Generalizations and Applications of Hypercontractivity and Small-Set Expansion
Thesis Proposal

Yu Zhao

Abstract

The hypercontractivity inequalities and the small-set expansion are two fundamental topics very related to each other and play important roles in every fields and several recent breakthroughs in theoretical computer science. This thesis proposal is focused on generalizations and applications of hypercontractivity and small-set expansion such as (i) pseudorandom-set expansion, (ii) communication distillation, (iii) decouping, and (iv) property testing on k-wise uniformity. For each of these problems, we try to propose new algorithms, improve complexity measures or give better bounds.

1 Introduction

The hypercontractivity inequalities and the small-set expansion are two fundamental topics very related to each other and play important roles in every fields and several recent breakthroughs in theoretical computer science. In this thesis proposal, we seek to establish more generalizations and applications of hypercontractivity and small-set expansion in several aspects.

In most parts of this proposal, we discuss the hypercontractivity on Boolean function and the small-set expansion on Boolean hypercube. We present some basic preliminaries here before further discussion. These notations are consistent with [O'D14].

The domain of a Boolean function

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

is the Hamming cube $\{-1, 1\}^n$. Such a boolean function can be represented as a unique multilinear polynomial

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

where $x^S = \prod_{i \in S} x_i$. This is called Fourier expansion of function f and $\widehat{f}(S)$ is the Fourier coefficient of f on set S. We define the usual norm $\|f\|_p = (\mathbf{E}[f(\boldsymbol{x})^p])^{1/p}$.

Let $\rho \in [0, 1]$ and fixed $x \in \{-1, 1\}^n$. We write $\boldsymbol{y} \sim N_\rho(x)$ to denote random Boolean string $\boldsymbol{y}$ is drawn as following:

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

for each $i \in [n]$ independently. If $\boldsymbol{x} \sim \{-1, 1\}^n$ is drawn uniformly and $\boldsymbol{y} \sim N_\rho(\boldsymbol{x})$, we say that $(\boldsymbol{x}, \boldsymbol{y})$ is ρ-correlated.

We define the noise operator T_ρ on functions $f : \{-1, 1\}^n \to \mathbb{R}$ be

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

1.1 Hypercontractivity inequality

In 1970, Bonami proved the full Hypercontractivity Theorem for uniform ± 1 in [Bon70]:

Theorem 1.1 (The Hypercontractivity Theorem for uniform distribution)**.** *Let $f : \{-1, 1\}^n \to \mathbb{R}$, and let $1 \le p \le q \le \infty$. Then $\|\mathrm{T}_\rho f\|_p \le \|f\|_q$ for $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$.*

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

The Hypercontractivity Theorem is crucial and its applications appears in every fields of theoretical computer science, such as expander graphs [HLW06], probability theory [BLM13], circuit complexity [LMN89], coding theory [CCH10], hardness of approximation [KKMO07, DS05], etc.

One branch of applications is based on Generalized Bonami Lemma [Bon68] and Level-k Inequalities [KKL88] proved via hypercontrctivity. These theorems show that low-degree polynomials are reasonable. Some highlights using these results are Kahn-Kalai-Linial Theorem [KKL88] and Bourgain's Sharp Threshold Theorem [FB99]. We study some new applications along this track in Section 4 and 5.

Neveu [Nev76] shows that there is an equivalent two-function version of hypercontractivity which has its own interests:

Theorem 1.2 (Two-Function Hypercontractivity Theorem for uniform distribution)**.** *Let $f, g : \{-1, 1\}^n \to \mathbb{R}$, and let $1 \le p \le q \le \infty$. Then*

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

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

1.2 Small-set expansion

Small set expansion is another nice application from hypercontractivity. $(p, 2)$-Hypercontracitvity Theorem says for any function $f : \{-1, 1\}^n \to \mathbb{R}$ and $1 \le p \le 2$,

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

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

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

By focusing on function $f : \{-1, 1\}^n \to \{0, 1\}$, a very interesting meaning is shown in [KKL88]:

Theorem 1.3 (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 $0 \le \rho \le 1$,*

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

Consider the 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. Then Theorem 1.3 suggests that for any subset A with 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, $1 - \alpha^{\frac{2}{1+\rho}}$. Therefore this hypercube graph is a good small-set expander.

Similarly, by focusing on $f, g : \{-1, 1\}^n \to \{0, 1\}$, Two-Function Hypercontractivity Theorem (Theorem 1.2) is also interpreted into a two-set generalization of the Small-Set Expansion Theorem due to [MOR+06]:

Theorem 1.4. *Let $A, B \subseteq \{-1, 1\}^n$ have volume α and β. Then for any $0 \le \rho \le 1$,*

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x} \sim \{-1,1\}^n \\ \boldsymbol{y} \sim N_\rho(\boldsymbol{x})}} [\boldsymbol{x} \in A, \boldsymbol{y} \in B] \le \alpha^{\frac{1}{p}} \beta^{\frac{1}{q'}},$$

for any p, q' which satisfies $\boldsymbol{x}$ is (p, q, ρ)-hypercontractive.

