(\alpha_i y + \beta_i z) = \mathbf{E}[f(\boldsymbol{x}^{(i)}) \mid (\boldsymbol{y}, \boldsymbol{z}) = (y, z)]. \tag{4.5}$$

Thus

$$\begin{aligned}\mathbf{E}[\Phi(C_k\|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})\|)] &= \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{z}}\left[\Phi\left(C_k\left\|\mathop{\mathbf{E}}_{\boldsymbol{x}^{(i)}|\boldsymbol{y},\boldsymbol{z}}\left[f(\boldsymbol{x}^{(i)})\right]\right\|\right)\right] \\ &\le \mathop{\mathbf{E}}_{\boldsymbol{y},\boldsymbol{z}}\mathop{\mathbf{E}}_{\boldsymbol{x}^{(i)}}\left[\Phi\big(C_k\|f(\boldsymbol{x}^{(i)})\|\big)\right] = \mathbf{E}[\Phi(C_k\|f(\boldsymbol{x})\|)],\end{aligned}$$

as claimed, where we used convexity again.

As for the tail-bound comparison, recall that we are now assuming f has real coefficients. As in the Gaussian case there is at least one $i \in [m]$ with

$$\mathbf{Pr}[|f(\alpha_i \boldsymbol{y} + \beta_i \boldsymbol{z})| > t] \ge \frac{1}{O(k)}\mathbf{Pr}[|\breve{f}(\boldsymbol{y}, \boldsymbol{z})| > C_k t].$$

Now suppose y, z are such that $|f(\alpha_i y + \beta_i z)| > t$ and consider the conditional distribution on $\boldsymbol{x}^{(i)}$. If we can show that, conditionally, $\mathbf{Pr}[|f(\boldsymbol{x}^{(i)})| > t] \ge k^{-O(k)}$ then we are done. But from (4.5) we have that $\left|\mathbf{E}[f(\boldsymbol{x}^{(i)})]\right| > t$; therefore the desired result follows from Theorem 4.1.4 and the fact that $\min(|\alpha_i|, |\beta_i|) \ge 1/O(C_k) = 1/\mathrm{poly}(k)$. □

4.4.1 Proof of Lemma 4.4.1

The proof of Lemma 4.4.1 involves minimizing $\sum_{i=1}^m |c_i|$ by carefully setting the ratios of α_i and β_i to be a hyperharmonic progression.

Proof of Lemma 4.4.1. The main work involves treating the homogeneous case.

Homogeneous case. Our goal for homogeneous f is to write

$$\breve{f}(y, z) = \sum_{i=1}^{k+1} c_i f(\alpha_i y + \beta_i z).$$

Comparing the expressions term by term, it is equivalent to say that for any $S \subseteq [n]$ with $|S| = k$,

$$\sum_{j \in S} y_j z_{S/j} = \sum_{i=1}^{k+1} c_i \prod_{j \in S}(\alpha_i y_j + \beta_i z_j).$$

We can further simplify this to the conditions

$$\sum_{i=1}^{k+1} c_i \alpha_i^{k-t} \beta_i^t = \begin{cases} 1 & \text{if } t = k-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

for all integers $0 \le t \le k$. Let us write $\Delta_i = \frac{\beta_i}{\alpha_i}$ and introduce the Vandermonde matrix

$$V = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \Delta_1 & \Delta_2 & \cdots & \Delta_{k+1} \\ \dots & \dots & \dots & \dots \\ \Delta_1^{k-1} & \Delta_2^{k-1} & \dots & \Delta_{k+1}^{k-1} \\ \Delta_1^k & \Delta_2^k & \cdots & \Delta_{k+1}^k \end{bmatrix}.$$

We will also write A for the diagonal matrix $\mathrm{diag}(\alpha_1^k, \alpha_2^k, \dots, \alpha_{k+1}^k)$, and write e_k for the indicator vector of the kth coordinate, $e_k = (0, 0, \dots, 0, 1, 0)$. Then the necessary conditions (4.6) are equivalent to the matrix equation $VAc = e_k$. Assuming all the Δ_i's are different, V is invertible and there is an explicit formula for its inverse [MS58]. This yields the following expression for the $c_1, \dots, c_{k+1}$ in terms of α and β:

$$c_i = (A^{-1}V^{-1}e_k)_i = \frac{1}{\alpha_i^k} \cdot \frac{\Delta_i - \sum_{j=1}^{k+1} \Delta_j}{\prod_{j=1, j\neq i}^{k+1} (\Delta_i - \Delta_j)}. \tag{4.7}$$

The main illustrative case: Hypothesis H1 and k odd. We will now assume that k is odd; this assumption will be easily removed later. It will henceforth be convenient to replace our indices $1, \dots, k+1$ with the following slightly peculiar but symmetric set of indices:

$$I = \left\{ \pm 1, \pm 2, \dots, \pm \tfrac{k-1}{2}, \pm \tfrac{1}{2} \right\}.$$

Now under Hypothesis **H1**, we will choose

$$\alpha_i = \frac{i}{\sqrt{k^2 + i^2}}, \quad \beta_i = \frac{k}{\sqrt{k^2 + i^2}} \quad \Longrightarrow \quad \Delta_i = \frac{k}{i}$$

for all $i \in I$. These choices satisfy $\alpha_i^2 + \beta_i^2 = 1$ and $|\alpha_i|, |\beta_i| \ge 1/O(C_k)$, so it remains to prove that for c defined by (4.7) we have $\sum |c_i| \le O(k)$.

Let us upper-bound all $|c_i|$. Since it easy to see that $|c_i| = |c_{-i}|$ for all $i \in I$, it will suffice

for us to consider the positive $i \in I$. For $1 \le i \le \frac{k-1}{2}$, we have

$$\begin{aligned}
\left|\prod_{j\in I, j\neq i}(\Delta_i - \Delta_j)\right| &= (\Delta_{1/2} - \Delta_i)(\Delta_i - \Delta_{-1/2}) \cdot \prod_{j=1, j\neq i}^{(k-1)/2} |\Delta_i - \Delta_j| \cdot \prod_{j=-(k-1)/2}^{-1} (\Delta_i - \Delta_j) \\
&= \left(2k - \frac{k}{i}\right)\left(2k + \frac{k}{i}\right) \cdot \prod_{j=1, j\neq i}^{(k-1)/2} \left|\frac{k}{i} - \frac{k}{j}\right| \cdot \prod_{j=1}^{(k-1)/2} \left(\frac{k}{i} + \frac{k}{j}\right) \\
&= k^k\left(4 - \frac{1}{i^2}\right) \cdot \prod_{j=1, j\neq i}^{(k-1)/2} \frac{|j-i|}{ij} \cdot \prod_{j=1}^{(k-1)/2} \frac{j+i}{ij} \\
&= \frac{k^k}{i^{k-2}}\left(4 - \frac{1}{i^2}\right) \frac{\left(\frac{k-1}{2}+i\right)!\left(\frac{k-1}{2}-i\right)!}{\left(\frac{k-1}{2}\right)!^2}.
\end{aligned}$$

Thus from (4.7),

$$\begin{aligned}
|c_i| &= \left(\frac{\sqrt{k^2+i^2}}{i}\right)^k \cdot \frac{k}{i} \cdot \frac{i^{k-2}}{k^k} \cdot \frac{1}{4 - 1/i^2} \cdot \frac{\left(\frac{k-1}{2}\right)!^2}{\left(\frac{k-1}{2}+i\right)!\left(\frac{k-1}{2}-i\right)!} \\
&= \frac{k}{i^3}\left(1 + \frac{i^2}{k^2}\right)^{k/2} \frac{1}{4 - 1/i^2} \frac{\left(\frac{k-1}{2}\right)!^2}{\left(\frac{k-1}{2}+i\right)!\left(\frac{k-1}{2}-i\right)!}.
\end{aligned}$$

When $1 \le i \le \sqrt{k}$, we have

$$|c_i| = \frac{k}{i^3}\left(1 + \frac{i^2}{k^2}\right)^{k/2} \frac{1}{4 - 1/i^2} \frac{\left(\frac{k-1}{2}\right)!^2}{\left(\frac{k-1}{2}+i\right)!\left(\frac{k-1}{2}-i\right)!} \le \frac{k}{i^3}\left(1 + \frac{1}{k}\right)^{k/2} \le \frac{\sqrt{e}k}{i^3}.$$

For $\sqrt{k} \le i \le \frac{k-1}{2}$, consider the ratio between $(i+1)^3|c_{i+1}|$ and $i^3|c_i|$; it satisfies

$$\begin{aligned}
\frac{(i+1)^3|c_{i+1}|}{i^3|c_i|} &\le \frac{(k^2 + (i+1)^2)^{k/2}}{(k^2+i^2)^{k/2}} \cdot \frac{\frac{k-1}{2} - i}{\frac{k-1}{2} + i + 1} \\
&= \left(1 + \frac{2i+1}{k^2+i^2}\right)^{k/2} \cdot \frac{k-1-2i}{k+1+2i} \\
&\le \left(1 + \frac{2i+1}{k^2}\right)^{k/2} \cdot \frac{k-1-2i}{k} \\
&\le e^{\frac{2i+1}{2k}}\left(1 - \frac{2i+1}{k}\right) \le 1.
\end{aligned}$$

The last inequality holds since $e^{x/2}(1-x) \le 1$ for all $0 \le x \le 1$. Thus we have $(i+1)^3|c_{i+1}| \le i^3|c_i|$, and hence by induction that

$$|c_i| \le \frac{\sqrt{e}k}{i^3} \quad \forall\, 1 \le i \le \tfrac{k-1}{2}. \tag{4.8}$$

We now need to bound $c_{1/2}$. Similarly to the above, we have

$$\begin{aligned}\left|\prod_{j\in I, j\neq\frac{1}{2}}(\Delta_{1/2}-\Delta_j)\right| &= (\Delta_{\frac{1}{2}}-\Delta_{-1/2})\cdot\prod_{j=1}^{(k-1)/2}(\Delta_{1/2}-\Delta_j)\cdot\prod_{j=-(k-1)/2}^{-1}(\Delta_{\frac{1}{2}}-\Delta_j)\\ &= 4k\cdot\prod_{j=1}^{(k-1)/2}\left(2k-\frac{k}{j}\right)\cdot\prod_{j=1}^{(k-1)/2}(2k+\frac{k}{j})\\ &= 4k^k\cdot\prod_{j=1}^{(k-1)/2}\frac{2j-1}{j}\cdot\prod_{j=1}^{(k-1)/2}\frac{2j+1}{j}\\ &= 4k^k\frac{(k-2)!!k!!}{\left(\frac{k-1}{2}\right)!^2}\end{aligned}$$

Thus from (4.7) we get

$$\begin{aligned}|c_{1/2}| &= \frac{(\sqrt{k^2+(1/2)^2})^k}{(1/2)^k}\cdot 2k\cdot\frac{1}{4k^k}\cdot\frac{\left(\frac{k-1}{2}\right)!^2}{(k-2)!!k!!}\\ &= \left(1+\frac{1}{4k^2}\right)^{k/2}\left(\frac{(k-1)!!}{(k-2)!!}\right)^2\leq 4k. \qquad (4.9)\end{aligned}$$

Now combining (4.8), (4.9), we obtain

$$\sum_i|c_i| = 2\sum_{i=1}^{(k-1)/2}|c_i|+2|c_{1/2}|\leq 2\sqrt{e}\sum_{i=1}^{(k-1)/2}\frac{k}{i^3}+8k\leq 20k,$$

as needed.

Handling even k. If k is even, we define our index set to be

$$I=\left\{0,\pm1,\pm2,\ldots,\pm\tfrac{k-2}{2},\pm\tfrac{1}{2}\right\}.$$

For $i\in I\setminus\{0\}$ we define α_i and β_i as before; we also define $\alpha_0=1$, $\beta_0=0$, and hence $\Delta_0=0$. It is easy to check that $c_0=0$ (and hence we haven't actually violated $|\beta_i|\geq 1/O(C_k)$), and the upper bounds for the other $|c_i|$ still hold. This completes the proof of the homogeneous case under Hypothesis **H1**.

Hypotheses H3. We explain the case of k odd; the same trick as before can be used for even k. For Hypothesis **H3** we use

$$\alpha_i=\frac{i}{k^{3/2}+|i|},\quad \beta_i=\frac{k^{3/2}}{k^{3/2}+|i|}\quad\Longrightarrow\quad \Delta_i=\frac{k^{3/2}}{i},$$

which satisfy $|\alpha_i|+|\beta_i|=1$ and $|\alpha_i|,|\beta_i|\geq 1/O(k^{3/2})$. Analysis similar to before shows that $\sum_i|c_i|\leq O(k^{3/2})$. This completely finishes the proof under Hypothesis **H3**.

Hypothesis H2, the homogeneous case. Here we do something slightly different. For even or odd k we let the index set be $I = \{1, 2, \ldots, k, \frac{1}{2}\}$ and then define

$$\alpha_i = \frac{i^2}{k^2 + i^2}, \quad \beta_i = \frac{k^2}{k^2 + i^2} \quad \Longrightarrow \quad \Delta_i = \frac{k^2}{i^2}.$$

Now we have $|\alpha_i| + |\beta_i| = \alpha_i + \beta_i = 1$ and $|\alpha_i|, |\beta_i| \geq 1/O(k^2)$. Again, similar analysis shows that $\sum_i |c_i| \leq O(k^2)$.

Extending to the non-homogeneous case under H2. Now we need to be concerned with the terms at degree $k' < k$. Here a key observation is that, since $\alpha_i + \beta_i = 1$ for all i, the following holds for all $k' < k$:

$$\sum_i c_i \alpha_i^{k'-t} \beta_i^t = \sum_i c_i \alpha_i^{k'-t} \beta_i^t (\alpha_i + \beta_i) = \sum_i c_i \alpha_i^{k'-t+1} \beta_i^t + \sum_i c_i \alpha_i^{k'-t} \beta_i^{t+1}.$$

Thus an induction shows that in fact

$$\sum_i c_i \alpha_i^{k'-t} \beta_i^t = \begin{cases} k - k' & \text{if } t = k' \\ 1 & \text{if } t = k' - 1 \\ 0 & \text{otherwise} \end{cases}$$

for all $k' \leq k$. This is almost exactly what we need to treat the non-homogeneous case using all the same choices for c, α, β, except for the $t = k'$ case. But we can use a simple trick to fix this:

$$\frac{1}{2} \sum_i c_i \alpha_i^{k'-t} \beta_i^t - \frac{1}{2} \sum_i c_i (-\alpha_i)^{k'-t} \beta_i^t = \frac{1 - (-1)^{k'-t}}{2} \sum_i c_i \alpha_i^{k'-t} \beta_i^t = \begin{cases} 1 & \text{if } t = k' - 1 \\ 0 & \text{otherwise} \end{cases}$$

From this we get

$$\breve{f}(y, z) = \sum_{i=1}^{m} c_i f(\alpha_i y + \beta_i z)$$

even in the non-homogeneous case, with all the desired conditions and $m = 2(k + 1)$.

