TOWARDS UNDERSTANDING THE APPROXIMATION OF BOOLEAN FUNCTIONS BY NONCLASSICAL POLYNOMIALS

by

ABHISHEK BHRUSHUNDI

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Swastik Kopparty

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2020

ABSTRACT OF THE DISSERTATION

Towards understanding the approximation of Boolean functions by nonclassical polynomials

By ABHISHEK BHRUSHUNDI

Dissertation Director:

Swastik Kopparty

The representation and approximation of Boolean functions by polynomials is an important area of research in theoretical computer science, having numerous applications in circuit complexity, communication complexity, pseudorandomness, quantum computation, learning theory, algorithm design, and explicit combinatorial constructions. Results of Green and Tao [GT09], Lovett et al. [LMS11], and Tao and Ziegler [TZ12], on the *Inverse Conjecture for the Gowers norm* over finite fields of low characteristic, and the subsequent work of Bhowmick and Lovett [BL15], suggest that a potential barrier to the resolution of some of the outstanding open problems in this area is the class of *nonclassical polynomials* and its ability to nontrivially represent and approximate Boolean functions.

Motivated by these works, in this dissertation, we investigate the ability of nonclassical polynomials to approximate Boolean functions with respect to both previously studied and new notions of approximation:

- We introduce and study an agreement-based notion of approximation by polynomials over $\mathbb{Z}/2^k\mathbb{Z}$. Investigating this notion serves as a proxy for understanding the maximum possible agreement between nonclassical polynomials and Boolean functions. We prove several new results that shed light on this new notion of approximation, and these results help us answer some questions left open in the work of Bhowmick and Lovett [BL15] concerning the approximation of Boolean functions by nonclassical polynomials in the agreement sense.

- We propose a new notion of point-wise approximation by nonclassical polynomials. Using a result of Green et al. [GKT92], which itself is an extension of the classic work of Beigel and Tarui [BT91], we observe that Boolean functions computable by ACC^0 circuits (constant-depth circuits of polynomial size, containing AND, OR, NOT, and MOD_q gates) are amenable to point-wise approximation by low-degree nonclassical polynomials. Motivated by this new observation, we then explore how well can low-degree nonclassical polynomials point-wise approximate the majority function, in the hope of resolving the long-standing open problem of proving that majority is not computable by ACC^0 circuits.

Our results suggest several interesting and promising directions of research. We explore some of these directions and state concrete open problems along with plausible approaches to solving them.

Acknowledgements

I would like to thank my advisor Swastik Kopparty for his guidance and patience, and for introducing me to the problem of proving correlation bounds for polynomials through Viola's survey on the topic [Vio09]. I am grateful to Swastik for encouraging me during my early years in the program to pursue my research interests and to collaborate with others.

I am grateful to Prahladh Harsha, Shachar Lovett, and Srikanth Srinivasan for hosting and mentoring me during research visits. Most of the research in this dissertation is a direct result of their mentoring and guidance. I would also like to thank my collaborators Kaave Hosseini and Sankeerth Rao.

Special thanks to Eric Allender, Suryateja Gavva, Sivakanth Gopi, and Hamed Hatami for their comments and for suggesting corrections.

Finally, I thank my wife Tara for being patient and supportive through the writing of this dissertation.

Published work

This dissertation is based on the original work of the dissertator that appears in [BHS17] and [BHLR19].

Dedication

To my wife Tara, my parents, and my sister Madhura

Table of Contents

Chapter 1
Introduction

The representation and approximation of Boolean functions by polynomials is a fairly well-studied topic in theoretical computer science, having applications in several areas such as explicit combinatorial constructions (e.g., [Gro00, Gop14]), learning theory (e.g., [LMN93, KS04]), design of algorithms for combinatorial problems (e.g., [Wil14a, AWY15]), circuit complexity (e.g., [Raz87, Smo87, AB01]), quantum computation (e.g., [BNRW05, BKT18]), pseudorandomess (e.g., [BV10, CHLT19]), communication complexity (e.g., [BW01, She12]), etc.

In this dissertation, we focus on the approximation of Boolean functions by *nonclassical polynomials*, an extension of standard polynomials in higher order Fourier analysis introduced by Tao and Ziegler [TZ12], with respect to three different notions of approximation: agreement-based, correlation-based, and point-wise approximation. Before stating the contributions of this dissertation, we discuss these notions of approximation along with the notion of nonclassical polynomials.

1.1 Three notions of approximation

1.1.1 Agreement-based approximation

For functions $F, G : D \to R$ for some finite domain D and range R, the *agreement* between F and G, denoted by $\mathrm{agr}(F, G)$, is defined to be the fraction of inputs where F and G agree (or take the same value), i.e.,

$$\mathrm{agr}(F, G) = \Pr_{x \sim D}[F(x) = G(x)],$$

where $x \sim D$ means that x is sampled uniformly at random from D.

We say that F *ϵ-approximates G in the agreement sense* if $\mathrm{agr}(F, G) \geq \epsilon$, for some $\epsilon > 0$.

To study the agreement between a Boolean function $F : \{0,1\}^n \to \{0,1\}$ and functions of the form $P : \{0,1\}^n \to R$, where R is an arbitrary set containing at least two elements, we define an R-valued version of F as follows. Let $r_0, r_1 \in R$ be two distinct elements in the range R. We use r_0 and r_1 to represent the elements 0 and 1 from the set $\{0,1\}$. In particular, we define the function $F' : \{0,1\}^n \to R$ defined as

$$F'(x) = \begin{cases} r_0 & \text{if } F(x) = 0 \\ r_1 & \text{if } F(x) = 1, \end{cases}$$

for all $x \in \{0,1\}^n$. F' is an R-valued version of the function F.

Let us assume that the choice of $r_0, r_1 \in R$ has been fixed. Then we say that the function $P : \{0,1\}^n \to R$ *ϵ-approximates the Boolean function* $F : \{0,1\}^n \to \{0,1\}$ if $\mathrm{agr}(F', P) \geq \epsilon$, where F' is the R-valued version of F defined above. For the sake of convenience, we often refer to $\mathrm{agr}(F', P)$ as the agreement between F and P.

We remark that it is often the case that the range R contains the elements 0 and 1. In this case, it makes sense to choose $r_0 = 0, r_1 = 1$.

1.1.2 Correlation-based approximation

Let $F, G : D \to \mathbb{C}$ be two complex-valued functions defined over a finite domain D. Then the *correlation* between F and G, denoted by $\mathrm{Corr}(F, G)$, is defined as

$$\mathrm{Corr}(F, G) = \left| \mathbb{E}_{x \in D} \left[F(x) \cdot \overline{G(x)} \right] \right|. \tag{1.1.1}$$

We remark that $\mathrm{Corr}(F, G)$ is essentially the *normalized inner product* of F and G if we view them as vectors in $\mathbb{C}^{|D|}$. For $\epsilon > 0$, we say that F *ϵ-approximates G in the correlation sense* if $\mathrm{Corr}(F, G) \geq \epsilon$.

Let R be either the nontrivial finite abelian group $\mathbb{Z}/m\mathbb{Z}$ (for an integer $m > 1$), or the infinite abelian group $\mathbb{R}/\mathbb{Z}$, and let $\exp(\cdot)$ denote the exponential function. As

in the case of agreement-based approximation, we choose and fix $r_0, r_1 \in R$ so that we can consider R-valued versions of Boolean functions. Then, for a Boolean function $F : \{0,1\}^n \to \{0,1\}$ and a function $P : \{0,1\}^n \to R$, we say P *ϵ-approximates F in the correlation sense* if

$$\mathrm{Corr}\left(\phi \circ F', \phi \circ P\right) \geq \epsilon,$$

where F' denotes the R-valued version of F defined earlier, and $\phi : R \to \mathbb{C}^*$ is the character $\phi : \mathbb{Z}/m\mathbb{Z} \to \mathbb{C}^*$ defined as $\phi(x) = \omega^x$ (ω is the m-th root of unity $\exp\left(2\pi i/m\right)$), if $R = \mathbb{Z}/m\mathbb{Z}$, and the character $\phi : \mathbb{R}/\mathbb{Z} \to \mathbb{C}^*$ defined as $\phi(x) = \exp(2\pi i x)$, if $R = \mathbb{R}/\mathbb{Z}$.

Abusing notation, for $F, G : D \to R$, by *correlation between F and G*, denoted by $\mathrm{Corr}(F, G)$, we mean the correlation between $\phi \circ F$ and $\phi \circ G$:

$$\mathrm{Corr}(F, G) = \mathrm{Corr}(\phi \circ F, \phi \circ G).$$

It can be verified that, for $F, G : D \to R$,

$$\mathrm{Corr}(F, G) = \mathrm{Corr}(\phi \circ F, \phi \circ G) \geq \left|2 \cdot \mathrm{agr}(F, G) - 1\right|.$$

When $R = \mathbb{Z}/2\mathbb{Z}$, the inequality becomes an equality, i.e.,

$$\mathrm{Corr}(F, G) = \mathrm{Corr}(\phi \circ F, \phi \circ G) = \left|2 \cdot \mathrm{agr}(F, G) - 1\right|.$$

Thus, nontrivial approximability in the agreement-sense implies nontrivial approximability in the correlation-sense. We remark that, barring the case of $R = \mathbb{Z}/2\mathbb{Z}$, the converse is not necessarily true.

1.1.3 Point-wise approximation

Let R be a metric space with the metric $d_R : R \times R \to \mathbb{R}$ defined on it, and let $F, G : D \to R$ be two functions defined on a finite domain D. We say that F *ϵ-approximates G in the point-wise sense* if for all $x \in D$,

$$d_R\left(F(x), G(x)\right) \leq \epsilon.$$

Again, as before, in order to study the approximation of a Boolean function $F : \{0,1\}^n \to \{0,1\}$ by a function $P : \{0,1\}^n \to R$, we consider the R-valued version $F' : \{0,1\}^n \to R$ defined as above, with respect to a fixed choice of $r_0, r_1 \in R$. In a similar spirit as before, we say that the function P *ϵ-approximates the Boolean function F in the point-wise sense* if for all $x \in \{0,1\}^n$,

$$d_R(F'(x), P(x)) \leq \epsilon.$$

Typically, point-wise approximation isn't directly related to agreement-based approximation. However, in certain cases, (e.g., Section 3.2.4), it can be related to correlation-based approximation.

1.2 Nonclassical polynomials

We remark that, although the following discussion concerns correlation between Boolean functions and polynomials over $\mathbb{F}_2$, most of the statements and results mentioned below hold true even for polynomials over arbitrary finite fields $\mathbb{F}_p$, for a prime p. Similarly the notion of nonclassical polynomials can be defined over $\mathbb{F}_p$ (see [TZ12]). We restrict ourselves to the $\mathbb{F}_2$ case for the sake of simplicity.

1.2.1 The problem of proving correlation bounds

A long-standing open problem in the area of polynomial-based representations and approximations of Boolean functions is that of proving correlation bounds against polynomials over $\mathbb{F}_2$. In particular, the problem asks for the following: for $d \leq \alpha \cdot n$, where $\alpha > 0$ is an absolute constant, find an explicit[1] Boolean function F in n variables that cannot be ϵ-approximated in the correlation sense by degree d polynomials in $\mathbb{F}_2[x_1, \ldots, x_n]$ for $\epsilon = 2^{-c \cdot n}$, where $c > 0$ is an absolute constant. In other words, the explicit function F is such that for any degree d polynomial P over

[1]By explicit, we mean a function that belongs to a complexity class such as P or NP.

$\mathbb{F}_2$,

$$\mathrm{Corr}(F, P) < 2^{-c \cdot n}.$$

If we restrict to $d \leq \beta \cdot n$ for $\beta \approx 10^{-4}$, it can be shown using a counting argument that most Boolean functions cannot even be $2^{-\delta \cdot n}$-approximated by degree d polynomials over $\mathbb{F}_2$ for $\delta \approx 1/2$. The challenge of course is to find an explicit function with this property. The survey by Viola [Vio09] is an excellent source for more information on the problem and its applications.

1.2.2 The Gowers norm

An important tool used in studying the correlation between a function and low-degree polynomials is the *Gowers norm*, introduced in the work of Gowers [Gow01]. For the sake of convenience, let us think of Boolean functions as both $\mathbb{F}_2$-valued and $\mathbb{R}/\mathbb{Z}$-valued functions, by mapping their codomain $\mathbb{F}_2$ to the set $\{0, 1/2\} \subset \mathbb{R}/\mathbb{Z}$ via the group homomorphism from $(\mathbb{F}_2, +)$ to $(\mathbb{R}/\mathbb{Z}, +)$. For $h \in \mathbb{F}_2^n$ and a function $F : \mathbb{F}_2^n \to \mathbb{R}/\mathbb{Z}$, define the *derivative of F in the direction h*, denoted by $D_h F$, as the function

$$D_h F(x) := F(x + h) - F(x),$$

and, for k vectors $h_1, \ldots, h_k \in \mathbb{F}_2^n$, define the *derivative of F in the directions $h_1, \ldots, h_k$*, denoted by $D_{h_1,\ldots,h_k} F$, to be the function

$$D_{h_1,\ldots,h_k} F(x) := D_{h_k} \left(D_{h_1,\ldots,h_{k-1}} F(x) \right).$$

Then the d-th Gowers norm of F, denoted by $\|F\|_{U^d}$, is defined as

$$\|F\|_{U^d} := \left(\mathbb{E}_{x,h_1,\ldots,h_d \sim \mathbb{F}_2^n} \left[\phi \left(D_{h_1,\ldots,h_d} F(x) \right) \right] \right)^{1/2^d},$$

where $\phi : \mathbb{R}/\mathbb{Z} \to \mathbb{C}^*$ is the character $\phi(x) = \exp(2\pi i x)$ of $\mathbb{R}/\mathbb{Z}$.

It can be observed that a Boolean function $F : \mathbb{F}_2^n \to \{0, 1/2\}$ can be represented as a polynomial over $\mathbb{F}_2$ of degree at most d if and only if for all $h_1, \ldots, h_{d+1} \in \mathbb{F}_2^n$,

$$D_{h_1,\ldots,h_{d+1}} F \equiv 0.$$

Equivalently, we can rewrite this as

$$\mathrm{Corr}(F, d) = 1 \Leftrightarrow \|F\|_{U^{d+1}} = 1,$$

where $\mathrm{Corr}(F, d)$ is the maximum possible correlation between F and degree d polynomials over $\mathbb{F}_2$. Implicit in the work of Gowers [Gow01] and Green and Tao [GT08], is the following "robust" version of the forward direction of the above statement: for all $F : \mathbb{F}_2^n \to \mathbb{R}/\mathbb{Z}$ and $d > 0$,

$$\mathrm{Corr}(F, d) > \epsilon \implies \|F\|_{U^{d+1}} > \epsilon.$$

This can be used to upper-bound the correlation between F and degree d polynomials over $\mathbb{F}_2$ by upper-bounding the $(d+1)$-th Gowers norm of F, and the work of Viola and Wigderson [VW08] employs this idea to show the existence of explicit functions in P that cannot be ϵ-approximated by polynomials over $\mathbb{F}_2$ of degree d, for some $\epsilon = \exp(-\Omega(n/2^d))$.

1.2.3 Inverse Conjecture for the Gowers Norm

A natural question to ask is whether there is a robust version of the reverse direction, i.e., if $\|F\|_{U^{d+1}} > \epsilon$ for a Boolean function F, does it mean that $\mathrm{Corr}(F, d) > \epsilon$? If true, this would mean that the Gowers norm of a Boolean function F precisely captures the correlation between F and degree d polynomials over $\mathbb{F}_2$. This problem was formalized as the *Inverse Conjecture for the Gowers Norm (ICGN)* in the works of Samorodnitsky [Sam07] and Green and Tao [GT08].

Unfortunately, the conjecture was disproved by Green and Tao in a subsequent work [GT09], and independently, by Lovett et al. [LMS11]. In particular, both the works give a counterexample to the ICGN for $d = 4$; they show that the Boolean function S_4, the elementary symmetric polynomial over $\mathbb{F}_2$ of degree 4, satisfies $\|S_4\|_{U^4} = \Omega(1)$, but for every polynomial P over $\mathbb{F}_2$ of degree at most 3, $\mathrm{Corr}(S_4, P) = o(1)$, due to a result of Alon and Beigel [AB01]. In fact, Green and Tao [GT09] show that

S_{2^ℓ}, the elementary symmetric polynomial of degree 2^ℓ, is a counterexample to the ICGN for all $\ell \geq 2$.

1.2.4 The work of Tao and Ziegler

In their work, Tao and Ziegler [TZ12] explain why the ICGN (as stated above) is false, and prove a modified version of the conjecture. To do this, they introduce the notion of *nonclassical polynomials*.

Recall that if, for a Boolean function $F : \mathbb{F}_2^n \to \{0, 1/2\}$,

$$D_{h_1,\ldots,h_{d+1}} F \equiv 0 \tag{1.2.1}$$

for all $h_1, \ldots, h_{d+1} \in \mathbb{F}_2^n$, then F can be represented as a polynomial of degree at most d over $\mathbb{F}_2$. We refer to such functions and their polynomial representations as *classical polynomials* of degree at most d.

Since the derivative operator is defined for all $\mathbb{R}/\mathbb{Z}$-valued functions defined on $\mathbb{F}_2^n$, we can consider arbitrary functions $F : \mathbb{F}_2^n \to \mathbb{R}/\mathbb{Z}$ that satisfy Eq. (1.2.1). Tao and Ziegler show that the set of such functions is a strict superset of the set of the aforementioned classical polynomials of degree d, and refer to such functions as *nonclassical polynomials of degree d*. They also prove that, structurally, nonclassical polynomials of degree d can be thought of as real polynomials of degree d evaluated modulo 1, having the following structure[2]:

$$P(x_1, \ldots, x_n) = \left(\alpha + \sum_{S \subseteq [n]; k \geq 0 : |S| + k \leq d} \frac{c_{S,k}}{2^{k+1}} \prod_{i \in S} x_i \right) \mod 1,$$

where $\alpha \in [0, 1)$ and $c_{S,k} \in \{0, 1\}$. Conversely, they show that every function that can be written in the above form is a nonclassical polynomial of degree d, i.e., it satisfies Eq. (1.2.1).

Tao and Ziegler aruge that the reason ICGN is false is that the $(d+1)$-th Gowers norm of a Boolean function F captures the correlation between F and *nonclassical*

[2]Below, we identify $\mathbb{F}_2$ with the set $\{0, 1\} \subset \mathbb{R}$ without explicitly using any inclusion maps.

polynomials of degree d, and not just classical polynomials. In other words, their results imply that the following modified version of the ICGN is true: if $\|F\|_{U^{d+1}}$ is "large" for some Boolean function $F : \mathbb{F}_2^n \to \{0, 1/2\}$, then there is a nonclassical polynomial $P : \mathbb{F}_2^n \to \mathbb{R}/\mathbb{Z}$ of degree at most d that has "large" correlation with F, or equivalently, there is a nonclassical polynomial of degree d that approximates F well in the correlation sense.

We remark that the counterexamples to the ICGN given by Green and Tao [GT09] and Lovett et al. [LMS11] (e.g., S_4) were subsequently shown to have nontrivial correlation with specific nonclassical polynomials.

1.2.5 Nonclassical polynomials as a barrier

The discussion in the previous section suggests that nonclassical polynomials are, in some sense, a barrier to studying the correlation between explicit Boolean functions and low-degree classical polynomials using the Gowers norm because of their ability to approximate in the correlation sense Boolean functions that are "hard" for classical polynomials. The work of Bhowmick and Lovett [BL15] provides concrete examples of such functions.

In fact, Bhowmick and Lovett suggest that nonclassical polynomials are a barrier to a more general set of techniques for proving correlation bounds that are based on iterative schemes involving squaring and the Cauchy-Schwarz inequality (see, e.g., [Vio09, Section 2.2] and [VW08]). They refer to such techniques as "derivative-based" techniques, and argue that, since the derivative operator is at the heart of these techniques, they should generalize to nonclassical polynomials, which would then mean that such techniques cannot "separate" classical polynomials from the nonclassical ones.

To illustrate this more clearly, consider any explicit Boolean function F that is believed to have correlation $o(1)$ with classical polynomials of degree d, but has correlation $\Omega(1)$ with a nonclassical polynomial of the same degree (e.g, the MOD_3 function).

Then, any "derivative-based" technique that proves an upper bound of γ on the correlation between F and classical polynomials of degree d, must, presumably, imply the same upper bound on the correlation between F and nonclassical polynomials of the same degree, which would mean that γ must be $\Omega(1)$.

In their work, Bhowmick and Lovett [BL15] also give examples of other problems, such as the problem of constructing *weak representations* of the OR function using polynomials over $\mathbb{Z}/6\mathbb{Z}$, where the same phenomenon occurs because of the ability of nonclassical polynomials to nontrivially represent or approximate "classically hard" Boolean functions.

1.3 Contributions of this dissertation

Motivated by the results discussed in the previous section, we continue the study of nonclassical polynomials and their ability to approximate Boolean functions. We go beyond the paradigm of correlation-based approximation and focus on agreement-based and point-wise approximation by nonclassical polynomials. We now give a high-level overview of our main results.

1.3.1 Agreement-based approximation by nonclassical polynomials

The work of Bhowmick and Lovett [BL15] mentioned above initiates the study of agreement between nonclassical polynomials and Boolean functions. In particular, they focus on the problem of approximating the majority function on n bits, Maj_n, in the agreement sense by nonclassical polynomials. They conjecture that, when it comes to approximating the majority function, nonclassical polynomials should not be any more powerful than classical polynomials. In fact, they also conjecture that this should be true when considering the approximation of any arbitrary Boolean function, and not just the majority function.

In Chapter 2, we show that their first conjecture (concerning the majority function) is, in fact, true, by showing that the agreement between Maj_n and any nonclassical polynomial P of degree d is bounded from above as follows,

$$\mathrm{agr}(\mathrm{Maj}_n, P) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}.$$

Juxtaposing this bound with those of Szegedy [Sze89] and Smolensky [Smo93] on the agreement between Maj_n and classical polynomials of degree d confirms their first conjecture.

We, however, disprove their second conjecture for general Boolean functions by showing examples of functions that can be nontrivially approximated by nonclassical polynomials in the agreement sense but are inapproximable by classical polynomials of the same degree.

To compute the maximum possible agreement between Boolean functions and nonclassical polynomials and prove the above results, we study the agreement between Boolean functions and polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$. In particular, we study the quantity $\gamma_{d,k}(F)$, the maximum possible agreement between a Boolean function F and degree d polynomials over $\mathbb{Z}/2^k\mathbb{Z}$, and prove several interesting properties of $\gamma_{d,k}(F)$ as a function of d and k.

We remark that the material and results in Chapter 2 are based on the original work of the author of this dissertation that appears in [BHS17].

1.3.2 Point-wise approximation by nonclassical polynomials

In Chapter 3, we introduce a notion of point-wise approximation by nonclassical polynomials by defining a metric on $\mathbb{R}/\mathbb{Z}$. The main motivation for the results in this chapter comes from our observation that a result from the work of Green et al. [GKT92] implies that any function computable by ACC^0 circuits[3] can be approximated by a nonclassical polynomial of degree $\mathrm{polylog}(n)$ with respect to the notion

[3]Circuits of polynomial size and constant depth, containing AND, OR, NOT, and MOD_m gates, where a MOD_m gate outputs 1 if and only if the number of ones in its inputs is non-zero modulo m.

of point-wise approximation introduced by us.

This observation naturally leads to the problem of showing that an explicit Boolean function is inapproximable by low-degree nonclassical polynomials in the point-wise sense, in the hope of proving an ACC^0 lower bound for the function. We study the point-wise approximability of the majority function, Maj_n, and show that low-degree symmetric torus polynomials cannot approximate Maj_n. While this does not prove $\mathrm{Maj}_n \notin \mathrm{ACC}^0$, we believe that it is an important first step towards achieving such a separation.

We prove several other upper and lower bound results demonstrating both the power and limitations of nonclassical polynomials, when it comes to approximating Boolean functions in the point-wise sense. Our results open up several promising directions of research, and we discuss them along with plausible approaches to attacking them in Chapter 4.

The material and results in Chapters 3 and 4 are based on the original work of the author of this dissertation that appears in [BHLR19].

Chapter 2

Agreement-based approximation by polynomials over $\mathbb{Z}/2^k\mathbb{Z}$

2.1 Introduction

In this chapter, we introduce and study a notion of agreement-based approximation of Boolean functions by polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$. We motivate our setup by recalling Razborov's [Raz87] result on approximating Boolean functions in the agreement sense by polynomials over $\mathbb{Z}/2\mathbb{Z}$. Given a Boolean function $F : \{0,1\}^n \to \{0,1\}$ and degree $d \leq n$, Razborov considers the largest γ such that there is a polynomial $Q \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ of degree d that has agreement γ with F, i.e.,

$$\Pr_{x \sim \{0,1\}^n}[Q(x) = F(x)] = \gamma.$$

Call this $\gamma_d(F)$.

We consider a generalization of $\gamma_d(F)$ to rings $\mathbb{Z}/2^k\mathbb{Z}$ in the following simple manner. For illustration, let us consider the ring $\mathbb{Z}/4\mathbb{Z}$. Given a Boolean function $F : \{0,1\}^n \to \{0,1\}$, let $F_2 : \{0,1\}^n \to \{0,2\}$, where $\{0,2\}$ is thought of as a subset of $\mathbb{Z}/4\mathbb{Z}$, be the 2-*lift* of F defined as

$$F_2(x) := \begin{cases} 0 & \text{if } F(x) = 0, \\ 2 & \text{if } F(x) = 1. \end{cases}$$

Building upon the notation introduced earlier, we can define $\gamma_{d,2}(F)$ to be the largest γ such that there exists a polynomial $Q_2 \in \mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n]$ of degree d that has agreement γ with F_2. Note that $\gamma_{d,2}(F) \geq \gamma_d(F)$ since if, for instance, $Q(x) = x_1x_2 + x_3 \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ has agreement γ with F, then $Q_2 := 2(x_1x_2 + x_3) \in$

$\mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n]$ also has the same agreement γ with F_2.

More generally, we can extend these definitions to define $\gamma_{d,k}(F)$: the largest possible agreement that $F_k : \{0,1\}^n \to \{0, 2^{k-1}\}$, which we call the *k-lift of F*, defined as

$$F_k(x) := \begin{cases} 0 & \text{if } F(x) = 0, \\ 2^{k-1} & \text{if } F(x) = 1, \end{cases}$$

can have with polynomials of degree d in $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$.

Our study of this notion of approximation of Boolean functions by polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ is motivated by the work of Bhowmick and Lovett [BL15] that studies the maximum possible agreement between nonclassical polynomials over $\mathbb{F}_2$ of degree d and a Boolean function F. Informally speaking, nonclassical polynomials of degree d can be thought of as a subset of the degree d polynomials in $\mathbb{Z}/2^d\mathbb{Z}[x_1, \ldots, x_n]$, and so the maximum possible agreement between nonclassical polynomials of degree d and a Boolean function F is upper bounded by $\gamma_{d,d}(F)$ (we formalize this relationship in Section 2.2). One of the results from their work shows that low-degree nonclassical polynomials can only have "small" agreement with the majority function on n bits. To do this, they essentially prove the following bound, which can be stated using our notation as:

$$\gamma_{d,d}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d \cdot 2^d)}{\sqrt{n}}.$$

If $d = \omega(\log n)$, this result unfortunately does not give any nontrivial upper bound on the maximum possible agreement between nonclassical polynomials of degree d and the majority function. This is in contrast to the case of classical polynomials of degree d (i.e, polynomials of degree d in $\mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$) where a much stronger upper bound on the agreement with the majority function is known due to results of Szegedy [Sze89] and Smolensky [Smo93]:

$$\gamma_{d,1}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}};$$

this gives a nontrivial upper bound even when d is as large as $O(\sqrt{n})$.

Bhowmick and Lovett, however, conjectured that their result could be improved, and left open the question of whether nonclassical polynomials can have larger agreement with the majority function than their classical counterparts.
More generally, they informally conjectured that although nonclassical polynomials can approximate some Boolean functions in the correlation sense better than classical polynomials of the same degree do, this is not true in the case of agreement-based approximation. Our work stems from trying to answer these questions.

2.1.1 Our results

We prove the following results about agreement of Boolean functions with polynomials over the ring $\mathbb{Z}/2^k\mathbb{Z}$.

1. We explore whether there exist Boolean functions for which agreement can increase by increasing k. In particular, are there Boolean functions F such that $\gamma_{d,k}(F) > \gamma_{d,1}(F)$?

   We begin by investigating this question for $d \leq 1$, and prove that there are no such functions in this case.