Until recently, the small-set expansion is seen as a corollary of hypercontractivity. However in [Nai14] they showed that these two theorems are equivalent. More discussion of the relationship between small-set expansion and hypercontractivity will be mentioned in Section 3.

A recent breakthrough of proving 2-to-2 games conjecture is related to the pseudomrandom-set expansion in Grassmann graphs [KMS18]. We will study pseudorandom-set expansion and its relationship to classical small-set expansion and hypercontractivity in Section 2.

1.3 Problem studied

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. 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 DFKO Theorems.

Property testing on k-wise uniformity. 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 satisfy that all low-degree Fourier coefficients of its density function is equal to zero. Using the hypercontractivity inequalities to study the properties of low-degree Fourier weights of Boolean function, we show better bounds for the Closeness and Testing problems of k-wise uniformity.

1.4 Organization

In Section 2, we discuss the pseudorandom-set expansion on Grassmann graphs, Johnson graphs and biased Boolean hypercube as a generalization of small-set expansion. In Section 3, we study the communication distillation problem and its underlying relationship to small-set expansion and extreme points of hypercontractivity domain. In Section 4, we discuss the decoupling method and tight bound of DFKO Theorem as an application of hypercontractivity. In Section 5, we study property testing on k-wise uniformity as another application of hypercontractivity.

2 Pseudorandom-set expansion

2.1 Pseudorandom-set expansion on Grassmann and Johnson graphs

A recent exciting breakthrough is proving the 2-to-2-games conjecture which is a milestone of attacking the unique games conjecture. [KMS18] completed the last missing piece of the proof by proving *Grassmann expansion hypothesis*. The reduction from Grassmann expansion hypothesis to 2-to-2-games conjecture is established along [DKK$^+$18, KMS17, BKS18]. We will not explain about the details of the reduction here, and concentrate more on the statement of Grassmann Expansion Hypothesis itself.

Grassmann expansion hypothesis says that pseudorandom sets in Grassmann graph have near-perfect expansion. Here is the definition of Grassmann graph and pseudorandom set on it:

Definition 2.1 (Grassmann graph)**.** *Grassmann graph* $\mathrm{Gr}_{n,k}$ is the graph on vertex set of all k-dimensional subspaces of vector space $\mathbb{F}_2^n$. There is an edge between subspaces H, H' if and only if $\dim(H \wedge H') = k-1$. For any subspaces $A \leq B \leq \mathbb{F}_2^n$, we define $\mathrm{Gr}_{n,k}[A, B]$ be the subset of vertices H satisfying $A \leq H \leq B$.

A subset of vertices $S \subseteq \mathrm{Gr}_{n,k}$ is called (r, ϵ)-*pseudorandom* if for any subspace $A \leq B \leq \mathbb{F}_2^n$ with $\dim(A) + \mathrm{codim}(B) \leq r$, we have

$$\frac{|S \cap \mathrm{Gr}_{n,k}[A, B]|}{|\mathrm{Gr}_{n,k}[A, B]|} \leq \epsilon.$$

A precise statement of Grasssmann expansion hypothesis appears below (proved as Theorem 1.8 in [KMS18]):

Theorem 2.2. *For every constant $0 < \eta < 1$, there exists a constant $\epsilon > 0$ and a non-negative integer r such that for large enough k and (after fixing k) sufficiently large n, the following holds. If subset of veritices $S \subseteq \mathrm{Gr}_{n,k}$ is (r, ϵ)-pseudorandom, then its edge expansion $\Phi(S) \geq 1 - \eta$.*

The proof of Grassmann expansion hypothesis uses Fourier analysis on a selected subset of Boolean cube (Cayley graph). This irregular setting makes it hard to figure out an exact orthogonal basis so the proof of Grassmann expansion hypothesis in [KMS18] uses a lot of approximations and is long and not easy to comprehend.

A similar *Johnson expansion hypothesis* is proved in [KMMS18] and the technical insight therein is similar to the proof technique of Grassmann expansion hypothesis in [KMS18]. Here is the definition of generalized Johnson graph and pseudorandom set on it.

Definition 2.3 (Johnson graph)**.** *Johnson graph* $J_{n,k,t}$ is the graph whose vertices are subsets of size k in $[n]$. There is an edge between vertices u, v if and only if the intersection of u and v is of size t. For any subset $R \subseteq [n]$, we define $J_{n,k,t}[R]$ be the subset of vertices u in $J_{n,k,t}$ satisfying $u \supseteq R$.