Extending to the non-homogeneous case under H1. The trick here for handling degree $k' < k$ is similar. Using the fact that $\alpha_i^2 + \beta_i^2 = 1$ for all i, we get that for all $k' < k$,

$$\sum_i c_i \alpha_i^{k'-t} \beta_i^t = \sum_i c_i \alpha_i^{k'-t} \beta_i^t (\alpha_i^2 + \beta_i^2) = \sum_i c_i \alpha_i^{k'-t+2} \beta_i^t + \sum_i c_i \alpha_i^{k'-t} \beta_i^{t+2}.$$

Then by induction, the we conclude that

$$\sum_{i=1}^{k+1} c_i \alpha_i^{k'-t} \beta_i^t = \begin{cases} 1 & \text{if } t = k' - 1 \\ 0 & \text{otherwise} \end{cases}$$

holds for all $0 \leq k' \leq k$ such that $k - k'$ is even. We are therefore almost done: we have established the **H1** case of Lemma 4.4.1 for all polynomials with only odd-degree terms or only

even-degree terms. Finally, for a general polynomial f we can decompose it as $f = f_{\text{odd}} + f_{\text{even}}$, where f_{odd} (respectively, f_{even}) contains all the terms in f with odd (respectively, even) degree. We know that there exist some vectors α, β, c and α', β', c' satisfying

$$\breve{f}_{\text{odd}}(y,z) = \sum_{i=1}^{k+1} c_i f_{\text{odd}}(\alpha_i y + \beta_i z), \quad \breve{f}_{\text{even}}(y,z) = \sum_{i=1}^{k+1} c'_i f_{\text{even}}(\alpha'_i y + \beta'_i z),$$

and $\sum_i |c_i|, \sum_i |c'_i| \le 20k$. Thus

$$\begin{aligned}
\breve{f}(y,z) &= \breve{f}_{\text{odd}}(y,z) + \breve{f}_{\text{even}}(y,z) \\
&= \sum_{i=1}^{k+1} c_i f_{\text{odd}}(\alpha_i y + \beta_i z) + \sum_{i=1}^{k+1} c'_i f_{\text{even}}(\alpha'_i y + \beta'_i z) \\
&= \sum_{i=1}^{k+1} \frac{1}{2} c_i (f(\alpha_i y + \beta_i z) - f(-\alpha_i y - \beta_i z)) + \sum_{i=1}^{k+1} \frac{1}{2} c'_i (f(\alpha'_i y + \beta'_i z) + f(-\alpha'_i y - \beta'_i z)) \\
&= \sum_{i=1}^{4(k+1)} c''_i f(\alpha''_i y + \beta''_i z),
\end{aligned}$$

where $c'' = (c/2, -c/2, c'/2, c'/2), \alpha'' = (\alpha, -\alpha, \alpha', -\alpha'), \beta'' = (\beta, -\beta, \beta', -\beta')$ and $\sum_i |c''_i| \le 40k$. □

Chapter 5

On Closeness to k-wise Uniformity

5.1 Introduction

5.1.1 k-wise uniformity and almost k-wise uniformity

We say that a probability distribution over $\{-1,1\}^n$ is *k-wise uniform* if its marginal distribution on every subset of k coordinates is the uniform distribution. For Fourier analysis of the Hamming cube, it is convenient to identify the distribution with its density function $\varphi : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ satisfying

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}[\varphi(\boldsymbol{x})] = 1.$$

We write $\boldsymbol{x} \sim \varphi$ to denote that $\boldsymbol{x}$ is a random variable drawn from the associated distribution with density φ:

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[\boldsymbol{x} = x] = \frac{\varphi(x)}{2^n}$$

for any $x \in \{-1,1\}^n$. Then a well-known fact is that a distribution is k-wise uniform if and only if the Fourier coefficient of φ is 0 on every subset $S \subseteq [n]$ of size between 1 and k:

$$\widehat{\varphi}(S) = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\varphi}\left[\prod_{i\in S} \boldsymbol{x}_i\right] = 0.$$

k-wise uniformity is an essential tool in theoretical computer science. Its study dates back to work of Rao [Rao47]. He studied k-wise uniform sets, which are special cases of k-wise uniform distribution. A subset of $\{-1,1\}^n$ is a *k-wise uniform set* if the uniform distribution on this subset is k-wise uniform. Rao gave constructions of a pairwise-uniform set of size $n+1$ (when $n = 2^r - 1$ for any integer r), a 3-wise uniform set of size $2n$ (when $n = 2^r$ for any integer r), and a lower bound (reproved in [ABI86, CGH$^+$85]) that a k-wise uniform set on $\{-1,1\}^n$ requires size at least $\Omega(n^{\lfloor k/2 \rfloor})$. An alternative proof of the lower bound for even k is shown in [AGM03] using a hypercontractivity-type technique, as opposed to the linear algebra method. Coding theorists have also heavily studied k-wise uniformity, since MacWilliams and Sloane showed that linear codes with dual minimum distance $k+1$ correspond to k-wise uniform sets in [MS77]. The importance in theoretical computer science of k-wise independence for derandomization arose

simultaneously in many papers, with [KW85, Lub86] emphasizing derandomization via the most common pairwise-uniformity case, and [ABI86, CGH$^+$85] emphasizing derandomization based on k-wise independence more generally.

A distribution is "almost k-wise uniform" if its marginal distribution on every k coordinates is very close to the uniform distribution. Typically we say two distributions φ, ψ are *δ-close*, if the total variation distance between φ and ψ is at most δ; and we say they are *δ-far*, if the total variation distance between them is more than δ. However the precise notion of "close to uniform" has varied in previous work. Suppose ψ is the density function for the marginal distribution of φ restricted to some specific k coordinates and $\mathbf{1}$ is the density function for the uniform distribution. Several standard ways are introduced in [AGM03, AAK$^+$07] to quantify closeness to uniformity, corresponding to the L_1, L_2, L_∞ norms:

- (L_1 norm): $\|\psi - \mathbf{1}\|_1 = 2d_{\mathrm{TV}}(\psi, \mathbf{1}) \leq \epsilon$, where d_{TV} denotes total variation distance;
- (L_2 norm): $\|\psi - \mathbf{1}\|_2 = \sqrt{\chi^2(\psi, \mathbf{1})} = \sqrt{\sum_{S \neq \emptyset} \widehat{\psi}(S)^2} \leq \epsilon$, where $\chi^2(\psi, \mathbf{1})$ denotes the χ^2-divergence of ψ from the uniform distribution;
- (L_∞ norm): $\|\psi - \mathbf{1}\|_\infty \leq \epsilon$, or in other words, for any $x \in \{-1, 1\}^n$,

$$\left| \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \psi}[\boldsymbol{x} = x] - 2^{-k} \right| \leq 2^{-k}\epsilon.$$

Note the following: First, closeness in L_1 norm is the most natural for algorithmic derandomization purposes: it tells us that the algorithm cannot tell ψ is different from the uniform distribution up to ϵ error. Second, these definitions of closeness are in increasing order of strength. On the other hand, we also have that $\|\psi - \mathbf{1}\|_1 \leq \|\psi - \mathbf{1}\|_\infty \leq 2^k \|\psi - \mathbf{1}\|_1$; thus all of these notions are within a factor of 2^k. We generally consider k to be constant (or at worst, $O(\log n)$), so that these notions are roughly the same.

A fourth reasonable notion, proposed by Naor and Naor in [NN93], is that the distribution has a small bias over every non-empty subset of at most k coordinates. We say density function φ is (ϵ, k)-*wise uniform* if for every non-empty set $S \subseteq [n]$ with size at most k,

$$|\widehat{\varphi}(S)| = \left| \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \varphi}\left[\prod_{i \in S} \boldsymbol{x}_i = 1\right] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \varphi}\left[\prod_{i \in S} \boldsymbol{x}_i = -1\right] \right| \leq \epsilon.$$

Here we also have $\epsilon = 0$ if and only if φ is exactly k-wise uniform. Clearly if the marginal density of φ over every k coordinates is ϵ-close to the uniform distribution in total variation distance, then φ is (ϵ, k)-wise uniform. On the other hand, if φ is (ϵ, k)-wise uniform, then the marginal density of φ over every k coordinates is $2^{k/2}\epsilon$-close to uniform distribution in total variation distance. Again, if k is considered constant, this bias notion is also roughly the same as previous notions. In the rest of this chapter we prefer this (ϵ, k)-wise uniform notion for "almost k-wise uniform" because of its convenience for Fourier analysis.

The original paper about almost k-wise uniformity, [NN93], was concerned with derandomization; e.g., they use (ϵ, k)-wise uniformity for derandomizing the "set balancing (discrepancy)" problem. Alon et al. gave a further discussion of the relationship between almost k-wise uniformity and derandomization in [AGM03]. The key idea is the following: In many cases of

randomized algorithms, the analysis only relies on the property that the random bits are k-wise uniform, as opposed to fully uniform. Since there exists an efficiently samplable k-wise uniform distribution on a set of size at most $O(n^{\lfloor k/2 \rfloor})$, one can reduce the number of random unbiased bits used in the algorithm down to $O(k \log n)$. To further reduce the number of random bits used, a natural line of thinking is to consider distributions which are "almost k-wise uniformity". Alon et al. [AGHP92] showed that we can deterministically construct (ϵ, k)-wise uniform sets that are of size poly$(2^k, \log n, 1/\epsilon)$, much smaller than exact k-wise uniform ones (roughly $\Omega(n^{\lfloor k/2 \rfloor})$ size). Therefore we can use substantially fewer random bits by taking random strings from an almost k-wise uniform distribution.

However we need to ensure that the original analysis of the randomized algorithm still holds under the almost k-wise uniform distribution. This is to say that if the randomized algorithm behaves well on a k-wise uniform distribution, it may or may not also work as well with an (ϵ, k)-wise uniform distribution, when the parameter ϵ is small enough.

5.1.2 The Closeness Problem

For the analysis of derandomization, it would be very convenient if (ϵ, k)-wise uniformity – which means that "every k-local view looks close to uniform" – implies global δ-closeness to k-wise uniformity. A natural question that arises, posed in [AGM03], is the following:

How small can δ be such that the following is true? For every (ϵ, k)-wise uniform distribution φ on $\{-1, 1\}^n$, φ is δ-close to some k-wise uniform distribution.

In this chapter, we will refer to this question as *the Closeness Problem*.

Previous work and applications

On one hand, the main message of [AGM03] is a lower bound: For every even constant $k > 4$, they gave an (ϵ, k)-wise uniform distribution with $\epsilon = O(1/n^{k/4-1})$, yet which is $\frac{1}{2}$-far from every k-wise uniform distribution in total variation distance.

On the other hand, [AGM03] proved a very simple theorem that $\delta \leq O(n^k \epsilon)$ always holds. Despite its simplicity, this upper bound has been used many times in well known results.

One application is in circuit complexity. [AGM03]'s upper bound is used for fooling disjunctive normal formulas (DNF) [Baz09] and AC^0 [Bra10]. In these works, once the authors showed that k-wise uniformity suffices to fool DNF/AC^0, they deduced that $(O(1/n^k), k)$-uniform distributions suffice, and hence $O(1/n^k)$-biased sets sufficed trivially. [AGM03]'s upper bound is also used as a tool for the construction of two-source extractors for a similar reason in [CZ16, Li16].

Another application is for hardness of constraint satisfactory problems (CSPs). Austrin and Mossel [AM09] show that one can obtain integrality gaps and UGC-hardness for CSPs based on k-wise uniform distributions of small support size. If a predicate is k-wise uniform, Kothari et al. [KMOW17] showed that one can get SOS-hardness of refuting random instances of it when there are around $n^{(k+1)/2}$ constraints. Indeed, [KMOW17] shows that if we have a predicate that is δ-close to k-wise uniform, then with roughly $n^{(k+1)/2}$ random constraints, SOS cannot refute that a $(1 - O(\delta))$-fraction of constraints are satisfiable. This also motivates studying δ-closeness

to k-wise uniformity and how it relates to Fourier coefficients. δ-closeness to k-wise uniformity is also relevant for hardness of random CSP, as shown in [AOW15].

Alon et al. [AAK$^+$07] investigated the Closeness Problem further by improving the upper bound to $\delta = O((n \log n)^{k/2}\epsilon)$. Indeed, they showed a strictly stronger fact that a distribution is $O\Big(\sqrt{\mathbf{W}^{1\dots k}[\varphi]}\log^{k/2} n\Big)$-close to some k-wise uniform, where $\mathbf{W}^{1\dots k}[\varphi] = \sum_{1\le|S|\le k}\widehat{\varphi}(S)^2$. Rubinfeld and Xie [RX13] generalized some of these results to non-uniform k-wise independent distributions over larger product spaces.

Let us briefly summarize the method [AAK$^+$07] used to prove their upper bounds. Given an (ϵ, k)-wise uniform φ, they first try to generate a k-wise uniform "pseudo-distribution" φ' by forcing all Fourier coefficients at degree at most k to be zero. It is a "pseudo-distribution" because some points might have negative density. After this, they use a fully uniform distribution and k-wise uniform distributions with small support size to try to mend all points to be nonnegative. They bound the weight of these mending distributions to upper-bound the distance incurred by the mending process. This mending process uses the fully uniform distribution to mend the small negative weights and uses k-wise uniform distributions with small support size to correct the large negative weights point by point. By optimizing the threshold between small and large weights it introduces a factor of $(\log n)^{k/2}$.

Though they did not mention it explicitly, they also give a lower bound for the Closeness Problem of $\delta \ge \Omega\Big(\frac{n^{(k-1)/2}}{\log n}\epsilon\Big)$ for $k > 2$ by considering the uniform distribution on a set of $O(n^k)$ random chosen strings. No previous work gave any lower bound for the most natural case of $k = 2$.

Our result

In this chapter, we show sharper upper and lower bounds for the Closeness Problem, which are tight for k even and $k = 1$. Comparing to the result in [AAK$^+$07], we get rid of the factor of $(\log n)^{k/2}$.

Theorem 5.1.1. *Any density φ over $\{-1, 1\}^n$ is δ-close to some k-wise uniform distribution, where*

$$\delta \le e^k\sqrt{\mathbf{W}^{1\dots k}[\varphi]} = e^k\sqrt{\sum_{1\le|S|\le k}\widehat{\varphi}(S)^2}.$$

Consequently, if φ is (ϵ, k)-wise uniform, i.e., $|\widehat{\varphi}(S)| \le \epsilon$ for every non-empty set S with size at most k, then

$$\delta \le e^k n^{k/2}\epsilon.$$

For the special case $k = 1$, the corresponding δ can be further improved to $\delta \le \epsilon$.

Our new technique is trying to mend the original distribution to be k-wise uniform all at once. We want to show that some mixture distribution $(\varphi + w\psi)$ is k-wise uniform with small mixture weight w. The distance between the final mixture distribution and the original distribution φ is bounded by $O(w)$. Therefore we only need to show that the mending distribution ψ exists for some small weight w. Showing the existence of such a distribution ψ can be written as the feasibility of a linear program (LP). We upper bound w by bounding the dual LP, using the hypercontractivity inequality.

Our result is sharp for all even k, and is also sharp for $k = 1$. We state the matching lower bound for even k:

Theorem 5.1.2. *For any n and even k, and small enough ϵ, there exists some (ϵ, k)-wise uniform distribution φ over $\{-1, 1\}^n$, such that φ is δ-far from every k-wise uniform distribution in total variation distance, where*

$$\delta \geq \Omega\left(\frac{1}{k}\right)^k n^{k/2}\epsilon.$$

Our method for proving this lower bound is again LP duality. Our examples in the lower bound are symmetric distributions with Fourier weight only on level k. The density functions then can be written as binary Krawtchouk polynomials which behave similar to Hermite polynomials when n is large. Our dual LP bounds use various properties of Krawtchouk and Hermite polynomials.

Interestingly both our upper and lower bound utilize LP-duality, which we believe is the most natural way of looking at this problem.

We remark that we can derive a lower bound for odd k from Theorem 5.1.2 trivially by replacing k by $k - 1$. There exists a gap of $\sqrt{n}$ between the resulting upper and lower bounds for odd k. We believe that the lower bound is tight, and the upper bound may be improvable by a factor of $\sqrt{n}$, as it is in the special case $k = 1$. We leave it as a conjecture for further work:

Conjecture 5.1.3. *Suppose the distribution φ over $\{-1, 1\}^n$ is (ϵ, k)-wise uniform. Then φ is δ-close to some k-wise uniform distribution in total variation distance, where*

$$\delta \leq O(n^{\lfloor k/2 \rfloor}\epsilon).$$

5.1.3 The Testing Problem

Another application of the Closeness Problem is to property testing of k-wise uniformity. Suppose we have sample access from an unknown and arbitrary distribution; we may wonder whether the distribution has a certain property. This question has received tremendous attention in the field of statistics. The main goal in the study of property testing is to design algorithms that use as few samples as possible, and to establish lower bound matching these sample-efficient algorithms. In particular, we consider the property of being k-wise uniform:

Given sample access to an unknown and arbitrary distribution φ on $\{-1, 1\}^n$, how many samples do we need to distinguish between the case that φ is k-wise uniform versus the case that φ is δ-far from every k-wise uniform distribution?

