Electronic Colloquium on Computational Complexity, Report No. 37 (2012)

MAKING POLYNOMIALS ROBUST TO NOISE

ALEXANDER A. SHERSTOV*

ABSTRACT. A basic question in any computational model is how to reliably compute a given function when the inputs or intermediate computations are subject to noise at a constant rate. Ideally, one would like to use at most a constant factor more resources compared to the noise-free case. This question has been studied for decision trees, circuits, automata, data structures, broadcast networks, communication protocols, and other models.

Buhrman et al. (2003) posed the noisy computation problem for *real polynomials*. We give a complete solution to this problem. For any polynomial $p\colon \{0,1\}^n \to [-1,1]$, we construct a polynomial $p_{\text{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\deg p + \log \frac{1}{\epsilon})$ that ϵ-approximates p and is additionally robust to noise in the inputs: $|p(x) - p_{\text{robust}}(x+\delta)| < \epsilon$ for all $x \in \{0,1\}^n$ and all $\delta \in [-1/3, 1/3]^n$. This result is optimal with respect to all parameters. We construct p_{robust} explicitly for each p. Previously, it was open to give such a construction even for $p = x_1 \oplus x_2 \oplus \cdots \oplus x_n$ (Buhrman et al., 2003). The proof contributes a technique of independent interest, which allows one to force partial cancellation of error terms in a polynomial.

1. INTRODUCTION

Noise is a well studied phenomenon in the computing literature. It arises naturally in several ways. Most obviously, the input to a computation can be noisy due to imprecise measurement or human error. In addition, both the input and the intermediate results of a computation can be corrupted to some extent by a malicious third party. Finally, even in a setting with correct input and no third-party interference, errors can be introduced by using a randomized algorithm as a subroutine in the computation. In all these settings, one would like to compute the correct answer with high probability despite the presence of noise. A matter of both theoretical and practical interest is *how many* additional resources are necessary to combat the noise. Research has shown that the answer depends crucially on the computational model in question. Models studied in this context include decision trees [25, 49, 23, 20, 43], circuits [46, 26, 21, 34, 59, 60], broadcast networks [27, 40, 24, 43, 29, 17, 18], and communication protocols [51, 52, 10, 28]. Some computational models exhibit a surprising degree of robustness to noise, in that one can compute the correct answer with probability 99% with only a constant-factor increase in cost relative to the noise-free setting. In other models, even the most benign forms of noise increase the computational complexity by a superconstant factor.

In most cases, one can overcome the noise by brute force, with a logarithmic-factor increase in computational complexity. In a noisy decision tree, for example, one can repeat each query a logarithmic number of times and use the majority answer. Assuming independent corruption of the queries, this strategy results in a correct computation with high probability. Similarly, in a noisy broadcast network one can repeat each broadcast a logarithmic number of times and take the majority of the received bits. It may seem,

* Computer Science Department, UCLA, Los Angeles, California 90095. ✉ sherstov@cs.ucla.edu Supported by NSF CAREER award CCF-1149018.

ISSN 1433-8092

then, that noise is an issue of minor numerical interest. This impression is incorrect on several counts. First, in some settings such as communication protocols [51, 52, 10, 28], it is nontrivial to perform any computation at all in the presence of noise. Second, even a logarithmic-factor increase in complexity can be too costly for some applications; see, e.g., the analysis in [48]. Third and most important, the question at hand is a *qualitative* one: is it possible to arrange the steps in a computation so as to cause the intermediate errors to almost always cancel? Put differently, in studying the robustness of a computational model to noise, one aims first and foremost to understand a fundamental property of the model rather than make numerical improvements. This study frequently reveals aspects of the model that would otherwise be overlooked.

This last point is nicely illustrated by *noisy broadcast networks*, in which n processors have bits $x_1, x_2, \ldots, x_n$, respectively, and communicate in broadcast mode to compute some function $f(x_1, x_2, \ldots, x_n)$. A bit transmitted from one processor to another arrives corrupted with a small constant probability, independent of any other bit transmissions. In a surprising 1988 paper, Gallager [27] proved that $O(n \log\log n)$ broadcasts are enough for all processors to learn the string $(x_1, x_2, \ldots, x_n)$ with constant probability and thus to compute any function f. Despite sustained efforts, it was unknown until recently whether Gallager's result is optimal. Goyal, Kindler, and Saks [29] solved this problem, showing that $\Omega(n \log\log n)$ broadcasts are necessary for all processors to learn $(x_1, x_2, \ldots, x_n)$ with constant probability. The work in [29] contributed a novel entropy-based view of broadcast networks and related them to an intermediate model of interest in its own right, the generalized noisy decision tree, which may not have been discovered otherwise. Remarkably, it is open to this day whether Gallager's upper bound is tight for computing any function $f(x_1, \ldots, x_n)$ with *Boolean* range.