A subset of vertices $S \subseteq J_{n,k,t}$ is called (r, ϵ)-*pseudorandom* if for any subset $R \subseteq [n]$ of size at most r, we have

$$\left| \mathop{\mathbf{Pr}}_{u \supseteq R}[u \in S] - \mathop{\mathbf{Pr}}_{u}[u \in S] \right| \leq \epsilon.$$

If we represent a subset $[n]$ as a Boolean string of length n, then we can treat the vertices of Johnson graph $J_{n,k,t}$ as a slice of the Boolean cube of dimension n. Johnson graphs and slices of the Boolean hypercube are the subject of considerable interest recently, see [Fil14, FM16, FKMW18]. Johnson graphs and slices of the Boolean hypercube have been used in numerous applications: for the study of sharp thresholds of graph properties [FB99], for direct product tests [IKW12] and for a recent candidate hard unique game [KM16].

A precise statement of Johnson expansion hypothesis appears below (proved as Theorem 1.3 in [KMMS18]):

Theorem 2.4. *For every constant $0 < \alpha < 1$ and $0 < \eta < 1$, there exists a constant $\epsilon > 0$ and a non-negative integer r such that for large enough k and (after fixing k) sufficiently large n, the following holds. If subset of veritices $S \subseteq J_{n,k,\alpha k}$ is (r, ϵ)-pseudorandom, then its edge expansion $\Phi(S) \geq 1 - \eta$.*

The proof of Johnson graph hypothesis is slightly easier comparing to the proof of Grassmann graph hypothesis due to the easier structure of Johnson graph. However it is still hard to get the exact orthogonal basis so the proof still need a lot of approximation and complicated calculation.

2.2 Our current progress and goal

As mentioned above, Johnson graph can be seen as a slice of the Boolean hypercube. This connection leads us to consider using biased distribution on Boolean hypercube to simplify the proof.

Consider p-biased Fourier analysis on $\{-1,1\}^n$ with $p = k/n$. This distribution concentrates on Boolean strings with k bits of -1 which is similar to Johnson Graph $J_{n,k,t}$. Then we consider the noise operator with parameter $\rho = t/k$, which means that for each bit with probability ρ we keep it, and with probability $1-\rho$ we redraw the bit for p-biased distribution. This is an analog of edges in $J_{n,k,t}$ since for any Boolean strings with k bits of -1's, it is most likely that we keep t bits of -1's in the string choose other $k-t$ of -1's uniformly randomly among the rest bits.

We define a subset $A \subseteq \{-1,1\}^n$ be (r,ϵ)-*pseudorandom* if for any subcube $R \subseteq \{-1,1\}^n$ with codimension at most r,

$$\left| \mathop{\mathbf{Pr}}_{\boldsymbol{x}\in R}[\boldsymbol{x} \in A] - \mathop{\mathbf{Pr}}_{\boldsymbol{x}}[\boldsymbol{x} \in A] \right| \leq \epsilon.$$

We currently prove the following analog of Johnson graph hypothesis.

Theorem 2.5. *For every constant $0 < \rho < 1$ and $0 < \eta < 1$, there exists a constant $\epsilon > 0$ and a non-negative integer r such that for any p-biased distribution, the following holds. If subset of Boolean cube $A \subseteq \{-1,1\}^n$ is (r,ϵ)-pseudorandom, then*

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x}\sim\pi^{\otimes n}\\ \boldsymbol{y}\sim N_\rho(\boldsymbol{x})}}[\boldsymbol{y} \in A | \boldsymbol{x} \in A] \leq \eta.$$

Theorem 2.5 is proposed as an open problem in [KMMS18]. We prove this theorem and also remove the assumption that p should be tiny (k should be much smaller than n in Johnson graph). The proof of Theorem 2.5 is simple and clean comparing to the proof of Grassmann and Johnson graph hypothesis in [KMS18, KMMS18]. The proof uses some tricks of randomization/symmetrization of Boolean functions which is also used in the proof of Bourgain's Sharp Threshold Theorem in [FB99].

Our ultimate goal is to give an analog of Grassmann graph hypothesis. This may simplify the proof in [KMS18] and may give some inspirations on the further direction of attacking the unique games conjecture. One possible approach is to construct a q-analog of Fourier analysis on the Boolean hypercube and a q-analog of hypercontractivity, since Grassmann graph is a q-analog of Johnson graph with $q = 2$.

3 Communication distillation

3.1 Hypercontractivity and small-set expansion on general finite domain

In fact we can generalize our definition of hypercontractivity and small-set expansion to any joint finite probability space $((\mathcal{X},\mathcal{Y}),\mu)$, and its product $((\mathcal{X},\mathcal{Y})^n,\mu^{\otimes n})$. We say that $(\boldsymbol{X},\boldsymbol{Y})\sim\mu$ is (p,q)-hypercontractive if for any function $f:\mathcal{X}\to\mathbb{R}$ and $g:\mathcal{Y}\to\mathbb{R}$,

$$\mathop{\mathbf{E}}_{(\boldsymbol{X},\boldsymbol{Y})\sim\mu}[f(\boldsymbol{X})g(\boldsymbol{Y})]\le\|f\|_p\|g\|_{q'}.$$

In fact even with this generalized definition, Small-Set Expansion Theorem is not a weaker statement than Hypercontractivity Theorem. They are equivalent as well as similar statements in other measures, like KL divergence and mutual information.