   (a) For all Boolean functions F and $d \leq 1, k > 1$,

   $$\gamma_{d,k}(F) = \gamma_{d,1}(F).$$

   Keeping this in mind, the first place where we can expect larger k to give better agreement is $\gamma_{2,2}$ vs. $\gamma_{2,1}$. Our next result shows that there are indeed separating examples in the regime.

   (b) For infinitely many n, there exists a Boolean function $F : \{0,1\}^n \to \{0,1\}$ such that

   $$\gamma_{2,1}(F) \leq \frac{1}{2} + \frac{1}{8} + o(1),$$

   whereas

   $$\gamma_{2,2}(F) \geq \frac{1}{2} + \frac{1}{4} - o(1).$$

Since $\gamma_{d,k}(F)$ is a quantity based on agreement of polynomials with a Boolean function F, it can be shown that $\gamma_{d,k}(F) \geq 1/2$ for any $d \geq 0, k \geq 1$. We then ask if there exist Boolean functions F such that $\gamma_{d,1}(F)$ is more or less equal to the trivial bound of $1/2$, while $\gamma_{d',k}(F)$ is significantly larger for $d' \leq d$ and some $k > 1$. In this context, we show the following result.

(c) Fix $\ell \geq 2$. There is a Boolean function $F : \{0,1\}^n \to \{0,1\}$ such that

$$\gamma_{2^\ell - 1,1}(F) \leq \frac{1}{2} + o(1),$$

whereas

$$\gamma_{d,3}(F) \geq \frac{1}{2} + \frac{1}{16} - o(1)$$

for $d = 2^{\ell-1} + 2^{\ell-2} \leq 2^\ell - 1$.

2. Next, we turn to the majority function on n bits and show that for $d \geq 0, k \geq 1$,

$$\gamma_{d,k}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}},$$

by adapting a proof due to Green [Gre00] of a result on the inapproximability of the parity function by low-degree polynomials over the ring $\mathbb{Z}/p^k\mathbb{Z}$ for prime $p \neq 2$.

Coupled with the observation that polynomials of degree d over the ring $\mathbb{Z}/2^d\mathbb{Z}$ "subsume" nonclassical polynomials over $\mathbb{F}_2$ of the same degree, part (c) of the first result from above provides a counterexample to the informal conjecture of Bhowmick and Lovett [BL15] mentioned above, that for any Boolean function F, nonclassical polynomials of degree d do not approximate F in the agreement sense any better than classical polynomials of the same degree do. The second result from above confirms their conjecture that nonclassical polynomials do not do any better than classical polynomials of the same degree as far as agreement with the majority function is concerned.

2.1.2 Organisation

We start with some preliminaries in Section 2.2. In Section 2.3, we study the behaviour of $\gamma_{d,k}$ as a function of k (Part 1 of the results from above). Next, in Section 2.4, we prove an upper bound on $\gamma_{d,k}(\mathrm{Maj}_n)$ (Part 2 of the results from above), and finally, in Section 2.5, we discuss the applications of our results in answering questions raised by Bhowmick and Lovett [BL15].

2.2 Preliminaries

We use $[n]$ to denote the set $\{1, \ldots, n\} \subset \mathbb{N}$. We will naturally identify $\{0, 1\} \subset \mathbb{Z}$, $\mathbb{F}_2$, and $\mathbb{Z}/2\mathbb{Z}$ with each other, without explicitly using inclusion or embedding maps. Similarly, we will identify $\mathbb{Z}/2^k\mathbb{Z}$ with the set $\{0, \ldots, 2^k - 1\}$.
For $x \in \{0,1\}^n$ and $i \geq 0$, we use $|x|$ to denote the Hamming weight of x, and $|x|_i$ to denote the $(i+1)$-th least significant bit of $|x|$ in base 2.
For an $x \in \{0,1\}^n$, by the *Hamming ball of radius d* around x, we mean the set

$$\{y \in \{0,1\}^n | \text{ the Hamming distance between } x \text{ and } y \text{ is at most } d\}.$$

For a finite set S, by $x \sim S$, we mean an x sampled uniformly at random from S.

Since we always restrict the domain of polynomials to $\{0,1\}^n$, it will suffice for us to only consider *multilinear polynomials*, i.e., polynomials in which the individual degree of any variable is at most one. Thus, unless otherwise specified, "polynomials" will refer to "multilinear polynomials".

2.2.1 Elementary symmetric polynomials

Recall that for $t \geq 1$, the elementary symmetric polynomial of degree t over $\mathbb{F}_2$ is defined as

$$S_t(x_1, \ldots, x_n) := \sum_{S \subseteq [n]; |S| = t} \prod_{i \in S} x_i.$$

It may be noted that for all $x \in \{0,1\}^n$,

$$S_t(x_1, \ldots, x_n) = \binom{|x|}{t} \bmod 2. \tag{2.2.1}$$

A direct consequence of Lucas's theorem and Eq. (2.2.1) is the following:

Lemma 2.1. *Fix $\ell \geq 0$. Then, for every $x \in \{0,1\}^n$, we have that $S_{2^\ell}(x) = |x|_\ell$. More generally, for $t \geq 1$, $S_t(x) = \prod_i |x|_i$ where the product runs over all $i \geq 0$ such that the $(i+1)^{th}$ least significant bit of the binary expansion of t is 1.*

The following result follows from the work of Green and Tao [GT09, Theorem 11.3], who build upon the ideas of Alon and Beigel [AB01].

Theorem 2.2 (Green-Tao [GT09], Alon-Beigel [AB01])**.** *Fix $\ell \geq 0$. Then, for every polynomial $P \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ of degree at most $2^\ell - 1$, we have that*

$$\Pr_{x \sim \{0,1\}^n}[S_{2^\ell}(x) = P(x)] \leq \frac{1}{2} + o(1).$$

Theorem 2.2 has a useful corollary that immediately follows from it.

Corollary 2.3. *For every fixed $\ell \geq 0$, the functions $\{S_{2^i}(x_1, \ldots, x_n)\}_{0 \leq i \leq \ell}$ are almost balanced and almost uncorrelated, i.e.,*

- $\forall\ 0 \leq i \leq \ell$,

$$\left| \Pr_{x \sim \{0,1\}^n}[S_{2^i}(x) = 0] - \Pr_{x \sim \{0,1\}^n}[S_{2^i}(x) = 1] \right| = o(1).$$

- $\forall\ a_0, \ldots, a_\ell \in \{0,1\}$,

$$\left| \Pr_{x \sim \{0,1\}^n} \left[\bigwedge_{0 \leq i \leq \ell} (S_{2^i}(x) = a_i) \right] - \frac{1}{2^{\ell+1}} \right| = o(1).$$

Combining Corollary 2.3 with Lemma 2.1, we get another useful fact:

Lemma 2.4. *For every fixed $r \geq 1$, if x is a random variable that is uniformly distributed in $\{0,1\}^n$ then the random variables $\{|x|_i\}_{0 \leq i \leq r-1}$ are almost uniform and almost r-wise independent i.e.*

- $\forall\ 0 \leq i \leq r-1$,

$$\left| \Pr_{x \sim \{0,1\}^n}[|x|_i = 0] - \Pr_{x \sim \{0,1\}^n}[|x|_i = 1] \right| = o(1).$$

- $\forall\ (a_0, \ldots, a_{r-1}) \in \{0,1\}^r$,

$$\left| \Pr_{x \sim \{0,1\}^n}[(|x|_0, \ldots, |x|_{r-1}) = (a_0, \ldots, a_{r-1})] - \frac{1}{2^r} \right| = o(1).$$

2.2.2 Polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ and Boolean functions

For $d \geq 0$ and $k \geq 1$, $\mathcal{P}_{d,k}$ will denote the set of multilinear polynomials of degree at most d over the ring $\mathbb{Z}/2^k\mathbb{Z}$, i.e., polynomials $P \in \mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$ of the form

$$P(x) = \sum_{S \subseteq [n]; |S| \leq d} c_S \prod_{i \in S} x_i,$$

where $c_S \in \mathbb{Z}/2^k\mathbb{Z}$. We say that $P \in \mathcal{P}_{d,k}$ is the *zero polynomial* if $c_S = 0$ for all $S \subseteq [n]$.

Following [Gop08], we call a set $I \subseteq \{0,1\}^n$ an *interpolating set*[1] for $\mathcal{P}_{d,k}$ if the only polynomial $P \in \mathcal{P}_{d,k}$ that vanishes at all points in I is zero everywhere. Formally speaking, $I \subseteq \{0,1\}^n$ is an interpolating set for $\mathcal{P}_{d,k}$ if for any $P \in \mathcal{P}_{d,k}$,

$$(\forall x \in I \ \ P(x) = 0) \Rightarrow (\forall y \in \{0,1\}^n \ \ P(y) = 0).$$

We now state a number of basic facts about polynomials in $\mathcal{P}_{d,k}$.

Lemma 2.5. *Let $d \geq 0, k \geq 1$. Then, any polynomial $Q \in \mathcal{P}_{d,k}$ satisfies the following:*

1. *If Q is a non-zero polynomial, then*

$$\Pr_{x \sim \{0,1\}^n}[Q(x) \neq 0] \geq \frac{1}{2^d}.$$

2. *Q is the zero polynomial iff $Q(x) = 0$ for all $x \in \{0,1\}^n$.*

[1] This is also called a *hitting set* in the literature.

3. *Say* $Q(x) = \sum_{S \subseteq [n]; |S| \leq d} c_S x_S$*, where* $c_S \in \mathbb{Z}/2^k\mathbb{Z}$ *and* x_S *denotes* $\prod_{i \in S} x_i$*. Then, for all* $S \subseteq [n], |S| \leq d$*,*

$$c_S = \sum_{T \subseteq S} (-1)^{|S|-|T|} Q(1_T)$$

where $1_T \in \{0,1\}^n$ *is the characteristic vector of* T*.*

4. Q *vanishes at all points in* $\{0,1\}^n$ *iff* Q *vanishes on the Hamming ball of radius* d *centered around the origin, i.e., on all points* x *with* $|x| \leq d$*. By shifting the origin to any point of* $\{0,1\}^n$*, the same is true of* any *Hamming ball of radius* d *in* $\{0,1\}^n$*.*

Proof. Part 1: Write Q as $Q(x) = 2^\ell \cdot Q'(x)$, where $\ell < k$ is the largest power of 2 that divides the GCD of the coefficients of Q. Projecting Q' to a non-zero polynomial Q'' over $\mathbb{Z}/2\mathbb{Z}$ by dropping all its coefficients modulo 2, and applying the DeMillo-Lipton-Schwartz-Zippel lemma for $\mathbb{F}_2$ (see, e.g., [CT15, Appendix C]), gives us that

$$\Pr_{x \in \{0,1\}^n} [Q''(x) \neq 0] \geq \frac{1}{2^d}.$$

This means that $Q'(x)$ is not a multiple of 2 on at least a 2^{-d} fraction of points in $\{0,1\}^n$. Noting that $Q(x) = 2^\ell \cdot Q'(x)$ cannot be zero unless $Q'(x)$ is even, then completes the proof.

Part 2 follows from Part 1.

Part 3 is follows from the Möbius inversion formula (see, e.g., [GSL10, Section 2]).

Part 4 follows immediately from parts 2 and 3. □

Given a Boolean function $F : \{0,1\}^n \to \{0,1\}$ and $k \geq 1$, we define the *k-lift of* F to be the function $F_k : \{0,1\}^n \to \mathbb{Z}/2^k\mathbb{Z}$ defined as follows[2]. For any $x \in \{0,1\}^n$,

$$F_k(x) := \begin{cases} 0 & \text{if } F(x) = 0, \\ 2^{k-1} & \text{if } F(x) = 1. \end{cases}$$

[2]This is a $\mathbb{Z}/2^k\mathbb{Z}$-valued version of F; see Section 1.1.1.

We will consider how well polynomials in $\mathcal{P}_{d,k}$ can approximate k-lifts of Boolean functions in the agreement sense. More precisely, for any Boolean function $F : \{0,1\}^n \to \{0,1\}$, we define

$$\gamma_{d,k}(F) := \max_{Q \in \mathcal{P}_{d,k}} \operatorname{agr}(F_k, Q) = \max_{Q \in \mathcal{P}_{d,k}} \left[\Pr_{x \sim \{0,1\}^n} [F_k(x) = Q(x)] \right].$$

Note that $\gamma_{d,k}(F)$ is always at least $1/2$, since one can consider the *constant* polynomials to achieve an agreement of at least $1/2$ with the k-lift of a Boolean function F.

Lemma 2.6. *Let F be a Boolean function and $k \geq 1, d \geq 0$ be integers. Then $\gamma_{d,k}(F) \geq 1/2$.*

2.2.3 Polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ and nonclassical polynomials

In this section, we relate the agreement between a Boolean function F and nonclassical polynomials over $\mathbb{F}_2$, to the agreement between F and polynomials in $\mathcal{P}_{d,k}$. This section will also shed some light on why we study agreement-based approximation by polynomials over $\mathbb{Z}/2^k\mathbb{Z}$.

For $x \in \mathbb{R}$, x modulo one, denoted by $x \bmod 1$, is equal to the fractional part of x given by $x - \lfloor x \rfloor$. We think of $\mathbb{R}/\mathbb{Z}$ as the set $[0,1) \subset \mathbb{R}$ equipped with addition modulo 1.

We will now state a characterization of nonclassical polynomials over $\mathbb{F}_p$ due to Tao and Ziegler [TZ12]. Here we state a slightly modified version of the original statement for $p = 2$.

Theorem 2.7 (Tao and Ziegler [TZ12])**.** *A function $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is a nonclassical polynomial over $\mathbb{F}_2$ of degree d if and only if it has the following form:*

$$P(x_1, \ldots, x_n) = \left(\alpha + \sum_{S \subseteq [n]; k \geq 0 : |S| + k \leq d} \frac{c_{S,k}}{2^{k+1}} \prod_{i \in S} x_i \right) \bmod 1$$

Here $\alpha \in [0,1) \subset \mathbb{R}$ and $c_{S,k} \in \{0,1\} \subset \mathbb{R}$ are uniquely determined. α is called the shift of P, and the largest k such that $c_{S,k} \neq 0$ for some $S \subseteq [n]$ is called the depth of P.

Following [BL15], we define the agreement between a nonclassical $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ over $\mathbb{F}_2$ and a Boolean function $F : \{0,1\}^n \to \{0,1\}$ as

$$\operatorname{agr}\left(\frac{F}{2}, P\right) = \Pr_{x \sim \{0,1\}^n}\left[\frac{F(x)}{2} = P(x)\right].$$

Note that $F/2$ is a $\mathbb{R}/\mathbb{Z}$-valued version of F (see Section 1.1.2), since it always takes values in $\{0, 1/2\} \subset \mathbb{R}/\mathbb{Z}$.

We remark that nonclassical polynomials $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ whose shift and depth are both zero, are called *classical polynomials*. It's not hard to verify that a classical polynomial $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ of degree d can be written as

$$P(x) = \frac{P'(x)}{2} \bmod 1,$$

where $P'(x)$ is some polynomial in $\mathcal{P}_{d,1}$ (see [BL15, Section 2]). Thus, we will refer to polynomials in $\mathcal{P}_{d,1}$ as classical polynomials.

Also note that, since we are studying agreement between nonclassical polynomials over $\mathbb{F}_2$ and Boolean functions (and thus, $\{0, 1/2\}$-valued functions), it suffices to restrict ourselves to nonclassical polynomials whose shift α is 0.

The following facts immediately follow from Theorem 2.7.

Lemma 2.8. *Let $d, k \geq 1$.*

1. *Let $d \geq k$ and $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ be a nonclassical polynomial of degree d and depth k over $\mathbb{F}_2$, and suppose the shift α of P is zero. Then there is a polynomial $Q(x) \in \mathbb{Z}[x_1, \ldots, x_n]$ of degree d with coefficients in $\{0, \ldots, 2^k - 1\}$ such that for all $x \in \{0,1\}^n$,*

$$P(x) = \frac{Q(x)}{2^k} \bmod 1.$$

2. *Let $Q(x) \in \mathbb{Z}[x_1, \ldots, x_n]$ be a polynomial of degree d. If $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is such that for all $x \in \{0,1\}^n$,*

$$P(x) = \frac{Q(x)}{2^k} \ mod\ 1$$

then P is a nonclassical polynomial over $\mathbb{F}_2$ of degree at most $d + k - 1$ and depth at most k.

Fix $d, k \geq 1$, and suppose that $F : \{0,1\}^n \to \{0,1\}$ is Boolean function such that there is a polynomial $Q \in \mathcal{P}_{d,k}$ with $\mathrm{agr}(F_k, Q) = \gamma$. We can naturally think of Q as a polynomial of degree d in $\mathbb{Z}[x_1, \ldots, x_n]$ with coefficients in $\{0, \ldots, 2^k - 1\} \subset \mathbb{Z}$, and define the function $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ as follows. For all $x \in \{0,1\}^n$,

$$P(x) := \frac{Q(x)}{2^k} \bmod 1.$$

It now follows from the definition of F_k that for all $x \in \{0,1\}^n$,

$$F_k(x) = Q(x) \Leftrightarrow P(x) = \frac{F(x)}{2} \bmod 1.$$

Combining this with Part 2 of Lemma 2.8, we conclude that P is a nonclassical polynomial over $\mathbb{F}_2$ of degree $\leq d + k - 1$ and depth $\leq k$ such that

$$\mathrm{agr}\left(\frac{F}{2}, P\right) = \mathrm{agr}(F_k, Q) = \gamma.$$

Conversely, suppose that $d \geq k$, and $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is a nonclassical polynomial of degree d and depth k such that

$$\mathrm{agr}\left(\frac{F}{2}, P\right) = \gamma.$$

By Part 1 of Lemma 2.8, there is a polynomial Q of degree d in $\mathbb{Z}[x_1, \ldots, x_n]$ such that for all $x \in \{0,1\}^n$,

$$P(x) = \frac{Q(x)}{2^k} \bmod 1.$$

Furthermore, it is guaranteed that all the coefficients of Q are in $\{0, \ldots, 2^k - 1\}$, and so we can naturally think of Q as a polyomial in $\mathcal{P}_{d,k}$. Using the definition of F_k and

a similar argument as before, we can conclude that

$$\operatorname{agr}(F_k, Q) = \operatorname{agr}\left(\frac{F}{2}, P\right) = \gamma.$$

We can summarize this discussion as follows.

Lemma 2.9. *Let F be a Boolean function and $d, k \geq 1$. Then there is a nonclassical polynomial P over $\mathbb{F}_2$ of degree $\leq d + k - 1$ and depth $\leq k$ such that*

$$\operatorname{agr}\left(\frac{F}{2}, P\right) = \gamma_{d,k}(F).$$

Additionally, if $d \geq k$, and γ is the maximum possible agreement between F and nonclassical polynomials over $\mathbb{F}_2$ of degree d and depth k, then

$$\gamma_{d,k}(F) \geq \gamma.$$

Lemma 2.9 motivates the study of the quantity $\gamma_{d,k}$ in order to bound the agreement between nonclassical polynomials and Boolean functions.

2.3 Understanding the behavior of $\gamma_{d,k}(F)$ as a function of k

Our goal in this section is to understand how $\gamma_{d,k}(F)$ behaves as k increases and d remains fixed. The first observation one can make, which was alluded to in Section 2.1, is that $\gamma_{d,k}(F)$ cannot *decrease* as k increases.

Lemma 2.10. *For every Boolean function F and $k \geq 1, d \geq 0$, $\gamma_{d,k+1}(F) \geq \gamma_{d,k}(F)$.*

Proof. Suppose that $P \in \mathcal{P}_{d,k}$ has agreement α with F_k. Then, $2 \cdot P$ (interpreted naturally as a polynomial in $\mathcal{P}_{d,k+1}$) also has agreement α with F_{k+1}. □

This naturally motivates the question as to whether there are Boolean functions F for which $\gamma_{d,k'}(F) > \gamma_{d,k}(F)$ for some $d \geq 0$ and $k' > k \geq 1$.

2.3.1 Behavior of $\gamma_{d,k}(F)$ for $d \leq 1$

Let F be any Boolean function. For $d = 0$, the only polynomials in $\mathcal{P}_{d,k}$ that can have non-zero agreement with F_k are the constant polynomials $Q(x) = 0$ and $Q(x) = 2^{k-1}$. For some $k > 1$, if $Q(x) = 2^{k-1}$ (resp. $Q(x) = 0$) in $\mathcal{P}_{0,k}$ has agreement γ with F_k then the polynomial $Q(x) = 1$ (resp. $Q(x) = 0$) in $\mathcal{P}_{0,1}$ has agreement γ with F (this follows from the definition of F_k), and so $\gamma_{0,k}(F) \leq \gamma_{0,1}(F)$, which together with Lemma 2.10 implies that $\gamma_{0,k}(F) = \gamma_{0,1}(F)$.

We now investigate this question for $d = 1$, i.e., for degree one polynomials.

Recall that Lemma 2.6 tells us that $\gamma_{1,k}(F) \geq 1/2$. In particular, this means that $\gamma_{1,1}(F) \geq 1/2$. Thus, if there were a $k > 1$ for which $\gamma_{1,k}(F) > \gamma_{1,1}(F)$ then it must be the case that $\gamma_{1,k}(F) > 1/2$. We will now use the Schwartz-Zippel lemma for $\mathbb{Z}/2^k\mathbb{Z}$ polynomials (see Lemma 2.5) to show that if $\gamma_{1,k}(F) > 1/2$ then $\gamma_{1,k}(F) = \gamma_{1,1}(F)$, which rules out the existence of a $k > 1$ for which $\gamma_{1,k}(F) > \gamma_{1,1}(F)$. In fact, we prove a more general result that holds for all $d \geq 1$.

Lemma 2.11. *For any Boolean function F and $d \geq 1, k > 1$,*

$$\gamma_{d,k}(F) > 1 - \frac{1}{2^d} \Rightarrow \gamma_{d,k}(F) = \gamma_{d,1}(F).$$

Proof. Fix an F and $d \geq 1, k > 1$ such that $\gamma_{d,k}(F) = \gamma > 1 - 2^{-d}$. This means that there must be a $Q \in \mathcal{P}_{d,k}$ such that $\mathrm{agr}(F_k, Q) = \gamma$. Since F_k is a $\{0, 2^{k-1}\}$-valued function, it follows that Q must take a value in $\{0, 2^{k-1}\}$ on at least a γ fraction of points in $\{0,1\}^n$. Let $Q' \in \mathcal{P}_{d,k-1}$ be the polynomial obtained from Q by dropping all its coefficients modulo 2^{k-1}. Then, since Q is $\{0, 2^{k-1}\}$-valued on at least a γ fraction of points in $\{0,1\}^n$, it follows that Q' is 0 on at least a $\gamma > 1 - 2^{-d}$ fraction of points. Lemma 2.5 then implies that Q' must be the zero polynomial, and so $Q'(x) = 0$ for all $x \in \{0,1\}^n$.

Since Q' was obtained from Q by dropping the coefficients of Q modulo 2^{k-1}, it follows that every coefficient of Q is a multiple of 2^{k-1}, and so Q can be naturally identified with $2^{k-1} \cdot Q''$ for some $Q'' \in \mathcal{P}_{d,1}$. Furthermore, whenever Q agrees with F_k

it must be the case that Q'' agrees with F — this just follows from the definition of F_k. Thus, $\mathrm{agr}(F, Q'') \geq \gamma$, and so $\gamma_{d,1}(F) \geq \gamma_{d,k}(F)$. We also know from Lemma 2.10 that $\gamma_{d,k}(F) \geq \gamma_{d,1}(F)$ and so we conclude that $\gamma_{d,k}(F) = \gamma_{d,1}(F)$. □

We can now conclude that for every Boolean function F and $k > 1$:

- either $\gamma_{1,k}(F) \leq 1/2$, in which case Lemmas 2.6 and 2.10 together imply that

$$\frac{1}{2} \leq \gamma_{1,1}(F) \leq \gamma_{1,k}(F),$$

  and so $\gamma_{1,k}(F) = \gamma_{1,1}(F)$, or

- $\gamma_{1,k}(F) > 1/2$, in which case Lemma 2.11 implies that $\gamma_{1,k}(F) = \gamma_{1,1}(F)$.

We can summarize the results for $d \leq 1$ as follows.

Theorem 2.12. *For every Boolean function F and $d \leq 1, k > 1$, $\gamma_{1,k}(F) = \gamma_{1,1}(F)$.*

2.3.2 Examples of F with $\gamma_{2,2}(F) > \gamma_{2,1}(F)$

In light of Theorem 2.12, the next question to ask is whether we can find examples of F for which $\gamma_{2,k}(F) > \gamma_{2,1}(F)$ for some $k > 1$, and in particular for $k = 2$. Somewhat surprisingly, it turns out there are examples of this kind:

Theorem 2.13. *For infinitely many n, there exist Boolean functions $F : \{0,1\}^{2n} \to \{0,1\}$ such that*

$$\gamma_{2,2}(F) \geq \frac{1}{2} + \frac{1}{4} - o(1),$$

whereas

$$\gamma_{2,1}(F) \leq \frac{1}{2} + \frac{1}{8} + o(1).$$

Before we go into the formal details of the proof of Theorem 2.13, we will give a sketch of the proof.

We consider Boolean functions F in $2n$ variables $x_1, \ldots x_n, y_1, \ldots y_n$. Lemma 2.11 implies that for any F, $\gamma_{2,2}(F) = \gamma_{2,1}(F)$ unless $\gamma_{2,2}(F) \leq 3/4$. We then restrict our

attention to Boolean functions F for which this bound is almost tight i.e., $\gamma_{2,2}(F) = 3/4 - o(1)$. Any quadratic polynomial Q over $\mathbb{Z}/4\mathbb{Z}$ that has agreement $3/4 - o(1)$ with F_2 (the 2-lift of F) must be $\{0,2\}$-valued on a $3/4 - o(1)$ fraction of points in $\{0,1\}^n$, and so, by dropping its coefficients modulo 2, we obtain a quadratic polynomial Q' over $\mathbb{F}_2$ that is zero on a $3/4 - o(1)$ fraction of points in $\mathbb{F}_2^n$.

Results on the structure of polynomials over $\mathbb{F}_2$ (see, e.g., [LN97, Chapter 6]) then dictate that Q' must be of the form $L_1(x,y)L_2(x,y) + L_3(x,y)$ where L_1, L_2, and L_3 are independent linear polynomials over $\mathbb{F}_2$. This then implies that the polynomial Q must have been of the form $L_1'(x,y)L_2'(x,y) + 2Q''(x,y)$ for some linear polynomials L_1', L_2' and quadratic polynomial Q'' over $\mathbb{Z}/4\mathbb{Z}$.

Guided by the above, we try to work our way backwards: we begin with a quadratic polynomial P from $\mathbb{Z}/4\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_n]$ of the above form, and construct a Boolean function F whose 2-lift has agreement $3/4 - o(1)$ with P, hoping that the agreement of F with quadratic polynomials over $\mathbb{Z}/2\mathbb{Z}$ is significantly less than $3/4$. In particular, we choose

$$P(x,y) = \left(\sum_{1 \le i \le n} x_i\right)\left(\sum_{1 \le i \le n} y_i\right),$$

and define F as

$$F(x,y) := \begin{cases} 0 & \text{if } P(x,y) = 0 \\ 1 & \text{if } P(x,y) = 2 \\ H(x,y) & \text{otherwise,} \end{cases}$$

where $H(x,y)$ is some Boolean function that we later choose carefully.

By construction, $\gamma_{2,2}(F) \ge 3/4 - o(1)$. Suppose $S = \{(x,y) \mid P(x,y) \notin \{0,2\}\}$. We first ensure that F has agreement $1/2 + o(1)$ with quadratic polynomials over $\mathbb{Z}/2\mathbb{Z}$ when restricted to S, by carefully choosing $H(x,y)$ to be a function that is "hard" for these polynomials (e.g., a random function). Using the fact that F restricted to the complement of S is a quadratic polynomial of *high rank*, we then show that, over the complement of S, F has "sufficiently" low agreement with quadratic polynomials,

thus concluding that $\gamma_{2,1}(F) \leq 5/8 + o(1)$.

Let us begin the formal proof of Theorem 2.13. We first define a family of Boolean functions on $\{0,1\}^{2n}$, where $n \equiv 1 \pmod 2$. As mentioned before, we denote the $2n$ variables by $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

Define the set S to be

$$S := \{(x,y) \in \{0,1\}^{2n} \mid |x|, |y| \equiv 1 \pmod 2\}.$$

Given any function $H : S \to \{0,1\}$, we define the Boolean function $F_H : \{0,1\}^{2n} \to \{0,1\}$ as follows:

$$F_H(x,y) := \begin{cases} 0 & \text{if } |x||y| \equiv 0 \pmod 4, \\ 1 & \text{if } |x||y| \equiv 2 \pmod 4, \\ H(x,y) & \text{if } (x,y) \in S. \end{cases} \tag{2.3.1}$$

First of all, let us note that for any choice of H, $\gamma_{2,2}(F_H) \geq 3/4$.