In this chapter, we will refer to this question as the *Testing Problem*.

We say a testing algorithm is *a δ-tester for k-wise uniformity* if the algorithm outputs "Yes" with high probability when the distribution φ is k-wise uniform, and the algorithm outputs "No" with high probability when the distribution φ is δ-far from any k-wise uniform distribution (in total variation distance).

Property testing is well studied for Boolean functions and distributions. Previous work studied the testing of related properties of distribution, including uniformity [GR11, BFR$^+$00, RS09] and independence [BFF$^+$01, BKR04, ADK15, DK16].

The papers [AGM03, AAK$^+$07, Xie12] discussed the problem of testing k-wise uniformity. [AGM03] constructed a δ-tester for k-wise uniformity with sample complexity $O(n^{2k}/\delta^2)$, and [AAK$^+$07] improved it to $O(n^k \log^{k+1} n/\delta^2)$. As for lower bounds, [AAK$^+$07] showed that $\Omega(n^{(k-1)/2}/\delta)$ samples are necessary, albeit only for $k > 2$. This lower bound is in particular for distinguishing the uniform distribution from δ-far-from-k-wise distributions.

We show a better upper bound for sample complexity:

Theorem 5.1.4. *There exists a δ-tester for k-wise uniformity of distributions on $\{-1,1\}^n$ with sample complexity $O\big(\frac{1}{k}\big)^{k/2}\frac{n^k}{\delta^2}$. For the special case of $k = 1$, the sample complexity is $O\big(\frac{\log n}{\delta^2}\big)$.*

A natural δ-tester of k-wise uniformity is mentioned in [AAK$^+$07]: Estimate all Fourier coefficients up to level k from the samples; if they are all smaller than ϵ then output "Yes". In fact this algorithm is exactly attempting to check whether the distribution is (ϵ, k)-wise uniform. Hence the sample complexity depends on the upper bound for the Closeness Problem. Therefore we can reduce the sample complexity of this algorithm down to $O\Big(\frac{n^k \log n}{\delta^2}\Big)$ via our improved upper bound for the Closeness Problem. One $\log n$ factor remains because we need to union-bound over the $O(n^k)$ Fourier coefficients up to level k. To further get rid of the last $\log n$ factor, we present a new algorithm that estimates the Fourier weight up to level k, $\sum_{1\le|S|\le k}\widehat{\varphi}^2(S)$, rather than estimating these Fourier coefficients one by one.

Unfortunately, a lower bound for the Closeness Problem does not imply a lower bound for the Testing Problem directly. In [AAK$^+$07], they showed that a uniform distribution over a random subset of $\{-1,1\}^n$ of size $O(\frac{n^{k-1}}{\delta^2})$, is almost surely δ-far from any k-wise uniform distribution. On the other hand, by the Birthday Paradox, it is hard to distinguish between the fully uniform distribution on all strings of length n and a uniform distribution over a random set of such size. This gives a lower bound for the Testing Problem as $\Omega(n^{(k-1)/2}/\delta)$. Their result only holds for $k > 2$; there was no previous non-trivial lower bound for testing pairwise uniformity. We show a lower bound for the pairwise case.

Theorem 5.1.5. *Any δ-tester for pairwise uniformity of distributions on $\{-1,1\}^n$ needs at least $\Omega(\frac{n}{\delta^2})$ samples.*

For this lower bound we analyze a symmetric distribution with non-zero Fourier coefficients only on level 2. We prove that it is hard to distinguish a randomly shifted version of this distribution from the fully uniform distribution. This lower bound is also better than [AAK$^+$07] in that we have a better dependence on the parameter δ ($\frac{1}{\delta^2}$ rather than $\frac{1}{\delta}$). Unfortunately we are unable to generalize our lower bound for higher k.

Notice that for our new upper and lower bounds for k-wise uniformity testing, there still remains a quadratic gap for $k \ge 2$, indicating that the upper bound might be able to be improved. Both the lower bound in this chapter and that in [AAK$^+$07] show that it is hard to distinguish between the fully uniform distribution and some specific sets of distributions that are far from k-wise uniform. We show that if one wants to improve the lower bound, one will need to use a distribution in the "Yes" case that is *not* fully uniform, because we give a sample-efficient algorithm for distinguishing between fully uniform and δ-far from k-wise uniform:

Theorem 5.1.6. *For any constant k, for testing whether a distribution is fully uniform or δ-far from every k-wise uniform distribution, there exists an algorithm with sample complexity $O(k)^k \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \big(\log\frac{n}{\delta}\big)^{k/2}$.*

In fact, for testing whether a distribution is αk-wise uniform or δ-far from k-wise uniform with $\alpha > 4$, there exists an algorithm with sample complexity $O(\alpha)^{k/2} \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\frac{n^k}{\delta^4}\right)^{1/(\alpha-2)}$.

We remark that testing full uniformity can be treated as a special case of testing αk-wise uniformity approximately, by setting $\alpha = \log \frac{n}{\delta}$.

Testing full uniformity has been studied in [GR11, BFR+00]. Paninski [Pan08] showed that testing whether an unknown distribution on $\{-1, 1\}^n$ is $\Theta(1)$-close to fully uniform requires $2^{n/2}$ samples. Rubinfeld and Servedio [RS09] studied testing whether an unknown monotone distribution is fully uniform or not.

The fully uniform distribution has the nice property that every pair of samples is different in $\frac{n}{2} \pm O(\sqrt{n})$ bits with high probability when the sample size is small. Our algorithm first rejects those distributions that disobey this property. We show that the remaining distributions have small Fourier weight up to level $2k$. Hence by following a similar analysis as the tester in Theorem 5.1.4, we can get an improved upper bound when these lower Fourier weights are small.

The lower bound remains the same as testing k-wise vs. far from k-wise. Our tester is tight up to a logarithmic factor for the pairwise case, and is tight up to a factor of $\tilde{O}(\sqrt{n})$ when $k > 2$.

We compare our results and previous best known bounds from [AAK+07] in Table 5.1. (We omit constant factors depending on k.)

	Upper bound		Lower bound	
	[AAK+07]	Our results	[AAK+07]	Our results
Closeness Problem	$O(n^{k/2}(\log n)^{k/2}\epsilon)$	$O(n^{k/2}\epsilon)$ $O(\epsilon)$ for $k = 1$	$\Omega\left(\frac{n^{(k-1)/2}}{\log n}\epsilon\right)$	$\Omega(n^{\lfloor k/2 \rfloor}\epsilon)$
Testing k-wise vs. far from k-wise	$O\left(\frac{n^k(\log n)^{k+1}}{\delta^2}\right)$	$O\left(\frac{n^k}{\delta^2}\right)$ $O\left(\frac{\log n}{\delta^2}\right)$ for $k = 1$	$\Omega\left(\frac{n^{(k-1)/2}}{\delta}\right)$ for $k > 2$	$\Omega\left(\frac{n}{\delta^2}\right)$ for $k = 2$
Testing n-wise vs. far from k-wise	$O\left(\frac{n^k(\log n)^{k+1}}{\delta^2}\right)$	$O\left(\frac{n^{k/2}}{\delta^2}(\log \frac{n}{\delta})^{k/2}\right)$ $O\left(\frac{\log n}{\delta^2}\right)$ for $k = 1$	$\Omega\left(\frac{n^{(k-1)/2}}{\delta}\right)$ for $k > 2$	$\Omega\left(\frac{n}{\delta^2}\right)$ for $k = 2$

Table 5.1: Comparison of our results to [AAK+07]

5.1.4 Organization

Section 5.2 contains definitions and notations. We will discuss upper and lower bounds for the Closeness Problem in Section 5.3. We will discuss the sample complexity of testing k-wise uniformity in Section 5.4. We present a tester for distinguishing between αk-wise uniformity (or fully uniformity) and far-from k-wise uniformity in Section 5.5.

5.2 Preliminaries

5.2.1 Fourier analysis of Boolean functions

We use $[n]$ to denote the set $\{1, \ldots, n\}$. We denote the symmetric difference of two sets S and T by $S \oplus T$. For Fourier analysis we use notations consistent with [O'D14]. Every function $f : \{-1,1\}^n \to \mathbb{R}$ has a unique representation as a multilinear polynomial

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x^S \quad \text{where} \quad x^S = \prod_{i \in S} x_i.$$

We call $\widehat{f}(S)$ the Fourier coefficient of f on S. We use $\boldsymbol{x} \sim \{-1,1\}^n$ to denote that $\boldsymbol{x}$ is uniformly distributed on $\{-1,1\}^n$. We can represent Fourier coefficients as

$$\widehat{f}(S) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n} \big[f(\boldsymbol{x}) \boldsymbol{x}^S \big].$$

We define an inner product $\langle \cdot, \cdot \rangle$ on pairs of functions $f, g : \{-1,1\}^n \to \mathbb{R}$ by

$$\langle f, g \rangle = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{-1,1\}^n} [f(\boldsymbol{x}) g(\boldsymbol{x})] = \sum_{S \subseteq [n]} \widehat{f}(S) \widehat{g}(S).$$

We introduce the following p-norm notation: $\|f\|_p = (\mathbf{E}[|f(\boldsymbol{x})|^p])^{1/p}$, and the Fourier ℓ_p-norm is $\|\widehat{f}\|_p = \left(\sum_{S \subseteq [n]} |\widehat{f}(S)|^p \right)^{1/p}$.

We say the *degree* of a Boolean function, $\deg(f)$ is k if its Fourier polynomial is degree k. We denote $f^{=k}(x) = \sum_{|S|=k} \widehat{f}(S) x^S$, and $f^{\leq k}(x) = \sum_{|S| \leq k} \widehat{f}(S) x^S$. We denote the *Fourier weight* on level k by $\mathbf{W}^k[f] = \sum_{|S|=k} \widehat{f}(S)^2$. We denote $\mathbf{W}^{1 \ldots k}[\varphi] = \sum_{1 \leq |S| \leq k} \widehat{\varphi}(S)^2$.

We define the convolution $f * g$ of a pair of functions $f, g : \{-1,1\}^n \to \mathbb{R}$ to be

$$(f * g)(x) = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim \{-1,1\}^n} [f(x) g(x \circ \boldsymbol{y})],$$

where $\circ$ denotes entry-wise multiplication. The effect of convolution on Fourier coefficients is that $\widehat{f * g}(S) = \widehat{f}(S) \widehat{g}(S)$.

5.2.2 Densities and distances

When working with probability distribution on $\{-1,1\}^n$, we prefer to define them via *density* function. A density function $\varphi : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ is a nonnegative function satisfying $\widehat{\varphi}(\emptyset) = \mathbf{E}_{\boldsymbol{x} \sim \{-1,1\}^n}[\varphi(\boldsymbol{x})] = 1$. We write $\boldsymbol{y} \sim \varphi$ to denote that $\boldsymbol{y}$ is a random variable drawn from the distribution φ, defined by

$$\mathop{\mathbf{Pr}}_{\boldsymbol{y} \sim \varphi} [\boldsymbol{y} = y] = \frac{\varphi(y)}{2^n},$$

for all $y \in \{-1,1\}^n$. We identify distributions with their density functions when there is no risk of confusion.

We denote $\varphi^{+t}(x) = \varphi(x \circ t)$. We denote by $\mathbf{1}_A$ the density function for the uniform distribution on support set A. The density function associated to the fully uniform distribution is the constant function $\mathbf{1}$.

The following lemma about density functions of degree at most k derives from Fourier analysis and hypercontractivity.

Lemma 5.2.1. *Let $\varphi : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ be a density function of degree at most k. Then*

$$\|\widehat{\varphi}\|_2 = \sqrt{\sum_S \widehat{\varphi}(S)^2} \leq e^k.$$

Proof.

$$\|\widehat{\varphi}\|_2 = \|\varphi\|_2 \leq e^k \|\varphi\|_1 = e^k.$$

The first equality holds by Parseval's Theorem (see Section 1.4 in [O'D14]). The inequality holds by hypercontractivity (see Theorem 9.22 in [O'D14]). The last equality holds since φ is a density function. □

A distribution φ over $\{-1,1\}^n$ is *k-wise uniform* if and only if $\widehat{\varphi}(S) = 0$ for all $1 \leq |S| \leq k$ (see Chapter 6.1 in [O'D14]). We say that distribution φ over $\{-1,1\}^n$ is *(ϵ, k)-wise uniform* if $|\widehat{\varphi}(S)| \leq \epsilon$ for all $1 \leq S \leq k$.

The most common way to measure the distance between two probability distributions is via their *total variation distance*. If the distributions have densities φ and ψ, then the total variation distance is defined to be

$$d_{\mathrm{TV}}(\varphi, \psi) = \sup_{A \subseteq \{-1,1\}^n} \left| \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \varphi}[\boldsymbol{x} \in A] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \psi}[\boldsymbol{x} \in A] \right| = \frac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{x}}[|\varphi(\boldsymbol{x}) - \psi(\boldsymbol{x})|] = \frac{1}{2}\|\varphi - \psi\|_1.$$

We say that φ and ψ are *δ-close* if $d_{\mathrm{TV}}(\varphi, \psi) \leq \delta$.

Supposing H is a set of distributions, we denote

$$d_{\mathrm{TV}}(\varphi, H) = \min_{\psi \in H} d_{\mathrm{TV}}(\varphi, \psi).$$

In particular, we denote the set of k-wise uniform densities by kWISE. We say that density φ is *δ-close to k-wise uniform* if $d_{\mathrm{TV}}(\varphi, \text{kWISE}) \leq \delta$, and is *$\delta$-far* otherwise.

5.2.3 Krawtchouk and Hermite polynomials

Krawtchouk polynomials were introduced in [Kra29], and arise in the analysis of Boolean functions as shown in [Lev95, Kal02]. Consider the following Boolean function of degree k and input length n: $f(x) = \sum_{|S|=k} x^S$. It is symmetric and therefore only depends on the Hamming weight of x. Let t be the number of -1's in x. Then the output of f is exactly the same as the Krawtchouk polynomial $K_k^{(n)}(t)$.

Definition 5.2.2. We denote by $K_k^{(n)}(t)$ the Krawtchouk polynomial:

$$K_k^{(n)}(t) = \sum_{j=0}^{k} (-1)^j \binom{t}{j} \binom{n-t}{k-j},$$

for $k = 0, 1, \ldots, n$.

We will also use Hermite polynomials in our analysis.

Definition 5.2.3. We denote by $h_k(z)$ the normalized Hermite polynomial:

$$h_k(z) = \frac{1}{\sqrt{k!}}(-1)^k e^{\frac{1}{2}z^2} \frac{d^k}{dz^k} e^{-\frac{1}{2}z^2}.$$

Its explicit formula is

$$h_k(z) = \sqrt{k!} \cdot \left(\frac{z^k}{0!! \cdot k!} - \frac{z^{k-2}}{2!! \cdot (k-2)!} + \frac{z^{k-4}}{4!! \cdot (k-4)!} - \frac{z^{k-6}}{6!! \cdot (k-6)!} + \cdots \right).$$

One useful fact is that the derivative of a Hermite polynomial is a scalar multiple of a Hermite polynomial (see Exercise 11.10 in [O'D14]):

Fact 5.2.4. *For any integer $k \geq 1$, we have*

$$\frac{d}{dz} h_k(z) = \sqrt{k} h_{k-1}(z).$$

The relationship between Krawtchouk and Hermite polynomials is that we can treat Hermite polynomials as a limit version of Krawtchouk polynomials when n goes to infinity (see Exercise 11.14 in [O'D14]).