Theorem 3.1. *The following statements are equivalent:*

1) $(\boldsymbol{X},\boldsymbol{Y})\sim\mu$ *is* (p,q)*-hypercontractive;*

2) $D(\nu\|\mu)\ge\frac{1}{p}D(\nu_X\|\mu_X)+\frac{1}{q'}D(\nu_Y\|\mu_Y)$ *for any distribution* ν *on space* $\mathcal{X}\times\mathcal{Y}$*;*

3) $\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x}\in A,\boldsymbol{y}\in B]\le|A|^{\frac{1}{p}}|B|^{\frac{1}{q'}}$ *for all* n*,* $A\subseteq\mathcal{X}^n$*,* $B\subseteq\mathcal{Y}^n$*;*

4) $I(\boldsymbol{U};\boldsymbol{X},\boldsymbol{Y})\ge\frac{1}{p}I(\boldsymbol{U};\boldsymbol{X})+\frac{1}{q'}I(\boldsymbol{U};\boldsymbol{Y})$ *for any random variable* $\boldsymbol{U}$*, where* $(\boldsymbol{X},\boldsymbol{Y})\sim\mu$*.*

Theorem 3.1 is first mentioned and proved in [Nai14]. They showed that 1), 2) and 4) are equivalent. They did not mention about small-set expansion explicitly, but the 3) is hidden in the proof to be equal to all other statements. We refined the proof and pointed out the equivalence of small-set expansion and hypercontractivity. As far as we know, [Nai14] is the very first study showing that small-set expansion is as strong as hypercontractivity.

If we fix subsets $A\subseteq\mathcal{X}^n$, $B\subseteq\mathcal{Y}^n$, and try to optimize the bound in Small-Set Expansion Theorem, we get the following corollary:

Corollary 3.2. *For any* n*, and subsets* $A\subseteq\mathcal{X}^n$*,* $B\subseteq\mathcal{Y}^n$*,*

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

This corollary is tight because of Theorem 3.1, the equivalence of small-set expansion and hypercontractivity. That is to say, for (p,q) not hypercontractive on μ, there exists some n, $A\subseteq\mathcal{X}^n$, $B\subseteq\mathcal{Y}^n$, such that

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

We propose a stronger conjecture that Small-Set Expansion Theorem is still tight even when we fix ratio $\frac{\log|B|}{\log|A|}$ to any constant.

Conjecture 3.3. *If we fix* $\frac{\log|B|}{\log|A|}$ *to be some constant* c*, then Corollary 3.2 is still tight—i.e.,*

$$\lim_{n\to\infty}\inf_{\frac{\log|B|}{\log|A|}=c}\left\{\frac{\log\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes n}}[\boldsymbol{x}\in A,\boldsymbol{y}\in B]}{\log|A|}\right\}=\sup_{(p,q)\text{-hypercontractive on }\mu}\left\{\frac{1}{p}+\frac{c}{q'}\right\};$$

Or the equivalent KL-divergence form is

$$\inf_{\frac{D(\nu_y\|\mu_y)}{D(\nu_x\|\mu_x)}=c}\left\{\frac{D(\nu\|\mu)}{D(\nu_x\|\mu_x)}\right\}=\sup_{(p,q)\text{-hypercontractive on }\mu}\left\{\frac{1}{p}+\frac{c}{q'}\right\}.$$

Conjecture 3.3 is directly related to communication-assisted agreement distillation mentioned in the next subsection.

3.2 Communication-assisted agreement distillation

Consider the following problem. Suppose Alice holds string $\boldsymbol{x} \in \mathcal{X}^n$ and Bob holds string $\boldsymbol{y} \in \mathcal{Y}^n$ where each bit pair $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is drawn from the joint distribution μ independently. Alice and Bob want to extract a common uniformly random string with length k. We aim to maximize the agreement probability with limited communication between Alice and Bob.

In [BM11], they studied this scenario with zero communication, for the motivation from the problem of extracting a unique identification string from process variations. Furthermore, this communication-free scenario with noisy Boolean strings $\boldsymbol{x}$ and $\boldsymbol{y}$, is also related to a widely interested information-theoretic conjecture proposed in [CK14]: the dictator function maximize mutual information on noisy Boolean inputs. The communication-assisted agreement version of this distillation problem is related to the question of communication with imperfectly shared randomness in [CGMS17].

In this work we care less about the length of string $\boldsymbol{x}$ and $\boldsymbol{y}$, assuming $n \to \infty$.