Lemma 2.14. *For all $H : S \to \{0,1\}$, $\gamma_{2,2}(F_H) \geq 3/4$.*

Proof. Consider the following polynomial in $\mathcal{P}_{2,2}$,

$$P(x,y) = \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{j=1}^{n} y_j\right).$$

It may be noted that if $|x||y| \equiv 0 \pmod 2$, then $P(x,y) = F_{H,2}(x,y)$, where $F_{H,2}$ is the 2-lift of F_H. Thus, the probability that $P(x,y) \neq F_{H,2}(x,y)$ is the probability that $|x|$ and $|y|$ are both odd, which is $1/4$. This finishes the proof. □

We will now prove some lemmas that will help use prove $\gamma_{2,1}(F_H) \leq 5/8 + o(1)$ for an appropriate choice of H. We first show how to choose H so that no quadratic function over $\mathbb{Z}/2\mathbb{Z}$ can agree with H on too many points in S.

Lemma 2.15. *There is an $H : S \to \{0,1\}$ such that for all quadratic polynomials $q \in \mathcal{P}_{2,1}$ we have that*

$$\Pr_{(x,y)\sim S}[H(x,y) \neq q(x,y)] \geq \frac{1}{2} - o(1).$$

Proof. Sample H uniformly at random from the set of all Boolean functions defined on S. For each such q, the expected number of locations $(x, y) \in S$ where $H(x, y) \neq q(x, y)$ is $|S|/2$. By a standard Chernoff bound, the probability that this number is less than $|S|/2 - |S|^{2/3}$ is $\exp(-\Omega(|S|^{1/3}))$, which is at most $\exp(-2^{\Omega(n)})$ using the fact that $|S| = \Omega(2^{2n})$. Observing that the number of quadratic polynomials $q \in \mathcal{P}_{2,1}$ is at most $2^{O(n^2)}$, a union bound over all possible q tells us that with probability $1 - o(1)$ over the choice of H,

$$\Pr_{(x,y)\sim S}[H(x, y) \neq q(x, y)] \geq \frac{1}{2} - o(1).$$

This establishes the existence of an H with the desired property. □

We will now show that for any $H : S \to \{0, 1\}$, F_H cannot agree with a quadratic polynomial on too many points in $\overline{S}$, i.e., the complement of S. Noting that $\overline{S}$ is essentially the union of the subspaces given by $\sum_i x_i = 0$ and $\sum_i y_i = 0$, it will be helpful to work with respect to the following alternate basis for the space of linear functions on $\mathbb{F}_2^n$ (recall that $n \equiv 1 \pmod 2$):

$$u_1(x) = \sum_{i=1}^{n} x_i$$

$$\begin{aligned}
u_2(x) &= x_1 + \sum_{i\geq 3} x_i & u_3(x) &= x_2 + \sum_{i\geq 3} x_i \\
u_4(x) &= x_3 + \sum_{i\geq 5} x_i & u_5(x) &= x_4 + \sum_{i\geq 5} x_i \\
&\vdots & &\vdots \\
u_{2j}(x) &= x_{2j-1} + \sum_{i\geq 2j+1} x_i & u_{2j+1}(x) &= x_{2j} + \sum_{i\geq 2j+1} x_i \\
&\vdots & &\vdots \\
u_{n-1}(x) &= x_{n-2} + x_n & u_n(x) &= x_{n-1} + x_n.
\end{aligned}$$

We can similarly define linear functions $v_1(y), \ldots, v_n(y)$ that span the set of all linear functions over $\mathbb{F}_2$ in the y variables.

To see that these linear functions are indeed linearly independent and form a basis for all linear functions on $\mathbb{F}_2^n$, it suffices to note that for every $i \in [n]$,

$$x_i = \begin{cases} \sum_{j=1}^{i} u_j(x) & \text{if } i \text{ is even,} \\ \sum_{j=1}^{i} u_j(x) + u_{i+2}(x) & \text{if } i \text{ is odd.} \end{cases}$$

We can similarly write the y variables in terms of $v_1(y), \ldots, v_n(y)$. We will use u and v to denote the variables $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ respectively. These variables will always be assumed to be related to the variables $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ via the equations given above.

It may be noted that for any Boolean function $F(x_1, \ldots, x_n, y_1, \ldots, y_n)$, there is a unique degree d polynomial $P \in \mathbb{F}_2[u_1, \ldots, u_n, v_1, \ldots, v_n]$ such that for all $(x, y) \in \{0,1\}^{2n}$,

$$P(u_1(x), \ldots, u_n(x), v_1(y), \ldots, v_n(y)) = F(x_1, \ldots, x_n, y_1, \ldots, y_n),$$

where $u_1(x), \ldots, u_n(x), v_1(y), \ldots, v_n(y)$ are the linear functions defined above. We will say that P *represents* F as a polynomial in u, v.

Lemma 2.16. *The degree 2 elementary symmetric polynomials $S_2(x)$ and $S_2(y)$ are represented as polynomials in u, v over $\mathbb{F}_2$ by*

$$u_2u_3 + \ldots + u_{n-1}u_n + L'(u)$$

and

$$v_2v_3 + \ldots v_{n-1}v_n + L''(v)$$

respectively, for some linear functions $L'(u)$ and $L''(v)$.

Proof. Observe that

$$\begin{aligned} &u_2(x)u_3(x) + u_4(x)u_5(x) + \cdots + u_{n-1}(x)u_n(x) \\ &= \left(\sum_{1 \le i < j \le n} x_i x_j \right) + \sum_{i=1}^{\frac{n-1}{2}} \left(\sum_{j=2i+1}^{n} x_j \right)^2 \\ &= S_2(x) + \sum_{i=1}^{\frac{n-1}{2}} \left(\sum_{j=2i+1}^{n} x_j \right). \end{aligned}$$

Thus, the polynomial $S_2(x)$ is represented as a polynomial in u, v by the quadratic $u_2u_3 + \ldots + u_{n-1}u_n + L'(u)$ for some linear function L'. A similar argument can be used to prove the statement for $S_2(y)$. □

We will also need some basic preliminaries about $\mathbb{F}_2$-quadratics, i.e., polynomials from $\mathcal{P}_{2,1}$. Given $Q(x_1, \ldots, x_n) \in \mathcal{P}_{2,1}$, we define the *rank of* Q to be the least r such that we can write Q in the form

$$Q(x) = \sum_{i=1}^{r} L_i(x)L'_i(x) + L(x)$$

where L_i, L'_i ($i \in [r]$) and L are linear polynomials in $x_1, \ldots, x_n$. We use $\mathrm{rk}(Q)$ to denote the rank of Q.

The following are standard facts about the rank of quadratic polynomials (see, e.g., [LN97, Chapter 6]).

Lemma 2.17. *Let Q, Q_1, and Q_2 be polynomials in $\mathcal{P}_{2,1}$ such that*

$$Q(x) = \sum_{i=1}^{r} L_i(x)L'_i(x) + L(x), \tag{2.3.2}$$

where L_i, L'_i ($i \in [r]$) and L are linear polynomials.

1. *For any linear function L'', $\mathrm{rk}(Q + L'') = \mathrm{rk}(Q)$.*
2. *$\mathrm{rk}(Q_1 + Q_2) \leq \mathrm{rk}(Q_1) + \mathrm{rk}(Q_2)$.*
3. *If the L_i, L'_i in Eq. (2.3.2) form a set of $2r$ linearly independent polynomials, then $\mathrm{rk}(Q) = r$.*
4. *If Q has rank r, then*

$$\Pr_{x \sim \{0,1\}^n}[Q(x) \neq 0] \geq \frac{1}{2} - \frac{1}{2^{r+1}}.$$

We are now ready to prove the main technical lemma of this section that bounds the agreement between F_H and quadratic polynomials on the set $\overline{S}$.

Lemma 2.18. *Let $H : S \to \{0,1\}$ be any function. Then for every $q \in \mathcal{P}_{2,1}$, we have*

$$\Pr_{(x,y)\sim\overline{S}}[F_H(x,y) \neq q(x,y)] \geq \frac{1}{3} - o(1).$$

Proof. Fix H, and let $G(u,v)$ be the unique polynomial over $\mathbb{F}_2$ in the variables u, v that represents F_H. Define the set Ω as

$$\Omega := \{(u,v) \in \{0,1\}^{2n} |\ u_1 = 0 \text{ or } v_1 = 0\}.$$

We can rewrite the definition of $\overline{S}$ as

$$\overline{S} = \{(x,y) \in \{0,1\}^{2n} \mid\ u_1(x) = 0 \text{ or } v_1(y) = 0\},$$

and so if (x,y) is uniformly distributed in $\overline{S}$ then (u,v) is uniformly distributed in Ω. Thus, proving the statement of the lemma is equivalent to proving that for every quadratic $q(u,v)$,

$$\Pr_{(u,v)\sim\Omega}[G(u,v) \neq q(u,v)] \geq \frac{1}{3} - o(1).$$

For any $\alpha, \beta \in \{0,1\}$ such that $\alpha\beta \neq 1$, define $\Omega_{\alpha,\beta}$ as

$$\Omega_{\alpha,\beta} := \{(u,v) \in \{0,1\}^{2n} |\ u_1 = \alpha, v_1 = \beta\}.$$

Also let $G_{\alpha,\beta}(u_2, \ldots, u_n, v_2, \ldots, v_n)$ denote the polynomial obtained by setting $u_1 = \alpha, v_1 = \beta$ in $G(u,v)$, i.e., $G_{\alpha,\beta} = G|_{\Omega_{\alpha,\beta}}$, and for any quadratic $q(u,v)$, define

$$\Delta_{\alpha,\beta}(q) := \Pr_{(u,v)\sim\Omega_{\alpha,\beta}}[G_{\alpha,\beta}(u,v) + q(u,v) \neq 0]. \tag{2.3.3}$$

Noting that

$$\Omega = \Omega_{0,0} \cup \Omega_{0,1} \cup \Omega_{1,0},$$

and that

$$|\Omega_{0,0}| = |\Omega_{0,1}| = |\Omega_{1,0}| = \frac{|\Omega|}{3},$$

it follows that for every quadratic $q(u,v)$,

$$\Pr_{(u,v)\sim\Omega}[G(u,v) \neq q(u,v)] = \frac{1}{3}(\Delta_{0,0}(q) + \Delta_{0,1}(q) + \Delta_{1,0}(q)). \tag{2.3.4}$$

So, it suffices to prove that $\Delta_{0,0}(q) + \Delta_{0,1}(q) + \Delta_{1,0}(q) \geq 1 - o(1)$ for every q. To do this, we begin by analyzing the structure of the polynomials $G_{0,0}, G_{1,0}, G_{0,1}$.

The condition $u_1(x) = v_1(y) = 0$ in equivalent to both $|x|$ and $|y|$ being even, which implies that $|x| \cdot |y|$ is a multiple of 4. Eq. (2.3.1) tells us that F_H is the zero function when $|x| \cdot |y|$ is a multiple of 4, and so whenever $u_1(x) = v_1(y) = 0$, F_H must be zero. Since $G(u, v)$ represents F_H as a polynomial in u, v, $G(u, v)$ must be zero for any (u, v) with $u_1 = v_1 = 0$, i.e., for any $(u, v) \in \Omega_{0,0}$. Thus, $G_{0,0} = G|_{\Omega_{0,0}}$ is the zero polynomial.

Let us now consider points (x, y) which satisfy $u_1(x) = 0, v_1(y) = 1$. In this case, we know that $|x|$ must be even and $|y|$ must be odd. Thus, if $|x| \equiv 2 \pmod 4$ then $|x| \cdot |y| \equiv 2 \pmod 4$, and so Eq. (2.3.1) implies that $F_H(x, y) = 1$. Similarly, if $|x| \equiv 0 \pmod 4$ we can conclude that $F_H(x, y) = 0$. Furthermore, it may be noted that restricted to x such that $|x|$ is even,

$$|x| \equiv \begin{cases} 2 \pmod 4 & \text{if } |x|_1 = 1, \\ 0 \pmod 4 & \text{if } |x|_1 = 0. \end{cases}$$

Combining all these observations, we can conclude that

$$F_H|_{u_1(x)=0, v_1(y)=1}(x, y) = \begin{cases} 1 & \text{if } |x|_1 = 1, \\ 0 & \text{if } |x|_1 = 0. \end{cases}$$

Using Lemma 2.1, it follows that $F_H|_{u_1(x)=0, v_1(y)=1}(x, y) = S_2(x)$. Recalling that the points $(u, v) \in \Omega_{0,1}$ correspond to points (x, y) with $u_1(x) = 0, v_1(y) = 1$, and using Lemma 2.16, we can conclude that

$$G_{0,1} = G|_{\Omega_{0,1}} = u_2u_3 + u_4u_5 + \ldots + u_{n-1}u_n + L'(u)$$

for some linear function L'. A similar argument yields that

$$G_{1,0} = G|_{\Omega_{1,0}} = v_2v_3 + v_4v_5 + \ldots + v_{n-1}v_n + L''(v).$$

Part 3 of Lemma 2.17 then implies that $G_{0,1}$ and $G_{1,0}$ are rank $(n-1)/2$ quadratics in u and v respectively.

Note that any quadratic $q(u, v)$ can be written as

$$q(u, v) = c + L(u, v) + B(u, v) + Q(u) + R(v),$$

where c is the constant term, $L(u, v)$ is the homogeneous degree 1 part, $B(u, v)$ is the bilinear[3] part, and $Q(u)$ and $R(v)$ are homogeneous degree 2 polynomials in u and v respectively.

We can further expand $Q(u)$ as $u_1 L_1(u) + Q'(u)$ where $Q'(u)$ is the sum of all the terms in Q that don't involve u_1, and $L_1(u_2, \ldots, u_n)$ is a linear function. Similarly, we can expand $R(v)$ as $v_1 L_2(u) + R'(v)$. Thus, q can be expanded as

$$q(u, v) = c + L(u, v) + B(u, v) + u_1 L_1(u) + Q'(u) + v_1 L_2(v) + R'(v) \tag{2.3.5}$$

For any $\alpha, \beta \in \{0, 1\}$ such that $\alpha\beta \neq 1$, define $P_{\alpha,\beta}(u, v)$ to be the polynomial

$$P_{\alpha,\beta}(u, v) := G_{\alpha,\beta}(u, v) + q|_{u_1=\alpha, v_1=\beta}(u, v).$$

By substituting values of u_1 and v_1 in the expansion of q in Eq. (2.3.5), and recalling our observations about the structure of $G_{0,0}, G_{0,1}$, and $G_{1,0}$, it may be noted that

$$\begin{aligned}
P_{0,0}(u, v) &= 0 + c + L|_{u_1=0,v_1=0}(u, v) + B|_{u_1=0,v_1=0}(u, v) + Q'(u) + R'(v) \\
P_{0,1}(u, v) &= G_{0,1}(u) + c + L|_{u_1=0,v_1=0}(u, v) + B|_{u_1=0,v_1=1}(u, v) + Q'(u) + L_2(v) + R'(v), \\
P_{1,0}(u, v) &= G_{1,0}(v) + c + L|_{u_1=0,v_1=0}(u, v) + B|_{u_1=1,v_1=0}(u, v) + L_1(u) + Q'(u) + R'(v).
\end{aligned}$$

We can further simplify these expression as

$$\begin{aligned}
P_{0,0}(u, v) &= Q'(u) + R'(v) + B_{0,0}(u, v) \\
P_{0,1}(u, v) &= G_{0,1}(u) + Q'(u) + R'(v) + B_{1,0}(u, v), \\
P_{1,0}(u, v) &= G_{1,0}(v) + Q'(u) + R'(v) + B_{0,1}(u, v),
\end{aligned} \tag{2.3.6}$$

where $B_{0,0}, B_{0,1}$, and $B_{1,0}$ are quadratics in which every monomial contains at most one u variable and at most one v variable.

[3] degree 2 terms that contain exactly one variable from each of u and v, i.e., terms of the form $u_i v_j$

We are now ready to prove that $\Delta_{0,0}(q) + \Delta_{0,1}(q) + \Delta_{1,0}(q) \geq 1 - o(1)$ for every quadratic $q(u, v)$.

Fix $q(u, v)$ to be any quadratic. Noting that $P_{\alpha,\beta}$ is essentially a polynomial in the variables $u_2, \ldots, u_n, v_2, \ldots, v_n$, we can rewrite the definition of $\Delta_{\alpha,\beta}(q)$ given in Eq. (2.3.3) in the following two ways:

$$\begin{aligned}\Delta_{\alpha,\beta}(q) &= \mathbb{E}_{(u_2,\ldots,u_n)\sim\{0,1\}^{n-1}}\left[\Pr_{(v_2,\ldots,v_n)\sim\{0,1\}^{n-1}}[P_{\alpha,\beta}(u_2,\ldots,u_n,v_2,\ldots,v_n)\neq 0]\right], \\ &= \mathbb{E}_{(v_2,\ldots,v_n)\sim\{0,1\}^{n-1}}\left[\Pr_{(u_2,\ldots,u_n)\sim\{0,1\}^{n-1}}[P_{\alpha,\beta}(u_2,\ldots,u_n,v_2,\ldots,v_n)\neq 0]\right].\end{aligned}$$

The remaining proof is a case analysis based on the tuple $(\mathrm{rk}(Q'), \mathrm{rk}(R'))$ (recall that $\mathrm{rk}(Q)$ denotes the rank of the quadratic Q). Without loss of generality, we may assume that $\mathrm{rk}(Q') \leq \mathrm{rk}(R')$.

- **Case 1** ($\mathrm{rk}(Q') \geq (n-1)/4$): In particular, this implies that $\mathrm{rk}(R') \geq \mathrm{rk}(Q') \geq (n-1)/4$, and hence both Q' and R' are high-rank quadratics. Note that for every setting of the u variables, Eq. (2.3.6) tells us that $P_{0,0}$ simplifies to $R'(v)$ plus a linear function in v, and so its rank is at least $(n-1)/4$ by part 1 of Lemma 2.17. Part 4 of Lemma 2.17 combined with the alternate definition of $\Delta_{\alpha,\beta}(q)$ given above then implies that $\Delta_{0,0}(q) \geq (1/2) - o(1)$. A similar argument yields that both $\Delta_{1,0}(q)$ and $\Delta_{0,1}(q)$ are also at least $(1/2) - o(1)$, and so $\Delta_{0,0}(q) + \Delta_{1,0}(q) + \Delta_{0,1}(q) \geq 1 - o(1)$ in this case.

- **Case 2** ($\mathrm{rk}(Q') \leq (n-1)/4 \leq \mathrm{rk}(R')$): In this case, for every setting of the u variables, an analysis similar to that of case 1 yields that the quadratics $P_{0,0}$ and $P_{0,1}$ have rank at least $(n-1)/4$. Hence part 4 of Lemma 2.17 along with the alternate definition of $\Delta_{\alpha,\beta}(q)$ implies that $\Delta_{0,0}, \Delta_{0,1} \geq \frac{1}{2} - o(1)$, which implies that $\Delta_{0,0} + \Delta_{0,1} + \Delta_{1,0} \geq 1 - o(1)$.

- **Case 3** ($\mathrm{rk}(R') \leq (n-1)/4$): This means that $\mathrm{rk}(Q') \leq (n-1)/4$. Recall that we argued earlier that both $G_{0,1}$ and $G_{1,0}$ are quadratics of rank $(n-1)/2$. For any setting of the v variables, $P_{0,1}$ simplifies to the quadratic $G_{0,1}(u) + Q'(u)$

plus a linear function in u. Using part 2 of Lemma 2.17, and the fact that $Q' + Q' \equiv 0$ (we are working over $\mathbb{Z}/2\mathbb{Z}$), we have that

$$\begin{aligned} &\mathrm{rk}\left((G_{0,1} + Q') + Q'\right) \leq \mathrm{rk}(G_{0,1} + Q') + \mathrm{rk}(Q') \\ \implies &\mathrm{rk}(G_{0,1} + Q') \geq \mathrm{rk}(G_{0,1}) - \mathrm{rk}(Q') \\ \implies &\mathrm{rk}(G_{0,1} + Q') \geq \frac{n-1}{2} - \frac{n-1}{4} = \frac{n-1}{4} \end{aligned}$$

Part 1 of Lemma 2.17 then implies that $P_{0,1}$ is a quadratic of rank at least $(n-1)/4$ in the u variables, for every setting of the v variables. The alternate definition of $\Delta_{\alpha,\beta}(q)$ given above together with part 4 of Lemma 2.17 then gives us that $\Delta_{0,1}(q) \geq (1/2) - o(1)$.

A similar argument can be used to show that $P_{1,0}$ is a quadratic of rank at least $(n-1)/4$ in the v variables, for every setting of the u variables, implying that $\Delta_{1,0}(q) \geq (1/2) - o(1)$. Once again, we can conclude that $\Delta_{0,0}(q) + \Delta_{0,1}(q) + \Delta_{1,0}(q) \geq 1 - o(1)$.

Combining the fact that $\Delta_{0,0}(q) + \Delta_{0,1}(q) + \Delta_{1,0}(q) \geq 1 - o(1)$ for every q with Eq. (2.3.4) then implies the statement of Lemma 2.18. □

We now have all the tools needed to prove Theorem 2.13.

Proof of Theorem 2.13. Let $H : S \to \{0,1\}$ be any function that satisfies the statement of Lemma 2.15, and use H to define F_H as per Eq. (2.3.1).

Let $q \in \mathcal{P}_{2,1}$ be an arbitrary quadratic polynomial. Then Lemma 2.18 implies that

$$\Pr_{(x,y)\sim\bar{S}}[F_H(x,y) \neq q(x,y)] \geq \frac{1}{3} - o(1).$$

Also, our choice of H guarantees that

$$\Pr_{(x,y)\sim S}[F_H(x,y) \neq q(x,y)] = \Pr_{(x,y)\sim S}[H(x,y) \neq q(x,y)] \geq \frac{1}{2} - o(1).$$

Using the fact that $|S| = 2^{2n}/4$, we can combine the above equations to conclude that

$$\begin{aligned}
&\Pr_{(x,y)\sim\{0,1\}^{2n}}[F_H(x,y) \neq q(x,y)] \\
&= \left(\frac{3}{4} \cdot \Pr_{(x,y)\sim\bar{S}}[F_H(x,y) \neq q(x,y)]\right) + \left(\frac{1}{4} \cdot \Pr_{(x,y)\sim S}[F_H(x,y) \neq q(x,y)]\right) \\
&\geq \frac{3}{4} \cdot \left(\frac{1}{3} - o(1)\right) + \frac{1}{4} \cdot \left(\frac{1}{2} - o(1)\right) \geq \frac{3}{8} - o(1),
\end{aligned}$$

and so $\mathrm{agr}(F_H, q) \leq 5/8 + o(1)$. Since this holds for any $q \in \mathcal{P}_{2,1}$, it follows that

$$\gamma_{2,1}(F_H) \leq \frac{5}{8} + o(1) = \frac{1}{2} + \frac{1}{8} + o(1). \tag{2.3.7}$$

On the other hand, Lemma 2.14 implies that

$$\gamma_{2,2}(F_H) \geq \frac{3}{4} = \frac{1}{2} + \frac{1}{4}. \tag{2.3.8}$$

Eqs. (2.3.7) and (2.3.8) together give us the statement of Theorem 2.13.

□

2.3.3 Behavior of $\gamma_{d,k}(F)$ for functions with $\gamma_{d,1}(F) \approx 1/2$

In Section 2.3.2, we showed that there are examples of Boolean functions $F : \{0,1\}^n \to \{0,1\}$ for which $\gamma_{2,2}(F) > \gamma_{2,1}(F)$. However, these examples all had nontrivial agreement with polynomials from $\mathcal{P}_{2,1}$, i.e., from $\mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$; $\gamma_{2,1}(F)$ was approximately 5/8.

Recall that Lemma 2.6 says that $\gamma_{d,k}(F)$ is trivially at least 1/2. We say "trivially" because the constant polynomials in $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$, for all $k \geq 1$, can achieve agreement at least 1/2 with any Boolean function F. Thus, it makes sense to ask the following question: are there Boolean functions F such that degree d polynomials over $\mathbb{Z}/2\mathbb{Z}$ can't approximate F in the agreement sense any better than constant polynomials do, i.e., $\gamma_{d,1}(F) \approx 1/2$, whereas, for some $k > 1$ there are polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ of the same degree that have nontrivial agreement with F, i.e., $\gamma_{d,k}(F) \gg 1/2$.

To guide our search for a suitable function $F : \{0,1\}^n \to \{0,1\}$, we can make the following observations:

- Clearly, F must satisfy $\gamma_{d,1}(F) \leq (1/2) + o(1)$. This means that for every degree d polynomial P over $\mathbb{F}_2$, we have that $\mathrm{agr}(F, P) \leq (1/2) + o(1)$. It may be recalled that the *correlation* between F and P is related to the agreement in the following manner,

$$\mathrm{Corr}(F, P) = \left|\mathbb{E}_{x\sim\{0,1\}^n}\left[(-1)^{F(x)-G(x)}\right]\right| = \left|2 \cdot \mathrm{agr}(F, P) - 1\right|,$$

  and this implies that

$$\mathrm{Corr}(F, P) \leq o(1).$$

  In other words, F is uncorrelated with degree d *classical* polynomials over $\mathbb{F}_2$.

- F must also satisfy $\gamma_{d',k}(F) \geq \frac{1}{2} + \Omega(1)$ for some $d' < d$ and $k > 1$. Lemma 2.9 then implies that there must a nonclassical polynomial $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ of degree $\leq d' + k - 1$ and depth $\leq k$ such that

$$\mathrm{agr}\left(\frac{F}{2}, P\right) = \gamma_{d',k}(F) \geq \frac{1}{2} + \Omega(1).$$

  In this case, we can recall that

$$\mathrm{Corr}\left(\frac{F}{2}, P\right) \geq \left|2 \cdot \mathrm{agr}\left(\frac{F}{2}, P\right) - 1\right| \geq \Omega(1).$$

  If we further constrain ourselves to d', k such that $d' + k - 1 < d$ then the above equation implies that there is a nonclassical polynomial over $\mathbb{F}_2$ of degree at most d that has nontrivial correlation with F.

Combining the two observations, we can note that if we constrain ourselves to choosing d, d' and k that satisfy $d' + k - 1 < d$, then F must be a counterexample to the Inverse Conjecture for the Gowers norm. As mentioned before, it was established by Lovett et al. [LMS11] and Green and Tao [GT09] that $S_{2^\ell}(x)$, the elementary symmetric polynomial over $\mathbb{F}_2$ of degree 2^ℓ, is a counterexample to the conjecture for $\ell \geq 2$, and so we work with this function.

We know from Theorem 2.2 that

$$\gamma_{2^\ell-1,1}(S_{2^\ell}) \leq \frac{1}{2} + o(1), \tag{2.3.9}$$

We observe that $S_{2^\ell,3}(x)$, the 3-lift of S_{2^ℓ}, has nontrivial agreement with degree 3 polynomials over $\mathbb{Z}/8\mathbb{Z}$. This is the main technical result of this section.

Lemma 2.19. *Let $\ell \geq 2$ and $d = 2^{\ell-1} + 2^{\ell-2}$. Then*

$$\gamma_{d,3}(S_{2^\ell}) \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

For $\ell = 2$, Lemma 2.19 gives us that

$$\gamma_{3,3}(S_4) \geq \frac{1}{2} + \frac{1}{8} - o(1).$$

While Eq. (2.3.9) instantiated for $\ell = 2$ gives

$$\gamma_{3,1}(S_4) \leq \frac{1}{2} + o(1).$$

Together, the two equations answer the question raised at the beginning of this section:

Theorem 2.20. *There is a Boolean function $F : \{0,1\}^n \to \{0,1\}$ such that*

$$\gamma_{3,1}(F) \leq \frac{1}{2} + o(1)$$

while

$$\gamma_{3,3}(F) \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

We now give a brief sketch of the main idea behind the proof of Lemma 2.19. For the sake of clarity, we focus on the $\ell = 2$ case. Recall that for $x \in \{0,1\}^n$, $S_4(x)$ is equal to $|x|_2$, the 3^{rd} least significant bit of $|x|$. Consider the quantity $\binom{|x|}{3}$. It can be shown that modulo 2, $\binom{|x|}{3}$ only depends on $|x|_1$ and $|x|_0$, and is uncorrelated with $|x|_2$ (see Section 2.2.1 for more details). In other words, it has no "information" about $|x|_2$.