Fact 5.2.5. *For all $k \in \mathbb{N}$ and $z \in \mathbb{R}$ we have*

$$\binom{n}{k}^{-1/2} \cdot K_k^{(n)}\left(\frac{n - z\sqrt{n}}{2}\right) \xrightarrow{n\to\infty} h_k(z).$$

Instead of analyzing Krawtchouk polynomials, it is easier to study Hermite polynomials when n is large because Hermite polynomials have a more explicit form. We present some basic properties of Hermite polynomials with brief proofs.

Lemma 5.2.6. *The following are properties of $h_k(z)$:*

1. *$|h_k(z)| \leq h_k(k)$ for any $|z| \leq k$;*
2. *$h_k(z)$ is positive and increasing when $z \geq k$;*
3. *$h_k(Ck) \leq (Ck)^k/\sqrt{k!}$ for any constant $C \geq 1$.*

Proof. We will treat the case of $k = 4i + 2$ for some integer i. The proof for the general case is similar. When $k = 4i + 2$, we can group adjacent terms into pairs:

$$h_k(z) = \sqrt{k!} \cdot \sum_{i=0}^{(k-2)/4} \frac{z^{k-4i-2}}{(4i+2)!! \cdot (k-4i)!}((4i+2)z^2 - (k-4i)(k-4i-1)).$$

1. Notice that $|(4i+2)z^2 - (k-4i)(k-4i-1)|$ is always between $-(k-4i)(k-4i-1)$ and $(4i+2)k^2 - (k-4i)(k-4i-1)$ when $|z| \leq k$. Both the upper and lower bound have absolute value at most $(4i+2)k^2 - (k-4i)(k-4i-1)$. Therefore by the triangle inequality we have $|h_k(z)| \leq h_k(k)$.
2. It is easy to check that $((4i+2)z^2 - (k-4i)(k-4i-1))$ is positive when $z \geq k$. Then by Fact 5.2.4, $\frac{d}{dz} h_k(z) = \sqrt{k} h_{k-1}(z) > 0$ when $z \geq k$.
3. This is trivial from the explicit formula since each term is exactly smaller than the previous term when $z \geq k$. □

5.3 The Closeness Problem

In this section, we prove the upper bound in Theorem 5.1.1 and the lower bound in Theorem 5.1.2. One interesting fact is that we use duality of linear programming (LP) in both the upper and lower bound. We think this is the proper perspective for analyzing these questions.

5.3.1 Upper bound

The key idea for proving the upper bound is mixture distributions. Given an (ϵ, k)-wise uniform density φ, we try to mix it with some other distribution ψ using mixture weight w, such that the mixture distribution $\frac{1}{1+w}(\varphi + w\psi)$ is k-wise uniform and is close to the original distribution. The following lemma shows that the distance between the original distribution and the mixture distribution is bounded by the weight w.

Lemma 5.3.1. *If $\varphi' = \frac{1}{1+w}(\varphi + w\psi)$ for some $0 \le w \le 1$ and density functions φ, ψ, then $d_{\mathrm{TV}}(\varphi, \varphi') \le w$.*

Proof. $d_{\mathrm{TV}}(\varphi, \varphi') = \frac{1}{2}\|\varphi' - \varphi\|_1 = \frac{1}{2}\|\varphi' - ((1+w)\varphi' - w\psi))\|_1 = \frac{1}{2}w\|\varphi' - \psi\|_1 \le w.$ □

Therefore we only need to show the existence of an appropriate ψ for some small w. The constraints on ψ can be written as an LP feasibility problem. Therefore by Farkas' Lemma we only need to show that its dual is not feasible. The variables in the dual LP can be seen as a density function of degree at most k.

Proof of Theorem 5.1.1 (general k case). Given density function φ, we try to find another density function ψ with constraints

$$\widehat{\psi}(S) = -\frac{1}{w}\widehat{\varphi}(S)$$

for all $1 \le |S| \le k$. Suppose such a density function ψ exists. Then it is trivial that $\frac{\varphi + w\psi}{1+w}$ is also a density function and is k-wise uniform. By Lemma 5.3.1, we conclude that $d_{\mathrm{TV}}(\varphi, \text{kWISE}) \le w$.

The rest of proof is to show that such a ψ exists when $w = e^k\sqrt{\mathbf{W}^{1\dots k}[\varphi]}$. We can write the existence as an LP feasibility problem with variables $\psi(x)$ for $x \in \{-1,1\}^n$ and constraints:

$$\begin{aligned}
&\widehat{\psi}(\emptyset) = 1, \\
&\widehat{\psi}(S) = -\frac{1}{w}\widehat{\varphi}(S), && \forall 1 \le |S| \le k, \\
&\psi(x) \ge 0, && \forall x \in \{-1,1\}^n,
\end{aligned}$$

where $\widehat{\psi}(S) = \mathbf{E}[\psi(\boldsymbol{x})\boldsymbol{x}^S]$ is a linear combination of variables $\psi(x)$.

The dual LP has variables $\psi'(x)$ for $x \in \{-1,1\}^n$ with constraints:

$$\begin{aligned}
\widehat{\psi'}(\emptyset) &= 1, \\
\widehat{\psi'}(S) &= 0, && \forall |S| > k, \\
\psi'(x) &\ge 0, && \forall x \in \{-1,1\}^n, \\
\frac{1}{w}\sum_{1 \le |S| \le k} \widehat{\varphi}(S)\widehat{\psi'}(S) &> 1.
\end{aligned}$$

The original LP is feasible if and only if its dual LP is infeasible, by Farkas' Lemma. This completes the proof, since when $w = e^k\sqrt{\mathbf{W}^{1\ldots k}[\varphi]}$, for any density function ψ' with degree k we have

$$\frac{1}{w}\sum_{1\le|S|\le k}\widehat{\varphi}(S)\widehat{\psi'}(S) \le \frac{1}{e^k\sqrt{\mathbf{W}^{1\ldots k}[\varphi]}}\sum_{1\le|S|\le k}|\widehat{\varphi}(S)||\widehat{\psi'}(S)| \le \frac{1}{e^k}\|\widehat{\psi'}\|_2 \le 1,$$

where the second inequality holds by Cauchy–Schwarz, and the last inequality holds by Lemma 5.2.1 since ψ' has degree at most k. □

For $k = 1$, further improvement can be achieved. We still try to use mixture distributions. Here we want to mix the distribution φ with indicator distributions on subsets of coordinates that have opposite biases to those of the original distribution.

Proof of Theorem 5.1.1 (case $k = 1$). By identifying each x_i with $-x_i$ if necessary, we may assume without loss of generality that $\widehat{\varphi}(\{i\}) \ge 0$ for all i. In addition, by reordering the coordinates, we may assume without loss of generality that $0 \le \widehat{\varphi}(\{1\}) \le \cdots \le \widehat{\varphi}(\{n\}) = \epsilon$. Define ψ_j to be the density of the distribution over $\{-1,1\}^n$ which is uniform on coordinates $x_1,\ldots,x_{j-1}$, and has x_i constantly fixed to be -1 for $j \le i \le n$. It is easy to check $\widehat{\psi_j}(\{i\}) = 0$ for $i < j$ and $\widehat{\psi_j}(\{i\}) = -1$ for $i \ge j$.

We define φ' as

$$\varphi' = \frac{1}{1+\epsilon}\left(\varphi + \sum_{j=1}^{n} w_j\psi_j\right),$$

where

$$w_1 = \widehat{\varphi}(\{1\}), \qquad w_j = \widehat{\varphi}(\{j\}) - \widehat{\varphi}(\{j-1\}) \quad \forall 1 < j \le n.$$

It is easy to check that φ' is a density function and

$$\widehat{\varphi'}(\{i\}) = \frac{1}{1+\epsilon}\left(\widehat{\varphi}(\{i\}) + \left(\sum_{j=1}^{i} w_j\right)(-1)\right) = 0.$$

Therefore φ' is 1-wise uniform. Then by Lemma 5.3.1,

$$d_{TV}(\varphi, \text{1WISE}) \le \frac{1}{2}\|\varphi - \varphi'\|_1 \le \sum_{j=1}^{n} w_j = \epsilon.$$ □

5.3.2 Lower bound

Interestingly, our proof of the lower bound also utilizes LP duality. We can write the Closeness Problem in the form of linear programming with variables $\varphi'(x)$ for $x \in \{-1,1\}^n$, as follows:

$$\begin{aligned}
\text{minimize} \qquad & d_{\text{TV}}(\varphi,\varphi') = \frac{1}{2}\|\varphi - \varphi'\|_1 & \\
\text{subject to:} \qquad & \widehat{\varphi'}(\emptyset) = 1, & \\
& \widehat{\varphi'}(S) = 0, & \forall 1 \le |S| \le k, \\
& \varphi'(x) \ge 0, & \forall x \in \{-1,1\}^n.
\end{aligned}$$

We ignore the factor of $1/2$ in the minimization for convenience in the following analysis. The dual LP, which has variables $p(x), q(x)$ for $x \in \{-1,1\}^n$, is the following:

$$\begin{array}{lrl} \text{maximize} & \langle \varphi, q\rangle - \widehat{p}(\emptyset) & \\ \text{subject to:} & p(x) - q(x) \geq 0, & \forall x \in \{-1,1\}^n, \\ & q(x) \leq 1, & \forall x \in \{-1,1\}^n, \\ & p(x) \geq -1, & \forall x \in \{-1,1\}^n, \\ & \deg(p) \leq k. & \end{array}$$

Thus given a pair of Boolean functions p, q satisfying the constraints, the quantity $\langle \varphi, q\rangle - \widehat{p}(\emptyset)$ is a lower bound for our Closeness Problem. Our distribution φ achieving the lower bound is a symmetric polynomial, homogeneous of degree k (except that it has a constant term of 1, as is necessary for every density function). We can use Krawtchouk and Hermite polynomials to simplify the analysis.

Proof of Theorem 5.1.2. We define

$$\varphi(x) = 1 + \mu\binom{n}{k}^{-1/2} \sum_{|S|=k} x^S, \quad p(x) = \mu\binom{n}{k}^{-1/2} \sum_{|S|=k} x^S, \quad q(x) = \min(p(x), 1),$$

where μ is a small parameter to be chosen later that will ensure $\varphi(x) \geq 0$ and $p(x) \geq -1$ for all $x \in \{-1,1\}^n$. We have $\epsilon = \max_{1\leq|S|\leq k} |\widehat{\varphi}(S)| = \mu\binom{n}{k}^{-1/2}$.

Since $\widehat{p}(\emptyset) = 0$, the objective function of the dual LP is

$$\begin{aligned} \langle \varphi, q\rangle &= \langle \varphi, \min(p,1)\rangle = \langle \varphi, 1_{p>1}\rangle + \langle \varphi, p1_{p\leq 1}\rangle = \langle \varphi, p\rangle - \langle \varphi, (p-1)1_{p>1}\rangle \\ &\geq \langle \varphi, p\rangle - \sqrt{\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[p(\boldsymbol{x}) > 1] \cdot \langle \varphi, (p-1)^2\rangle}, \end{aligned}$$

where the last inequality holds by Cauchy–Schwarz. It is easy to calculate the inner products $\langle \varphi, p\rangle = \mu^2$, and

$$\begin{aligned} \langle \varphi, (p-1)^2\rangle &= \langle \varphi, p^2\rangle - 2\langle \varphi, p\rangle + 1 \\ &= \mu^2 + \mu^3\binom{n}{k}^{-1/2}\binom{k}{k/2}\binom{n-k}{k/2} - 2\mu^2 + 1 \\ &\leq 1 + \mu^3\binom{k}{k/2}^{3/2} - \mu^2. \end{aligned}$$

Assuming $\mu < 2^{-\frac{3}{2}k}$, we have $\langle \varphi, (p-1)^2\rangle < 1$.

Now we need to upper bound $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}[p(\boldsymbol{x}) > 1]$. Define $\boldsymbol{z}$ satisfying $(n - \boldsymbol{z}\sqrt{n})/2 = \sum_i \boldsymbol{x}_i$. Then

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[p(\boldsymbol{x}) > 1] = \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}\left[\mu\binom{n}{k}^{-1/2} \cdot K_k\left(\frac{n - \boldsymbol{z}\sqrt{n}}{2}, n\right) > 1\right].$$

By Fact 5.2.5, we know that when $z \le k$, for sufficient large n,

$$\binom{n}{k}^{-1/2} \cdot K_k\left(\frac{n - z\sqrt{n}}{2}, n\right) < 2h_k(z).$$

Now we set $\mu = \frac{\sqrt{k!}}{2(Ck)^k}$ with some constant $C \ge 1$. It is easy to check that $\mu < 2^{-\frac{3}{2}k}$. Using the properties in Lemma 5.2.6, we get

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}\left[\mu\binom{n}{k}^{-1/2} \cdot K_k\left(\frac{n - \boldsymbol{z}\sqrt{n}}{2}, n\right) > 1\right] &\le \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[2\mu h_k(\boldsymbol{z}) > 1] \\ &\le \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[h_k(\boldsymbol{z}) > h_k(Ck)] \\ &= \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[|\boldsymbol{z}| > Ck].\end{aligned}$$

Then using Cauchy–Schwarz again, we get

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[|\boldsymbol{z}| > Ck] &\le \sqrt{\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}[\varphi(\boldsymbol{x})^2]}\sqrt{\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\{-1,1\}^n}[|\boldsymbol{z}| > Ck]} \\ &\le \sqrt{1+\mu^2}\sqrt{2\exp(-C^2k^2/2)} \\ &\le 2\exp(-(Ck)^2/4).\end{aligned}$$

Therefore we get that the objective function is at least

$$\langle\varphi, p\rangle - \sqrt{\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[p(\boldsymbol{x}) > 1] \cdot \langle\varphi, (p-1)^2\rangle} \ge \mu^2 - \sqrt{2\exp(-(Ck)^2/4)} \ge \Omega\left(\frac{1}{k}\right)^k.$$

The last inequality holds when we choose a sufficiently large constant C.

This completes the proof, because φ is at least δ-far from k-wise uniform with $\delta = \Omega\big(\frac{1}{k}\big)^k$, and we have $\epsilon = \mu\binom{n}{k}^{-1/2} \le \frac{n^{-k/2}}{2^{\Omega(k)}}$. Therefore we have $\delta \ge \Omega\big(\frac{1}{k}\big)^k n^{k/2}\epsilon$. □

5.4 The Testing Problem

In this section, we study the problem of testing whether a distribution is k-wise uniform or δ-far from k-wise uniform. These bounds are based on new bounds for the Closeness Problem. We present a new testing algorithm for general k in Section 5.4.1. We give a lower bound for the pairwise case in Section 5.4.2.

5.4.1 Upper bound

Given m samples from φ, call them $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$, we will first show that

$$\Delta(\boldsymbol{X}) = \mathop{\mathrm{avg}}_{1\le s<t\le m}\left(\sum_{1\le|S|\le k} \boldsymbol{x}_s^S \boldsymbol{x}_t^S\right)$$

is a natural estimator of $\mathbf{W}^{1\ldots k}[\varphi]$.

Lemma 5.4.1. *It holds that*

$$\begin{aligned}\mu = \mathbf{E}[\Delta(\boldsymbol{X})] &= \mathbf{W}^{1\dots k}[\varphi];\\ \mathbf{Var}[\Delta(\boldsymbol{X})] &\le \frac{4}{m^2}L_k(\varphi) + \frac{4}{m}\sqrt{L_k(\varphi)}\mu,\end{aligned} \tag{5.1}$$

where $L_k(\varphi) = \sum_{1\le|S_1|,|S_2|\le k}\widehat{\varphi}(S_1\oplus S_2)^2$.