We focus on the trade-off between communication and success probability in this agreement distillation problem. I.e., what is the exact maximum agreement probability for any fixed number of bits in communication? In [GR16], they proved the following result.

Theorem 3.4 (Generalized version of Theorem 4.1, [GR16])**.** *Suppose Alice holds string $\boldsymbol{x}$ and Bob holds string $\boldsymbol{y}$ as in the above setting ($\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{y} \in \mathcal{Y}^n$ with sufficiently large length n, $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is drawn from joint distribution μ independently). Alice decides a uniformly distributed Boolean string $g_A(X) \in \{0,1\}^k$. Bob wants to agree on $g_A(X)$ after two-way communication with Alice. For any communication protocol exchanging ck bits, the maximum success probability will be at most $2^{-\gamma k}$ where*

$$\gamma = \sup_{(p,q)\text{-hypercontractive on } \mu} \left\{ \frac{1}{p} + \frac{1-c}{q'} - 1 \right\}.$$

In [GR16], they focused on binary symmetric channel and binary erasure channel distortion, but their proof works for general joint distribution. We also rephrase the proof in a more comprehensive way using Small-Set Expansion Theorem and Theorem 3.1.

On the other side, we want to construct a protocol to meet the upper bound in Theorem 3.4. Theorem 4.1 in [AC98] indicates a way to construct a protocol, which is also related to the expansion property.

Theorem 3.5 (Theorem 4.1 in [AC98])**.** *Suppose $A \subseteq \mathcal{X}^m$, $B \subseteq \mathcal{Y}^m$ satisfies $|B| = |A|^{1-c}$ and*

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mu^{\otimes m}}[\boldsymbol{y} \in B | \boldsymbol{x} \in A] = |A|^{\gamma},$$

Then we can construct a protocol such that Alice first proposes uniformly random Boolean string $g_A(\boldsymbol{x})$ of length k, and then sends ck bits to Bob. Bob can guess $g_A(\boldsymbol{x})$ successfully with probability at least $2^{-\gamma k}$.

Therefore if we wants to maximize the success probability using the protocol proposed in Theorem 3.5, it is equal to say that we want to maximize γ among all sets A, B satisfying $|B| = |A|^{1-c}$. This is exactly what states in Conjecture 3.3. Therefore Conjecuture 3.3 induces tightness of the bound in Theorem 3.4:

Corollary 3.6. *If Conjecture 3.3 holds, then Theorem 3.4 is tight. I.e., there exists an one-way communication protocol with ck bits, and the probability of Bob guessing $g_A(\boldsymbol{x})$ successfully is $2^{-\gamma k}$, for any*

$$\gamma > \sup_{(p,q)\text{-hypercontractive on } \mu} \left\{ \frac{1}{p} + \frac{1-c}{q'} - 1 \right\}.$$

3.3 Further discussion on communication distillation and hypercontractivity

The whole discussion above gives us a general scheme for communication distillation in any finite probability space $((\mathcal{X}, \mathcal{Y}), \mu)$. When we try to calculate the exact bound for some interesting special cases, like BEC or BSC channels focused in [GR16], it remains to calculate the quantity

$$\sup_{(p,q)\text{-hypercontractive on } \mu} \left\{ \frac{1}{p} + \frac{1-c}{q'} - 1 \right\}$$

for these specific probability spaces. The underlying general discussion is a fundamental question: In some specific probability spaces, like BEC and BSC channels, can we calculate the exact hypercontractivity domains?

For BEC channel, this is just the classical case that $\boldsymbol{x}$ and $\boldsymbol{y}$ are ρ-correlated Boolean strings and we know the exact hypercontractivity domain. For BSC channel, Nair and Wang shows partial result in [NW16].

Theorem 3.7 (Theorem 3, [NW16])**.** *Consider a uniform Boolean random variable $\boldsymbol{X}$ passed through a binary erasure channel* BEC(ϵ) *producing the ternary output $\boldsymbol{Y}$. If*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(q' - 1),$$

then $\boldsymbol{X}$ and $\boldsymbol{Y}$ is (p, q)-hypercontractive if and only if

$$1 - \epsilon \leq (p - 1)(q' - 1).$$

Exact computation of the hypercontractivity parameters has been a challenging task with very few exact characterizations. But if we only want to know how many bits in communication is needed to get the agreement distillation with high probability, then we do not really need to calculate the exact hypercontractivity domain.