The key insight here is that modulo 8, $\binom{|x|}{3}$ has nontrivial "information" about $|x|_2$. This then can be used to show that the function $\binom{|x|}{3}$ mod 8 has good agreement with

$S_{4,3}(x)$, the 3-lift of S_4, since the latter is completely determined by $|x|_2$. Note that the function $\binom{|x|}{3}$ mod 8 can be represented as a polynomial of degree 3 over $\mathbb{Z}/8\mathbb{Z}$; if

$$P(x) = \sum_{T\subseteq[n];|T|=3} \prod_{i\in T} x_i,$$

i.e. the elementary symmetric polynomial of degree 3 over $\mathbb{Z}/8\mathbb{Z}$, then for all $x \in \{0,1\}^n$,

$$P(x) = \binom{|x|}{3} \bmod 8.$$

Understanding what "information" $P(x)$ has about $|x|_2$, and thus about $S_{4,3}(x)$, comes down to computing the largest power of 2 that divides $\binom{|x|}{3}$, and we use a classic theorem of Kummer to do this analysis[4].

We now give the formal details of the proof.

Proof of Lemma 2.19. Fix $\ell \geq 2$ and let $d = 2^{\ell-1} + 2^{\ell-2}$. Assume n is much larger than 2^ℓ. Lemma 2.1 from Section 2.2 tells us that $S_{2^\ell}(x) = |x|_\ell$. Thus, $S_{2^\ell,3}(x) \in \mathbb{Z}/8\mathbb{Z}[x_1,\ldots,x_n]$, the 3-lift of $S_{2^\ell}(x)$, is given by

$$S_{2^\ell,3}(x) = \begin{cases} 4 & \text{if } |x|_\ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.3.10}$$

Let P be the elementary symmetric polynomial of degree d in $\mathbb{Z}/8\mathbb{Z}[x_1,\ldots,x_n]$, i.e.,

$$P(x) = \sum_{T\subseteq[n];|T|=d} \prod_{i\in T} x_i.$$

To prove the theorem, it suffices to show that

$$\mathrm{agr}(S_{2^\ell,3}, P) = \Pr_{x\sim\{0,1\}^n}[P(x) = S_{2^\ell,3}(x)] \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

[4]It is known that the linear polynomial $\sum_{i=1}^n x_i$ in $\mathcal{P}_{1,3}$ has nontrivial correlation with $S_{4,3}(x)$ (see, e.g., [LMS11, Section 1.3]). However, this polynomial cannot have agreement better than 1/4 with $S_{4,3}(x)$ since it is $\{0,4\}$-valued on only about a 1/4 fraction of the points in $\{0,1\}^n$, whereas $S_{4,3}(x)$ takes values in $\{0,4\}$ on all the points in $\{0,1\}^n$ by the definition of a 3-lift. This is why we cannot simply use $\sum_{i=1}^n x_i$ and need to find another suitable polynomial over $\mathbb{Z}/8\mathbb{Z}$.

It may be noted that

$$P(x) = \binom{|x|}{d} \bmod 8,$$

and so

$$P(x) = \begin{cases} 0 & \text{if 8 divides } \binom{|x|}{d} \\ 4 & \text{if 4 divides } \binom{|x|}{d} \text{ but 8 does not} \end{cases} \tag{2.3.11}$$

To understand the behavior of P on different inputs, we use the following theorem due to Kummer (see, e.g., [Knu97, Section 1.2.6, Ex. 11]) that determines the largest power of a prime p that divides a binomial coefficient (we state it for $p = 2$).

Theorem 2.21 (Kummer)**.** *Let $0 \leq b \leq a$. Suppose r is the largest integer such that 2^r divides $\binom{a}{b}$. Then r is equal to the number of borrows required when subtracting b from a in base 2.*

Let $B_d(x)$ be the number of borrows required when subtracting d from $|x|$. Rewriting Eq. (2.3.11) in terms of $B_d(x)$ using Kummer's theorem, we get

$$P(x) = \begin{cases} 4 & \text{if } B_d(x) = 2 \\ 0 & \text{if } B_d(x) \geq 3 \end{cases} \tag{2.3.12}$$

The following lemma uses Eqs. (2.3.10) and (2.3.12) to provide sufficient conditions for $P(x)$ to agree with $S_{2^\ell,3}(x)$.

Lemma 2.22. *$P(x) = S_{2^\ell,3}(x)$ if either*

1. *$|x|_{\ell-2} = 0$, or*
2. *$(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1, 0, 0, 0)$.*

Proof. Since $d = 2^{\ell-1} + 2^{\ell-2}$, all the bits of d except $d_{\ell-1}$ and $d_{\ell-2}$ are zero. Thus, when subtracting d from $|x|$, no borrows are required by the bits $|x|_i$, $0 \leq i \leq \ell - 3$.

Using the above observation, it immediately follows that when

$$(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1, 0, 0, 0),$$

the number of borrows required is at least 3 i.e., $B_d(x) \geq 3$, which in turn implies that $P(x) = 0$. Since $|x|_\ell = 0$, it must be the case that $S_{2^\ell,3}(x) = 0$, which proves that the second condition in the statement of the lemma implies that $P(x) = S_{2^\ell,3}(x)$.

To prove that the first condition also implies agreement between P and $S_{2^\ell,3}$, suppose $|x|_{\ell-2} = 0$. Since $d_{\ell-1} = d_{\ell-2} = 1$, it follows that both $|x|_{\ell-2}$ and $|x|_{\ell-1}$ will need to borrow when subtracting d from $|x|$. As argued before, no borrows are required by the bits before (i.e., less significant than) $|x|_{\ell-2}$, and thus the total number of borrows required by the bits $|x|_i$ for $0 \leq i \leq \ell - 1$ is 2.
Noting that the bit $|x|_{\ell-1}$ always borrows from $|x|_\ell$, consider the following case analysis:

- Case $|x|_\ell = 1$: $|x|_\ell$ will not need to borrow since $d_\ell = 0$. In fact, none of the bits after (i.e., more significant than) $|x|_\ell$ will need to borrow, and thus $B_d(x) = 2$. This implies that $P(x) = 4$. We also have $S_{2^\ell,3}(x) = 4$ (since $|x|_\ell = 1$) and so $P(x) = S_{2^\ell,3}(x)$.

- Case $|x|_\ell = 0$: $|x|_\ell$ will require a borrow and this means $B_d(x) \geq 3$. This implies that $P(x) = 0$. Since $|x|_\ell = 0$, we have that $S_{2^\ell,3}(x) = 0$ and it follows that $P(x) = S_{2^\ell,3}(x)$.

This completes the proof. □

By Lemma 2.22, we have

$$\begin{aligned} &\mathrm{agr}(S_{2^\ell,3}, P) = \Pr[P(x) = S_{2^\ell,3}(x)] \\ &\geq \Pr[|x|_{\ell-2} = 0] + \Pr\left[(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1,0,0,0)\right] \end{aligned} \tag{2.3.13}$$

We now recall that if x is uniformly distributed in $\{0,1\}^n$ then for large enough n, Lemma 2.4 from Section 2.2 tells us that the bits $\{|x|_i\}_{0 \leq i \leq \ell+1}$ are almost uniformly and almost independently distributed in $\{0,1\}$. This gives us that

$$\begin{aligned} &\Pr_{x \sim \{0,1\}^n}[|x|_{\ell-2} = 0] \geq \frac{1}{2} - o(1) \\ &\Pr_{x \sim \{0,1\}^n}[(|x|_{\ell-2}, |x|_{\ell-1}, |x|_\ell, |x|_{\ell+1}) = (1,0,0,0)] \geq \frac{1}{16} - o(1), \end{aligned}$$

which, together with Eq. (2.3.13), imply that

$$\mathrm{agr}(S_{2^\ell,3}, P) = \Pr[P(x) = S_{2^\ell,3}(x)] \geq \frac{1}{2} + \frac{1}{16} - o(1).$$

□

2.4 Bounds on $\gamma_{d,k}(\mathrm{Maj}_n)$

Recall that Maj_n denotes the *majority function* on n bits: $\mathrm{Maj}_n(x_1, \ldots, x_n)$ takes the value 1 if strictly more than half of the n bits $x_1, \ldots, x_n$ are equal to 1, otherwise it takes the value 0. The goal of this section is to study how well polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ can approximate Maj_n in the agreement sense. We begin by discussing what we know for $k = 1$.

The classic works of Szegedy [Sze89] and Smolensky [Smo93] show that for every degree d polynomial $P \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ the agreement between P and Maj_n is bounded as follows.

Theorem 2.23 (Szegedy [Sze89], Smolensky [Smo93])**.** *For any polynomial $P \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ of degree $d \geq 0$,*

$$\mathrm{agr}(\mathrm{Maj}_n, P) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}.$$

An equivalent way of stating the above statement is that for all $d \geq 0$,

$$\gamma_{d,1}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d)}{\sqrt{n}}. \tag{2.4.1}$$

Although there are polynomials of degree $\Theta(\sqrt{n})$ over $\mathbb{Z}/2\mathbb{Z}$ that have nontrivial agreement with Maj_n (see, e.g., [Vio09, Section 2.3]), Eq. (2.4.1) implies that no polynomial over $\mathbb{Z}/2\mathbb{Z}$ of degree $d \ll \sqrt{n}$ can have nontrivial agreement with Maj_n. Then, a natural question to ask is whether there is a $k > 1$ and $d \ll \sqrt{n}$ such that the k-lift of Maj_n has good agreement with a degree d polynomial over $\mathbb{Z}/2^k\mathbb{Z}$. We will later prove that this is impossible and the same bound as in Eq. (2.4.1) holds for

$\gamma_{d,k}(\mathrm{Maj}_n)$ for $k > 1$. We begin by establishing a weaker yet nontrivial upper bound for the majority function by proving a more general statement that upper-bounds $\gamma_{d,k}(F)$ for all F.

2.4.1 A nontrivial upper bound on $\gamma_{d,k}(F)$

Let F be any Boolean function. Implicit in the work of Bhowmick and Lovett [BL15, Proof of Lemma 2.2] is an interesting observation that the agreement between F_k and polynomials of degree d over $\mathbb{Z}/2^k\mathbb{Z}$ can be no larger than the maximum possible agreement between F and polynomials of degree $d(2^k - 1)$ over $\mathbb{Z}/2\mathbb{Z}$. More formally,

Theorem 2.24 (Implicit in Bhowmick and Lovett [BL15])**.** *For $d \geq 0, k \geq 1$, and any Boolean function F, $\gamma_{d,k}(F) \leq \gamma_{d',1}(F)$ for $d' = d(2^k - 1)$.*

We now present a proof of the result based on ideas from [BL15, Proof of Lemma 2.2].

Proof of Theorem 2.24. Suppose $\gamma_{d,k}(F) = \gamma$. Then there must be a polynomial $P \in \mathcal{P}_{d,k}$ such that $\mathrm{agr}(F_k, P) = \gamma$. Let $\mathcal{A} \subseteq \{0,1\}^n$ denote the set of points where F_k and P agree; $|\mathcal{A}| = \gamma \cdot 2^n$. We can naturally think of $P(x_1, \ldots, x_n)$ as an integer polynomial with coefficients in $\{0, \ldots, 2^k - 1\}$. Furthermore, we can write P as

$$P(x) = M_1(x) + \ldots M_t(x),$$

where each $M_i(x)$ is a monomial of degree d of the form $M_i(x) = \prod_{j \in S_i} x_j$ for some $S_i \subseteq [n]$ with $|S_i| = d$. It is possible that $M_i(x) = M_{i'}(x)$ for $i \neq i'$, since we might have to use multiple copies of the same monomial if it originally appeared in P with a coefficient larger than 1. It may be noted that for all $x \in \mathcal{A}$,

$$M_1(x) + \ldots + M_t(x) \equiv F_k(x) \pmod{2^k}. \tag{2.4.2}$$

Let us define a Boolean function $H : \{0,1\}^t \to \{0,1\}$ as follows. For all $y =$

$(y_1, \ldots, y_t) \in \{0,1\}^t$,

$$H(y_1, \ldots, y_t) := \begin{cases} 1 & \text{if } |y| \equiv 2^{k-1} \pmod{2^k}, \\ 0 & \text{if } |y| \equiv 0 \pmod{2^k}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4.3}$$

where $|y|$ denotes the Hamming weight of y.

Note that for every $x \in \{0,1\}^n$ and $i \in [t]$, $M_i(x)$ is either 0 or 1, and so we can define the Boolean function $H' : \{0,1\}^n \to \{0,1\}$ as

$$H'(x_1, \ldots, x_n) := H(M_1(x), \ldots, M_t(x)) \tag{2.4.4}$$

for all $x \in \{0,1\}^n$.

It follows from the definition of H' and Eqs. (2.4.2) and (2.4.3) that for all $x \in \mathcal{A}$,

$$H'(x_1, \ldots, x_n) := \begin{cases} 1 & \text{if } F_k(x) \equiv 2^{k-1} \pmod{2^k}, \\ 0 & \text{if } F_k(x) \equiv 0 \pmod{2^k}. \end{cases}$$

Using the definition of F_k, we may conclude that $H'(x) = F(x)$ for all $x \in \mathcal{A}$, and so $\mathrm{agr}(F, H') \geq \gamma$ (recall that $|\mathcal{A}| = \gamma \cdot 2^n$). We will now show that $H'(x)$ can be represented as a polynomial of degree $d' = d(2^k - 1)$ over $\mathbb{Z}/2\mathbb{Z}$, which would imply that F has agreement at least γ with a polynomial in $\mathcal{P}_{d',1}$ and thus $\gamma_{d',1}(F) \geq \gamma_{d,k}(F)$. This would complete the proof.

Observe from Eq. (2.4.3) that $H(y_1, \ldots, y_n)$ only depends on the k least significant bits of the base 2 representation of $|y|$. Let $G : \{0,1\}^k \to \{0,1\}$ be the Boolean function such that for all $y \in \{0,1\}^t$,

$$H(y_1, \ldots, y_t) = G(|y|_0, |y|_1, \ldots, |y|_{k-1}).$$

We know from Lemma 2.1 that for any $\ell \geq 0$, $S_{2^\ell}(y) = |y|_\ell$ for all $y \in \{0,1\}^t$. Thus,

$$H(y_1, \ldots, y_t) = G(S_{2^0}(y), S_{2^1}(y), \ldots, S_{2^{k-1}}(y))$$

for all $y \in \{0,1\}^t$. Combining this equation with the definition of H' in Eq. (2.4.4), we get that for all $x \in \{0,1\}^n$,

$$H'(x_1, \ldots, x_n) = G\left(S_{2^0}\left(M_1(x), \ldots M_t(x)\right), \ldots, S_{2^{k-1}}\left(M_1(x), \ldots M_t(x)\right)\right).$$

For every $\ell \geq 0$, $S_{2^\ell}(M_1(x), \ldots, M_t(x))$ can be represented as a polynomial $Q_\ell(x)$ of degree $d \cdot 2^\ell$ over $\mathbb{Z}/2\mathbb{Z}$. We can also trivially represent $G(z_0, \ldots, z_{k-1})$ as a polynomial $Q(z_0, \ldots, z_{k-1})$ of degree at most k over $\mathbb{Z}/2\mathbb{Z}$. In the worst case, the monomial of largest degree in Q is $\prod_{i=0}^{k-1} z_i$, and hence in this case, the largest-degree monomial in the polynomial

$$Q'(x_1, \ldots, x_n) = Q(Q_0(x), Q_1(x), \ldots, Q_{k-1}(x))$$

would come from $\prod_{i=0}^{k-1} Q_i(x)$. The degree of any monomial in $\prod_{i=0}^{k-1} Q_i(x)$ is at most

$$d' = \sum_{i=0}^{k-1} d \cdot 2^i = d(2^k - 1),$$

and so the degree of $Q'(x)$ is at most d'. Noting that $Q'(x)$ represents $H'(x)$ then completes the proof. □

We can combine Theorem 2.24 with Eq. (2.4.1) to obtain the following upper bound for the majority function.

Theorem 2.25 (Implicit in Bhowmick and Lovett [BL15])**.** *For all $d \geq 0, k \geq 1$,*

$$\gamma_{d,k}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{O(d2^k)}{\sqrt{n}}.$$

Even though this bound is weaker than the one in Eq. (2.4.1), it can still establish that there is no polynomial in $\mathcal{P}_{d,k}$ that has nontrivial agreement with the k-lift of Maj_n for $k \ll \log n$ and $d \ll \sqrt{n}$. However, for $k = \Omega(\log n)$, the bound becomes trivial, and leaves open the possibility that there is a $k \geq c \cdot \log n$ and $d \ll \sqrt{n}$ such that some polynomial in $\mathcal{P}_{d,k}$ has nontrivial agreement with the k-lift of Maj_n. We now show that this is impossible.

2.4.2 A better upper bound on $\gamma_{d,k}(\mathrm{Maj}_n)$

The main result of this section is the following.

Theorem 2.26. *For any $k \geq 1, d \geq 0$,*

$$\gamma_{d,k}(\mathrm{Maj}_n) \leq \frac{1}{2} + \frac{10d}{\sqrt{n}}.$$

This matches the bound in Eq. (2.4.1) and shows that no matter how large k is, the agreement between polynomials in $\mathcal{P}_{d,k}$ and F_k cannot exceed the trivial bound of $1/2 + o(1)$. We now give a sketch of the proof. For the sake of clarity, we assume that n is even in the proof sketch.

We start by recalling a variant[5] of the arguments of Szegedy [Sze89] and Smolensky [Smo93] for upper bounding $\gamma_{d,1}(\mathrm{Maj}_n)$. Say that a polynomial P over $\mathbb{Z}/2\mathbb{Z}$ of degree d agrees with Maj_n on the points in $S_P \subseteq \{0,1\}^n$ where $|S_P| \geq ((1/2) + \varepsilon)\, 2^n$. We first find a non-zero degree D (D as small as possible) polynomial Q over $\mathbb{Z}/2\mathbb{Z}$ such that Q is zero at all points of $\overline{S_P}$. To be able to do this, we need to ensure that $\overline{S_P}$ is not an interpolating set for polynomials of degree D (see Section 2.2.2 for a definition). This can be done by choosing D so that the Hamming ball of radius D is larger than $|\overline{S_P}|$ (see part 4 of Lemma 2.5); in particular, choosing $D = (n/2) - \Theta(\varepsilon\sqrt{n})$ works.

Consider the polynomial $R = Q \cdot P$. On any input $x \in \mathrm{Maj}_n^{-1}(0)$, $R(x) = 0$ since either $x \notin S_P$, and hence $Q(x) = 0$, or $x \in S_P$, which implies that $P(x) = \mathrm{Maj}_n(x) = 0$. Secondly, since the Hamming ball of radius $(n/2) - 1$ around the all 1s vector is an interpolating set for Q (this follows from part 4 of Lemma 2.5), and Q is a non-zero polynomial, there must a point $x_0 \in \mathrm{Maj}_n^{-1}(1)$, such that $Q(x_0) \neq 0$. Also, since Q is zero on $\overline{S_P}$ it must be the case that $x_0 \in S_P$, and so $P(x_0) = \mathrm{Maj}_n(x_0) = 1$. Hence we have

$$R(x_0) = Q(x_0)P(x_0) \neq 0. \tag{2.4.5}$$

[5] This is essentially a "dual" view of their argument.

Therefore, R is a non-zero polynomial of degree at most $\deg(Q)+\deg(P)$ that vanishes at all points in $\mathrm{Maj}_n^{-1}(0)$, i.e., on all $x \in \{0,1\}^n$ of Hamming weight at most $n/2$. Appealing to the fact that Hamming balls of radius $n/2$ are interpolating sets for polynomials of degree at most $n/2$, this implies that $\deg(Q) + \deg(P) > n/2$, and hence $\varepsilon \leq O(d/\sqrt{n})$, finishing the proof of the theorem.

Let $\mathrm{Maj}_{n,k}$ denote the k-lift of the majority function. If we were to try and use the same approach as above for polynomials over $\mathbb{Z}/2^k\mathbb{Z}$, the problem we may run into is that Eq. (2.4.5) may not hold any more, since the product of two non-zero elements in $\mathbb{Z}/2^k\mathbb{Z}$ can be zero. In particular, it could be the case that $Q(x_0)$ is non-zero and even, and $P(x_0) = \mathrm{Maj}_{n,k}(x_0) = 2^{k-1}$, in which case their product is 0.

To overcome this, we instead try to find a Q that vanishes on $\overline{S}_P$ and moreover $Q(x_0)$ is *odd* for some $x_0 \in S_P$. We say that $\overline{S}_P$ is *forcing for degree D* polynomials if such a polynomial Q of degree D does not exist. Note that this notion is different from the notion of interpolating sets: every interpolating set is of course forcing, but the converse is not true (see Remark 2.27 below).

The main techinical lemma (Lemma 2.28) of this section gives a tight lower bound on the size of forcing sets for polynomials of degree D over $\mathbb{Z}/2^k\mathbb{Z}$, which lets us carry out the above argument and prove Theorem 2.26. Our proof of this lemma is an adaptation of techniques appearing in a work of Green [Gre00], who proved a similar result on the approximability of the parity function by polynomials over the ring $\mathbb{Z}/p^k\mathbb{Z}$, for prime $p \neq 2$. We now give formal details of the proof of Theorem 2.26.

We will need some more definitions and facts about $\mathcal{P}_{d,k}$. We use π to denote the unique ring homomorphism from $\mathbb{Z}/2^k\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z}$. Its kernel

$$\pi^{-1}(0) = \{a \in \mathbb{Z}/2^k\mathbb{Z} \mid 2^{k-1}a = 0\}$$

is the set of non-invertible elements in $\mathbb{Z}/2^k\mathbb{Z}$.

We call a set $S \subseteq \{0,1\}^n$ *forcing* for $\mathcal{P}_{d,k}$ if any polynomial $P \in \mathcal{P}_{d,k}$ that vanishes over S is forced to take a value in $\pi^{-1}(0)$ at all points $x \in \{0,1\}^n$. Formally, for all

$P \in \mathcal{P}_{d,k}$,

$$(\forall x \in S \;\; P(x) = 0) \Rightarrow (\forall y \in \{0,1\}^n \;\; \pi(P(y)) = 0).$$

Define the polynomial $\pi(P) \in \mathbb{Z}/2\mathbb{Z}[x_1, \ldots, x_n]$ to be the polynomial obtained by applying the map π to each of the coefficients of P. Since a multilinear polynomial in $\mathbb{Z}/2^k\mathbb{Z}[x_1, \ldots, x_n]$ is the zero polynomial iff it vanishes at all points of $\{0,1\}^n$ (see part 2 of Lemma 2.5), we see that S is forcing for $\mathcal{P}_{d,k}$ iff for all $P \in \mathcal{P}_{d,k}$,

$$(\forall x \in S \;\; P(x) = 0) \Rightarrow \pi(P) = 0.$$

Note that any interpolating set for $\mathcal{P}_{d,k}$ is forcing for $\mathcal{P}_{d,k}$, but the converse need not be true.

Remark 2.27. *Let $n = 3$, $d = 1$, and $k = 2$. Consider the set S consisting of all inputs of Hamming weight exactly 2 and the all 0s input. We first argue that S is not an interpolating set for degree 1 polynomials over $\mathbb{Z}/4\mathbb{Z}$. Consider the polynomial $P(x) = 2(x_1 + x_2 + x_3) \in \mathcal{P}_{1,2}$; P is a non-zero polynomial since $P(1,1,1) = 2$. Further, it can be verified that P vanishes on all points in S. This means that S is not an interpolating set for $\mathcal{P}_{1,2}$.*

We will now argue that S is forcing for $\mathcal{P}_{1,2}$. Note that for $a \in \mathbb{Z}/4\mathbb{Z}$, $\pi(a) = 0$ iff $2a = 0$ over $\mathbb{Z}/4\mathbb{Z}$. Consider an arbitrary $P \in \mathcal{P}_{1,2}$. We can write $P = a_0 + a_1x_1 + a_2x_2 + a_3x_3$. Since P vanishes over S, we have the following.

$$\begin{aligned} a_0 &= 0 \\ a_0 + a_1 + a_2 &= 0 \\ a_0 + a_1 + a_3 &= 0 \\ a_0 + a_2 + a_3 &= 0. \end{aligned} \tag{2.4.6}$$

Adding the last three equations and setting $a_0 = 0$ tells us that $2(a_1 + a_2 + a_3) = 2P(1,1,1) = 0$. This implies that $\pi_2(P(1,1,1)) = 0$. We can also easily derive that each of $2P(0,0,1), 2P(0,1,0)$, and $2P(1,0,0)$ are 0 as well. For example, for the case

of $P(0,0,1)$, we can do the following.

$$\begin{aligned} 2P(0,0,1) = 2a_3 &= 2a_3 + 4(a_1 + a_2) \\ &= 2(a_1 + a_2 + a_3) + 2(a_1 + a_2) \\ &= 2P(1,1,1) + 2(a_1 + a_2). \end{aligned}$$

We already showed that $2P(1,1,1) = 0$ and also know that $2(a_1 + a_2) = 0$ from Eq. (2.4.6), and thus, it follows that $2P(0,0,1) = 0$. A similar argument works for $2P(0,1,0)$ and $2P(1,0,0)$.
This shows that $2P(x) = 0$ for all $x \in \{0,1\}^3$ which implies that $\pi(P(x)) = 0$ for all x, and so S is a forcing set.

We now adapt the proof of Lemma 11 in [Gre00] to bound the size of forcing sets for $\mathcal{P}_{d,k}$.

Lemma 2.28. *Fix $d \geq 0, k \geq 1$. If S is forcing for $\mathcal{P}_{d,k}$, then $|S| \geq \binom{n}{\leq d}$.*

Proof. Assume for the sake of contradiction that $S \subseteq \{0,1\}^n$ is forcing for $\mathcal{P}_{d,k}$ and $|S| < \binom{n}{\leq d}$. Note that there must be a non-zero polynomial $Q(x) \in \mathbb{Q}[x_1, \ldots, x_n]$ of degree at most d satisfying $Q(x) = 0$ for all $x \in S$. To see why, recall that a multilinear polynomial Q of degree d over $\mathbb{Q}$ looks like

$$Q(x) = \sum_{T \subseteq [n]; |T| = d} c_T \prod_{i \in T} x_i,$$

where $c_T \in \mathbb{Q}$ for all T. We want to find a nontrivial assignment of values to the coefficients c_T such that $Q(s) = 0$ for all $s \in S$. Each $s \in S$ gives us an equation in the unknowns c_T, and so we have a homogeneous system of $|S| < \binom{n}{\leq d}$ equations over $\mathbb{Q}$ in $\binom{n}{\leq d}$ variables. This implies that there is a nontrivial assignment to the coefficients c_T such that $Q(s) = 0$ for all $s \in S$ though Q is a non-zero polynomial.

Having obtained the polynomial Q as above, we can rewrite it as

$$Q(x) = \frac{\alpha \cdot \tilde{Q}(x)}{\beta},$$

for some $\alpha, \beta \in \mathbb{Z}$, and an integer polynomial $\tilde{Q}(x) = \sum_{T \subseteq [n]; |T| \leq d} c'_T \prod_{i \in T} x_i$ such that the GCD of its coefficients $\{c'_T\}$ is 1.

Let $P(x) = \sum_{T \subseteq [n]; |T| \leq d} c''_T \prod_{i \in T} x_i$ be the polynomial in $\mathcal{P}_{d,k}$ obtained by choosing $c''_T \in \{0, \ldots, 2^k - 1\}$ for all T, such that $c''_T \equiv c'_T \pmod{2^k}$. It follows that P is a non-zero polynomial of degree at most d such that $\pi(P)$ is a non-zero polynomial, since if $\pi(P) = 0$ then every coefficent of P, and thus, every coefficent of $\tilde{Q}$, is divisible by two, which is impossible since the coefficients of $\tilde{Q}$ have no common divisors.

To complete the proof, observe that $P(x) = 0$ for all $x \in S$, and since S is forcing for $\mathcal{P}_{d,k}$, this implies that $\pi(P) = 0$, which is a contradiction.

□

We now use Lemma 2.28 to prove Theorem 2.26.