Proof. We denote $F(x,y) = \sum_{1\le|S|\le k} x^S y^S$. We know that

$$\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[\boldsymbol{x}^S\boldsymbol{y}^S] = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\varphi}[\boldsymbol{x}^S]\mathop{\mathbf{E}}_{\boldsymbol{y}\sim\varphi}[\boldsymbol{y}^S] = \widehat{\varphi}(S)^2,$$

when $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent samples drawn from φ. Therefore by linearity of expectation, $\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[F(\boldsymbol{x},\boldsymbol{y})] = \mathbf{W}^{1\dots k}[\varphi]$, and clearly by taking the average,

$$\mu = \mathbf{E}[\Delta(\boldsymbol{X})] = \mathbf{E}[\mathrm{avg}_{s<t}F(\boldsymbol{x}_s,\boldsymbol{x}_t)] = \mathrm{avg}_{s<t}\,\mathbf{E}[F(\boldsymbol{x}_s,\boldsymbol{x}_t)] = \mathbf{W}^{1\dots k}[\varphi].$$

We need to expand the variance:

$$\mathbf{Var}\left[\mathop{\mathrm{avg}}_{s<t}(F(\boldsymbol{x}_s,\boldsymbol{x}_t))\right] = \frac{1}{\binom{m}{2}^2}\sum_{\substack{s<t\\ s'<t'}}\mathbf{Cov}[F(\boldsymbol{x}_s,\boldsymbol{x}_t),F(\boldsymbol{x}_{s'},\boldsymbol{x}_{t'})]. \tag{5.2}$$

We will discuss these covariances in three cases.

Case 1: $|\{s,t\}\cap\{s',t'\}| = 2$. Let $\boldsymbol{x},\boldsymbol{y}\sim\varphi$ be independent random variables.

$$\mathbf{Cov}[F(\boldsymbol{x},\boldsymbol{y}),F(\boldsymbol{x},\boldsymbol{y})] = \mathop{\mathbf{Var}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[F(\boldsymbol{x},\boldsymbol{y})] \le \mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[F(\boldsymbol{x},\boldsymbol{y})^2] = \mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{1\le|S|\le k}\boldsymbol{x}^S\boldsymbol{y}^S\right)^2\right].$$

Notice here all $\boldsymbol{x}_i$'s are Rademacher variables with $\boldsymbol{x}_i^2 = 1$, and similarly for the $\boldsymbol{y}_i$'s. Therefore

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{1\le|S|\le k}\boldsymbol{x}^S\boldsymbol{y}^S\right)^2\right] &= \sum_{1\le|S_1|,|S_2|\le k}\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\boldsymbol{x}^{S_1\oplus S_2}\boldsymbol{y}^{S_1\oplus S_2}\right]\\ &= \sum_{1\le|S_1|,|S_2|\le k}\widehat{\varphi}(S_1\oplus S_2)^2 = L_k(\varphi).\end{aligned}$$

Case 2: $|\{s,t\}\cap\{s',t'\}| = 1$. Let $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\varphi$ be independent random variables. Similar to

Case 1, we have:

$$\begin{aligned}
\mathbf{Cov}[F(\boldsymbol{x},\boldsymbol{y}),F(\boldsymbol{x},\boldsymbol{z})] &\le \mathbf{E}[F(\boldsymbol{x},\boldsymbol{y})F(\boldsymbol{x},\boldsymbol{z})] \\
&= \mathbf{E}\left[\left(\sum_{1\le|S_1|\le k}\boldsymbol{x}^{S_1}\boldsymbol{y}^{S_1}\right)\left(\sum_{1\le|S_2|\le k}\boldsymbol{x}^{S_2}\boldsymbol{z}^{S_2}\right)\right] \\
&= \mathbf{E}\left[\sum_{1\le|S_1|,|S_2|\le k}\boldsymbol{x}^{S_1\oplus S_2}\boldsymbol{y}^{S_1}\boldsymbol{z}^{S_2}\right] \\
&= \sum_{1\le|S_1|,|S_2|\le k}\widehat{\varphi}(S_1\oplus S_2)\widehat{\varphi}(S_1)\widehat{\varphi}(S_2) \\
&\le \sqrt{\sum_{1\le|S_1|,|S_2|\le k}\widehat{\varphi}(S_1\oplus S_2)^2}\sqrt{\sum_{1\le|S_1|,|S_2|\le k}\widehat{\varphi}(S_1)^2\widehat{\varphi}(S_2)^2} \\
&= \sqrt{L_k(\varphi)}\mu,
\end{aligned}$$

where the inequality comes from Cauchy–Schwarz.

Case 3: $|\{s,t\}\cap\{s',t'\}|=0$. Let $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{w}\sim\varphi$ be independent random variables. Clearly $F(\boldsymbol{x},\boldsymbol{y})$ and $F(\boldsymbol{z},\boldsymbol{w})$ are independent and therefore $\mathbf{Cov}[F(\boldsymbol{x},\boldsymbol{y}),F(\boldsymbol{z},\boldsymbol{w})]=0$.

Plugging all these cases into eq. (5.2), we get

$$\begin{aligned}
\mathbf{Var}[\Delta(\boldsymbol{X})] &= \mathbf{Var}\left[\operatorname*{avg}_{s<t}(F(\boldsymbol{x}_s,\boldsymbol{x}_t)\right] \\
&= \frac{1}{\binom{m}{2}^2}\left(\binom{m}{2}L_k(\varphi)+m(m-1)(m-2)\sqrt{L_k(\varphi)}\mu\right) \\
&\le \frac{4}{m^2}L_k(\varphi)+\frac{4}{m}\sqrt{L_k(\varphi)}\mu. \qquad \square
\end{aligned}$$

Given Lemma 5.4.1 we can bound the samples we need for estimating $\mathbf{W}^{1\dots k}[\varphi]$.

Theorem 5.4.2 ($\mathbf{W}^{1\dots k}$ Estimation Test)**.** *Let $\varphi:\{-1,1\}^n\to\mathbb{R}^{\ge0}$ be a density function, promised to satisfy $\mathbf{W}^i[\varphi]\le An^{i/2}$ for all $i=0,1,\dots,2k$. There is an algorithm that, given*

$$m\ge 1000\frac{2^k\sqrt{A}n^{k/2}}{\theta} \tag{5.3}$$

samples, distinguishes with probability at least $3/4$ whether $\mathbf{W}^{1\dots k}[\varphi]\le\frac{1}{2}\theta$ or $\mathbf{W}^{1\dots k}[\varphi]>\theta$.

Proof. The algorithm is simple: we report "$\mu\le\frac{1}{2}\theta$" if $\Delta(\boldsymbol{X})\le\frac{3}{4}\theta$ and report "$\mu>\theta$" if $\Delta(\boldsymbol{X})>\frac{3}{4}\theta$.

Now we need to bound $L_k(\varphi)$ to bound the variance of $\Delta(\boldsymbol{X})$. For a fixed subset $|S|\le 2k$, how many pairs of $1\le|S_1|,|S_2|\le k$ are there satisfying $S=S_1\oplus S_2$? We denote $S_1=S_1'\cup T$, $S_2=S_2'\cup T$, where S_1',S_2',T are disjoint. Then $S=S_1'\cup S_2'$. For a fixed set S, there are at most $2^{|S|}$ different ways to split it into two sets S_1',S_2'. Because $\max\{S_1',S_2'\}\ge\lceil|S|/2\rceil$ and

$|S_1|, |S_2| \le k$, we have $|T| \le k - \lceil |S|/2 \rceil$. Therefore there are at most

$$\sum_{j=0}^{k-\lceil |S|/2 \rceil} \binom{n-|S|}{j} \le \frac{2n^{k-\lceil |S|/2 \rceil}}{(k-\lceil |S|/2 \rceil)!}$$

ways to choose the set T for any fixed S_1', S_2'. Hence,

$$\begin{aligned}
L_k(\varphi) &= \sum_{1 \le |S_1|, |S_2| \le k} \widehat{\varphi}(S_1 \oplus S_2)^2 \\
&= \sum_{|S| \le 2k} \sum_{\substack{S_1' \cap S_2' = \emptyset \\ S_1' \cup S_2' = S}} \sum_{\substack{T \cap S_1' = \emptyset, T \cap S_2' = \emptyset \\ |T| + \max\{|S_1'|, |S_2'|\} \le k}} \widehat{\varphi}(S)^2 \\
&\le \sum_{|S| \le 2k} 2^{|S|} \frac{2n^{k-\lceil |S|/2 \rceil}}{(k-\lceil |S|/2 \rceil)!} \widehat{\varphi}(S)^2 \\
&= \sum_{i=0}^{2k} 2^i \frac{2n^{k-\lceil i/2 \rceil}}{(k-\lceil i/2 \rceil)!} \mathbf{W}^i[\varphi].
\end{aligned}$$

Plugging in $\mathbf{W}^i[\varphi] \le An^{i/2}$, we get

$$L_k(\varphi) \le \sum_{i=0}^{2k} 2^i \frac{2n^{k-\lceil i/2 \rceil}}{(k-\lceil i/2 \rceil)!} \mathbf{W}^i[\varphi] \le 2^{2k+2} A n^k. \tag{5.4}$$

By substituting eq. (5.4) and eq. (5.3) into eq. (5.1), we have

$$\mathbf{Var}[\Delta(\boldsymbol{X})] \le \frac{4}{500^2}\theta^2 + \frac{4}{500}\theta\mu \le \frac{1}{64}\max\{\theta^2, \mu^2\}.$$

Then we conclude our proof by Chebyshev's inequality:

$$\begin{aligned}
\mathbf{Pr}\left[|\Delta(\boldsymbol{X}) - \mu| \le \frac{1}{4}\max\{\theta, \mu\}\right] &\ge \mathbf{Pr}\left[|\Delta(\boldsymbol{X}) - \mu| \le 2\sqrt{\mathbf{Var}[\Delta(\boldsymbol{X})]}\right] \\
&\ge 1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4}.
\end{aligned}$$

□

This $\mathbf{W}^{1\ldots k}$ Estimation Test is just what we need for testing k-wise uniformity with the upper bound of the Closeness Problem.

Proof of Theorem 5.1.4. From Theorem 5.1.1 we know that if density φ is δ-far from k-wise uniform, then $\mathbf{W}^{1\ldots k}[\varphi] > \left(\frac{\delta}{e^k}\right)^2$. On the other hand if φ is k-wise uniform, by definition we have $\mathbf{W}^{1\ldots k}[\varphi] = 0$. Therefore distinguishing between k-wise uniform and δ-far from k-wise uniform can be reduced to distinguishing between $\mathbf{W}^{1\ldots k}[\varphi] > \left(\frac{\delta}{e^k}\right)^2$ and $\mathbf{W}^{1\ldots k}[\varphi] = 0$.

For any density function φ, $|\widehat{\varphi}(S)| = \left|\mathbf{E}[\varphi(\boldsymbol{x})\boldsymbol{x}^S]\right| \le 1$ for any $S \subseteq [n]$. Therefore assigning $A = n^k$, we have

$$\mathbf{W}^i[\varphi] = \sum_{|S|=i} \widehat{\varphi}(S)^2 \le n^i \le An^{i/2}$$

for every $i = 0, 1, \ldots, 2k$.

Hence we can run the $\mathbf{W}^{1\ldots k}$ Estimator Test in Theorem 5.4.2 with parameter $\theta = \left(\frac{\delta}{e^k}\right)^2$ and $A = n^k$, thereby we solve the Testing Problem with sample complexity $2^{O(k)} n^k / \delta^2$.

In fact by mroe precise calculation we can further improve the constant factor involving k to $O\left(\frac{1}{k}\right)^{k/2}$, but we will omit the proof here for the sake of brevity. □

5.4.2 Lower bound for the pairwise case

An upper bound for the Closeness Problem implies an upper bound for the Testing Problem. But a lower bound for Closeness does not obviously yield a lower bound for the Testing Problem. The function used to show the lower bound for the Closeness Problem is far from k-wise uniform, but it is not sufficient to say that it is hard to distinguish between it and some k-wise uniform distribution. In [AAK+07], they show that it is hard to distinguish between the fully uniform distribution and the uniform distribution on a random set of size around $O(n^{k-1}/\delta^2)$; this latter distribution is far from k-wise uniform with high probability for $k > 2$.

We show that the density function φ we used for the lower bound for the Closeness Problem is a useful density to use for a testing lower bound in the pairwise case. However it is not hard to distinguish between the fully uniform distribution and φ. Our trick is shifting φ by a random "center". We remind the reader that we denote by $\varphi^{+t}(x) = \varphi(x \circ t)$ the distribution φ shifted by vector t. We claim that with $m = o(n/\delta^2)$ samples, it is hard to distinguish the fully uniform distribution from φ^{+t} with a uniformly randomly chosen t.

Lemma 5.4.3. *Let φ be the density function defined by $\varphi(x) = 1 + \frac{\delta}{n}\sum_{i<j} x_i x_j$. Assume $m < n/\delta^2$. Let $\Phi : (\{-1,1\}^n)^m \to \mathbb{R}^{\geq 0}$ be the density associated to the distribution on m-tuples of strings defined as follows: First, choose $\boldsymbol{t}$ in $\{-1,1\}^n$ uniformly; then choose m strings independently from $\varphi^{+\boldsymbol{t}}$. Let $\mathbf{1}$ denote the constantly 1 function on $(\{-1,1\}^n)^m$, the density associated to the uniform distribution. Then the χ^2-divergence between Φ and $\mathbf{1}$, $\|\Phi - \mathbf{1}\|_2^2$, is bounded by*

$$\|\Phi - \mathbf{1}\|_2^2 \leq O\left(\frac{m\delta^2}{n}\right).$$

Proof. We need to show that $\mathbf{E}[(\Phi - \mathbf{1})^2] = \mathbf{E}[\Phi^2] - 1 \leq O(m\delta^2/n)$. For uniform and independent $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$,

$$\begin{aligned}
\mathbf{E}[\Phi(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)})^2] &= \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\left(\mathop{\mathbf{E}}_{\boldsymbol{t}}\left[\prod_{i=1}^{m} \varphi^{+\boldsymbol{t}}(\boldsymbol{x}^{(i)})\right]\right)^2\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{t},\boldsymbol{t}'}\left[\prod_{i=1}^{m} \varphi^{+\boldsymbol{t}}(\boldsymbol{x}^{(i)})\varphi^{+\boldsymbol{t}'}(\boldsymbol{x}^{(i)})\right] \\
&= \mathop{\mathbf{E}}_{\boldsymbol{t},\boldsymbol{t}'}[\langle \varphi^{+\boldsymbol{t}}, \varphi^{+\boldsymbol{t}'}\rangle^m].
\end{aligned}$$

It is a trivial fact that $\langle \varphi^{+t}, \varphi^{+t'}\rangle = \varphi * \varphi(t + t')$. Therefore

$$\mathbf{E}[\Phi(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)})^2] = \mathbf{E}[(\varphi * \varphi)^m].$$

We know that $\widehat{\varphi * \varphi}(S) = \varphi(S)^2$. Therefore

$$\varphi * \varphi = 1 + \frac{\delta^2}{n^2} \sum_{i<j} x_i x_j.$$

To compute $\mathbf{E}[(\varphi * \varphi)^m]$, we just need to calculate the constant term of $(1 + \frac{\delta^2}{n^2} \sum_{i<j} x_i x_j)^m$ since $x_i^2 = 1$. Suppose that when expanding this out, we take l terms of $x_i x_j$; we think these as l (possibly parallel) edges in the complete graph on n vertices. Then if these l terms "cancel out", the associated edges form a collection of cycles, since each vertex has even degree. There are at most n^l collections of cycles with l edges. Considering choosing those l terms (edges) in order, we get an upper bound of $(mn)^l$ for the number of ways of choosing l terms of $x_i x_j$ to get canceled. Therefore we have

$$\mathbf{E}\left[\left(1 + \frac{\delta^2}{n^2} \sum_{i \neq j} \boldsymbol{x}_i \boldsymbol{x}_j\right)^m\right] \leq \sum_{l=0}^{m} (mn)^l \left(\frac{\delta^2}{n^2}\right)^l \leq \sum_{l=0}^{m} \left(\frac{m\delta^2}{n}\right)^l \leq 1 + O\left(\frac{m\delta^2}{n}\right),$$

which completes the proof. □

Now we are ready to give the lower bound for sample complexity of testing fully uniform vs. far-from-pairwise uniform.