Theorem 3.4 can be rephrased as following: if we want to achieve success probability $2^{-\gamma k}$ in the communication distillation problem, we needs to exchange at least ck bits in the protocol where

$$c \geq \sup_{(p,q)\text{-hypercontractive on } \mu} \left\{ 1 - q' \left(1 + \gamma - \frac{1}{p} \right) \right\}.$$

If we want to achieve $\Omega(1)$ probability for any k, then we need $\gamma \to 0$. Then want to calculate the minimum of $\frac{q'}{p'}$. If we define $q^*(p)$ as the infimum value of q such that μ is (p, q)-hypercontractive with fixed p. Then $\frac{q^*(p)'}{p'}$ is a monotone increasing in p as mentioned in Theorem 1, [AGKN13]. Therefore we only need to calculate

$$\lim_{p \to 1} \frac{q^*(p)'}{p'} = \lim_{p \to 1} \frac{p - 1}{q^*(p) - 1}.$$

.

We propose the following conjecture:

Conjecture 3.8. *Suppose*

$$q^*(p) = \inf_{(p,q)\text{-hypercontractive on } \mu} q.$$

If μ is the joint distribution that $\boldsymbol{Y}$ is a uniform random Boolean variable and $\boldsymbol{X}$ is the ternary output of $\boldsymbol{Y}$ passed through a binary erasure channel BEC(ϵ)*, then*

$$\lim_{p \to 1} \frac{p - 1}{q^*(p) - 1} = \log_{\frac{1-\epsilon}{2}} \frac{1}{2},$$

when $\epsilon > \frac{1}{2}$.

This gives the tight bound for communication distillation problem with high probability and reverse BEC channel.

4 Decoupling

4.1 Introduction

In this section, we focus on decoupling. 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. (A recent work of Makarychev and Sidirenko [MS14] provides an exception, though they use a much different kind of decoupling than the one studied in this section.) 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 the section, 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.

Let f denote a multilinear polynomial of degree at most k in n variables $x = (x_1, \dots, x_n)$, with coefficients a_S from a separable Banach space:

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

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

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

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 4.2. The *(fully) decoupled* version of f, which we denote by $\widetilde{f}$, is a multilinear polynomial over k *blocks* $x^{(1)}, \dots, x^{(k)}$ of n variables; each $x^{(i)}$ is $x^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$. 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,

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

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, AA15]. 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 section 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 poly(k) parameter losses, whereas full decoupling in general suffers exponential losses in k.

Remark 4.3. 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 $\mathbf{Var}[\widetilde{f}] = \frac{1}{(k-1)!}\mathbf{Var}[\breve{f}]$ for $f : \{\pm 1\}^n \to \mathbb{R}$.

4.1.1 Classical decoupling inequalities

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

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

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

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 4.5. *In the setting of Theorem 4.4, for all $t > 0$,*

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

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

4.1.2 DFKO theorems

A key theme in analysis of Boolean functions 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 [JOW12, 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 4.6. *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 $\mathbf{Pr}[|f(\boldsymbol{z})| > t] \geq \exp(-O(t^2 k^2 \log k))$.

Even though the Gaussian results in Corollary 4.6 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 4.6 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.[1]

In the next subsection 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.

4.2 Our results

We now state our new versions of Theorems 4.4, 4.5 which apply only to one-block decoupling, but that have *polynomial* dependence of C_k on k.

[1] 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.

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

H1: $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \sim \mathrm{N}(0,1)$ are standard Gaussians.

H2: $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are uniformly random ± 1 values.

H3: $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are uniformly random ± 1 values and f is homogeneous.

Theorem 4.7. *If $\Phi : \mathbb{R}^{\ge 0} \to \mathbb{R}^{\ge 0}$ is convex and nondecreasing, then*

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

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

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

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 4.8. The bound $C_k = O(k)$ under **H1** is best possible (assuming that $D_k \le \exp(O(k^2))$).

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).

4.2.1 Main application: tight versions of the DFKO theorems

One main application of Theorem 4.7 is the tight version of DFKO theorems. we first get an optimal version of the DFKO Inequality in the Gaussian setting.

Theorem 4.9. *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 $\mathbf{Var}[f(\boldsymbol{x})] \ge 1$. Then for $t \ge 1$ it holds that $\mathbf{Pr}[|f(\boldsymbol{x})| > t] \ge \exp(-O(t^2 k^2))$. Furthermore, if f is multilinear and homogeneous then the lower bound may be improved to $\exp(-O(t^2 k))$.*

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].

Corollary 4.10. *Theorem 4.9 holds when $\boldsymbol{x} \sim \{\pm 1\}^n$ is uniform and we additionally assume that $\mathbf{MaxInf}[f] \le \exp(-Ct^2 k^2)$, or just $\exp(-Ct^2 k)$ in the homogeneous case. Here C is a universal constant.*

Then we obtain the sharp DFKO Fourier Tail Theorem.

Corollary 4.11. *Suppose $f : \{\pm 1\}^n \to [-1, +1]$ is not an $(\epsilon, 2^{O(k^2/\epsilon)})$-junta. Then*

$$\sum_{|S| > k} \widehat{f}(S)^2 > \exp(-O(k^2)/\epsilon). \tag{1}$$

A similar (but easier) proof can be used to derive the following Gaussian version of Corollary 4.11; 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 4.12. *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 4.6.