Proof of Theorem 2.26. We assume throughout that $1 \leq d \leq \sqrt{n}/10$; otherwise, there is nothing to prove. Let $\mathrm{Maj}_{n,k} : \{0,1\}^n \to \mathbb{Z}/2^k\mathbb{Z}$ be the k-lift of the majority function. Let $P \in \mathcal{P}_{d,k}$ be arbitrary and let $S_P = \{x \in \{0,1\}^n \mid P(x) = \mathrm{Maj}_{n,k}(x)\}$. We want to show that

$$|S_P| \leq 2^n \cdot \left(\frac{1}{2} + \frac{10d}{\sqrt{n}}\right).$$

We will prove this by contradiction, and so we assume that

$$|S_P| > 2^n \cdot \left(\frac{1}{2} + \frac{10d}{\sqrt{n}}\right).$$

Let E_P be the complement of S_P, i.e., the set of points where P does not agree with $\mathrm{Maj}_{n,k}$. We have

$$|E_P| < 2^n \left(\frac{1}{2} - \frac{10d}{\sqrt{n}}\right). \tag{2.4.7}$$

We will try to find a degree D (for suitable $D \leq \lfloor n/2 \rfloor$) polynomial Q such that Q vanishes on all points in E_P but has the property that $Q(x)$ is a unit (i.e., $\pi(Q(x)) \neq 0$) for some $x \in \{0,1\}^n$. To be able to do this, we need the fact that E_P is not forcing for $\mathcal{P}_{D,k}$. By Lemma 2.28, if E_P is indeed forcing for $\mathcal{P}_{D,k}$, then

$$
\begin{aligned}
|E_P| \geq \binom{n}{\leq D} &= \sum_{i=0}^{D} \binom{n}{i} = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} - \sum_{i=D+1}^{\lfloor n/2 \rfloor} \binom{n}{i} \\
&\geq 2^{n-1} - (\lfloor n/2 \rfloor - D) \cdot \binom{n}{\lfloor n/2 \rfloor} \\
&\geq 2^n \cdot \left(\frac{1}{2} - \frac{2(\lfloor n/2 \rfloor - D)}{\sqrt{n}} \right) = 2^n \cdot \left(\frac{1}{2} - \frac{4d}{\sqrt{n}} \right)
\end{aligned}
$$

where the last equality follows if we choose $D = \lfloor n/2 \rfloor - 2d$. This contradicts the upper bound on the size of $|E_P|$ in Eq. (2.4.7). Hence, E_P cannot be forcing for $\mathcal{P}_{D,k}$, and in particular, we can find a Q that vanishes on E_P and also satisfies $\pi(Q(x)) \neq 0$ for some $x \in \{0,1\}^n$.

We now claim that $\pi(Q(x_0)) \neq 0$ for some x_0 of Hamming weight greater than $n/2$. To see this, consider the polynomial $Q_1 = \pi(Q)$ in $\mathcal{P}_{D,1}$. Q_1 must be a non-zero polynomial, since otherwise it would mean that all the coefficients of Q are even and so $Q(x)$ is even for all $x \in \{0,1\}^n$, which would be a contradiction — Q is guaranteed to be odd (i.e., a unit) on at least one point in $\{0,1\}^n$. It now follows that, by part 4 of Lemma 2.5, Q_1 is non-zero when restricted to the Hamming ball of radius $D < n/2$ around the all 1s vector. In particular, this implies that there is an x_0 of Hamming weight greater than $n/2$, where $Q_1(x_0)$ is non-zero, and hence $\pi(Q(x_0)) \neq 0$, or equivalently, $Q(x_0)$ is a unit in $\mathbb{Z}/2^k\mathbb{Z}$. Fix this x_0 for the remainder of the proof. Note that $x_0 \notin E_P$, and thus $x_0 \in S_P$, since Q vanishes on E_P. This implies that $P(x_0) = \mathrm{Maj}_{n,k}(x_0) = 2^{k-1}$.

Now, consider the polynomial $R(x) = Q(x) \cdot P(x)$. We first show that $R(x) = 0$ for all x of Hamming weight at most $n/2$. Consider any x such that $|x| \leq n/2$. If $x \in E_P$, then $R(x) = 0$ since $Q(x) = 0$. On the other hand, if $x \notin E_P$, then $x \in S_P$, and so $P(x) = \mathrm{Maj}_{n,k}(x) = 0$. Thus, R vanishes at all inputs of Hamming weight at most $n/2$.

Since the degree of R is at most

$$
\deg(Q) + \deg(P) = D + d = \left(\left\lfloor \frac{n}{2} \right\rfloor - 2d \right) + d \leq \left\lfloor \frac{n}{2} \right\rfloor - d,
$$

and R vanishes on all x of Hamming weight at most $n/2$, part 4 of Lemma 2.5 implies that R must be 0 everywhere. However, at x_0 we have that

$$R(x_0) = Q(x_0)P(x_0) = Q(x_0)\mathrm{Maj}_{n,k}(x_0) = 2^{k-1}Q(x_0) \neq 0,$$

and this yields the desired contradiction. □

2.5 On the conjectures of Bhowmick and Lovett

As mentioned earlier, Bhowmick and Lovett [BL15] were the first to study agreement between nonclassical polynomials over $\mathbb{F}_2$ and Boolean functions. In their work, they establish the following bound for the majority function.

Theorem 2.29 (Bhowmick and Lovett [BL15])**.** *For every $d \geq 0, k \geq 1$, the agreement between a nonclassical polynomial P over $\mathbb{F}_2$ of degree $\leq d$ and depth $\leq k$ and Maj_n, the majority function on n bits, is upper-bounded as follows.*

$$\mathrm{agr}\left(\frac{\mathrm{Maj}_n}{2}, P\right) \leq \frac{1}{2} + \frac{O(d2^k)}{\sqrt{n}}.$$

On the other hand, a much stronger bound is known for classical polynomials due to Szegedy [Sze89] and Smolensky [Smo93] as stated earlier in Theorem 2.23. Bhowmick and Lovett conjectured that their bound in Theorem 2.29 could be improved to match the bound in Theorem 2.23. Our results imply that their conjecture is true: Lemma 2.9 implies that any nonclassical polynomial of degree d and depth k can agree with the majority function on at most $\gamma_{d,k}(\mathrm{Maj}_n)$ points, and combining this with Theorem 2.26 gives us the following result.

Theorem 2.30. *For $d \geq 0, k \geq 1$, let P be any nonclassical polynomial over $\mathbb{F}_2$ of degree $\leq d$ and depth $\leq k$. Then,*

$$\mathrm{agr}\left(\frac{\mathrm{Maj}_n}{2}, P\right) \leq \frac{1}{2} + \frac{10d}{\sqrt{n}}.$$

Recall that there are Boolean functions F that cannot be nontrivially approximated by classical polynomials of degree d in the correlation sense but can be approximated in this sense by nonclassical polynomials of the same degree. For example, from the works of Alon and Beigel [AB01], Tao et al. [GT09], and Lovett et al. [LMS11], it is known that any classical polynomial P of degree at most 3 satisfies

$$\operatorname{Corr}(S_4, P) \leq o(1),$$

where S_4 is the elementary symmetric polynomial of degree 4. On the other hand, the work of Tao and Ziegler [TZ12] (see, e.g., Lovett et al. [LMS11, Section 1.3]) implies the existence of a nonclassical polynomial P' over $\mathbb{F}_2$ of degree at most 3 such that

$$\operatorname{Corr}\left(\frac{S_4}{2}, P'\right) \geq \Omega(1).$$

Bhowmick and Lovett [BL15] conjecture that this is not the case when considering agreement-based approximation: for every Boolean function F, the maximum possible agreement between nonclassical polynomials of degree d and F is the same as the maximum possible agreement between F and classical polynomials of the same degree. We now show that our results imply that this conjecture is false.

Consider the Boolean function $F = S_{2^4} = S_{16}$, i.e., the elementary symmetric polynomial of degree 16 over $\mathbb{Z}/2\mathbb{Z}$. Theorem 2.2 can be used to conclude the following bound on the agreement between F and classical polynomials: if P is a classical polynomial of degree at most 15 then

$$\operatorname{agr}(S_{16}, P) \leq \frac{1}{2} + o(1). \tag{2.5.1}$$

On the other hand, Lemma 2.19 instantiated for $\ell = 4$, together with Lemma 2.9 implies that there is a nonclassical polynomial P' over $\mathbb{F}_2$ of degree

$$2^{\ell-1} + 2^{\ell-2} + 3 - 1 = 2^3 + 2^2 + 2 = 14$$

such that

$$\operatorname{agr}\left(\frac{S_{16}}{2}, P'\right) \geq \frac{1}{2} + \frac{1}{16} - o(1). \tag{2.5.2}$$

We can now use Eqs. (2.5.1) and (2.5.2) to conclude that there is a counterexample to the conjecture of Bhowmick and Lovett discussed in the previous paragraph.

Theorem 2.31. *There is a Boolean function F such that for every classical polynomial P of degree at most* 15*, we have*

$$\operatorname{agr}(F, P) \leq \frac{1}{2} + o(1),$$

but there is a nonclassical polynomial P' of degree at most 15 *satisfying*

$$\operatorname{agr}\left(\frac{F}{2}, P'\right) \geq \frac{1}{2} + \Omega(1).$$

Chapter 3

Point-wise approximation by $\mathbb{R}/\mathbb{Z}$-valued polynomials

3.1 Introduction

In this chapter, we introduce and study a new notion of point-wise approximation by a generalization of nonclassical polynomials that we call *torus* polynomials. These polynomials are a generalization of nonclassical polynomials in the sense that they are real polynomials modulo one whose coefficients can be arbitrary real numbers, whereas the coefficients of nonclassical polynomials are of the form q/p^k for some integer $k > 0$, prime p, and integer $0 \leq q \leq p^k - 1$.

To introduce this new notion of point-wise approximation, it is imperative to first define a notion of distance between two points in $\mathbb{R}/\mathbb{Z}$. Recall that there is an isomorphism $\phi : \mathbb{R}/\mathbb{Z} \to \mathbb{T}$ between the groups $(\mathbb{R}/\mathbb{Z}, +)$ and $(\mathbb{T}, \times)$, where the latter is the set of all complex numbers on the unit circle equipped with complex multiplication. In particular, we have that for all $x \in \mathbb{R}/\mathbb{Z}$,

$$\phi(x) = \exp(2\pi i x).$$

Informally speaking[1], our notion of the distance $d(x, y)$ between points $x, y \in \mathbb{R}/\mathbb{Z}$ can be thought of as being proportional to the *shortest distance* between $\phi(x)$ and $\phi(y)$ along the circumference of the unit circle in the complex plane. For example, let $x = 1/8$ and $y = 7/8$ be two points in $\mathbb{R}/\mathbb{Z}$, then the shortest distance between $\phi(x) = \exp(\pi i/4)$ and $\phi(y) = \exp(3\pi i/4)$ along the circumference of the unit circle is

[1]The notion of distance will be formalized in Section 3.2.2.

$\pi/4$, and so $d(x,y) = c \cdot (\pi/4)$ for some absolute constant $c \in \mathbb{R}$.

With this notion of distance in mind, we say that a torus polynomial $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ *ϵ-approximates a Boolean function F in the point-wise sense*[2]

$$\forall x \in \{0,1\}^n,\ d\left(\frac{F(x)}{2}, P(x)\right) \leq \epsilon.$$

While the class of torus polynomials might seem unnatural at first, we remark that for the purpose of studying point-wise approximation of Boolean functions over the torus in the regime of parameters we work with (polylog(n) degree, and $\epsilon \geq n^{-O(1)}$), torus polynomials and nonclassical polynomials are equivalent: if a Boolean function F can be ϵ-approximated in the point-wise sense by a nonclassical polynomial of degree d then, trivially, it is also ϵ-approximable by a torus polynomial of degree d, and conversely, it can be shown that if a degree d torus polynomial ϵ-approximates F in the point-wise sense then there is a nonclassical polynomial of degree d' that ϵ'-approximates F where $d' = O(d \log n)$ and $\epsilon' = O(\epsilon)$.

Thus, all the results of this chapter can be similarly phrased in terms of point-wise approximation by nonclassical polynomials instead of torus polynomials. However, because torus polynomials are easier to describe (they are *arbitrary* real polynomials evaluated modulo 1) and more elegant (they can be defined in a *field-independent* manner), we believe that they are a better choice for stating our results, and serve as a convenient proxy for nonclassical polynomials.

We now discuss the motivation behind introducing and studying this notion of point-wise approximation by torus polynomials.

A major goal of complexity theory is to prove Boolean circuit lower bounds, i.e., find explicit Boolean functions that cannot be computed by small size circuits of a given type. Over the years, three general approaches have been developed to achieve this. The first approach is based on random restrictions. It applies to circuit classes

[2]Recall from Section 1.1.3 that we have to define an $\mathbb{R}/\mathbb{Z}$-valued version of F in order to consider its point-wise approximability by an $\mathbb{R}/\mathbb{Z}$-valued function. This is typically done by considering the function $F(x)/2$, which can be thought of as a $\{0, 1/2\}$-valued function, where $\{0, 1/2\} \subset \mathbb{R}/\mathbb{Z}$.

in which functions simplify when most inputs are fixed to random values. Classic examples are the proofs by Håstad that AC^0, i.e., polynomial size circuit families of constant depth consisting of AND, OR, and NOT gates, cannot compute or approximate the parity function [Hås87], and the shrinkage of De Morgan formulas (Boolean circuits consisting of AND, OR, and NOT gates whose underlying graph is a tree) under random restrictions [Hås98]. However, random restrictions don't seem to be useful against more powerful circuit classes such as $AC^0[\oplus]$ — the class of AC^0 circuits equipped with parity gates.

The second approach is based on representation/approximation by low-degree polynomials. Razborov [Raz87] and Smolensky [Smo87] used this approach to prove lower bounds for $AC^0[\oplus] = AC^0[2]$, and more generally for $AC^0[p]$ for any prime p (This is the class of AC^0 circuits that are allowed to have MOD_p gates [3]). This technique is based on showing that any function in the circuit class can be approximated (in the agreement sense) by a low-degree polynomial over the finite field $\mathbb{F}_p$. Then, functions that do not admit such an approximation are provably outside the circuit class. A classic example here is that the majority function cannot be approximated by a low-degree polynomial over $\mathbb{F}_p$, and thus cannot be computed by $AC^0[p]$. However, this method also breaks down when considering more powerful circuit classes such as $AC^0[6]$, and more generally ACC^0, i.e., AC^0 circuits with MOD_m gates where m is composite.

The third method involves designing nontrivial satisfiability algorithms and then using them along with classical tools from structural complexity theory (among other techniques and results) to prove circuit lower bounds against ACC^0 for functions in high complexity classes such as NEXP. Williams [Wil14b] used this approach to prove that $NEXP \not\subseteq ACC^0$, and this was subsequently extended to show that $NQP \not\subseteq ACC^0$ by Williams and Murray [MW18].

[3] a MOD_p gate outputs 1 if and only if the sum of its inputs is congruent to a non-zero value modulo p.

While the works of Williams and Murray are groundbreaking and prove nontrivial lower bounds against ACC^0, their proofs are not purely combinatorial[4]. One of the reasons why *non-uniform* models of computation such as circuits are studied in the first place is because they are believed to be more amenable to purely combinatorial approaches for proving lower bounds than uniform models. Thus, the above works, while being a major breakthrough in theoretical computer science, leave us with the desire for an approach to proving ACC^0 lower bounds that's more similar to the first two approaches mentioned above. Motivated by this, we focus on trying to use the second approach, namely the framework of representation/approximation by low-degree polynomials, to prove ACC^0 lower bounds.

The classic works of Yao [Yao85] and Beigel and Tarui [BT91] were the first works to explore this direction. In particular, they showed that ACC^0 functions can be represented as low-degree integer polynomials composed with functions of the form $\mathbb{Z} \to \{0, 1\}$. In fact, Beigel and Tarui mention in their work the problem of using their representation to prove ACC^0 lower bounds in a purely combinatorial or algebraic manner. While their characterization of ACC^0 plays a fundamental role in the works of Williams and Murray mentioned above, the problem mentioned by Beigel and Tarui still remains open.

To make the second approach work, we first have to find a suitable class of polynomials and a notion of approximation that can be used to approximate ACC^0 functions. We show that the class of torus polynomials and the notion of point-wise approximation introduced above are concrete candidates to achieve this. In particular, we observe that a strengthened version of a result of Green et al. [GKT92] that extends the work of Beigel and Tarui [BT91] implies that functions in ACC^0 can be point-wise approximated by low-degree torus polynomials.

This new characterization of ACC^0 raises a host of questions, the most remarkable

[4]By a combinatorial proof, we mean a proof that doesn't involve uniform models of computation – algorithms, turing machines, etc.

being the problem of finding an explicit Boolean function that cannot be point-wise approximated by low-degree torus polynomials — an answer to this question would imply an ACC^0 lower bound for the function. In this chapter, we take steps towards trying to resolve this question by initiating the study of point-wise approximation of Boolean functions by torus polynomials and proving some interesting results along the way.

3.1.1 Our results

The following are the main results of this chapter:

1. (Approximation of general classes of functions) We show that point-wise approximation by torus polynomials can be used to characterize some general classes of Boolean functions:

   - Using the notion of *modulus-amplifying polynomials* from the works of Yao [Yao85], Toda [Tod91], and Beigel and Tarui [BT91], we prove that if a Boolean function can be computed by a degree d polynomial over the field $\mathbb{F}_p$ (for a prime p) then it can also be ϵ-approximated in the point-wise sense by a torus polynomial of degree $O(d \log(1/\epsilon))$ for any $\epsilon > 0$.
   - Building off of the previous result, and a classic result of Razborov [Raz87] and Smolensky [Smo87], we show that any Boolean function that can be computed by an $\mathrm{AC}^0[p]$ circuit (for a prime p) can also be ϵ-approximated in the point-wise sense by a torus polynomial of degree $\mathrm{polylog}(n/\epsilon)$ for all $\epsilon > 0$.
   - Finally, we show that a stronger version of a result of Green et al. [GKT92] which builds on the classic work of Beigel and Tarui [BT91] can be used to generalize our previous result: if a function can be computed by an ACC^0 circuit then it can be ϵ-approximated by a torus polynomial of degree $\mathrm{polylog}(n/\epsilon)$ for $\epsilon > 0$.

2. (Upper and lower bounds for concrete functions) The above results suggest that an approach to proving ACC^0 lower bounds is to find an explicit function that is not approximable by low-degree torus polynomials in the point-wise sense. The majority function Maj_n is believed to be not computable in ACC^0, and so we investigate the approximability of the majority function and some other related functions by torus polynomials:

   - Noting from the above results that any function computable by ACC^0 circuits should be $(1/20n)$-approximable by torus polynomials of degree $\mathrm{polylog}(n)$, it suffices to show that Maj_n does not admit such an approximation, in order to prove that $\mathrm{Maj}_n \notin \mathrm{ACC}^0$. As a step towards proving this, we show that if any *symmetric* torus polynomial of degree d $(1/20n)$-approximates the majority function in the point-wise sense then $d = \Omega(\sqrt{n/\log n})$.
   - En route to proving the previous result, we also show that the *delta* functions[5] $\Delta_{n,w}$ can be $(1/20n)$-approximated by torus polynomials of degree d only if $d = \Omega(\sqrt{n/\log n})$.
   - Somewhat surprisingly, for relatively large values of ϵ ($\epsilon \geq (\mathrm{polylog}(n))^{-1}$), we show that the delta functions can be nontrivially approximated by torus polynomials. In particular, we show that for every $\epsilon > 0$, any delta function can be ϵ-approximated in the point-wise sense by a *symmetric* torus polynomial of degree $\mathrm{polylog}(n/\epsilon)\epsilon^{-1}$.

3.1.2 Organization

We begin with some preliminaries in Section 3.2. In Section 3.3, we prove approximation results for various classes of Boolean functions: Boolean functions computable by finite field polynomials, and the classes $\mathrm{AC}^0[p]$ and ACC^0. Finally, in Section 3.4.1,

[5]For $w \geq 0$, the delta function $\Delta_{n,w}(x)$ is defined as $\Delta_{n,w}(x) = 1 \Leftrightarrow |x| = w$.

we discuss concrete upper and lower bounds for the majority and delta functions.

3.2 Preliminaries

We carry over the preliminaries mentioned at the beginning of Section 2.2. Building on them, we naturally identify $\mathbb{F}_p$ with $\{0, \ldots, p-1\} \subset \mathbb{Z}$ without explicitly using inclusion or embedding maps.

We use polylog(n) to mean an arbitrary function of the form $\log^{O(1)}(n)$, and poly(n) to mean a function of the form $n^{O(1)}$.

3.2.1 Metrics and norms on $\mathbb{R}/\mathbb{Z}$

For an $x \in \mathbb{R}$, recall that x modulo 1, denoted by $x \bmod 1$, denotes the fractional part of x:

$$x \bmod 1 = x - \lfloor x \rfloor.$$

It may be observed that the modulo 1 function satisfies the following properties: for all $x, y \in \mathbb{R}$,

$$(x+y) \bmod 1 = ((x \bmod 1) + (y \bmod 1)) \bmod 1,$$

$$(x-y) \bmod 1 = ((x \bmod 1) - (y \bmod 1)) \bmod 1.$$

We think of the group $\mathbb{R}/\mathbb{Z}$ as the set $[0,1) \subset \mathbb{R}$ equipped with addition modulo 1. We will now define the following metric, denoted by $d_{\mathbb{T}}(x,y)$, on $\mathbb{R}/\mathbb{Z}$: for all $x, y \in \mathbb{R}/\mathbb{Z}$,

$$d_{\mathbb{T}}(x,y) := \min\left(|x-y|, 1-|x-y|\right). \tag{3.2.1}$$

It immediately follows from the definition that

$$0 \le d_{\mathbb{T}}(x,y) \le \frac{1}{2}.$$

To see that this is indeed a metric, we observe that for all $x, y \in \mathbb{R}/\mathbb{Z}$,

- $d_{\mathbb{T}}(x,y) \ge 0$,

- $d(x,y) = 0 \Leftrightarrow x = y$,
- $d(x,y) = d(y,x)$.
- and finally, for all $z \in \mathbb{R}/\mathbb{Z}$,

$$d_{\mathbb{T}}(x,y) \leq d_{\mathbb{T}}(x,z) + d_{\mathbb{T}}(y,z).$$

We will now prove this is in the lemma below.

Lemma 3.1. *Let $x, y \in \mathbb{R}/\mathbb{Z}$. Then for all $z \in \mathbb{R}/\mathbb{Z}$, $d_{\mathbb{T}}(x,y) \leq d_{\mathbb{T}}(x,z) + d_{\mathbb{T}}(y,z)$.*

Proof. Let $\alpha = x - z$ and $\beta = z - y$. Note that $0 \leq |\alpha|, |\beta| < 1$. Further more, $\alpha + \beta = x - y$, and so $0 \leq |\alpha + \beta| < 1$.
With these definitions, it follows from Eq. (3.2.1) that proving the desired statement is equivalent to proving

$$\min(|\alpha+\beta|, 1-|\alpha+\beta|) \leq \min(|\alpha|, 1-|\alpha|) + \min(|\beta|, 1-|\beta|), \tag{3.2.2}$$

where $0 \leq |\alpha|, |\beta|, |\alpha+\beta| < 1$. We will prove this using a case analysis on the tuple $(|\alpha|, |\beta|)$.
Case 1 ($|\alpha| \leq 1/2, |\beta| \leq 1/2$): In this case,

$$\begin{aligned}&\min(|\alpha|, 1-|\alpha|) + \min(|\beta|, 1-|\beta|) = |\alpha| + |\beta| \\ &\geq |\alpha+\beta| \geq \min(|\alpha+\beta|, 1-|\alpha+\beta|),\end{aligned}$$

and so we are done.
Case 2 ($|\alpha| > 1/2, |\beta| \leq 1/2$, or $|\alpha| \leq 1/2, |\beta| > 1/2$): In the case of the former,

$$\begin{aligned}&\min(|\alpha|, 1-|\alpha|) + \min(|\beta|, 1-|\beta|) = 1 - |\alpha| + |\beta| \\ &= 1 - (|\alpha| - |\beta|) \geq 1 - |\alpha+\beta| \qquad (\text{since } |\alpha+\beta| \geq |\alpha| - |\beta|) \\ &\geq \min(|\alpha+\beta|, 1-|\alpha+\beta|),\end{aligned}$$

and so we are done. The $|\alpha| \leq 1/2, |\beta| > 1/2$ case can be dealt with in a similar manner.

Case 3 ($|\alpha| > 1/2, |\beta| > 1/2$): Here, it must be the case that α and β have opposite signs, otherwise $|\alpha+\beta| = |\alpha|+|\beta| > 1$, which is impossible. Without loss of generality, assume that $|\alpha| > |\beta|$. Then we can write

$$|\alpha + \beta| = |\alpha| - |\beta| < \frac{1}{2},$$

where the last inequality follows from the fact that $|\alpha| < 1$ and $|\beta| > \frac{1}{2}$.
Then it follows that

$$\min(|\alpha + \beta|, 1 - |\alpha + \beta|) = |\alpha + \beta| = |\alpha| - |\beta|.$$

We also have that

$$\min(|\alpha|, 1 - |\alpha|) = 1 - |\alpha|,$$

$$\min(|\beta|, 1 - |\beta|) = 1 - |\beta|.$$

Now assume for the sake of contradiction that Eq. (3.2.2) is violated in this case. Based on the above discussion, this would imply that

$$|\alpha| - |\beta| > 2 - |\alpha| - |\beta| \Rightarrow |\alpha| > 1,$$

which is impossible, and so Eq. (3.2.2) could not have been violated in this case. This completes the proof. □

For $z \in \mathbb{R}/\mathbb{Z}$, we define the *torus norm*[6] of z, denoted by $\|z\|_{\mathbb{T}}$, to be

$$\|z\|_{\mathbb{T}} = d(z, 0).$$

It follows from the definition of $d_{\mathbb{T}}(x, y)$ that $\|z\|_{\mathbb{T}} \geq 0$ with the equality being true if and only if $z = 0$.
We can also show that the norm satisfies the following property that will help us prove the triangle inequality.

[6]Note that this is not a norm in the formal sense since it does not satisfy the property $\|\alpha z\|_{\mathbb{T}} = |\alpha| \|z\|_{\mathbb{T}}$ for all $z \in \mathbb{R}/\mathbb{Z}$, $\alpha \in \mathbb{R}$. However, as we shall see, the norm does satisfy the triangle inequality and the non-degeneracy condition, i.e., $\|z\|_{\mathbb{T}} = 0 \Leftrightarrow z = 0$, much like the sparsity norm ℓ_0.

Lemma 3.2. *Let $x, y \in \mathbb{R}/\mathbb{Z}$. Then $d_{\mathbb{T}}(x, y) = \|(x - y) \bmod 1\|_{\mathbb{T}}$.*

Proof. Let $z = (x - y) \bmod 1$. Note that it suffices to prove that $d_{\mathbb{T}}(x, y) = d_{\mathbb{T}}(z, 0)$ which in turn is equivalent to showing that

$$d_{\mathbb{T}}(z, 0) = \min(|x - y|, 1 - |x - y|).$$

Case I ($x > y$): In this case,

$$z = (x - y) \bmod 1 = x - y,$$

and so

$$d_{\mathbb{T}}(z, 0) = \min(|x - y|, 1 - |x - y|).$$

Case II ($x < y$): It follows immediately that

$$z = (x - y) \bmod 1 = 1 + x - y.$$

This implies that

$$\begin{aligned} d_{\mathbb{T}}(z, 0) &= \min(|1 + x - y|, 1 - |1 + x - y|) \\ &= \min(1 + x - y, y - x) \\ &= \min(1 - |x - y|, |x - y|). \end{aligned}$$

This completes the proof. □

An immediate consequence of this lemma is the following useful fact: for all $x, y \in \mathbb{R}/\mathbb{Z}$,

$$\|(x - y) \bmod 1\|_{\mathbb{T}} = d_{\mathbb{T}}(x, y) = d_{\mathbb{T}}(y, x) = \|(y - x) \bmod 1\|_{\mathbb{T}},$$

and thus for $z \in \mathbb{R}/\mathbb{Z}$, we have that $\|z\|_{\mathbb{T}} = \|-z \bmod 1\|_{\mathbb{T}}$.
The above lemma also implies the triangle inequality for the torus norm.

Lemma 3.3. *Let $x, y \in \mathbb{R}/\mathbb{Z}$, then*

$$\|(x+y) \ mod\ 1\|_{\mathbb{T}} \leq \|x\|_{\mathbb{T}} + \|y\|_{\mathbb{T}}.$$

Proof. Let $z \in \mathbb{R}/\mathbb{Z}$ such that $z = 1 - y$. Then we can write

$$\begin{aligned}
\|(x+y) \bmod 1\|_{\mathbb{T}} &= \|(x-z) \bmod 1\|_{\mathbb{T}} \\
&= d_{\mathbb{T}}(x, z) \qquad \text{(By Lemma 3.2)} \\
&\leq d_{\mathbb{T}}(x, 0) + d_{\mathbb{T}}(z, 0) \qquad \text{(By Lemma 3.1 and symmetry)} \\
&= \|x\|_{\mathbb{T}} + \min(|1-y|, 1-|1-y|) \\
&= \|x\|_{\mathbb{T}} + \min(y, 1-y) \\
&= \|x\|_{\mathbb{T}} + \|y\|_{\mathbb{T}}
\end{aligned}$$

□

We also need the following useful lemma which we will use in our approximation results.