Proof of Theorem 5.1.5. If $m = o(n/\delta^2)$, by Lemma 5.4.3 we have $\|\Phi - \mathbf{1}\|_2^2 \leq o(1)$. Then any tester cannot distinguish, with more than $o(1)$ advantage, whether those m samples are fully uniform or they are drawn from $\varphi^{+\boldsymbol{t}}$ for some random $\boldsymbol{t}$.

On the other hand, the proof of Theorem 5.1.2 shows that φ is $\Omega(\delta)$-far from pairwise uniform, and from the Fourier characterization, we have that φ^{+t} is pairwise uniform whenever φ is. We can conclude that testing fully uniform versus δ-far-from-pairwise-uniform needs sample complexity at least $\Omega(n/\delta^2)$. □

Unfortunately, we do not see an obvious way to generalize this lower bound to $k > 2$.

5.5 Testing αk-wise/fully uniform vs. far from k-wise uniform

5.5.1 The algorithm

In this section we show a sample-efficient algorithm for testing whether a distribution is αk-wise/fully uniform or δ-far from k-wise uniform. As a reminder, Theorem 5.4.2 indicates that the sample complexity of estimating $\mathbf{W}^{1\dots k}[\varphi]$ is bounded by the Fourier weight up to level $2k$. This suggests using a filter test to try to "kick out" those distributions with noticeable Fourier weight up to degree $2k$.

Filter Test. Draw m_1 samples from φ. If there exists a pair of samples $\boldsymbol{x}, \boldsymbol{y}$ such that $|\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i| > t\sqrt{n}$, output "Reject"; otherwise, output "Accept".

The Overall Algorithm is combining the Filter Test and the $\mathbf{W}^{1\dots k}$ Estimation Test.

Overall Algorithm. Do the Filter Test with m_1 samples and parameter t. If it rejects, say "No". Otherwise, do the $\mathbf{W}^{1\dots k}$ Estimation Test with m_2 samples and $\theta = (\delta/e^k)^2$. Say "No" if it outputs "$\mathbf{W}^{1\dots k}[\varphi] > \theta$" and say "Yes" otherwise.

Here "Yes" means φ is αk-wise/fully uniform, and "No" means φ is δ-far from k-wise uniform. We will decide the parameters m_1, t, m_2 in the Overall Algorithm later.

For simplicity, we denote $\overline{k} = \alpha k$. We will focus on testing αk-wise uniform vs. far from k-wise uniform in the analysis. For fully uniformity, the analysis is almost the same, and we will discuss it at the end of this subsection.

First of all, we will prove that if φ is $\overline{k}$-wise uniform, it will pass the Filter Test with high probability, provided we choose m_1 and t properly.

Lemma 5.5.1. *If φ is $\overline{k}$-wise uniform (assuming $\overline{k}$ is even), the Filter Test will accept with probability at least .9 when $m_1^2 \le \frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}$.*

Proof. If φ is $\overline{k}$-wise uniform with $\overline{k}$ even, then by Markov's inequality on the $\overline{k}$-th moment, we have

$$\Pr_{\substack{\boldsymbol{x},\boldsymbol{y}\sim\varphi\\ \text{independent}}}\left[\left|\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right| > t\sqrt{n}\right] = \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right)^{\overline{k}} > (t\sqrt{n})^{\overline{k}}\right] \le \frac{\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right)^{\overline{k}}\right]}{t^{\overline{k}}n^{\overline{k}/2}}.$$

When we expand $\left(\sum_{i=1}^n x_iy_i\right)^{\overline{k}}$, each term is at most degree $\overline{k}$ in x or y. Because $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent random variables chosen from $\overline{k}$-wise uniform distribution φ, the whole polynomial behaves the same as if $\boldsymbol{x}$ and $\boldsymbol{y}$ were chosen from the fully uniform distribution:

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right)^{\overline{k}}\right] &= \mathop{\mathbf{E}}_{\boldsymbol{z}\sim\{-1,1\}^n}\left[\left(\sum_{i=1}^n \boldsymbol{z}_i\right)^{\overline{k}}\right]\\ &\le \overline{k}^{\overline{k}/2}\left(\mathop{\mathbf{E}}_{\boldsymbol{z}\sim\{-1,1\}^n}\left[\left(\sum_{i=1}^n \boldsymbol{z}_i\right)^{2}\right]\right)^{\overline{k}/2}\\ &= \overline{k}^{\overline{k}/2}n^{\overline{k}/2}.\end{aligned}$$

The inequality uses hypercontractivity; see Theorem 9.21 in [O'D14]. Hence we have

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left|\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right| > t\sqrt{n}\right] \le \frac{\overline{k}^{\overline{k}/2}}{t^{\overline{k}}}.$$

When drawing m_1 examples, there are at most $\binom{m_1}{2} \le \frac{1}{2}m_1^2$ pairs. Hence by the union bound, the probability of φ getting rejected is at most $\frac{m_1^2\overline{k}^{\overline{k}/2}}{2t^{\overline{k}}} \le \frac{1}{10}$. □

Secondly, we claim that for any distribution φ that does not get rejected by the Filter Test, it is close to a distribution φ' with upper bounds on the Fourier weights of each of its levels.

Lemma 5.5.2. *Any distribution φ either gets rejected by the Filter Test with probability at least .9, or there exists some distribution φ' such that:*

1. *φ' and φ are $\frac{8}{m_1}$-close in total variation distance;*
2. *$\mathbf{W}^i[\varphi'] \le \frac{10^7}{m_1^2} n^i + t^i n^{i/2}$ for all $i = 1, \ldots, n$.*

We will present the proof of Lemma 5.5.2 in the next subsection.

If φ is not rejected by the Filter Test, Lemma 5.5.2 tells us that it is close to some distribution φ' with bounded Fourier weights on each of its levels. Even though we are drawing samples from φ, we can "pretend" that we are drawing samples from φ' since they are close:

Claim 5.5.3. *Let $m_2 \le \frac{m_1}{200}$, and let $A(X^{(m_2)})$ be any event related to m_2 samples in $\{-1,1\}^n$, $X^{(m_2)} = \{x_1, \ldots, x_{m_2}\}$. Then we have*

$$\left| \mathop{\mathbf{Pr}}_{\boldsymbol{X}^{(m_2)} \sim \varphi} [A(\boldsymbol{X}^{(m_2)})] - \mathop{\mathbf{Pr}}_{\boldsymbol{X}^{(m_2)} \sim \varphi'} [A(\boldsymbol{X}^{(m_2)})] \right| \le .08,$$

when φ and φ' are $\frac{8}{m_1}$-close.

Proof. We denote by Φ(respectively, Φ') the joint distribution of m_2 samples from φ(respectively, φ'). Then by a union bound we know that Φ and Φ' are .04-close, since $m_2 \frac{8}{m_1} \le .04$. We denote $\mathbf{1}[A(\boldsymbol{X}^{(m_2)})]$ as the indicator function of event A happening on $\boldsymbol{X}^{(m_2)}$. Then we have

$$\begin{aligned}
\left| \mathop{\mathbf{Pr}}_{\boldsymbol{X}^{(m_2)} \sim \varphi} [A(\boldsymbol{X}^{(m_2)})] - \mathop{\mathbf{Pr}}_{\boldsymbol{X}^{(m_2)} \sim \varphi'} [A(\boldsymbol{X}^{(m_2)})] \right| &= \left| \sum_{X^{(m_2)}} \mathbf{1}[A(X^{(m_2)})] \big(\Phi(X^{(m_2)}) - \Phi'(X^{(m_2)}) \big) \right| \\
&\le \sum_{X^{(m_2)}} \left| \Phi(X^{(m_2)}) - \Phi'(X^{(m_2)}) \right| \\
&= 2 d_{\mathrm{TV}}(\Phi, \Phi') \le .08
\end{aligned}$$

which completes the proof. □

Now we are ready to analyze the Overall Algorithm.

Proof of Theorem 5.1.6. We discuss distinguishing between $\overline{k}$-wise uniform and δ-far from k-wise uniform first. In the Overall Algorithm, we set the parameters $t = \left(10^{11} (4e^4)^k \overline{k}^{\overline{k}/2} \frac{n^k}{\delta^4} \right)^{\frac{1}{\overline{k} - 2k}}$ and $m_1 = \sqrt{\frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}}$ in the Filter Test; and, we set $m_2 = \frac{1}{200} m_1$ and $\theta = \left(\frac{\delta}{e^k}\right)^2$ in the $\mathbf{W}^{1\ldots k}$ Estimation test.

In total we use $m_1 + m_2 = O\left(\sqrt{\frac{t^{\overline{k}}}{\overline{k}^{\overline{k}/2}}} \right)$ samples in the Overall Algorithm. By plugging in the definition of t and $\overline{k} = \alpha k$, we can simplify the sample complexity to $O(\alpha)^{k/2} \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\frac{n^k}{\delta^4} \right)^{1/(\alpha - 2)}$.

The rest of the proof is to show the correctness of this algorithm. We discuss the two cases.

"Yes" case: Suppose φ is $\overline{k}$-wise uniform. By Lemma 5.5.1 we know that φ will pass the Filter Test with probability at least .9 since $m_1^2 = \frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}$.

Now φ is $\overline{k}$-wise uniform with $\overline{k} > 2k$, which means $\widehat{\varphi}(S) = 0$ for any $1 \le |S| \le 2k$. Therefore by setting $\delta = \left(\frac{\theta}{e^k}\right)^2$ and $A = 1$, Theorem 5.4.2 tells us that m_2 samples are large enough for $\mathbf{W}^{1\dots k}$ Estimation Test to output "$\mathbf{W}^{1\dots k}[\varphi] \le \frac{1}{2}\theta$" with probability $3/4$.

The overall probability of the Overall Algorithm saying "Yes" is therefore at least $.9 \times \frac{3}{4} > \frac{2}{3}$.

"No" case: Suppose φ is δ-far from k-wise uniform. Either φ gets rejected by the Filter Test with probability .9, or according to Lemma 5.5.2, we know that there exists some distribution φ' which is $\frac{8}{m_1}$-close to φ and $\mathbf{W}^i[\varphi'] \le \frac{10^7}{m_1^2} n^i + t^i n^{i/2}$ for all $i = 1, \dots, n$.

The second stage is slightly tricky. As described in Claim 5.5.3, at the expense of losing $.08$ probability, we may pretend we are drawing samples from φ' rather than φ. Notice that $m_1^2 = \frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}} = \omega(n^k)$. We have

$$\mathbf{W}^i[\varphi'] \le \frac{10^7}{m_1^2} n^i + t^i n^{i/2} = (1 + o(1)) t^i n^{i/2} \le A n^{i/2}$$

for $i = 0, \dots, 2k$ with parameter $A = 1.01 t^{2k}$. Then plugging $A = 1.01 t^{2k}$ and $\theta = \left(\frac{\delta}{e^k}\right)^2$ into Theorem 5.4.2, we know that the $\mathbf{W}^{1\dots k}$ Estimation Test will say "$\mathbf{W}^{1\dots k}[\varphi] > \theta$" with probability at least $\frac{3}{4}$ when φ' is δ-far from k-wise uniform, provided we have at least $1005 \frac{(2e^2)^k t^k n^{k/2}}{\delta^2}$ samples. It is easy to check $m_2 = \frac{1}{200}\sqrt{\frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}}$ is sufficient.

However, in the real algorithm we are drawing samples from φ rather than φ'. From Claim 5.5.3, we know that the estimator will accept with probability at least $\frac{3}{4} - .08 > \frac{2}{3}$ when φ' is δ-far from k-wise uniform. Notice that φ and φ' are $\frac{8}{m_1}$-close, where $\frac{8}{m_1} = o\left(\frac{\delta^4}{n^k}\right)$. Hence if φ is δ-far from k-wise uniform, φ' is also δ-far from k-wise uniform, which completes the proof.

Finally, for distinguishing between a distribution being fully uniform and a distribution being δ-far from k-wise uniform, the modification we need is that in Lemma 5.5.1 we use Hoeffding's inequality to get

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left|\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{y}_i\right| > t\sqrt{n}\right] \le 2e^{-t^2/2},$$

and then we have the constraint $m_1^2 \le \frac{1}{10} e^{t^2/2}$. Following exactly the same analysis, we get the same algorithm with sample complexity $O(k)^k \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\log \frac{n}{\delta}\right)^{k/2}$. □

5.5.2 Proof of Lemma 5.5.2

The rest of this section is devoted to proving Lemma 5.5.2. We will use the following definition in the analysis.

Definition 5.5.4. For $x, y \in \{-1, 1\}^n$, we say (x, y) is *skewed* if $|\sum_{i=1}^n x_i y_i| > t\sqrt{n}$. We say that x is β*-bad for distribution* φ if $\mathbf{Pr}_{\boldsymbol{y}\sim\varphi}[(x, \boldsymbol{y}) \text{ is skewed}] > \beta$.

Claim 5.5.5. *If* $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad for } \varphi\right] > \frac{8}{m_1}$*, then* φ *will be rejected by the Filter Test with probability at least* .9.

Proof. Suppose $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad for } \varphi\right] > \frac{8}{m_1}$. We will divide the samples we draw for the Filter Test into two sets with size $m_1/2$ each. Then the probability of choosing an $\frac{8}{m_1}$-bad x among the first $m_1/2$ samples is at least

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{m/2}\sim\varphi}\left[\exists x \ \frac{8}{m_1}\text{-bad for } \varphi \text{ among } \boldsymbol{x}_1,\ldots,\boldsymbol{x}_{m/2}\right] > 1-\left(1-\frac{8}{m_1}\right)^{m_1/2} \geq 1-e^{-4}.$$

Now if we have such an $\frac{8}{m_1}$-bad x among the first $m_1/2$ samples, each $(x,\boldsymbol{x}_t)$ will be skewed with probability at least $\frac{8}{m_1}$ for any $t = m_1/2+1,\ldots,m$. Therefore

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}_{m/2+1},\ldots,\boldsymbol{x}_m}[(x,\boldsymbol{x}_t) \text{ is skewed for some } t = \frac{m}{2}+1,\ldots,m] \geq 1-\left(1-\frac{8}{m_1}\right)^{m_1/2} \geq 1-e^{-4}.$$

Combining the two inequalities together, we know that the probability of at least one pair being skewed is at least $(1-e^{-4})^2 \geq .9$. □

Now we only need to consider the case when the probability of drawing a bad x from φ is very small. We want to show a stronger claim that even the probability of drawing a skewed pair from φ is small. However this might not be true for φ itself. Thus we look at another distribution φ', which is defined to be φ conditioned on outcomes being not bad. Define φ' as

$$\varphi'(x) = \varphi(x)\frac{\mathbf{1}\left[x \text{ not } \frac{8}{m_1}\text{-bad for } \varphi\right]}{1-\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad for } \varphi\right]}. \tag{5.5}$$

We show that φ' is close to φ and that φ' has no bad samples:

Claim 5.5.6. *Suppose φ satisfies $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad for } \varphi\right] \leq \frac{8}{m_1}$ and $m_1 \geq 16$. Let φ' be defined as in eq.* (5.5)*. Then:*

1. *φ and φ' are $\frac{8}{m_1}$-close;*
2. *$\varphi'(x) = 0$ for any x that is $\frac{16}{m_1}$-bad for φ'.*

Proof. 1. Notice that $\varphi'(x) = 0 \leq \varphi(x)$ when x is $\frac{8}{m_1}$-bad for φ, and $\varphi'(x) \geq \varphi(x)$ otherwise. Hence,

$$\begin{aligned} d_{\text{TV}}(\varphi,\varphi') &= \frac{1}{2}\mathop{\mathbf{E}}_{\boldsymbol{x}}[|\varphi(\boldsymbol{x})-\varphi'(\boldsymbol{x})|] \\ &= \frac{1}{2^n}\sum_{\varphi'(x)<\varphi(x)}(\varphi(x)-\varphi'(x)) \\ &\leq \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \ \frac{8}{m_1}\text{-bad on } \varphi\right] \leq \frac{8}{m_1}. \end{aligned}$$

2. $\varphi'(x)$ is either 0 or at most $(1+\frac{16}{m_1})\varphi(x)$ given $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad for } \varphi\right] \le \frac{8}{m_1}$ and $m_1 \ge 16$. Therefore if $\varphi'(x) > 0$, x is not $\frac{8}{m_1}$-bad for φ. Hence,

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\boldsymbol{y}\sim\varphi'}[(x,\boldsymbol{y}) \text{ is skewed}] &\le \left(1+\frac{16}{m_1}\right)\mathop{\mathbf{Pr}}_{\boldsymbol{y}\sim\varphi}[(x,\boldsymbol{y}) \text{ is skewed}] \\ &\le \left(1+\frac{16}{m_1}\right)\frac{8}{m_1} \le \frac{16}{m_1}.\end{aligned}$$

□

Claim 5.5.7. *Suppose distribution φ satisfies $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad for } \varphi\right] \le \frac{8}{m_1}$. Let φ' be defined as in eq.* (5.5)*. If $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\boldsymbol{x},\boldsymbol{y}) \text{ is skewed}] > \frac{10^7}{m_1^2}$, then with probability at least .9, φ will be rejected by the Filter Test.*

We want to clarify that the constraint is about φ', but we are drawing samples from φ in the Filter Test.