4.2.2 Other applications

In a recent work comparing quantum query complexity to classical randomized query complexity, Aaronson and Ambainis [AA15] proved the following:

Theorem 4.13. *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 4.14. *Aaronson and Ambainis's Theorem 4.13 holds without the assumption of block-multilinearity or homogeneity.*

Another application of one-block decoupling is also on a conjecture proposed by Aaronson and Ambainis. 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 $\mathbf{MaxInf}_i[f] \geq \mathrm{poly}(\mathbf{Var}[f]/k)$.*

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].

We use our results to show that the assumption that f is one-block decoupled is completely without loss of generality.

Theorem 4.15. *The AA Conjecture holds if and only if it holds for one-block decoupled functions f.*

We also show that the best known result towards the conjecture can be proven extremely easily for one-block-decoupled functions.

5 Property testing on k-wise uniformity

5.1 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]. They 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. A 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 non-empty set $S \subseteq [n]$ with size at most k,

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

The original paper about almost k-wise uniformity is [NN93], which is concerned with derandomization; e.g., they use (ϵ, k)-wise uniformity for derandomizing the "set balancing (discrepancy)" problem. Alon et al. give 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 $\mathrm{poly}(2^k, \log n, 1/\epsilon)$, 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 also work as well with an (ϵ, k)-wise uniform distribution, when the parameter ϵ is small enough.

5.1.1 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?

We will refer to this question as *the Closeness Problem*.

On one hand, the main message of [AGM03] showed 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 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].

The notions of pairwise-uniformity, k-wise uniformity, and δ-closeness to k-wise uniformity are also important for hardness of constraint satisfactory problems (CSPs). Austrin and Mossel [AM09] shows 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 discussed in [AOW15] on hardness of random CSP.

Alon et al. [AAK+07] investigated the Closeness Problem further by improving the upper bound to $O((n \log n)^{k/2} \epsilon)$. 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 $\mathbf{W}^{1\dots k}[\varphi] = \sum_{1 \leq |S| \leq k} \widehat{\varphi}(S)^2$. Rubinfeld and Xie [RX13] generalized some of these results to non-uniform k-wise independent distributions over larger product spaces.

Though Alon et al. [AAK+07] did not mention it explicitly, they also give a lower bound for the Closeness Problem of $\delta \geq \Omega\Big(\frac{n^{(k-1)/2}}{\log n} \epsilon\Big)$ 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 $k = 2$.

5.1.2 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?

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 testing related properties of distribution, including uniformity [GR11, BFR$^+$00, RS09] and independence [BFF$^+$01, BKR04, ADK15, DK16].

[AGM03, AAK$^+$07, Xie12] discussed about 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] show 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.

5.2 Our results

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. *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 5.2. *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*

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

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 5.2 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 5.3. *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*

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

We show a better upper bound for sample complexity for the Testing Problem:

Theorem 5.4. *There exists a δ-tester for k-wise uniformity of distributions on $\{-1,1\}^n$ with sample complexity $O\left(\frac{1}{k}\right)^{k/2}\frac{n^k}{\delta^2}$. For the special case of $k = 1$, the sample complexity is $O\left(\frac{\log n}{\delta^2}\right)$.*

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 $O\left(\frac{n^k \log n}{\delta^2}\right)$ 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, $\sum_{1\leq|S|\leq k}\widehat{\varphi}^2(S)$, 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 5.5. *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 $k \geq 2$, indicating that the upper bound might be able to be improved. Both the lower bound in our result 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 5.6. *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 high probability ($> \frac{2}{3}$) with sample complexity $O(k)^k \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\log \frac{n}{\delta}\right)^{k/2}$.*

In fact, for testing whether a distribution is αk-wise uniform or δ-far from k-wise uniform with $\alpha > 4$ (assuming αk is an even integer), there exists an algorithm with high probability ($> \frac{2}{3}$) 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 fully 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 5.4, 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 logarithm factor for the pairwise case, and is tight up to a factor of $\tilde{O}(\sqrt{n})$ when $k > 2$.

We compare our result and previous best known bounds from [AAK$^+$07] in Table 1. (We omit 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 1: Summary of our results

References

[AA14] Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *Theory Of Computing*, 10(6):133–166, 2014.

[AA15] Scott Aaronson and Andris Ambainis. Forrelation: a problem that optimally separates quantum from classical computing. pages 307–316, 2015.

[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.

[Aar05] Scott Aaronson. Ten semi-grand challenges for quantum computing theory, 2005. `http://www.scottaaronson.com/writings/qchallenge.html`.

[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`.

[Aar10] Scott Aaronson. Updated version of "ten semi-grand challenges for quantum computing theory", 2010. `http://www.scottaaronson.com/blog/?p=471`.

[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.

[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.

[ADK15] Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Advances in Neural Information Processing Systems*, pages 3591–3599, 2015.