Lemma 3.4. *Let $x, y > 0$ be real numbers, then*

$$\|(x+y) \ mod\ 1\ \|_{\mathbb{T}} \leq |x| + |y|.$$

Proof. Let $x' = x \bmod 1$ and $y' = y \bmod 1$. Observe that, since $x, y \geq 0$,

$$\|x'\|_{\mathbb{T}} = \min(|x \bmod 1|, 1 - |x \bmod 1|) \leq |x \bmod 1| \leq |x|.$$
$$\|y'\|_{\mathbb{T}} = \min(|y \bmod 1|, 1 - |y \bmod 1|) \leq |y \bmod 1| \leq |y|.$$

Then, using the triangle inequality, it follows that

$$\|(x+y) \bmod 1\ \|_{\mathbb{T}} = \|x' + y'\|_{\mathbb{T}} \leq \|x'\|_{\mathbb{T}} + \|y'\|_{\mathbb{T}} \leq |x| + |y|.$$

□

3.2.2 Torus polynomials and Boolean functions

We say that a function $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is a torus polynomial of degree at most d if it can be written as a real multilinear polynomial of degree d modulo one, i.e., for all $x \in \{0,1\}^n$,

$$P(x) = \left(\sum_{S \subseteq [n], |S| \leq d} c_S \prod_{i \in S} x_i \right) \mod 1, \tag{3.2.3}$$

where $c_S \in \mathbb{R}$. We can prove that every torus polynomial has a unique representation in the following sense:

Lemma 3.5. *Let $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ be a torus polynomial of degree at most d. Then for all $x \in \{0,1\}^n$,*

$$P(x) = \left(\sum_{S \subseteq [n], |S| \leq d} c_S \prod_{i \in S} x_i \right) \textit{ mod } 1,$$

where $c_S \in [0,1)$ are uniquely determined.

Proof. To prove the statement, it suffices to show that if for all $x \in \{0,1\}^n$,

$$\left(\sum_{S \subseteq [n], |S| \leq d} c_S \prod_{i \in S} x_i \right) \mod 1 = 0, \tag{3.2.4}$$

then $c_S \equiv 0 \pmod 1$ for all S. Suppose this is not true, then there must be an $S \subseteq [n]$ such that $c_S \not\equiv 0 \pmod 1$, and for all $T \subset S$, $c_T \equiv 0 \pmod 1$. Let $1_S \in \{0,1\}^n$ denote the characteristic vector of such an S. Substituting $x = 1_S$ in Eq. (3.2.4), we get

$$\left(c_S + \sum_{T \subset S} c_T \right) \mod 1 = 0.$$

Using the fact that $c_T \equiv 0 \pmod 1$ for all $T \subset S$, we can rewrite this as

$$c_S \mod 1 = 0,$$

which leads to a contradiction. This completes the proof. □

Let $F, G : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ and $\epsilon > 0$. Then we say that G *ϵ-approximates F in the point-wise sense* if for all $x \in \{0,1\}^n$,

$$d_{\mathbb{T}}\left(F(x), G(x)\right) = \left\| (F(x) - G(x)) \bmod 1 \right\|_{\mathbb{T}} \leq \epsilon.$$

Let $F : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then, for any $\epsilon > 0$, a torus polynomial $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is said to *ϵ-approximate F in the point-wise sense* if P ϵ-approximates the function $F/2$, i.e., for all $x \in \{0,1\}^n$,

$$\left\| \left(\frac{F(x)}{2} - P(x) \right) \bmod 1 \right\|_{\mathbb{T}} \leq \epsilon.$$

Intuitively, this means that for every $x \in \{0,1\}^n$, "the angle between" $\phi\left(F(x)/2\right)$ and $\phi(P(x))$ on the unit circle in the complex plane is "small". Here $\phi : \mathbb{R}/\mathbb{Z} \to \mathbb{T}$ is the obvious isomorphism between $\mathbb{R}/\mathbb{Z}$ and $\mathbb{T}$, the unit circle in the complex plane. Recall that the function $F/2$ used in the definition above is the $\mathbb{R}/\mathbb{Z}$-valued version of the Boolean function F that takes values in $\{0, 1/2\} \subset \mathbb{R}/\mathbb{Z}$.

3.2.3 Torus polynomials and nonclassical polynomials

As mentioned before, torus polynomials generalize the class of nonclassical polynomials. To further illustrate this, we now recall the global definition of nonclassical polynomials from the work of Tao and Zielger [TZ12].

Theorem 3.6 (Tao and Ziegler [TZ12])**.** *A function $Q : \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ is a nonclassical polynomial over $\mathbb{F}_p$ of degree at most d and depth k if and only if*

$$Q(x) = \left(\alpha + \sum_{0 \leq e_1, \ldots, e_n \leq p-1, k \geq 0; \sum_i e_i + (p-1)k \leq d} \frac{c_{e_1,\ldots,e_n,k}}{p^{k+1}} \prod_{i \in [n]} x_i^{e_i} \right) \bmod 1$$

where $c_{e_1,\ldots,e_n,k} \in \{0, 1, \ldots, p-1\}$, and $\alpha \in [0,1)$, are uniquely determined. The largest k such that $c_{e_1,\ldots,e_n,k} \neq 0$ for some $0 \leq e_1, \ldots, e_n \leq p-1$ is the depth, and α is called the shift.

Note that we only care about restrictions of nonclassical polynomials to $\{0,1\}^n \subset \mathbb{F}_p^n$. Abusing terminology, we will call a function $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ a nonclassical polynomial of degree d over $\mathbb{F}_p$ if it is the restriction of some nonclassical polynomial $Q : \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ of degree d to $\{0,1\}^n$, i.e., for all $x \in \{0,1\}^n$, $P(x) = Q(x)$. A useful characterization of such functions immediately follows from Theorem 3.6:

Corollary 3.7 (Corollary of Theorem 3.6)**.** *A function $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is a nonclassical polynomial of degree d over $\mathbb{F}_p$ (i.e., a restriction of some nonclassical polynomial $Q : \mathbb{F}_p^n \to \mathbb{R}/\mathbb{Z}$ of degree d to $\{0,1\}^n$) if and only if*

$$P(x) = \left(\alpha + \sum_{S \subseteq [n]; k \geq 0; |S| + (p-1)k \leq d} \frac{c_{S,k}}{p^{k+1}} \prod_{i \in S} x_i \right) \ mod\ 1, \tag{3.2.5}$$

where $\alpha \in [0,1)$, and $c_{S,k} \in \{0, \ldots, p-1\}$.

Eq. (3.2.5) implies that every nonclassical polynomial $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ of degree d over $\mathbb{F}_p$ can be written as a real polynomial $\hat{P}$ modulo 1 such that all the coefficients of $\hat{P}$ except the constant term are of the form q/p^k for some integer $k > 0$ and $0 \leq q \leq p^k - 1$. Comparing this to the definition of torus polynomials in Eq. (3.2.3) tells us that every such nonclassical polynomial P is also a torus polynomial of degree d, and so torus polynomials generalize the class of nonclasical polynomials in this sense.

Note that the structure and coefficients of nonclassical polynomials have a dependence on the field over which they are defined, i.e., the parameter p, and on the degree, i.e., the parameter d. In contrast, torus polynomials have a significantly simpler definition because they can be defined in a field-independent manner and their coefficients do not have a dependence on the degree, and thus are more convenient to work with. However, the goal of this dissertation is to understand the approximation power of nonclassical polynomials, and so it is natural to ask how the study of point-wise approximation by torus polynomials brings us closer to this goal. We will now show that for the regime of parameters we work with, point-wise approximation by torus polynomials is equivalent to point-wise approximation by nonclassical polynomials.

The following lemma shows that torus polynomials can be point-wise approximated by nonclassical polynomials.

Lemma 3.8. *Let $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ be a torus polynomial of degree at most d.*

Then, for every prime p and $\epsilon > 0$, there exists a nonclassical polynomial Q over $\mathbb{F}_p$ of degree at most $O(dp \log p \log(n/\epsilon))$ that ϵ-approximates P in the point-wise sense.

Proof. Fix a prime p and $\epsilon > 0$. Suppose that

$$P(x) = \left(\sum_{S \subseteq [n], |S| \leq d} c_S \prod_{i \in S} x_i \right) \bmod 1.$$

We can assume without loss of generality that $c_S \in [0, 1)$ for all S. We now approximate each c_S using p-adic rationals.

Let $t \geq 1$ be a parameter that we will choose later, and let $S \subseteq [n]$ such that $|S| \leq d$. We can always choose non-negative integers $c_{S,k} \in \{0, \ldots, p-1\}$ for $1 \leq k \leq t$ such that

$$0 \leq c_S - \sum_{1 \leq k \leq t} \frac{c_{S,k}}{p^k} < p^{-t}.$$

Define the torus polynomial

$$Q(x) := \left(\sum_{S \subseteq [n]; |S| \leq d; 1 \leq k \leq t} \frac{c_{S,k}}{p^k} \prod_{i \in S} x_i \right) \bmod 1,$$

By comparing the form of $Q(x)$ to that of a nonclassical polynomial in Eq. (3.2.5), we can conclude that Q is a nonclassical polynomial of degree at most $d + (p-1)(t-1)$. Furthermore, using Lemma 3.4, for every $x \in \{0,1\}^n$,

$$\begin{aligned}
\|(P(x) - Q(x)) \bmod 1\|_{\mathbb{T}} &= \left\| \left(\sum_{S \subseteq [n]; |S| \leq d} \left(c_S - \sum_{1 \leq k \leq t} \frac{c_{S,k}}{p^k} \right) \prod_{i \in S} x_i \right) \bmod 1 \right\|_{\mathbb{T}} \\
&\leq \sum_{S \subseteq [n]; |S| \leq d} \left| \left(c_S - \sum_{1 \leq k \leq t} \frac{c_{S,k}}{p^k} \right) \prod_{i \in S} x_i \right| \\
&\leq \binom{n}{\leq d} p^{-t} \\
&\leq \epsilon \qquad \text{(By choosing } t = O\left(d \log(n/\epsilon) \log p\right).
\end{aligned}$$

Noting that the degree of Q is $O(dp \log p \log(n/\epsilon))$ completes the proof. □

For our purposes, it suffices to work with $d = O(\text{polylog}(n))$, $\epsilon \geq n^{-O(1)}$, and $p = O(1)$, and so we have the following result that follows from the above lemma and the triangle inequality.

Lemma 3.9. *For every $\epsilon \geq n^{-O(1)}$ and prime $p = O(1)$, if there a torus polynomial of degree $O(\text{polylog}(n))$ that ϵ-approximates a Boolean function $F : \{0,1\}^n \to \{0,1\}$ in the point-wise sense then there is also a nonclassical polynomial of degree $O(\text{polylog}(n))$ over $\mathbb{F}_p$ that 2ϵ-approximates F in the point-wise sense.*

Thus, for the regime our parameters we work with, point-wise approximation by torus polynomials serves as an elegant proxy for point-wise approximation by nonclassical polynomials.

3.2.4 Correlation and point-wise approximation

Recall that the correlation between two $\mathbb{R}/\mathbb{Z}$-valued functions $F, G : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is defined as

$$\text{Corr}(F,G) = \left|\mathbb{E}_{x\sim\{0,1\}^n}\left[e^{2\pi i F(x)} e^{-2\pi i G(x)}\right]\right|,$$

and we say that F ϵ-approximates G in the correlation-sense if $\text{Corr}(F,G) \geq \epsilon$. We will now show that point-wise approximation is stronger than correlation-based approximation in the sense that if F approximates G well in the point-wise sense then F also approximates G well in the correlation-sense.

Lemma 3.10. *Suppose that $F : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ ϵ-approximates $G : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ in the point-wise sense for some $\epsilon < 1/2\pi$. Then $\text{Corr}(F,G) \geq 1 - 4\pi^2\epsilon^2$.*

Proof. Let $H : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ be the function defined as

$$H(x) := (F(x) - G(x)) \bmod 1.$$

Since F ϵ-approximates G in the point-wise sense, for all $x \in \{0,1\}^n$, we have that $\|H(x)\|_{\mathbb{T}} \leq \epsilon$, and so by the definition of the torus norm,

$$\min\left(H(x), 1 - H(x)\right) \leq \epsilon. \tag{3.2.6}$$

We can write

$$\begin{aligned}\mathrm{Corr}(F,G) &= \left|\mathbb{E}_{x\sim\{0,1\}^n}\left[e^{2\pi i(F(x)-G(x))}\right]\right| \\ &= \left|\mathbb{E}_{x\sim\{0,1\}^n}\left[e^{2\pi iH(x)}\right]\right| \\ &\geq \left|\mathbb{E}_{x\sim\{0,1\}^n}[\cos(2\pi H(x))]\right| \qquad (3.2.7)\end{aligned}$$

Let $\theta_x = \min(H(x), 1-H(x))$. Since $\cos(2\pi H(x)) = \cos(2\pi(1-H(x)))$, we can rewrite Eq. (3.2.7) as

$$\mathrm{Corr}(F,G) \geq \left|\mathbb{E}_{x\sim\{0,1\}^n}\left[\cos(2\pi\theta_x)\right]\right|.$$

Note that, from Eq. (3.2.6) and the definition of θ_x, we have that for all $x \in \{0,1\}^n$, $0 \leq \theta_x \leq \epsilon < 1/2\pi$, and so $0 \leq 2\pi\theta_x < 1 < \pi/2$. This means that $\cos(2\pi\theta_x) \in (0,1]$, and then, using $\cos(x) \geq 1 - x^2$ in the above lower bound on $\mathrm{Corr}(F,G)$, we get

$$\begin{aligned}\mathrm{Corr}(F,G) &\geq \left|\mathbb{E}_{x\sim\{0,1\}^n}\left[\cos(2\pi\theta_x)\right]\right| \\ &= \mathbb{E}_{x\sim\{0,1\}^n}\left[\cos(2\pi\theta_x)\right] \\ &\geq 1 - \mathbb{E}_{x\sim\{0,1\}^n}\left[4\pi^2\theta_x^2\right]\end{aligned}$$

It then follows from the fact that $0 \leq \theta_x < \epsilon$ that

$$\mathrm{Corr}(F,G) \geq 1 - \mathbb{E}_{x\sim\{0,1\}^n}\left[4\pi^2\theta_x^2\right] \geq 1 - 4\pi^2\epsilon^2.$$

□

3.3 Approximation of some classes of Boolean functions

In this section, we illustrate how the framework of point-wise approximation by torus polynomials captures some fairly general classes of Boolean functions. We begin by showing that functions that are computable by low-degree polynomials over prime finite fields can be point-wise approximated by low-degree torus polynomials. For the remainder of this section, assume that all the prime numbers p we consider satisfy $p = O(1)$.

3.3.1 Functions computable by polynomials over finite fields

We will only consider finite fields $\mathbb{F}_p$ where p is a prime. We say a Boolean function $F : \{0,1\}^n \to \{0,1\}$ is computable by a polynomial of degree d over $\mathbb{F}_p$ if there exists a multilinear polynomial $P(x) \in \mathbb{F}_p[x_1, \ldots, x_n]$ of degree d such that $F(x) = P(x)$ for all $x \in \{0,1\}^n$.

We will prove the following result in this section:

Theorem 3.11. *Let $F : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a polynomial of degree d over $\mathbb{F}_p$. Then for every $\epsilon > 0$, there is a torus polynomial P of degree $O(d \log(1/\epsilon))$ that ϵ-approximates F in the point-wise sense, i.e., for all $x \in \{0,1\}^n$,*

$$\left\| \left(\frac{F(x)}{2} - P(x) \right) \ mod\ 1 \right\|_{\mathbb{T}} \leq \epsilon.$$

It is well known that if a Boolean function F is computable by a polynomial of degree d over $\mathbb{F}_p$ then there is a nonclassical polynomial P (and so a torus polynomial) of degree d that exactly computes the function $F(x)/p$, i.e., for all $x \in \{0,1\}^n$,

$$\frac{F(x)}{p} = P(x).$$

To see why, suppose $Q \in \mathbb{F}_p[x_1, \ldots, x_n]$ is the polynomial of degree d that computes F. We can think of Q as an integer polynomial whose coefficients are in $\{0, \ldots, p-1\} \subset \mathbb{Z}$ such that $F(x) \equiv Q(x) \pmod{p}$ for all $x \in \{0,1\}^n$. Now define the nonclassical polynomial Q' as[7]

$$Q'(x) := \frac{Q(x)}{p} \bmod 1.$$

Then it follows from the definition of Q that Q' computes F/p.

Our goal, on the other hand, is to show that, *no matter what the p prime is*, if a Boolean function F is computable a low-degree polynomial over $\mathbb{F}_p$, then the function $F/2$ (and not F/p) can be point-wise approximated by a low-degree torus polynomial. In fact, we will prove a stronger version of this statement in this section: not only

[7] Q' should not be confused with the derivative of Q.

can the function $F/2$ be point-wise approximated, but for every $0 < \alpha < 1$, we can find a low-degree torus polynomial that point-wise approximates αF.

Theorem 3.12. *Let $F : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a polynomial of degree d over $\mathbb{F}_p$. Then for every $\epsilon > 0$ and $0 < \alpha < 1$, there is a torus polynomial P of degree $O(d \log(1/\epsilon))$ that ϵ-approximates the function $\alpha F : \{0,1\}^n \to \{0, \alpha\}$ in the point-wise sense, i.e., for all $x \in \{0,1\}^n$,*

$$\|(\alpha F(x) - P(x)) \bmod 1\|_{\mathbb{T}} \leq \epsilon.$$

Clearly, Theorem 3.12 implies Theorem 3.11.

The approach we take to prove this stronger result is inspired by the discussion in the previous paragraph. Suppose a Boolean function F is computable by some polynomial Q of degree d over $\mathbb{F}_p$. Recall that we can think of Q as an integer polynomial that, modulo p, equals $F(x)$ on $\{0,1\}^n$. Our main idea is to "amplify" Q to get an integer polynomial Q' such that for all $x \in \{0,1\}^n$,

$$Q'(x) \equiv F(x) \pmod{p^k},$$

for a k of our choosing. Then, the torus polynomial

$$\frac{qQ'(x)}{p^k} \bmod 1$$

for a q such that $|\alpha - (q/p^k)| \leq \epsilon$, ϵ-approximates αF.

To do the amplification, we use the following theorem on *modulus-amplifying polynomials* of Beigel and Tarui [BT91], following previous results of Toda [Tod91] and Yao [Yao85].

Theorem 3.13 (Beigel and Tarui [BT91])**.** *For every $k \geq 1$, there exists a univariate polynomial $A_k : \mathbb{Z} \to \mathbb{Z}$ of degree $2k - 1$ such that the following holds. For every $m \geq 2$,*

- *If $x \in \mathbb{Z}$ satisfies $x \equiv 0 \pmod{m}$ then $A_k(x) \equiv 0 \pmod{m^k}$.*

- *If* $x \in \mathbb{Z}$ *satisfies* $x \equiv 1 \pmod{m}$ *then* $A_k(x) \equiv 1 \pmod{m^k}$.

We now give the formal details of the proof of Theorem 3.12.

Proof of Theorem 3.12. Since F is computable by a degree-d polynomial over $\mathbb{F}_p$, there must be an integer polynomial $Q(x)$ of degree d such that for all $x \in \{0,1\}^n$,

$$Q(x) \equiv F(x) \pmod{p}$$

Choose $k = O(\log(1/\epsilon))$ so that $p^{-k} \leq \epsilon$. Note that we can always find $0 \leq q \leq p^k - 1$ such that

$$0 \leq \alpha - \frac{q}{p^k} < p^{-k} \leq \epsilon.$$

Define the torus polynomial

$$G(x) := \frac{qA_k(Q(x))}{p^k} \bmod 1, \tag{3.3.1}$$

where A_k is the modulus-amplifying polynomial from Theorem 3.13. We claim that

$$\|\alpha F(x) - G(x)\|_{\mathbb{T}} \leq \epsilon$$

for all $x \in \{0,1\}^n$. To see this, fix $x \in \{0,1\}^n$, and note that, if $F(x) = 0$ then $A_k(Q(x)) \equiv 0 \pmod{p^k}$, and so $G(x) = 0 = \alpha F(x)$. On the other hand, if $F(x) = 1$ then $A_k(Q(x)) \equiv 1 \pmod{p^k}$, which would imply that

$$\|(\alpha F(x) - G(x)) \bmod 1\|_{\mathbb{T}} = \left\| \left(\alpha - \frac{q}{p^k} \right) \bmod 1 \right\|_{\mathbb{T}} \leq \left| \alpha - \frac{q}{p^k} \right| < \epsilon$$

by Lemma 3.4 and our choice of q and k.

Noting that the degree of G is $(2k-1)d \leq O(d \log(1/\epsilon))$ completes the proof. □

To summarize the results of this section, torus polynomials generalize finite field polynomials in that they provide a uniform way to capture computation of Boolean functions by polynomials over different finite fields — if a Boolean function can be computed by a low-degree polynomial over *any* finite field then it can be point-wise approximated by a low-degree torus polynomial.

3.3.2 Functions computable by $\mathrm{AC}^0[p]$ circuits

Recall that, for a fixed prime p, $\mathrm{AC}^0[p]$ is the class of functions computable by polynomial size circuits of constant depth, consisting of AND, OR, NOT, and MOD_p gates, for a prime p. Here a MOD_p gate is one that outputs 1 if and only if the sum of its inputs is congruent to a non-zero value modulo p.

Let $F : \{0,1\}^n \to \{0,1\}$ be any function in $\mathrm{AC}^0[p]$ for some prime p. We will show that F can also be approximated in the point-wise sense by a low-degree torus polynomial:

Theorem 3.14. *Let $\epsilon > 0$. If $F : \{0,1\}^n \to \{0,1\}$ is computable by an $\mathrm{AC}^0[p]$ circuit for some prime p then there is a torus polynomial P of degree $d = \mathrm{polylog}(n/\epsilon)$ that ϵ-approximates F in the point-wise sense.*

The starting point is the classic result of Razborov [Raz87] and Smolensky [Smo87] which shows that $\mathrm{AC}^0[p]$ circuits can be approximated by random low-degree polynomials over $\mathbb{F}_p$ in the following sense.

Theorem 3.15 (Razborov [Raz87] and Smolensky [Smo87])**.** *Let $F : \{0,1\}^n \to \{0,1\}$ be computable by an $\mathrm{AC}^0[p]$ circuit. Then for every $\epsilon > 0$, there exists a distribution ν supported on polynomials over $\mathbb{F}_p$ of degree $d = \mathrm{polylog}(n/\epsilon)$ such that for all $x \in \{0,1\}^n$,*

$$\Pr_{P \sim \nu}[P(x) = F(x)] \geq 1 - \epsilon.$$

We now give a brief sketch of the proof of Theorem 3.14. Given a Boolean function F computable by an $\mathrm{AC}^0[p]$ circuit, we first use the above result of Razborov and Smolensky to obtain the distribution ν over low-degree polynomials over $\mathbb{F}_p$. It can be shown that, without loss of generality, we can assume that all the polynomials in the support of ν are $\{0,1\}$-valued. Thus, we can now think of ν as a distribution over Boolean functions that are computable by low-degree $\mathbb{F}_p$-polynomials. Next, we draw a large enough sample Ω of Boolean functions from ν so that for every $x \in \{0,1\}^n$,

the fraction of Boolean functions in Ω that agree with F on x is large. Finally, by using Theorem 3.12, since each Boolean function G in the sample is computable by a low-degree polynomial over $\mathbb{F}_p$, we can obtain a low-degree torus polynomial P_G that point-wise approximates $(1/2m) \cdot G$ where m is the size of the sample. It then immediately follows that the torus polynomial defined as $\sum_{G \in \Omega} P_G$ point-wise approximates F.

We now give formal details of this proof. We prove the following lemma which, when combined with Theorem 3.15, implies Theorem 3.14.

Lemma 3.16. *Let $F : \{0,1\}^n \to \{0,1\}$ be a Boolean function and p be a prime. Assume that there exists a distribution ν supported on polynomials of degree d over $\mathbb{F}_p$ such that for all $x \in \{0,1\}^n$,*

$$\Pr_{P \sim \nu}[P(x) = F(x)] \geq 1 - \epsilon.$$

Then there is a torus polynomial of degree $O(d \log(n/\epsilon))$ that 2ϵ-approximates F in the point-wise sense.

Proof. We can assume without loss of generality that all the polynomials in the support of the distribution ν are $\{0,1\}$-valued. This is because for any $P(x)$ in the support we can transform it into the polynomial $(P(x))^{p-1}$ which has range $\{0,1\} \subset \mathbb{F}_p$ by Fermat's little theorem. Note that the degree of each polynomial after the transformation is at most $pd = O(d)$.

By using the standard Chernoff bound followed by a union bound, if we sample polynomials $P_1, \ldots, P_m \sim \nu$ independently for $m = O(n/\epsilon^2)$ then with high probability, for all $x \in \{0,1\}^n$,

$$|\{i \in [m] : P_i(x) \neq F(x)\}| \leq 2\epsilon m. \tag{3.3.2}$$

Fix such a sample. Recall that the polynomials P_i are $\{0,1\}$-valued and so we can think of them as Boolean functions $P_i : \{0,1\}^n \to \{0,1\}$ computable by degree d polynomials over $\mathbb{F}_p$. Thus, for each $i \in [m]$, we can apply Theorem 3.12 with

$\alpha = 1/2m$ and approximation parameter ϵm^{-1} to obtain a torus polynomial $Q_i : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ of degree $O(d \log(m/\epsilon))$ such that for all $x \in \{0,1\}^n$,

$$\left\| \left(\frac{P_i(x)}{2m} - Q_i(x) \right) \mod 1 \right\|_{\mathbb{T}} \leq \frac{\epsilon}{m} \tag{3.3.3}$$

Finally, define the torus polynomial $Q : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ as

$$Q(x) := \left(\sum_{i=1}^{m} Q_i(x) \right) \mod 1.$$

We claim that $Q(x)$ is a torus polynomial which 2ϵ-approximates $F(x)$ in the point-wise sense. To see this, fix $x \in \{0,1\}^n$, and observe that Eq. (3.3.3) and the triangle inequality together imply that

$$\begin{aligned}
\left\| \left(\sum_{i=1}^{m} \frac{P_i(x)}{2m} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} &= \left\| \left(\sum_{i=1}^{m} \left(\frac{P_i(x)}{2m} - Q_i(x) \right) \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \sum_{i=1}^{m} \left\| \left(\frac{P_i(x)}{2m} - Q_i(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \epsilon.
\end{aligned}$$

Also, we can observe that

$$\sum_{i=1}^{m} \frac{P_i(x)}{2m} = \frac{F(x)}{2} - \frac{|\{i \in [m] : P_i(x) \neq F(x)\}|}{2m}.$$

Then, it follows from Lemma 3.4 and Eq. (3.3.2) that

$$\left\| \left(\frac{F(x)}{2} - \sum_{i=1}^{m} \frac{P_i(x)}{2m} \right) \mod 1 \right\|_{\mathbb{T}} \leq \frac{|\{i \in [m] : P_i(x) \neq F(x)\}|}{2m} \leq \epsilon,$$

and so we conclude that

$$\begin{aligned}
&\left\| \left(\frac{F(x)}{2} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \left\| \left(\frac{F(x)}{2} - \sum_{i=1}^{m} \frac{P_i(x)}{2m} \right) \mod 1 \right\|_{\mathbb{T}} + \left\| \left(\sum_{i=1}^{m} \frac{P_i(x)}{2m} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} \leq 2\epsilon.
\end{aligned}$$

□

3.3.3 Functions computable by ACC^0 circuits

We now turn our attention to ACC^0 functions. Recall that a function F is in the class ACC^0 if it can be computed by polynomial size circuits of constant depth with AND, OR, NOT, and MOD_m gates, where m may be composite. Thus, the class ACC^0 generalizes the class $\mathrm{AC}^0[p]$ since it allows for MOD gates with composite moduli. The goal of this section is to show that ACC^0 functions can be point-wise approximated by low-degree torus polynomials, generalizing the main result of the previous section.

The pioneering works of Yao [Yao85] and Beigel and Tarui [BT91] were the first to study and propose polynomial representations of ACC^0 functions. Green et al. [GKT92] further extended these works and proved the following result.