Proof. We only consider the first $m_1' = \frac{m_1}{200}$ samples. From Claim 5.5.6 we know that φ and φ' are $\frac{8}{m_1}$-close. Therefore, we only need to show that if the samples are drawn from φ', the probability of appearing a skewed pair among these m_1' samples is at least .98. Then φ will be rejected by the Filter Test with probability at least $.98 - .08 \ge .9$ according to Claim 5.5.3.

Define random variable $\boldsymbol{U}_{s,t}$ to be the indicator associated with the event that $(\boldsymbol{x}_s, \boldsymbol{x}_t)$ is skewed, and $\boldsymbol{U} = \sum_{1\le s<t\le m_1'} \boldsymbol{U}_{s,t}$. We need to prove that $\mathbf{Pr}[\boldsymbol{U} = 0] \le .02$. (From now on, all probabilities and expectations are based on choosing samples from distribution φ'.) By Chebyshev's inequality, we know that $\mathbf{Pr}[\boldsymbol{U}=0] \le \frac{\mathbf{Var}[\boldsymbol{U}]}{\mathbf{E}[\boldsymbol{U}]^2}$, so we need to calculate $\mathbf{Var}[\boldsymbol{U}]$ and $\mathbf{E}[\boldsymbol{U}]$.

Denote $\mu = \mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\boldsymbol{x},\boldsymbol{y}) \text{ is skewed}]$. Then $\mathbf{E}[\boldsymbol{U}_{s,t}] = \mu$ for any $s<t$ and hence we have

$$\mathbf{E}[\boldsymbol{U}] = \sum_{s<t}\mathbf{E}[\boldsymbol{U}_{s,t}] = \binom{m_1'}{2}\mu.$$

It remains to calculate $\mathbf{E}[\boldsymbol{U}^2]$. We can expand it as

$$\mathbf{E}[\boldsymbol{U}^2] = \mathbf{E}\left[\left(\sum_{s<t}\boldsymbol{U}_{s,t}\right)^2\right] = \sum_{\substack{s<t\\ s'<t'}}\mathbf{E}[\boldsymbol{U}_{s,t}\boldsymbol{U}_{s',t'}].$$

Similar to the proof of Lemma 5.4.1, we discuss these expectations in three cases.

Case 1: $|\{s,t\}\cap\{s',t'\}| = 2$. Since $\boldsymbol{U}_{s,t}$ is a Bernoulli random variable, we know that

$$\mathbf{E}[\boldsymbol{U}_{s,t}^2] = \mathbf{E}[\boldsymbol{U}_{s,t}] = \mu.$$

Case 2: $|\{s,t\}\cap\{s',t'\}| = 1$. Without loss of generality we assume $s = s'$. We consider drawing $\boldsymbol{x}_s$ first. For any fixed x_s with $\varphi'(x_s) > 0$,

$$\mathop{\mathbf{E}}_{\boldsymbol{x}_{t'}}[\boldsymbol{U}_{s,t'}] = \mathop{\mathbf{Pr}}_{\boldsymbol{x}_{t'}}[(x_s,\boldsymbol{x}_{t'}) \text{ get skewed}] \le \frac{16}{m_1} = \frac{2}{25m_1'},$$

where the inequality comes from Claim 5.5.6. Therefore,

$$\mathbf{E}[\boldsymbol{U}_{s,t}\boldsymbol{U}_{s,t'}] = \mathop{\mathbf{E}}_{\boldsymbol{x}_s,\boldsymbol{x}_t}[\boldsymbol{U}_{s,t} \mathop{\mathbf{E}}_{\boldsymbol{x}_{t'}}[\boldsymbol{U}_{s,t'}]] \leq \frac{2\mu}{25m_1'}.$$

Case 3: $|\{s,t\} \cap \{s',t'\}| = 0$. Since s,t,s',t' are all distinct, we have

$$\mathbf{E}[\boldsymbol{U}_{s,t}\boldsymbol{U}_{s',t'}] = \mathbf{E}[\boldsymbol{U}_{s,t}]\,\mathbf{E}[\boldsymbol{U}_{s',t'}] = \mu^2.$$

Combining these cases together, we get

$$\mathbf{E}[\boldsymbol{U}^2] = \binom{m_1'}{2}\mu + m_1'(m_1'-1)(m_1'-2)\frac{2\mu}{25m_1'} + \binom{m_1'}{2}\binom{m_1'-2}{2}\mu^2.$$

Then we have

$$\frac{\mathbf{Var}[\boldsymbol{U}]}{\mathbf{E}[\boldsymbol{U}]^2} = \frac{\mathbf{E}[\boldsymbol{U}^2]}{\mathbf{E}[\boldsymbol{U}]^2} - 1 \leq \frac{58}{25m_1'^2\mu}.$$

By substituting $\mu \geq \frac{10^7}{m_1^2} = \frac{10^3}{4m_1'^2}$, we conclude $\mathbf{Pr}[\boldsymbol{U}=0] = \frac{\mathbf{Var}[\boldsymbol{U}]}{\mathbf{E}[\boldsymbol{U}]^2} \leq .02$, which completes the proof. □

Now we only need to consider those distributions φ where their corresponding φ' satisfies that $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\boldsymbol{x},\boldsymbol{y}) \text{ is skewed}] \leq \frac{10^7}{m_1^2}$. This gives us an upper bound on the Fourier weight on all levels of φ'.

Claim 5.5.8. *If* $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\boldsymbol{x},\boldsymbol{y}) \text{ is skewed}] \leq \frac{10^7}{m_1^2}$, *then*

$$\mathbf{W}^i[\varphi'] \leq \frac{10^7}{m_1^2}n^i + t^i n^{i/2}$$

for $i = 1,\ldots,n$.

Proof. We will first show that $\mathbf{W}^i[\varphi'] \leq \mathbf{E}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\sum_{j=1}^n \boldsymbol{x}_j\boldsymbol{y}_j)^i]$. Since $(\sum_{j=1}^n x_jy_j)^i$ is a symmetric function, we can expand it as

$$\left(\sum_{j=1}^n x_jy_j\right)^i = \sum_{\substack{0\leq k\leq i\\ i-k \text{ even}}} \alpha_k \left(\sum_{|S|=k} x^S y^S\right),$$

with positive integer coefficients α_k. Notice that

$$\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}\left[\sum_{|S|=k} x^S y^S\right] = \mathbf{W}^k[\varphi'].$$

Therefore

$$\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}\left[\left(\sum_{j=1}^n \boldsymbol{x}_j\boldsymbol{y}_j\right)^i\right] = \sum_{\substack{0\leq k\leq i\\ i-k \text{ even}}} \alpha_k \mathbf{W}^k[\varphi'] \geq \mathbf{W}^i[\varphi'].$$

The last inequality holds because the α_k's are positive integers and each $\mathbf{W}^k[\varphi']$ is non-negative.

The rest of the proof is devoted to bounding $\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\sum_{j=1}^n \boldsymbol{x}_j\boldsymbol{y}_j)^i]$. When (x, y) is not skewed, $\sum_j x_j y_j$ is at most n; otherwise by the definition of "being skewed", $\sum_j x_j y_j$ is at most $t\sqrt{n}$. Therefore,

$$\mathbf{E}\left[\left(\sum_{j=1}^{n} \boldsymbol{x}_j\boldsymbol{y}_j\right)^i\right] \leq \frac{10^7}{m_1^2}n^i + t^i n^{i/2}$$

for all $i = 1, \ldots, n$. □

Combining the above discussion, we get the proof of Lemma 5.5.2.

Proof of Lemma 5.5.2. We consider three cases for φ.

Case 1: If $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad on } \varphi\right] > \frac{8}{m_1}$, Claim 5.5.5 tells us that φ is rejected by the Filter Test with probability at least .9.

For the remaining two cases we know that $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad on } \varphi\right] \leq \frac{8}{m_1}$. We construct φ' as in eq. (5.5).

Case 2: If $\mathbf{Pr}_{\boldsymbol{x}\sim\varphi}\left[\boldsymbol{x} \text{ is } \frac{8}{m_1}\text{-bad on } \varphi\right] \leq \frac{8}{m_1}$ but $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\boldsymbol{x}, \boldsymbol{y}) \text{ is skewed}] > \frac{10^7}{m_1^2}$, Claim 5.5.7 tells us that φ also gets rejected with probability at least .9.

Case 3: If $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi'}[(\boldsymbol{x}, \boldsymbol{y}) \text{ is skewed}] > \frac{10^7}{m_1^2}$, then according to Claim 5.5.8, $\mathbf{W}^i[\varphi'] \leq \frac{10^7}{m_1^2}n^i + t^i n^{i/2}$ for all $i = 1, \ldots, n$. Also by Claim 5.5.6 we know that φ and φ' are $\frac{8}{m_1}$-close. □

Bibliography

[AA14] Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *Theory Of Computing*, 10(6):133–166, 2014. 4.1, 4.2.2, 4.2.2, 4.2.2

[AA18] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM Journal on Computing*, 47(3):982–1038, 2018. 4.1, 4.1.1, 4.2.1, 3

[AAB+21] Scott Aaronson, Andris Ambainis, Andrej Bogdanov, Krishnamoorthy Dinesh, and Cheung Tsun Ming. On quantum versus classical query complexity. *ECCC Report TR21-115*, 2021. 3

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 496–505, 2007. 1.3, 5.1.1, 5.1.2, 5.1.2, 5.1.3, 5.1.3, 5.1.3, 5.1.3, **??**, **??**, 5.1, 5.4.2

[Aar05] Scott Aaronson. Ten semi-grand challenges for quantum computing theory, 2005. `http://www.scottaaronson.com/writings/qchallenge.html`. 4.2.2

[Aar08] Scott Aaronson. How to solve longstanding open problems in quantum computing using only Fourier Analysis. Lecture at Banff International Research Station, 2008. `http://www.scottaaronson.com/talks/openqc.ppt`. 4.1, 4.2.2

[Aar10] Scott Aaronson. Updated version of "ten semi-grand challenges for quantum computing theory", 2010. `http://www.scottaaronson.com/blog/?p=471`. 4.2.2

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986. 5.1.1

[AC98] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. II. CR capacity. *IEEE Transactions on Information Theory*, 44(1):225–240, 1998. 1.3, 3.1.1, 3.4

[ADK15] Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*, pages 3591–3599, 2015. 5.1.3

[AG76] Rudolf Ahlswede and Peter Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The Annals of Probability*, pages 925–939, 1976. 3.1.2, 3.1.2, 3.1.3, 3.2.1

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. 5.1.1

[AGK76] Rudolf Ahlswede, Peter Gács, and János Körner. Bounds on conditional probabilities with applications in multi-user communication. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34(2):157–177, 1976. 3.1.2

[AGKN14] Venkat Anantharam, Amin Gohari, Sudeep Kamath, and Chandra Nair. On hypercontractivity and a data processing inequality. In *2014 IEEE International Symposium on Information Theory*, pages 3022–3026. IEEE, 2014. 3.1.3

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003. 5.1.1, 5.1.2, 5.1.2, 5.1.3

[AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009. 5.1.2

[AOW15] Sarah R. Allen, Ryan O'Donnell, and David Witmer. How to refute a random CSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 689–708, 2015. 5.1.2

[Bal13] Deepak Bal. On sharp thresholds of monotone properties: Bourgain's proof revisited. *arXiv preprint arXiv:1302.1162*, 2013. 2.1.3

[Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009. 5.1.2

[Bec75] William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975. 1.1.1

[BFF+01] Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 442–451, 2001. 5.1.3

[BFR+00] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 259–269, 2000. 5.1.3, 5.1.3

[BGI14] Mohammad Bavarian, Dmitry Gavinsky, and Tsuyoshi Ito. On the role of shared randomness in simultaneous communication. In *International Colloquium on Automata, Languages, and Programming*, pages 150–162. Springer, 2014. 3.1.1

[BGM16] Iddo Bentov, Ariel Gabizon, and Alex Mizrahi. Cryptocurrencies without proof of work. In *International conference on financial cryptography and data security*, pages 142–157. Springer, 2016. 1.2.2

[BKK+92] Jean Bourgain, Jeff Kahn, Gil Kalai, Yitzhak Katznelson, and Nathan Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77(1):55–64, 1992. 1.1.3, 2.1.3

[BKR04] Tugkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 381–390, 2004. 5.1.3

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013. 1.1.2

[BM11] Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *IEEE Transactions on information theory*, 57(10):6351–6355, 2011. 3.1.1

[BM13] Franck Barthe and Emanuel Milman. Transference principles for log-Sobolev and spectral-gap with applications to conservative spin systems. *Communications in Mathematical Physics*, 323(2):575–625, 2013. 4.1

[BMO+15] Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, page 110, 2015. 4.3

[Bon68] Aline Bonami. Ensembles $\Lambda(p)$ dans le dual de D^∞. In *Annales de l'institut Fourier*, volume 18, pages 193–204, 1968. 1.1.1, 1.1.2

[Bon70] Aline Bonami. Étude des coefficients de fourier des fonctions de $L^p(G)$. In *Annales de l'institut Fourier*, volume 20, pages 335–402, 1970. 1.1.1, 1.1.1, 1.1.1

[Bor79] Christer Borell. On the integrability of banach space valued walsh polynomials. *Séminaire de probabilités de Strasbourg*, 13:1–3, 1979. 1.1.2

[Bou79] Jean Bourgain. Walsh subspaces of L^p-product spaces. *Séminaire Analyse fonctionnelle (dit" Maurey-Schwartz")*, pages 1–14, 1979. 2.1.3, 2.2.2

[Bou02] Jean Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel Journal of Mathematics*, 131(1):269–276, 2002. 1.2.2, 4.3

[Bra10] Mark Braverman. Polylogarithmic independence fools ac^0 circuits. *Journal of the ACM*, 57(5):28:1–28:10, 2010. 5.1.2

[CCH10] Claude Carlet, Yves Crama, and Peter L Hammer. Boolean functions for cryptography and error correcting codes. *Boolean models and methods in mathematics, computer science, and engineering*, 2:257–397, 2010. 1.1.2

[CF63] Péter Csáki and János Fischer. On the general notion of maximal correlation. *Magyar Tud. Akad. Mat. Kutato Int. Kozl*, 8:27–51, 1963. 3.1.2

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 396–407, 1985. 5.1.1

[CGMS17] Clément L Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *IEEE Transactions on Information Theory*, 63(10):6799–6818, 2017. 3.1.1

[CMN14] Siu On Chan, Elchanan Mossel, and Joe Neeman. On extracting common random bits from correlated sources on large alphabets. *IEEE Transactions on Information Theory*, 60(3):1630–1637, 2014. 3.1.1