[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.

[AGKN13] Venkat Anantharam, Amin Gohari, Sudeep Kamath, and Chandra Nair. On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover. *arXiv preprint arXiv:1304.6133*, 2013.

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003.

[AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.

[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.

[Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009.

[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.

[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.

[BKR04] Tuğkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 381–390, 2004.

[BKS18] Boaz Barak, Pravesh K Kothari, and David Steurer. Small-set expansion in shortcode graph and the 2-to-2 conjecture. *arXiv preprint arXiv:1804.08662*, 2018.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[BM11] Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *IEEE Transactions on information theory*, 57(10):6351–6355, 2011.

[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. 2015.

[Bon68] Aline Bonami. Ensembles $\lambda(p)$ dans le dual de d^infty. In *Annales de l'institut Fourier*, volume 18, pages 193–204, 1968.

[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.

[Bou02] Jean Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel Journal of Mathematics*, 131(1):269–276, 2002.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the ACM*, 57(5):28:1–28:10, 2010.

[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.

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[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.

[CK14] Thomas A Courtade and Gowtham R Kumar. Which boolean functions maximize mutual information on noisy inputs? *IEEE Transactions on Information Theory*, 60(8):4515–4525, 2014.

[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.

[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.

[Din07] Irit Dinur. The PCP Theorem by gap amplification. *Journal of the ACM*, 54(3):1–44, 2007.

[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.

[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.

[dlP92] Victor de la Peña. Decoupling and Khintchine's inequalities for U-statistics. *Annals of Probability*, 20(4):1877–1892, 1992.

[dlPG99] Víctor de la Peña and Evarist Giné. *Decoupling: from dependence to independence*. Springer, 1999.

[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.

[DS05] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of mathematics*, pages 439–485, 2005.

[EFF12] David Ellis, Yuval Filmus, and Ehud Friedgut. Triangle-intersecting families of graphs. *Journal of the European Mathematical Society*, 14(3):841–885, 2012.

[FB99] Ehud Friedgut and Jean Bourgain. Sharp thresholds of graph properties, and the -sat problem. *Journal of the American mathematical Society*, 12(4):1017–1054, 1999.

[Fil14] Yuval Filmus. Friedgut–kalai–naor theorem for slices of the boolean cube. *arXiv preprint arXiv:1410.7834*, 2014.

[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.

[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.

[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.

[FM16] Yuval Filmus and Elchanan Mossel. Harmonicity and invariance on slices of the boolean cube. *Probability Theory and Related Fields*, pages 1–62, 2016.

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–36, 1998.

[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.

[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.

[GR16] Venkatesan Guruswami and Jaikumar Radhakrishnan. Tight bounds for communication-assisted agreement distillation. In *31st Conference on Computational Complexity (CCC 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

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

[IKW12] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query pcps. *SIAM Journal on Computing*, 41(6):1722–1768, 2012.

[JOW12] Jacek Jendrej, Krzysztof Oleszkiewicz, and Jakub Wojtaszczyk. On some extensions of the FKN theorem. Manuscript, 2012. To appear in *Theory of Computation*.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. pages 767–775, 2002.

[Kin02] Guy Kindler. *Property Testing, PCP, and juntas*. PhD thesis, Tel Aviv University, 2002.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science*, pages 68–80. IEEE, 1988.

[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.

[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.

[KM16] Subhash Khot and Dana Moshkovitz. Candidate hard unique game. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 63–76. ACM, 2016.

[KMMS18] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Pseudorandom sets in johnson graph have near-perfect expansion. *ECCC Report TR18-078*, 2018.

[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.

[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.

[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.

[KN06] Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. *Mathematische Annalen*, 334(4):821–852, 2006.

[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.

[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.

[KS02] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas. Manuscript, 2002.

[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.

[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.

[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.

[Lov10] Shachar Lovett. An elementary proof of anti-concentration of polynomials in Gaussian variables. Technical Report 182, Electronic Colloquium on Computational Complexity, 2010.

[Lub86] Michael Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986.

[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.

[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.

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*. Elsevier, 1977.

[MS14] Konstantin Makarychev and Maxim Sviridenko. Solving optimization problems with diseconomies of scale via decoupling. pages 571–580, 2014.

[Nai14] Chandra Nair. Equivalent formulations of hypercontractivity using information measures. *Proceedings of International Zurich Seminar on Communications*, 2014.

[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.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

[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.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[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.

[Rao47] Calyampudi Radhakrishna Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society*, 9(1):128–139, 1947.

[RS09] Ronitt Rubinfeld and Rocco A. Servedio. Testing monotone high-dimensional distributions. *Random Structures & Algorithms*, 34(1):24–44, 2009.

[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.

[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.

[Xie12] Ning Xie. *Testing k-wise independent distributions*. PhD thesis, Massachusetts Institute of Technology, 2012.