Theorem 3.17 (Green et al. [GKT92])**.** *Let $F : \{0,1\}^n \to \{0,1\}$ be computable by an ACC^0 circuit of depth ℓ and size* $\mathrm{poly}(n)$*. Then for any $e \geq 1$, there exists an integer $k \geq e$ and a polynomial $Q \in \mathbb{Z}[x_1, \ldots, x_n]$ of degree $d = e^{O(\ell)} \log^{O(\ell^2)}(n)$ which satisfies the following condition: for all $x \in \{0,1\}^n$,*

$$Q(x) \equiv F(x)2^k + E(x) \pmod{2^{k+e}}$$

for some error $0 \leq E(x) \leq 2^{k-1}$.

Informally speaking, the above theorem states that the $(k+1)^{\text{th}}$ least-significant bit of $Q(x)$ in its binary representation always equals $F(x)$. Furthermore, this bit is "padded" with $e-1$ zeros to its left, i.e the $(k+2)^{\text{th}}$, $(k+3)^{\text{th}}$, $\ldots$, $(k+e)^{\text{th}}$ least-significant bits are all guaranteed to be equal to 0.

Then, a natural approach to constructing a torus polynomial that point-wise approximates a Boolean function $F \in \mathrm{ACC}^0$ is as follows: we can use Theorem 3.17 to obtain the polynomial Q and define the torus polynomial Q' as

$$Q'(x) := \frac{Q(x)}{2^{k+1}} \bmod 1.$$

It can then be shown that for all $x \in \{0,1\}^n$,

$$\left\| \left(\frac{F(x)}{2} - Q'(x) \right) \bmod 1 \right\|_{\mathbb{T}} = \left| \frac{E(x)}{2^{k+1}} \right| \leq \frac{1}{4}.$$

Thus, this gives us a torus polynomoial that ϵ-approximates F in the point-wise sense for $\epsilon = 1/4$, which is a large approximation error. Note that a bottleneck in obtaining a better approximation is the error $E(x)$ in Theorem 3.17 that can be as high as 2^{k-1}. One way to lower this error is to pad enough zeros on *both sides* of the output bit (i.e., the $(k+1)^{th}$ least-significant bit) so that $E(x) \leq 2^{k-e}$ for an appropriately chosen value of e, which would then make the error as small as $2^{-(e+1)}$.

We observe that the following stronger version of Theorem 3.17 is implicit in the work of Green et al. [GKT92].

Theorem 3.18 (Implicit in Green et al. [GKT92])**.** *Let $F : \{0,1\}^n \to \{0,1\}$ be computable by an ACC^0 circuit of depth ℓ and size* $\mathrm{poly}(n)$*. Then for any $e \geq 1$, there exists an integer $k \geq e$ and a polynomial $Q \in \mathbb{Z}[x_1, \ldots, x_n]$ of degree $d = e^{O(\ell)} \log^{O(\ell^2)}(n)$ which satisfies the following: for all $x \in \{0,1\}^n$,*

$$Q(x) \equiv F(x)2^k + E(x) \pmod{2^{k+e}}$$

for some error $0 \leq E(x) \leq 2^{k-e}$.

Note the difference between the statements of Theorem 3.17 and Theorem 3.18: while the former upper-bounds the error $E(x)$ by 2^{k-1}, the latter upper-bounds it by 2^{k-e} by padding the output bit with $e-1$ zeros on both the sides. We remark that the proof of Theorem 3.18 is essentially the same as that of Theorem 3.17, and so we choose to omit it here.

Based on ideas discussed in the previous paragraphs, we will now use Theorem 3.18 to prove that ACC^0 functions can be point-wise approximated by low-degree torus polynomials.

Theorem 3.19. *Let $F : \{0,1\}^n \to \{0,1\}$ be a function in ACC^0. Then for every $\epsilon > 0$, there is a torus polynomial of degree* $\mathrm{polylog}(n/\epsilon)$ *that ϵ-approximates F in the point-wise sense.*

Proof. Let F be computable by an ACC^0 circuit of size $\mathrm{poly}(n)$ and depth $\ell = O(1)$. Let $Q(x)$ be the polynomial obtained by applying Theorem 3.18 to F with $e = \lceil \log(1/2\epsilon) \rceil$ such that for some $k \geq e$,

$$Q(x) \equiv F(x)2^k + E(x) \pmod{2^{k+e}} \tag{3.3.4}$$

for all $x \in \{0,1\}^n$. We are guaranteed that $0 \leq E(x) \leq 2^{k-e}$, and the degree of $Q(x)$ is $d = e^{O(\ell)} \log^{O(\ell^2)}(n) = \mathrm{polylog}(n/\epsilon)$. Define the following torus polynomial

$$P(x) := \frac{Q(x)}{2^{k+1}} \bmod 1.$$

It is evident that the degree of P is also d.

Fix $x \in \{0,1\}^n$. Then, it follows from Eq. (3.3.4), that

$$P(x) = \frac{Q(x)}{2^{k+1}} \bmod 1 = \left(\frac{F(x)}{2} + \frac{E(x)}{2^{k+1}} \right) \bmod 1.$$

Combining this with the fact that $0 \leq E(x) \leq 2^{k-e}$, Lemma 3.4 implies that

$$\left\| \left(\frac{F(x)}{2} - P(x) \right) \bmod 1 \right\|_{\mathbb{T}} = \left\| \left(-\frac{E(x)}{2^{k+1}} \right) \bmod 1 \right\|_{\mathbb{T}} \leq \left| \frac{E(x)}{2^{k+1}} \right| \leq \frac{1}{2^{e+1}}.$$

Recalling that $e = \lceil \log(1/2\epsilon) \rceil$ completes the proof. □

3.4 Upper and lower bounds for concrete functions

The characterization of ACC^0 discussed in the previous section suggests a new approach to proving ACC^0 lower bounds: show that an explicit Boolean function F, ideally in the class NP[8], cannot be point-wise approximated by low-degree torus polynomials. One such candidate for this approach is the majority function on n bits, Maj_n. In fact, showing that $\mathrm{Maj}_n \notin \mathrm{ACC}^0$ is a long-standing open problem in circuit complexity, the resolution of which would imply a separation between TC^0 and ACC^0. Here TC^0 denotes the class of functions computable by constant depth circuits of polynomial size with AND, OR, NOT, and threshold gates. In this section,

[8]The result of Murray and Williams [MW18] separates NQP from ACC^0, and so the next big challenge is to do the same for the class NP.

we will study how well low-degree torus polynomials can point-wise approximate the majority function.

Recall that Theorem 3.19 implies that any function in ACC^0 can be $(1/20n)$-approximated in the point-wise sense by a torus polynomial of degree $\mathrm{polylog}(n)$. Thus, to show $\mathrm{Maj}_n \notin \mathrm{ACC}^0$, it suffices to prove that any torus polynomial that $(1/20n)$-approximates Maj_n must have degree $\omega(\mathrm{polylog}(n))$. Noting that torus polynomials are real polynomials modulo 1, and that strong bounds are known for the point-wise approximation of Maj_n by real polynomials [NS92], we now investigate these bounds hoping to gain some insights into the problem at hand.

Given a Boolean function $F : \{0,1\}^n \to \{0,1\}$, we say that a multilinear polynomial $P \in \mathbb{R}[x_1, \ldots, x_n]$ of degree d ϵ-approximates F in the point-wise sense if for all $x \in \{0,1\}^n$,

$$|F(x) - P(x)| \leq \epsilon.$$

It immediately follows that if there is a real polynomial P that ϵ-approximates F then the torus polynomial

$$P'(x) = \frac{P(x)}{2} \bmod 1$$

ϵ-approximates F over $\mathbb{R}/\mathbb{Z}$ in the point-wise sense. Thus, when proving degree lower bounds for point-wise approximation over $\mathbb{R}/\mathbb{Z}$, we are also implicitly proving lower bounds for approximation over $\mathbb{R}$.

A beautiful result of Nisan and Szegedy [NS92] shows that for any $\epsilon < 1/2$, any real polynomial that ϵ-approximates the majority function on n bits must have degree $\Omega(\sqrt{n})$. Their proof proceeds in two stages: (i) first show that if a *symmetric* real polynomial ϵ-approximates Maj_n then it must have degree $\Omega(\sqrt{n})$; (ii) then show that any polynomial that ϵ-approximates Maj_n can be *symmetrized* and made into a symmetric polynomial with the same degree and approximation guarantee.

In an attempt to follow the same strategy in the case of torus polynomials, we will now show that any *symmetric* torus polynomial (namely, symmetric real polynomials evaluated modulo one) that $1/(20n)$-approximates Maj_n in the point-wise sense must

have degree $\Omega(\sqrt{n/\log n})$. En route to proving this lower bound we will also prove lower bounds for the *delta functions* $\Delta_{n,w}(x)$, showing that one needs symmetric torus polynomials of degree $\Omega(\sqrt{n/\log n})$ in order to be able to $(1/20n)$-approximate the delta functions.

3.4.1 Lower bounds for Maj_n and $\Delta_{n,w}$

For $x \in \{0,1\}^n$, recall that $|x|$ denotes the Hamming weight of x. A real polynomial $P \in \mathbb{R}[x_1, \ldots, x_n]$ is said to be symmetric if for all $x, y \in \{0,1\}^n$ such that $|x| = |y|$, $P(x) = P(y)$. Using univariate polynomial interpolation over $\mathbb{R}$, and the fact that every function $P : \{0,1\}^n \to \mathbb{R}$ has a unique representation as a real multilinear polynomial (see, e.g., [GSL10, Section 2]), it immediately follows that a symmetric real polynomial $P(x)$ of degree d can be written as

$$P(x_1, \ldots, x_n) = \sum_{j=0}^{d} c_j \left(\sum_{i=1}^{n} x_i \right)^j,$$

for $c_j \in \mathbb{R}$.

A torus polynomial $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is said to be symmetric if for all $x, y \in \{0,1\}^n$ such that $|x| = |y|$, $P(x) = P(y)$. We will now show that every symmetric torus polynomial of degree d also has a nice representation, similar to that of a symmetric real polynomial.

Lemma 3.20. *If $P : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ is a symmetric torus polynomial of degree d then*

$$P(x_1, \ldots, x_n) = \left(\sum_{j=0}^{d} c_j \left(\sum_{i=1}^{n} x_i \right)^j \right) \mod 1,$$

where $c_j \in [0,1)$.

Proof. For $j \in \{0, \ldots, n\}$, let $y_j \in [0,1)$ denote the value that P takes on inputs of Hamming weight j. Using univariate polynomial interpolation, we can find a polynomial $Q \in \mathbb{R}[z]$ of degree n such that for all $j \in \{0, \ldots, n\}$, $Q(j) = y_j$. Since

the generalized binomial coefficients $\left\{\binom{z}{j}\right\}_{j\in\mathbb{N}}$ form a basis for polynomials in $\mathbb{R}[z]$, we can write

$$Q(z) = \sum_{j=0}^{n} c_j \binom{z}{j},$$

for $c_j \in \mathbb{R}$. Observe that, from our construction of Q, for all $x \in \{0,1\}^n$,

$$P(x_1, \ldots, x_n) = \left(\sum_{j=0}^{n} c_j \binom{\sum_{i=1}^{n} x_i}{j}\right) \bmod 1. \tag{3.4.1}$$

Let $d' \in \{0, \ldots, n\}$ be the largest integer such that $c_{d'} \not\equiv 0 \pmod 1$. Expanding and *multilinearizing*[9] the polynomial on the right-hand side of the above equation, we get

$$P(x_1, \ldots, x_n) = \left(\sum_{S\subseteq[n];|S|\leq d'} c_S \prod_{i\in S} x_i\right) \bmod 1$$

for all $x \in \{0,1\}^n$, where $c_S \in \mathbb{R}$. It follows from the definition of generalized binomial coefficients that $c_S = c_{d'}$ for all $S \subseteq [n]$ satisfying $|S| = d'$, and so $c_S \not\equiv 0 \pmod 1$ for all such S. Using the fact that P is a torus polynomial of degree d, it follows from Lemma 3.5 that d' must be at most d. Thus, we can rewrite Eq. (3.4.1) as

$$P(x_1, \ldots, x_n) = \left(\sum_{j=0}^{d} c_j \binom{\sum_{i=1}^{n} x_i}{j}\right) \bmod 1.$$

Let $Q' \in \mathbb{R}[z]$ be the polynomial defined as

$$Q'(z) := \sum_{j=0}^{d} c_j \binom{z}{j}.$$

Then we can write Q' in the standard polynomial basis as $Q(z) = \sum_{j=0}^{d} c'_j z^j$ for some $c'_j \in \mathbb{R}$, which implies that

$$P(x_1, \ldots, x_n) = \left(\sum_{j=0}^{d} c'_j \left(\sum_{i=1}^{n} x_i\right)^j\right) \bmod 1.$$

The statement of the lemma now follows by dropping the coefficients c'_j modulo one. □

[9] This is the process of making the degree of each x_i at most one in every monomial followed by the combining of similar terms if necessary.

The goal of this section is to prove the following theorem:

Theorem 3.21. *Any symmetric torus polynomial that* $(1/20n)$*-approximates* $\mathrm{Maj}_n(x)$ *in the point-wise sense must have degree* $\Omega\left(\sqrt{n/\log n}\right)$.

We now give a sketch of the proof of Theorem 3.21.

The delta function $\Delta_{n,w} : \{0,1\}^n \to \{0,1\}$, for $0 \leq w \leq n$, is defined as

$$\Delta_{n,w}(x) = \begin{cases} 1 & |x| = w \\ 0 & \text{otherwise} \end{cases}.$$

Let $d \geq 0$ be an integer. In the first part of the argument, we observe that if all of the $n+1$ delta functions can be $(1/20n)$-approximated by symmetric torus polynomials of degree d then every symmetric Boolean function in n variables can be $(1/20)$-approximated by a degree d torus polynomial. By using an argument similar to the one in the proof of Lemma 3.8, we then "discretize" the coefficients of these polynomials, showing that every symmetric Boolean function can be $(1/10)$-approximated by symmetric torus polynomials of degree d whose coefficients are of the form $(q/2^k)$ for $q \in \{-(2^k-1), \ldots, 2^k-1\}$. Observing that each such discretized symmetric torus polynomial can $(1/10)$-approximate at most one symmetric Boolean function, and that there are 2^{n+1} such functions, then implies via a counting argument that d must be $\Omega\left(\sqrt{n/\log n}\right)$.

The second part of the argument begins by observing that the delta functions on $n' = \lfloor n/2 \rfloor$ bits can be written as linear combinations of "projections" of the majority function on n bits. This lets us show that if there are symmetric torus polynomials of degree $o(\sqrt{n/\log n})$ that $(1/20n)$-approximate Maj_n then there also are symmetric torus polynomials of degree $o(\sqrt{n'/\log n'})$ that $(1/20n')$-approximate the delta functions on n' bits. In conjunction with the first part of the argument, this then implies the statement of Theorem 3.21.

We now provide the formal details of the first part of the argument, i.e., proving degree lower bounds for the delta functions.

Theorem 3.22. *Let $d \geq 0$, and suppose that for every $0 \leq w \leq n$, there is a symmetric torus polynomial $Q_w : \{0,1\}^n \to \mathbb{R}/\mathbb{Z}$ of degree d that $(1/20n)$-approximates $\Delta_{n,w}(x)$ in the point-wise sense, then $d = \Omega\left(\sqrt{n/\log n}\right)$.*

Proof. Fix $F : \{0,1\}^n \to \{0,1\}$ to be any symmetric Boolean function. Abusing notation, we let $F^{-1}(1)$ denote the set of all integers $0 \leq m \leq n$ such that

$$|x| = m \implies F(x) = 1.$$

Note that

$$F(x) = \sum_{w \in F^{-1}(x)} \Delta_{n,w}(x)$$

for all $x \in \{0,1\}^n$.

Define the torus polynomial Q_F as

$$Q_F(x) = \left(\sum_{w \in F^{-1}(1)} Q_w(x) \right) \mod 1.$$

It follows that Q_F is a symmetric torus polynomial of degree d. Using the fact that each Q_w $(1/20n)$-approximates the delta function $\Delta_{n,w}(x)$, and that $|F^{-1}(1)| \leq n+1$, we have that for every $x \in \{0,1\}^n$,

$$\begin{aligned}
&\left\| \left(\frac{F(x)}{2} - Q_F(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&= \left\| \left(\sum_{w \in F^{-1}(1)} \left(\frac{\Delta_{n,w}(x)}{2} - Q_w(x) \right) \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \sum_{w \in F^{-1}(1)} \left\| \left(\frac{\Delta_{n,w}(x)}{2} - Q_w(x) \right) \mod 1 \right\|_{\mathbb{T}} \qquad \text{(Using triangle inequality)} \\
&\leq |F^{-1}(1)| \cdot \frac{1}{20n} \\
&\leq \frac{1}{20} + o(1), \qquad (3.4.2)
\end{aligned}$$

and so Q_F $(0.05 + o(1))$-approximates F in the point-wise sense.

Since Q_F is a symmetric torus polynomial, we can rewrite it as follows using Lemma 3.20,

$$Q_F(x) = \left(\sum_{j=0}^{d} c_j \left(\sum x_i \right)^j \right) \mod 1$$

for $c_j \in [0,1)$. Let $k \geq 0$ be an integer whose value we choose later. For $0 \leq j \leq d$, let $q_j \in \{-(2^k-1), \ldots, 2^k-1\}$ be such that

$$c_j - \frac{q_j}{2^k} < \frac{1}{2^k},$$

and define Q'_F to be the torus polynomial

$$Q'_F(x) = \left(\sum_{j=0}^{d} \frac{q_j}{2^k} \cdot \left(\sum x_i \right)^j \right) \mod 1.$$

Observe that for every $x \in \{0,1\}^n$,

$$\begin{aligned}
&\|(Q_F(x) - Q'_F(x)) \mod 1\|_{\mathbb{T}} \\
&\leq \left\| \left(\sum_{j=0}^{d} \left(c_j - \frac{q}{2^k} \right) \left(\sum x_i \right)^j \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \sum_{j=0}^{d} \left| \left(c_j - \frac{q}{2^k} \right) \left(\sum x_i \right)^j \right| \qquad \text{(Using Lemma 3.4)} \\
&\leq \frac{(d+1)n^d}{2^k} \leq \frac{1}{20},
\end{aligned}$$

where the last inequality follows by choosing $k = \Theta(d \log(n))$. Also recall from Eq. (3.4.2) that for all $x \in \{0,1\}^n$,

$$\left\| \left(\frac{F(x)}{2} - Q_F(x) \right) \mod 1 \right\|_{\mathbb{T}} \leq \frac{1}{20} + o(1),$$

and so using the triangle inequality we can conclude that

$$\left\| \left(\frac{F(x)}{2} - Q'_F(x) \right) \mod 1 \right\|_{\mathbb{T}} \leq \left(\frac{1}{20} + o(1) \right) + \frac{1}{20} \leq \frac{1}{10} + o(1).$$

Thus, Q'_F $(0.1 + o(1))$-approximates F in the point-wise sense.

Let $\mathrm{SymPoly}_{d,k}$ denote the set of symmetric torus polynomials in n variables of degree d whose coefficients are of the form $q/2^k$ for $q \in \{-(2^k-1), \ldots, 2^k-1\}$. So far we have shown that for every symmetric Boolean function $F : \{0,1\}^n \to \{0,1\}$, there is a torus polynomial $Q'_F \in \mathrm{SymPoly}_{d,k}$ that $(0.1 + o(1))$-approximates F, where $k = \Theta(d \log(n))$. In the other direction, every polynomial in $\mathrm{SymPoly}_{d,k}$ can

$(0.1 + o(1))$-approximate *at most* one symmetric Boolean function. This implies that $|\text{SymPoly}_{d,k}| \geq 2^{n+1}$, since the number of symmetric Boolean function in n variables is 2^{n+1}. Noting that $|\text{SymPoly}_{d,k}| = 2^{O(kd)}$ and $k = \Theta(d \log n)$, implies that d must be $\Omega(\sqrt{n/\log n})$.

□

Before we proceed, we recall that $\text{Maj}_n(x)$ denotes the majority function on n bits, defined as:

$$\text{Maj}_n(x) = \begin{cases} 1 & |x| > \frac{n}{2} \\ 0 & \text{otherwise} \end{cases}$$

for all $x \in \{0,1\}^n$. We will now give the formal details of the second part of the argument used in proving Theorem 3.21. The argument is based on the observation that the delta functions on $n' = \lfloor n/2 \rfloor$ bits can be written in terms of projections of the majority function Maj_n.

Lemma 3.23. *Let $n' = \lfloor n/2 \rfloor$, and suppose that Maj_n, the majority function on n bits, can be $(1/20n)$-approximated by a symmetric torus polynomial of degree d in the point-wise sense. Then for every $0 \leq w \leq n'$, there is a symmetric torus polynomial in n' variables of degree at most d that $(1/20n')$-approximates $\Delta_{n',w}$.*

Proof. Fix $w \in \{1, \ldots, n'-1\}$ (we deal with $w = 0$ and $w = n'$ later), and let $\Delta_{\geq w} : \{0,1\}^{n'} \to \{0,1\}$ denote the function that takes value 1 iff $|x| \geq w$, for all $x \in \{0,1\}^{n'}$. Observe that

$$\Delta_{\geq w}(x_1, \ldots, x_{n'}) = \text{Maj}_n(x_1, \ldots, x_{n'}, c_1, \ldots c_{n-n'}), \tag{3.4.3}$$

where $c \in \{0,1\}^{n-n'}$ is the binary string whose first $n' + 1 - w$ bits are set to 1, and the rest are set to 0.

Let $Q(x_1, \ldots x_n)$ be a symmetric torus polynomial of degree d that $(1/20n)$-approximates $\text{Maj}_n(x)$. Define $Q_{\geq w}(x_1, \ldots, x_{n'})$ to be the following symmetric torus

polynomial in n' variables,

$$Q_{\geq w}(x_1, \ldots x_{n'}) := Q(x_1, \ldots, x_{n'}, c_1, \ldots, c_{n-n'}),$$

where $c \in \{0,1\}^{n-n'}$ is as defined above. It follows from (3.4.3) that $Q_{\geq w}(x_1, \ldots, x_{n'})$ $(1/20n)$-approximates $\Delta_{\geq w}(x_1, \ldots, x_{n'})$. Furthermore, since $Q_{\geq w}$ is a projection of Q, its degree is at most d.

Similarly, we can obtain a symmetric torus polynomial $Q_{\geq w+1}(x_1, \ldots, x_{n'})$ of degree d that $(1/20n)$-approximates $\Delta_{\geq w+1}(x_1, \ldots, x_{n'})$.

Note that for all $x \in \{0,1\}^{n'}$,

$$\frac{\Delta_{n',w}(x)}{2} \bmod 1 = \left(\frac{\Delta_{\geq w}(x)}{2} - \frac{\Delta_{\geq w+1}(x)}{2}\right) \bmod 1.$$

Defining the degree d symmetric torus polynomial

$$Q_w(x_1, \ldots, x_{n'}) := (Q_{\geq w}(x_1, \ldots, x_{n'}) - Q_{\geq w+1}(x_1, \ldots, x_{n'})) \bmod 1,$$

it follows from the triangle inequality and the definitions of $Q_{\geq w}$ and $Q_{\geq w+1}$ that for all $x \in \{0,1\}^{n'}$,

$$\begin{aligned}
&\left\| \left(\frac{\Delta_{n',w}(x)}{2} - Q_w(x)\right) \bmod 1 \right\|_{\mathbb{T}} \\
&\leq \left\| \left(\frac{\Delta_{\geq w}(x)}{2} - Q_{\geq w}(x)\right) \bmod 1 \right\|_{\mathbb{T}} + \left\| \left(Q_{\geq w+1}(x) - \frac{\Delta_{\geq w+1}(x)}{2}\right) \bmod 1 \right\|_{\mathbb{T}} \\
&\leq \frac{1}{20n} + \frac{1}{20n} \leq \frac{1}{10n} \leq \frac{1}{20n'}.
\end{aligned}$$

This shows that for all $1 \leq w \leq n'-1$, the function $\Delta_{w,n'}$ can be $(1/20n')$-approximated in the point-wise sense by a symmetric torus polynomial in n' variables of degree d.

Observe that, for $w = n' - 1$, the function $\Delta_{\geq w+1}(x_1, \ldots, x_{n'})$ is the same as the function $\Delta_{n',w}$, and so the symmetric torus polynomial $Q_{\geq w+1}$ of degree d from above $(1/20n')$-approximates $\Delta_{n',w}$ for $w = n' - 1$.

Finally, note that for all $x \in \{0,1\}^{n'}$,

$$\frac{\Delta_{n',0}(x)}{2} \bmod 1 = \left(\frac{1}{2} - \frac{\Delta_{\geq 1}(x)}{2}\right) \bmod 1.$$

Since it was shown earlier that $Q_{\geq 1}(x_1, \ldots, x_{n'})$ $(1/20n)$-approximates $\Delta_{\geq 1}(x_1, \ldots, x_{n'})$, it follows that $\Delta_{n',0}$ can be $(1/20n')$-approximated by the symmetric torus polynomial

$$\left(\frac{1}{2} - Q_{\geq 1}(x_1, \ldots, x_{n'})\right) \bmod 1,$$

thus completing the proof. □

Theorem 3.21 immediately follows from Theorem 3.22 and Lemma 3.23.

Recalling the approach of [NS92] discussed at the beginning of Section 3.4, the next thing to do is to try and mimic their symmetrization step in the setting of torus polynomials. This essentially means showing that if there is a degree d torus polynomial Q that ϵ-approximates Maj_n then Q can be symmetrized to obtain a symmetric torus polynomial Q' of the same degree that ϵ-approximates Maj_n. Such a result, when combined with Theorem 3.21, would immediately imply that Maj_n cannot be $(1/20n)$-approximated by polylog(n) degree torus polynomials, thus implying $\mathrm{Maj}_n \notin \mathrm{ACC}^0$ via Theorem 3.19. Unfortunately, it is unclear how to make the idea of symmetrization work in the torus setting. We now explain why.

Let S_n denote the group of permutations on the set $[n]$, and for $x \in \{0,1\}^n$, we abuse notation and define

$$\pi(x) := (x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

Then, the symmetrized version of a real polynomial $P \in \mathbb{R}[x_1, \ldots x_n]$ is the real polynomial P^{sym}, defined as follows.

$$P^{sym}(x_1, \ldots x_n) := \sum_{\pi \in S_n} \frac{P(\pi((x_1, \ldots, x_n))}{n!}.$$

Note that P^{sym} is symmetric and has the same degree as P.

Now, it immediately follows that if $P(x_1, \ldots, x_n)$ is a real polynomial of degree d that ϵ-approximates Maj_n in the point-wise sense over $\mathbb{R}$, i.e., for all $x \in \{0,1\}^n$,

$$|\mathrm{Maj}_n(x) - P(x)| \leq \epsilon,$$

then $P^{sym}(x_1, \ldots, x_n)$ is also of degree d and ϵ-approximates F over $\mathbb{R}$ in the point-wise sense: for any $x \in \{0,1\}^n$,

$$\begin{aligned} |\mathrm{Maj}_n(x) - P^{sym}(x)| &\leq \sum_{\pi \in S_n} \left| \frac{\mathrm{Maj}_n(\pi(x)) - P(\pi(x))}{n!} \right| \\ &= \sum_{\pi \in S_n} \frac{|\mathrm{Maj}_n(\pi(x)) - P(\pi(x))|}{n!} \\ &\leq |S_n| \cdot \frac{\epsilon}{n!} \leq \epsilon \end{aligned} \tag{3.4.4}$$

In the case of torus polynomials and point-wise approximation over $\mathbb{R}/\mathbb{Z}$, all the above steps can be carried out in an analogous manner except for the equality in Eq. (3.4.4): it is not always true that, for all $x \in \{0,1\}^n$ and $\pi \in S_n$,

$$\left\| \left(\frac{\frac{\mathrm{Maj}_n(\pi(x))}{2} - P(\pi(x))}{n!} \right) \bmod 1 \right\|_{\mathbb{T}} \leq \frac{\left\| \left(\frac{\mathrm{Maj}_n(\pi(x))}{2} - P(\pi(x)) \right) \bmod 1 \right\|_{\mathbb{T}}}{n!}.$$

Consider the case when $\mathrm{Maj}_n(\pi(x)) = 0$ and $P(\pi(x)) = m + \epsilon$ for integer $0 \leq m < n!$. Assuming $\epsilon \ll 1$, we can see that

$$\left\| \left(\frac{\mathrm{Maj}_n(\pi(x))}{2} - P(\pi(x)) \right) \bmod 1 \right\|_{\mathbb{T}} = \epsilon,$$

whereas

$$\left\| \left(\frac{\frac{\mathrm{Maj}_n(\pi(x))}{2} - P(\pi(x))}{n!} \right) \bmod 1 \right\|_{\mathbb{T}} = \left(\frac{\epsilon}{n!} + \frac{m}{n!} \right) \bmod 1 > \frac{\epsilon}{n!}.$$

3.4.2 Upper bounds for $\Delta_{n,w}$

It is plausible that point-wise approximation by low-degree torus polynomials is too powerful a framework that not only characterizes functions computable by ACC^0 circuits, but also the majority function Maj_n, the delta functions $\Delta_{n,w}$, and possibly the whole of TC^0. Thus, it is worth exploring the possibility that the majority function and the delta functions are ϵ-approximable by polylog(n/ϵ) degree torus polynomials. We now investigate this by trying to prove upper bounds, i.e., trying to construct explicit low-degree polynomials that approximate Maj_n and $\Delta_{n,w}$. In

particular, we prove the following surprising result that demonstrates the power of point-wise approximation by torus polynomials.