[Cov99] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999. 3.2.2

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 670–683, 2016. 5.1.2

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. In *22nd International Conference on Randomization and Computation (RANDOM'2018)*, 2018. 2.1.3

[DDG+17] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. *SIAM Journal on Computing*, 46(4):1336–1369, 2017. 2.1.3

[DFKO07] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel Journal of Mathematics*, 160(1):389–412, 2007. (document), 4.1, 4.2.1, 4.2.2, 4.3, 4.3, 4.3, 5, 4.3.4, 4.3.1, 4.3.7

[Din07] Irit Dinur. The PCP Theorem by gap amplification. *Journal of the ACM*, 54(3):1–44, 2007. 4.3

[DK16] Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, pages 685–694. IEEE, 2016. 5.1.3

[DKK+18] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 376–389. ACM, 2018. 2.1.1, 2.1.3

[DKK+21] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in grassmann graphs. *Israel Journal of Mathematics*, pages 1–44, 2021. 2.1.1, 2.1.3

[dlP92] Victor de la Peña. Decoupling and Khintchine's inequalities for U-statistics. *Annals of Probability*, 20(4):1877–1892, 1992. 4.2.1

[dlPG99] Víctor de la Peña and Evarist Giné. *Decoupling: From Dependence to Independence*. Springer, 1999. 4.1, 4.2.1, 4.2.1

[dlPMS95] Victor de la Peña and Stephen Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Annals of Probability*, 23(2):806–816, 1995. 4.2.1

[DS05] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, pages 439–485, 2005. 1.1.2

[EFF12] David Ellis, Yuval Filmus, and Ehud Friedgut. Triangle-intersecting families of graphs. *Journal of the European Mathematical Society*, 14(3):841–885, 2012. 4.3

[FB99] Ehud Friedgut and Jean Bourgain. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the American mathematical Society*, 12(4):1017–1054, 1999. 1.1.3, 2.1.2, 2.1.3

[FI19] Yuval Filmus and Ferdinand Ihringer. Boolean constant degree functions on the slice are juntas. *Discrete Mathematics*, 342(12):111614, 2019. 2.1.3

[FI21] Rupert L Frank and Paata Ivanisvili. Hypercontractivity of the semigroup of the fractional Laplacian on the n-sphere. *Journal of Functional Analysis*, page 109145, 2021. 1.1.3

[Fil16] Yuval Filmus. An orthogonal basis for functions over a slice of the Boolean hypercube. *The Electronic Journal of Combinatorics*, pages P1–23, 2016. 2.1.3

[FK96] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10):2993–3002, 1996. 1.1.3, 2.1.3, 4.3

[FKLM20] Yuval Filmus, Guy Kindler, Noam Lifshitz, and Dor Minzer. Hypercontractivity on the symmetric group. *arXiv preprint arXiv:2009.05503*, 2020. 1.1.3

[FKMW18] Yuval Filmus, Guy Kindler, Elchanan Mossel, and Karl Wimmer. Invariance principle on the slice. *ACM Transactions on Computation Theory (TOCT)*, 10(3):11, 2018. 2.1.3

[FKN02] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose Fourier transform is concentrated on the first two levels and neutral social choice. *Advances in Applied Mathematics*, 29(3):427–437, 2002. 1.1.1, 1.2.2, 4.3

[FM19] Yuval Filmus and Elchanan Mossel. Harmonicity and invariance on slices of the Boolean cube. *Probability Theory and Related Fields*, 175(3):721–782, 2019. 2.1.3

[FOW18] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu. A log-Sobolev inequality for the multislice, with applications. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*, 2018. 1.1.3

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–36, 1998. 1.1.3, 2.1.3, 4.3

[Gin98] Evarist Giné. A consequence for random polynomials of a result of de la Peña and Montgomery-Smith. In *Probability in Banach Spaces 10*, volume 43 of *Progress in Probability*. Birkhäuser–Verlag, 1998. 4.2.1

[GJ18] Badih Ghazi and TS Jayram. Resource-efficient common randomness and secret-key schemes. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1834–1853. SIAM, 2018. 3.1.1

[GR11] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 68–75. Springer, 2011. 1.3, 5.1.3, 5.1.3

[GR16] Venkatesan Guruswami and Jaikumar Radhakrishnan. Tight bounds for communication-assisted agreement distillation. In *31st Conference on Computational Complexity (CCC 2016)*, 2016. 3.1.1, 3.1.1, 3.1.1, 3.1.1, 3.1.2, 3.3, 3.5.1, 3.5.1, 3.5.1

[Gro75] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. 1.1.1

[GS20] Noah Golowich and Madhu Sudan. Round complexity of common randomness generation: The amortized setting. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1076–1095. SIAM, 2020. 3.1.1

[HKL20] Max Hopkins, Tali Kaufman, and Shachar Lovett. High dimensional expanders: Random walks, pseudorandomness, and unique games. *arXiv preprint arXiv:2011.04658*, 2020. 2.1.3

[HLPS19] Uri Hadar, Jingbo Liu, Yury Polyanskiy, and Ofer Shayevitz. Communication complexity of estimating correlations. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 792–803, 2019. 3.1.1

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006. 1.1.2, 1.2

[Jan97] Svante Janson. *Gaussian Hilbert Spaces*. Number 129. Cambridge university press, 1997. 1.1.2, 1.1.2, 3.5.1

[JOW15] Jacek Jendrej, Krzysztof Oleszkiewicz, and Jakub O Wojtaszczyk. On some extensions of the FKN theorem. *Theory of Computing*, 11(1):445–469, 2015. 4.3

[KA12] Sudeep Kamath and Venkat Anantharam. Non-interactive simulation of joint distributions: The hirschfeld-gebelein-rényi maximal correlation and the hypercontractivity ribbon. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1057–1064. IEEE, 2012. 3.2.2

[KA16] Sudeep Kamath and Venkat Anantharam. On non-interactive simulation of joint distributions. *IEEE Transactions on Information Theory*, 62(6):3419–3435, 2016. 3.1.3

[Kah93] Jean-Pierre Kahane. *Some Random Series of Functions*, volume 5. Cambridge University Press, 1993. 2.1.3

[Kal02] Gil Kalai. A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem. *Advances in Applied Mathematics*, 29(3):412–426, 2002. 1.1.1, 5.2.3

[Kan11] Daniel M Kane. k-independent Gaussians fool polynomial threshold functions. In *Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity*, pages 252–261, 2011. 4.3.1

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, 2002. 4.3

[Kie69] Konrad Kiener. *Uber Produkte von quadratisch integrierbaren funktionen*

endlicher Vielfalt. PhD thesis, Universität Innsbruck, 1969. 1.1.1

[Kin02] Guy Kindler. *Property Testing, PCP, and juntas*. PhD thesis, Tel Aviv University, 2002. 4.3

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80. IEEE, 1988. 1.1.1, 1.2.1, 1.2.2, 2.1.2, 3.1.2

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007. 1.1.2

[KLLM21] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications. *arXiv preprint arXiv:2103.04604*, 2021. 1.1.3, 1.4, 2.1.1, 2.1.2, 2.1.2, 2.1.3

[KM13] Daniel Kane and Raghu Meka. A PRG for Lipschitz functions of polynomials with applications to Sparsest Cut. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 1–10, 2013. 4.1.1

[KMMS18] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Pseudorandom sets in Johnson graph have near-perfect expansion. *ECCC Report TR18-078*, 2018. (document), 1.1.3, 1.2.2, 1.3, 2.1.1, 2.1.1, 2.1.1, 2.1.3, 2.1.3, 2.1.3

[KMOW17] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 132–145, 2017. 5.1.2

[KMS17] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and Grassmann graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 576–589. ACM, 2017. 2.1.1, 2.1.3

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in Grassmann graph have near-perfect expansion. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 592–601. IEEE, 2018. (document), 1.1.3, 1.2.2, 1.3, 2.1.1, 2.1.1, 2.1.1, 2.1.3, 2.1.3

[KN06] Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Mathematische Annalen*, 334(4):821–852, 2006. 4.3

[KN08] Subhash Khot and Assaf Naor. Linear equations modulo 2 and the L_1 diameter of convex bodies. *SIAM Journal on Computing*, 38(4):1448–1463, 2008. 4.1.1

[KO12] Guy Kindler and Ryan O'Donnell. Gaussian noise sensitivity and Fourier tails. In *Proceedings of the 27th Annual IEEE Conference on Computational Complexity*, pages 137–147, 2012. 4.3, 4.3

[Kra29] Mikhail Krawtchouk. Sur une généralisation des polynomes d'hermite. *Comptes Rendus de l'Académie des sciences*, 189:620–622, 1929. 5.2.3

[KS02] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas. Manuscript, 2002. 4.3

[KW85] Richard M. Karp and Avi Wigderson. A fast parallel algorithm for the maximal independent set problem. *Journal of the ACM*, 32(4):762–773, 1985. 5.1.1

[Kwa87] Stanisław Kwapień. Decoupling inequalities for polynomial chaos. *Annals of Probability*, 15(3):1062–1071, 1987. 4.2.4, 4.2.1

[LCCV16] Jingbo Liu, Thomas A Courtade, Paul Cuff, and Sergio Verdu. Smoothing Brascamp-Lieb inequalities and strong converses for common randomness generation. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1043–1047. IEEE, 2016. 3.1.2

[LCV15] Jingbo Liu, Paul Cuff, and Sergio Verdú. Secret key generation with one communicator and a one-shot converse via hypercontractivity. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 710–714. IEEE, 2015. 3.1.2

[Lev95] Vladimir I. Levenshtein. Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces. *IEEE Transactions on Information Theory*, 41(5):1303–1321, 1995. 5.2.3

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, pages 168–177. IEEE, 2016. 5.1.2

[LM19] Noam Lifshitz and Dor Minzer. Noise sensitivity on the p-biased hypercube. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1205–1226. IEEE, 2019. 2.1.3

[LMN89] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. In *30th Annual Symposium on Foundations of Computer Science*, pages 574–579. IEEE, 1989. 1.1.2

[LO94] Rafał Latała and Krzysztof Oleszkiewicz. On the best constant in the Khinchin-Kahane inequality. *Studia Mathematica*, 109(1):101–104, 1994. 1.1.3, 2.1.2, 2.1.3

[Lov10] Shachar Lovett. An elementary proof of anti-concentration of polynomials in Gaussian variables. Technical Report 182, Electronic Colloquium on Computational Complexity, 2010. 4.1.1

[Lub86] Michael Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986. 5.1.1

[LY98] Tzong-Yow Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *The Annals of Probability*, 26(4):1855–1873, 1998. 2.1.3

[Mar06] Andrei Andreevich Markov. Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 15(135-156):18, 1906. 3.1.2

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010. 1.1.2, 1.1.2, 2.1.2, 4.3, 4.3.1, 4.3.1

[MOR+06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains,

and the reverse Bonami-Beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006. 1.2.3, 3.1.2

[MS58] Nathaniel Macon and Abraham Spitzbart. Inverses of Vandermonde matrices. *The American Mathematical Monthly*, 65:95–100, 1958. 4.4.1

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The Theory of Error-Correcting Codes*. Elsevier, 1977. 5.1.1

[MS14] Konstantin Makarychev and Maxim Sviridenko. Solving optimization problems with diseconomies of scale via decoupling. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 571–580. IEEE, 2014. 4.1

[Nai14] Chandra Nair. Equivalent formulations of hypercontractivity using information measures. *Proceedings of International Zurich Seminar on Communications*, 2014. 1.2.3, 3.1.1, 3.1.2, 3.1.6, 3.1.2, 3.6

[Nel73] Edward Nelson. The free Markoff field. *Journal of Functional Analysis*, 12(2):211–227, 1973. 1.1.1

[Nev76] Jacques Neveu. Sur l'espérance conditionnelle par rapport à un mouvement brownien. In *Annales de l'IHP Probabilités et statistiques*, volume 12, pages 105–109, 1976. 1.1.4, 1.1.4

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. 5.1.1

[NW16] Chandra Nair and Yan Nan Wang. Evaluating hypercontractivity parameters using information measures. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 570–574. IEEE, 2016. 1.4, 3.1.2, 3.1.3, 3.2.2, 3.5.2

[NW17] Chandra Nair and Yan Nan Wang. Reverse hypercontractivity region for the binary erasure channel. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 938–942. IEEE, 2017. 3.1.2, 3.1.3

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 1.1.1, 1.1.4, 1.2.3, 2.2.1, 3.1.2, 3.2.2, 3.5.1, 4.1.2, 1, 2, 4.2.2, 4.3, 4.3, 4.3.1, 5.2.1, 5.2.2, 5.2.3, 5.2.3, 5.5.1

[OW12] Ryan O'Donnell and John Wright. A new point of np-hardness for unique games. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 289–306, 2012. 3.1.1

[OZ16] Ryan O'Donnell and Yu Zhao. Polynomial bounds for decoupling, with applications. In *Proceedings of the 31st Conference on Computational Complexity*, pages 1–18, 2016. 1.4

[OZ18] Ryan O'Donnell and Yu Zhao. On closeness to k-wise uniformity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, 2018. 1.4

[Pal32] R.E.A.C. Paley. A remarkable series of orthogonal functions (I). *Proceedings of the London Mathematical Society*, 2(1):241–264, 1932. 1.1.1

[Pan08] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008. 5.1.3

[PZ78] Gilles Pisier and Joel Zinn. On the limit theorems for random variables with values in the spaces $L_p(2 \leq p < \infty)$. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 41(4):289–304, 1978. 1.1.2

[Rao47] Calyampudi Radhakrishna Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society*, 9(1):128–139, 1947. 5.1.1

[RS09] Ronitt Rubinfeld and Rocco A. Servedio. Testing monotone high-dimensional distributions. *Random Structures & Algorithms*, 34(1):24–44, 2009. 5.1.3, 5.1.3

[RX13] Ronitt Rubinfeld and Ning Xie. Robust characterizations of k-wise independence over product spaces and related testing results. *Random Structures & Algorithms*, 43(3):265–312, 2013. 5.1.2

[Sch69] Michel Schreiber. Fermeture en probabilité de certains sous-espaces d'un espace L^2. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(1):36–48, 1969. 1.1.1

[SGGB19] Madhu Sudan, Badih Ghazi, Noah Golowich, and Mitali Bafna. Communication-rounds tradeoffs for common randomness and secret key generation. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1861–1871. SIAM, 2019. 3.1.1

[Sha49] Claude E Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28(4):656–715, 1949. 3.1.1

[SHK72] Barry Simon and Raphael Høegh-Krohn. Hypercontractive semigroups and two dimensional self-coupled Bose fields. *Journal of Functional Analysis*, 9(2):121–180, 1972. 1.1.1

[ST18] K.R. Sahasranand and Himanshu Tyagi. Extra samples can reduce the communication for independence testing. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 2316–2320. IEEE, 2018. 3.1.1

[ST21] K.R. Sahasranand and Himanshu Tyagi. Communication complexity of distributed high dimensional correlation testing. *IEEE Transactions on Information Theory*, 2021. 3.1.1

[STW19] Madhu Sudan, Himanshu Tyagi, and Shun Watanabe. Communication for generating correlation: A unifying survey. *IEEE Transactions on Information Theory*, 66(1):5–37, 2019. 3.1.1

[Tal94] Michel Talagrand. On Russo's approximate zero-one law. *Annals of Probability*, 22(3):1576–1587, 1994. 1.1.3, 2.1.3

[Wim14] Karl Wimmer. Low influence functions over slices of the Boolean hypercube depend on few coordinates. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 120–131. IEEE, 2014. 2.1.3

[Wit75] Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. 3.1.1

[Wol07] Paweł Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, 3(180):219–236, 2007. 1.1.3, 2.1.3

[Xie12] Ning Xie. *Testing k-wise independent distributions*. PhD thesis, Massachusetts Institute of Technology, 2012. 5.1.3

[ZC11] Lei Zhao and Yeow-Kiang Chia. The efficiency of common randomness generation. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 944–950. IEEE, 2011. 3.1.1