Theorem 3.24. *For every $0 \leq w \leq n$ and $\epsilon > 0$, there is a symmetric torus polynomial of degree* $\mathrm{polylog}(n/\epsilon)\epsilon^{-1}$ *that ϵ-approximates $\Delta_{n,w}$ in the point-wise sense.*

Note that the degree of the symmetric torus polynomial in the statement of Theorem 3.24 has an extra multiplicative factor of ϵ^{-1} on top of $\mathrm{polylog}(n/\epsilon)$; without this factor, the statement of the theorem instantiated for $\epsilon = 1/20n$ would contradict Theorem 3.22. In fact, Theorems 3.22 and 3.24 together present an interesting contrast: if ϵ is "large enough", i.e., $\epsilon \geq 1/\mathrm{polylog}(n)$, then there are symmetric torus polynomials of degree $\mathrm{polylog}(n)$ that ϵ-approximate the delta functions in the point-wise, whereas if ϵ is "too small", i.e., $\epsilon \leq 1/20n$, then we need symmetric polynomials of degree $\Omega(\sqrt{n/\log n})$ to achieve the same error of approximation.
We remark that we are not able to obtain similar nontrivial upper bounds for the majority function.

We now give a proof sketch for Theorem 3.24. Fix a w. Note that we want to construct torus polynomials that can "detect" if $|x| = w$, or equivalently, if $|x|-w = 0$. The main observation we use is that if $|x| - w = 0$ then trivially, all of the first t primes (for some $t \geq 0$) divide $|x| - w$, but if $|x| - w \neq 0$ then the integer $|x| - w$ is divisible by at most $\log n$ primes out of the first t primes, since $||x| - w| \leq n$.
We use this criterion because it can be written in terms of the subcriteria of divisibility of $|x| - w$ by primes, and all these subcriteria can themselves be written in terms of finite field polynomials. In particular, note that for every prime p, the following $\{0, 1\}$-valued polynomial over $\mathbb{F}_p$ detects if $|x| - w$ is divisible by p:

$$F_p(x) = 1 - \left(\sum x_i - w\right)^{p-1}.$$

Using results from Section 3.3.1 (in particular, Theorem 3.12), we can then approximate scaled versions of these polynomials using low-degree torus polynomials. Finally,

it can be shown that these torus polynomials can be appropriately combined to detect whether all of the first t primes divide $|x| - w$, or only a small fraction of them divide $|x| - w$ (the fraction can be made small by choosing t to be large enough, i.e., $t = \Omega(\log n)$). Additionally, we have to make sure that the resulting polynomial is low-degree and symmetric. We now cover these details in the formal proof.

Proof of Theorem 3.24. Fix w. For any prime $p \geq 2$, let $F_p : \{0,1\}^n \to \{0,1\}$ denote the Boolean function:

$$F_p(x) = \begin{cases} 1 & |x| \equiv w \pmod{p} \\ 0 & \text{otherwise} \end{cases}.$$

Note that each F_p can be computed by a polynomial of degree $p-1$ over $\mathbb{F}_p$:

$$F_p(x) = 1 - \left(\sum x_i - w\right)^{p-1}.$$

Let $\mathcal{P} = \{p_1, \ldots, p_t\}$ be the first t primes, for t to be chosen later. For each $p \in \mathcal{P}$, we can apply Theorem 3.12 with $F = F_p$, $\alpha = 1/2t$, and error $\epsilon/2t$ to obtain a torus polynomial $Q_p : \{0,1\} \to \mathbb{R}/\mathbb{Z}$ of degree $O(p \log(t/\epsilon))$ that $(\epsilon/2t)$-approximates the function $F_p(x)/2t$, i.e., for all $x \in \{0,1\}^n$,

$$\left\| \left(\frac{F_p(x)}{2t} - Q_p(x) \right) \bmod 1 \right\|_{\mathbb{T}} \leq \frac{\epsilon}{2t}. \tag{3.4.5}$$

Define the torus polynomial

$$Q(x) := \left(\sum_{p \in \mathcal{P}} Q_p(x) \right) \bmod 1. \tag{3.4.6}$$

We claim that Q is a symmetric torus polynomial that ϵ-approximates $\Delta_{n,w}$ in the point-wise sense.

Let $x \in \{0,1\}^n$ such that $|x| = w$, i.e., $\Delta_{n,w}(x) = 1$. In this case, for each $p \in \mathcal{P}$, we have that $F_p(x) = 1$, and so Eq. (3.4.5) implies that

$$\left\| \left(\frac{1}{2t} - Q_p(x) \right) \bmod 1 \right\|_{\mathbb{T}} \leq \epsilon/2t.$$

Then, using the triangle inequality,

$$\begin{aligned}
&\left\| \left(\frac{\Delta_{n,w}(x)}{2} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&= \left\| \left(\frac{1}{2} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&= \left\| \left(\sum_{p \in \mathcal{P}} \left(\frac{1}{2t} - Q_p(x) \right) \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \sum_{p \in \mathcal{P}} \left\| \left(\frac{1}{2t} - Q_p(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \frac{\epsilon}{2}.
\end{aligned}$$

Now suppose $x \in \{0,1\}^n$ is such that $|x| \neq w$, i.e., $\Delta_{n,w}(x) = 0$. Let $\mathcal{P}_x \subset \mathcal{P}$ be the set of primes that divide $|x| - w$. Since $||x| - w| \leq n$, $|\mathcal{P}_x| \leq \log n$. For each $p \in \mathcal{P}$, we have that $F_p(x) = 1$ if and only if $p \in \mathcal{P}_x$. Then, using Eq. (3.4.5), we have that

$$\begin{aligned}
&\forall p \in \mathcal{P} \setminus \mathcal{P}_x \quad \| (-Q_p(x)) \mod 1 \|_{\mathbb{T}} \leq \epsilon/2t \\
&\forall p \in \mathcal{P}_x \quad \| ((1/2t) - Q_p(x)) \mod 1 \|_{\mathbb{T}} \leq \epsilon/2t.
\end{aligned} \tag{3.4.7}$$

The triangle inequality and Lemma 3.4 imply that

$$\begin{aligned}
&\left\| \left(\frac{\Delta_{n,w}(x)}{2} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} \\
&= \| (-Q(x)) \mod 1 \|_{\mathbb{T}} \\
&= \left\| \left(\frac{|\mathcal{P}_x|}{2t} - Q(x) - \frac{|\mathcal{P}_x|}{2t} \right) \mod 1 \right\|_{\mathbb{T}} \\
&\leq \left\| \left(\frac{|\mathcal{P}_x|}{2t} - Q(x) \right) \mod 1 \right\|_{\mathbb{T}} + \left\| \left(-\frac{|\mathcal{P}_x|}{2t} \right) \mod 1 \right\|_{\mathbb{T}} \\
&= \left\| \left(\sum_{p \in \mathcal{P}_x} \left(\frac{1}{2t} - Q_p(x) \right) + \sum_{p \in \mathcal{P} \setminus \mathcal{P}_x} (-Q_p(x)) \right) \mod 1 \right\|_{\mathbb{T}} + \frac{|\mathcal{P}_x|}{2t} \\
&\leq \sum_{p \in \mathcal{P}_x} \left\| \left(\frac{1}{2t} - Q_p(x) \right) \mod 1 \right\|_{\mathbb{T}} + \sum_{p \in \mathcal{P} \setminus \mathcal{P}_x} \| (-Q_p(x)) \mod 1 \|_{\mathbb{T}} + \frac{|\mathcal{P}_x|}{2t} \\
&\leq |\mathcal{P}_x| \cdot \frac{\epsilon}{2t} + |\mathcal{P} \setminus \mathcal{P}_x| \cdot \frac{\epsilon}{2t} + \frac{|\mathcal{P}_x|}{2t} \qquad \text{(Using Eq. (3.4.7))} \\
&\leq |\mathcal{P}| \cdot \frac{\epsilon}{2t} + \frac{\log n}{2t} \\
&= \frac{\epsilon}{2} + \frac{\log n}{2t} \leq \epsilon,
\end{aligned}$$

where the last inequality follows by choosing $t = O(\log n/\epsilon)$. This proves that Q ϵ-approximates $\Delta_{n,w}$ in the point-wise sense.

Eq. (3.4.6) implies that the degree of Q is equal to $O(p \log(t/\epsilon))$, where p is the largest prime in $\mathcal{P}$. Note that the largest prime in $\mathcal{P}$ is upper bounded by $O(t \log t) = \text{polylog}(n/\epsilon)\epsilon^{-1}$ by the prime number theorem, since $\mathcal{P}$ only contains the first t primes. It follows that the degree of Q is $\text{polylog}(n/\epsilon)\epsilon^{-1}$.

To see why Q is symmetric, first observe that for each $p \in \mathcal{P}$, the torus polynomial Q_p obtained by applying Theorem 3.12 to F_p is of the form

$$Q_p(x) = \frac{qA_k\left(1 - (\sum_i x_i)^{p-1}\right)}{p^k} \bmod 1,$$

where q is some integer, and $A_k\left(1 - (\sum_i x_i)^{p-1}\right)$ is a symmetric integer polynomial in $\mathbb{Z}[x_1, \ldots, x_n]$ obtained by composing the univariate modulus amplifying polynomial A_k with the symmetric integer polynomial $1 - (\sum_i x_i)^{p-1}$ (see Eq. (3.3.1) in the proof of Theorem 3.12). Thus, each Q_p is a symmetric torus polynomial. It then follows from Eq. (3.4.6) that Q is also symmetric, and this completes the proof. □

We remark that combining Theorem 3.24 with Lemma 3.9 implies the existence of symmetric *nonclassical* polynomials of degree $\text{polylog}(n)\epsilon^{-1}$ that ϵ-approximate the delta functions. This is because the proof of Lemma 3.9 simply "discretizes" the coefficients of a torus polynomial using diadic rationals in order to transform it into a nonclassical polynomial, and thus this process is symmetry preserving.

Chapter 4
Conclusion and open problems

The research in this dissertation was motivated by two main questions.

1. How much more powerful are nonclassical polynomials in comparison to classical polynomials, when it comes to approximating Boolean functions in the agreement-sense?

2. Is it possible to prove that some explicit Boolean function is not computable by ACC^0 circuits in a purely combinatorial manner, using only polynomial-based approximations/representations of Boolean functions?

As far as the first question is concerned, in Chapter 2, we made some progress towards answering it by giving examples of explicit Boolean functions which have good agreement with nonclassical polynomials but not with classical polynomials of the same degree. We also proved that this is not true for all Boolean functions, and in particular, showed that both classical and nonclassical polynomials of degree $o(\sqrt{n})$ have agreement at most $1/2 + o(1)$ with Maj_n, the majority function on n bits. These results resolved some of the open problems stated in the work of Bhowmick and Lovett [BL15].

We remark that the results in Chapter 2 were obtained by studying the behavior of the quantity $\gamma_{d,k}(F)$, which is the maximum possible agreement between polynomials over $\mathbb{Z}/2^k\mathbb{Z}$ of degree d and a Boolean function F, and several interesting properties of $\gamma_{d,k}(F)$ were observed en route to obtaining these results.

In Chapter 3, we attempted to answer the second question from above by using the polynomial-based representations of Green et al. [GKT92] and Beigel and

Tarui [BT91], to show that functions computable by ACC^0 circuits can be approximated by low-degree nonclassical polynomials with respect to a new notion of point-wise approximation introduced by us. In hopes of proving $\mathrm{Maj}_n \notin \mathrm{ACC}^0$ in a purely combinatorial or algebraic manner, we then tried to show that Maj_n cannot be point-wise approximated by low-degree nonclassical polynomials. While we were unsuccessful at this attempt, we did manage to show that Maj_n cannot be approximated by *symmetric* low-degree nonclassical polynomials. We also obtained nontrivial upper and lower bounds on the degree of nonclassical polynomials that point-wise approximate the delta functions $\Delta_{n,w}$.

Several interesting research directions are suggested by our results, and we now state some of them:

ACC^0 lower bounds via nonclassical polynomials

Theorem 3.19 in conjunction with Lemma 3.9 suggests a combinatorial approach to proving lower bounds against ACC^0. Concretely, we pose the following open problem.

Problem 4.1. *Find an explicit Boolean function $F : \{0,1\}^n \to \{0,1\}$, and an $\epsilon = \epsilon(n)$, such that any nonclassical polynomial that ϵ-approximates F in the point-wise sense must have degree $\omega(\mathrm{polylog}(n/\epsilon))$. Such an F is not in ACC^0.*

Recall that we know from the results of Williams [Wil14b], and Williams and Murray [MW18], that $\mathrm{NEXP} \not\subseteq \mathrm{ACC}^0$, and more generally, $\mathrm{NQP} \not\subseteq \mathrm{ACC}^0$. As mentioned before the tools and techniques used in proving these results are not purely combinatorial or algebraic, and so an immediate goal could be to solve Problem 4.1 by finding an $F \in \mathrm{NEXP}$ or $F \in \mathrm{NQP}$. A more ambitious goal is to find an $F \in \mathrm{NP}$.

Proving $\mathrm{Maj}_n \notin \mathrm{ACC}^0$ via degree lower bounds

In Section 3.4, we proved that no symmetric torus polynomial of degree $o\left(\sqrt{n/\log n}\right)$ can $(1/20n)$-approximate Maj_n in the point-wise sense. We conjecture that this should

hold even if we consider general torus polynomials:

Problem 4.2. *Prove that any torus polynomial that* $(1/20n)$*-approximates* Maj_n *must have degree* $\Omega(\sqrt{n/\log n})$.

We remark that the resolution of this problem will show that the majority function is not in ACC^0, thereby proving $\mathrm{TC}^0 \not\subseteq \mathrm{ACC}^0$. Note note that, in light of Lemma 3.9, it suffices to only prove that no nonclassical polynomial of degree $o\left(\sqrt{n/\log n}\right)$ can $(1/40n)$-approximate Maj_n.

Proving $\mathrm{Maj}_n \notin \mathbf{ACC}^0$ via correlation bounds

Recall that if $\mathrm{Maj}_n \in \mathrm{ACC}^0$ then there is a nonclassical polynomial P of degree $\mathrm{polylog}(n)$ that $(1/n)$-approximates Maj_n in the point-wise sense. By Lemma 3.10, it follows that

$$\mathrm{Corr}\left(\frac{\mathrm{Maj}_n}{2}, P\right) \geq 1 - O\left(\frac{1}{n^2}\right),$$

i.e., P ϵ-approximates Maj_n in the correlation sense for $\epsilon \geq 1 - O(1/n^2)$.

Thus, another way to show $\mathrm{Maj}_n \notin \mathrm{ACC}^0$ is to upper bound the correlation that Maj_n can have with degree $\mathrm{polylog}(n)$ nonclassical polynomials. In particular, we pose the following problem.

Problem 4.3. *Show that if* P *is any nonclassical polynomial of degree* $O(\mathrm{polylog}(n))$ *then it must be the case that*

$$\mathrm{Corr}\left(\frac{\mathrm{Maj}_n}{2}, P\right) \leq 1 - \omega\left(\frac{1}{n^2}\right).$$

We remark that no nontrivial upper bounds on the correlation between the majority function and nonclassical polynomials of degree $O(\mathrm{polylog}(n))$ are known[1]. On the contrary, a result of Bhowmick and Lovett [BL15] constructs a nonclassical polynomial over $\mathbb{F}_2$ of degree $O(\log n)$ that has correlation at least δ with Maj_n, for some absolute constant $\delta > 0$.

[1]Obviously, this correlation must be strictly less than 1 because otherwise it would imply that the majority function has agreement 1 with a nonclassical polynomial of degree $\mathrm{polylog}(n)$, which is impossible.

(In)approximability by torus polynomials as a natural property

The natural proofs barrier of Razborov and Rudich [RR97] isn't really a problem for the approach to proving ACC^0 lower bounds outlined in Section 3.4. This is because we are only trying to prove lower bounds against ACC^0, and pseudorandom generators are not believed to be contained in this class. It is also not clear whether the property in question, i.e., (in)approximability by torus polynomials, is *natural*, and, in particular, it will be interesting to investigate whether this property is *constructive*, i.e., whether one can efficiently distinguish between Boolean functions which can be approximated by low-degree torus polynomials and a random Boolean function.

Problem 4.4. *Given the truth table of a function $F : \{0,1\}^n \to \{0,1\}$ and an $\epsilon > 0$, decide in polynomial time (in 2^n and $1/\epsilon$) whether F is ϵ-approximable in the point-wise sense by a torus polynomial of degree* $\mathrm{polylog}(n/\epsilon)$.

We remark that if this property is indeed constructive (and thus, also natural) then the work of Carmosino et al. [CIKK16] would imply quasipolynomial time learning algorithms for ACC^0 — currently such learning algorithms are only known for $\mathrm{AC}^0[p]$.

Amplification for torus polynomials

An interesting property of point-wise approximation by polynomials over $\mathbb{R}$ is its amenability to amplification, namely the fact that, for any Boolean function F and $\epsilon < 1/3$, given a polynomial P over $\mathbb{R}$ of degree d that $1/3$-approximates F in the point-wise sense, it can be transformed into a polynomial P' of degree $d' = O(d \log(1/\epsilon))$ that ϵ-approximates F.

It is not clear whether such a transformation is possible in the case of point-wise approximation by torus polynomials. In the case of approximation by real polynomials, the transformation is symmetry preserving, but, given the upper and lower bounds for the delta functions discussed in Section 3.4 (see Theorems 3.22 and 3.24), we should not expect this in the case of torus polynomials. This motivates the following

problem.

Problem 4.5. *Suppose that a Boolean function F can be $1/3$-approximated in the point-wise sense by torus polynomials of degree d, and $(1/20n)$-approximated by torus polynomials of degree d'. Then how does d' compare to d?*

Nontrivial point-wise approximation of Maj_n

Theorem 3.24 shows the existence of symmetric torus polynomials of degree at most $\mathrm{polylog}(n)\epsilon^{-1}$ that ϵ-approximate the delta functions in the point-wise sense. We were not able to do the same in the case of the majority function, and leave this as an open problem.

Problem 4.6. *Is there a symmetric torus polynomial of degree at most* $\mathrm{polylog}(n)\epsilon^{-1}$ *that ϵ-approximates* Maj_n *in the point-wise sense?*

References

[AB01] Noga Alon and Richard Beigel. *Lower Bounds for Approximations by Low Degree Polynomials over $\mathbb{Z}_m$*. In *Proc. 16th IEEE Conf. on Computational Complexity (CCC)*, pages 184–187. 2001.

[AWY15] Amir Abboud, Richard Ryan Williams, and Huacheng Yu. *More Applications of the Polynomial Method to Algorithm Design*. In *Proc. 26th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 218–230. 2015. doi:10.1137/1.9781611973730.17.

[BHLR19] Abhishek Bhrushundi, Kaave Hosseini, Shachar Lovett, and Sankeerth Rao. *Torus Polynomials: An Algebraic Approach to ACC Lower Bounds*. In *Proc. 10th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 13:1–13:16. 2019. doi:10.4230/LIPIcs.ITCS.2019.13.

[BHS17] Abhishek Bhrushundi, Prahladh Harsha, and Srikanth Srinivasan. *On polynomial approximations over $Z/2^kZ$*. In *Proc. 34th Annual Symp. on Theoretical Aspects of Comp. Science (STACS)*, volume 66 of *LIPIcs*, pages 12:1–12:12. Schloss Dagstuhl, 2017. doi:10.4230/LIPIcs.STACS.2017.12.

[BKT18] Mark Bun, Robin Kothari, and Justin Thaler. *The Polynomial Method Strikes Back: Tight Quantum Query Bounds via Dual Polynomials*. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 297–310. Association for Computing Machinery, New York, NY, USA, 2018. doi:10.1145/3188745.3188784.

[BL15] Abhishek Bhowmick and Shachar Lovett. *Nonclassical Polynomials as a Barrier to Polynomial Lower Bounds*. In *Proc. 30th Computational Complexity Conf. (CCC)*, pages 72–87. 2015. doi:10.4230/LIPIcs.CCC.2015.72.

[BNRW05] Harry Buhrman, Ilan Newman, Hein Röhrig, and Ronald de Wolf. *Robust polynomials and quantum algorithms*. In *Proc. 22nd Annual Symp. on Theoretical Aspects of Comp. Science (STACS)*, pages 593–604. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[BT91] Richard Beigel and Jun Tarui. *On ACC (circuit complexity)*. In *Proceedings., 32nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 783–792. IEEE, 1991.

[BV10] Andrej Bogdanov and Emanuele Viola. *Pseudorandom bits for polynomials*. SIAM Journal on Computing, 39(6):2464–2486, 2010. doi:10.1137/070712109.

[BW01] Harry Buhrman and Ronald de Wolf. *Communication complexity lower bounds by polynomials*. In *Proceedings of the 16th Annual Conference on Computational Complexity (CCC)*, page 120. IEEE Computer Society, USA, 2001.

[CHLT19] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. *Pseudorandom Generators from the Second Fourier Level and Applications to AC0 with Parity Gates*. In *Proc. 10th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 124, pages 22:1–22:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019. doi:10.4230/LIPIcs.ITCS.2019.22.

[CIKK16] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. *Learning Algorithms from Natural Proofs*. In *Proceedings of the 31st Conference on Computational Complexity (CCC)*, volume 50, pages 10:1–10:24. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2016. doi:10.4230/LIPIcs.CCC.2016.10.

[CT15] Gil Cohen and Avishay Tal. *Two Structural Results for Low Degree Polynomials and Applications*. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 680–709. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2015.

[GKT92] Frederic Green, Johannes Kobler, and Jacobo Toran. *The power of the middle bit*. In *Proc. of the 7th Annual Structure in Complexity Theory Conference*, pages 111–117. IEEE, 1992.

[Gop08] Parikshit Gopalan. *Query-efficient algorithms for polynomial interpolation over composites*. SIAM J. Comput., 38(3):1033–1057, 2008. (Preliminary version in *17th SODA*, 2006). doi:10.1137/060661259.

[Gop14] ———. *Constructing Ramsey Graphs from Boolean Function Representations*. Combinatorica, 34(2):173–206, April 2014. doi:10.1007/s00493-014-2367-1.

[Gow01] William T. Gowers. *A new proof of Szemerédi's theorem*. GAFA, Geom. funct. anal., 11:465–588, 2001. doi:10.1007/s00039-001-0332-9.

[Gre00] Frederic Green. *A complex-number Fourier technique for lower bounds on the mod-m degree*. Comput. Complexity, 9(1):16–38, 2000. (Preliminary version in *12th STACS*, 1995). doi:10.1007/PL00001599.

[Gro00] Vince Grolmusz. *Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs*. Combinatorica, 20(1):71–86, 2000. doi:10.1007/s004930070032.

[GSL10] Parikshit Gopalan, Amir Shpilka, and Shachar Lovett. *The Complexity of Boolean Functions in Different Characteristics*. Computational complexity, 19(2):235–263, 2010.

[GT08] Ben Green and Terence Tao. *An Inverse Theorem for the Gowers $U^3(G)$ norm*. volume 51, page 73–153. Cambridge University Press, 2008. doi:10.1017/S0013091505000325.

[GT09] Ben Joseph Green and Terence Tao. *The distribution of polynomials over finite fields, with applications to the Gowers norms*. Contributions to Discrete Mathematics, 4(2), 2009.

[Hås87] Johan Håstad. *Computational Limitations of Small-Depth Circuits*. MIT Press, Cambridge, MA, USA, 1987.

[Hås98] ———. *The shrinkage exponent of De Morgan formulas is 2*. SIAM Journal on Computing, 27(1):48–64, 1998.

[Knu97] Donald Ervin Knuth. *Fundamental Algorithms*, volume I of *The Art of Computer Programming*. Addison-Wesley, 3rd edition, 1997.

[KS04] Adam R. Klivans and Rocco A. Servedio. *Learning DNF in time $2^{O(n^{1/3})}$*. J. Comput. Syst. Sci., 68(2):303–318, 2004. (Preliminary version in *33rd STOC*, 2001). doi:10.1016/j.jcss.2003.07.007.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. *Constant depth circuits, Fourier transform, and learnability*. J. ACM, 40(3):607–620, 1993. (Preliminary version in *30th FOCS*, 1989). doi:10.1145/174130.174138.

[LMS11] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. *Inverse Conjecture for the Gowers Norm is False*. Theory of Computing, 7(9):131–145, 2011. doi:10.4086/toc.2011.v007a009.

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite Fields*, volume 2 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2nd edition, 1997. doi:10.1017/CBO9780511525926.

[MW18] Cody D. Murray and R. Ryan Williams. *Circuit Lower Bounds for Nondeterministic Quasi-polytime from a New Easy Witness Lemma*. SIAM Journal on Computing, pages 300–322, 2018. doi:10.1137/18M1195887.

[NS92] Noam Nisan and Mario Szegedy. *On the degree of Boolean functions as real polynomials*. In *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, STOC '92, pages 462–467. ACM, 1992. doi:10.1145/129712.129757.

[Raz87] Alexander A. Razborov. *Нжние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения (Russian) [Lower bounds on the size of bounded depth circuits over a complete basis with logical addition]*. Mathematicheskie Zametki, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987). doi:10.1007/BF01137685.

[RR97] Alexander A Razborov and Steven Rudich. *Natural proofs*. Journal of Computer and System Sciences, 55(1):24 – 35, 1997. doi:10.1006/jcss.1997.1494.

[Sam07] Alex Samorodnitsky. *Low-degree tests at large distances*. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, page 506–515. Association for Computing Machinery, New York, NY, USA, 2007. doi:10.1145/1250790.1250864.

[She12] Alexander A. Sherstov. *Strong direct product theorems for quantum communication and query complexity*. SIAM Journal on Computing, 41(5):1122–1165, 2012. doi:10.1137/110842661.

[Smo87] Roman Smolensky. *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*. In *Proc. 19th ACM Symp. on Theory of Computing (STOC)*, pages 77–82. 1987. doi:10.1145/28395.28404.

[Smo93] ———. *On representations by low-degree polynomials*. In *Proc. 34th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 130–138. 1993. doi:10.1109/SFCS.1993.366874.

[Sze89] Mario Szegedy. *Algebraic Methods in Lower Bounds for Computational Models with Limited Communication*. Ph.D. thesis, University of Chicago, 1989.

[Tod91] Seinosuke Toda. *PP is as hard as the polynomial-time hierarchy*. SIAM Journal on Computing, 20(5):865–877, 1991.

[TZ12] Terence Tao and Tamar Ziegler. *The inverse conjecture for the Gowers norm over finite fields in low characteristic*. Ann. Comb., 16(1):121–188, 2012. doi:10.1007/s00026-011-0124-3.

[Vio09] Emanuele Viola. *Correlation Bounds for Polynomials over* $\{0,1\}$. SIGACT News, 40(1):27–44, February 2009. doi:10.1145/1515698.1515709.

[VW08] Emanuele Viola and Avi Wigderson. *Norms, XOR Lemmas, and Lower Bounds for Polynomials and Protocols*. Theory of Computing, 4(7):137–168, 2008. doi:10.4086/toc.2008.v004a007.

[Wil14a] Ryan Williams. *New algorithms and lower bounds for circuits with linear threshold gates*. In *Proc. 46th ACM Symp. on Theory of Computing (STOC)*, pages 194–202. 2014. arXiv:1401.2444, doi:10.1145/2591796.2591858.

[Wil14b] ———. *Nonuniform ACC circuit lower bounds*. Journal of the ACM (JACM), 61(1):2, 2014.

[Yao85] Andrew Chi-Chih Yao. *Separating the polynomial-time hierarchy by oracles*. In *Proc. 26th Annual Symposium on Foundations of Computer Science*, pages 1–10. IEEE, 1985.