Our problem. The computational model of interest to us is the *real polynomial.* In this model, the complexity measure of a Boolean function $f\colon \{-1,+1\}^n \to \{-1,+1\}$ is the least degree of a real polynomial that approximates f pointwise. Formally, the *approximate degree* of f, denoted $\widetilde{\deg}(f)$, is the least degree of a real polynomial p with $|f(x) - p(x)| \leqslant 1/3$ for every $x \in \{-1,+1\}^n$. The constant $1/3$ is chosen for aesthetic reasons and can be replaced by any other in $(0, 1)$ without changing the model. The contribution of this paper is to show that as a computational model, real polynomials are highly robust to noise.

The formal study of the approximate degree and of polynomial representations in general began in 1969 with the seminal work of Minsky and Papert [42], who famously proved that the parity function on n variables has approximate degree n. Since then, the approximate degree has been used to solve a vast array of problems in algorithm design and complexity theory. In this line of research, *upper* bounds on the approximate degree are used to obtain efficient algorithms, and *lower* bounds are used to prove hardness and impossibility results. For example, the approximate degree and its variants have yielded a variety of lower bounds in circuit complexity [45, 58, 9, 5, 38, 39, 56, 8]. Following the seminal work of Beals et al. [6], the approximate degree has been used many times to prove tight lower bounds on quantum query complexity [6, 12, 2, 1, 33]. Starting in the early 2000s, the approximate degree has enabled dramatic progress in communication complexity on problems that were previously thought to be beyond reach; see [15, 47, 14, 48] and the survey [53]. In computational learning theory, the approximate degree has played a central role in various lower bounds [36, 37, 55, 57] as well as algorithmic results, such as the fastest known algorithms for PAC-learning DNF formulas [61, 35] and agnostically learning disjunctions [32]. Earlier algorithmic applications include approximating the inclusion-exclusion formula [41, 31, 54, 62]. Most recently, the approximate degree has been used to prove lower bounds in proof complexity [7].

Despite these motivating applications, there has been little progress in understanding real polynomials on the Boolean hypercube, i.e., understanding the approximate degree *itself* as a complexity measure. This may be surprising given that approximation theory has existed in its modern form for over 150 years and is a very mature branch of analysis. However, approximation on the *Boolean hypercube* is rooted mostly in theoretical computer science and remains a relatively new topic. The only truly general result on the approximate degree, due to Nisan and Szegedy [44], is that it is polynomially related to *decision tree complexity* and *block sensitivity*. When a more precise estimate is needed, a common way to construct an approximating polynomial is to design a quantum query algorithm for the corresponding function, e.g., [30, 62, 22, 4]. However, the quantum query approach gives only *upper* bounds on the approximate degree, and even there its applicability is limited because quantum query algorithms are a less powerful computational model than real polynomials [3].

In this paper, we answer a question about real polynomials posed nine years ago by Buhrman et al. [13]. These authors asked whether real polynomials, as a computational model, are *robust to noise.* Robustness to noise becomes necessary when one wants to do anything nontrivial with approximating polynomials, e.g., compose them. To use a motivating example from [13], suppose that we have approximating polynomials p and q for Boolean functions $f\colon\{-1,+1\}^n \to \{-1,+1\}$ and $g\colon\{-1,+1\}^m \to \{-1,+1\}$, respectively. Having these two polynomials gives us no way whatsoever to approximate the composed function $f(g,g,\dots,g)$ on nm variables. In particular, the natural construction $p(q,q,\dots,q)$ does not work for this purpose because q can range anywhere in $[-4/3,-2/3]\cup[2/3,4/3]$ and the behavior of p on non-Boolean inputs can be arbitrary. In other words, the problem is that the output of q is inherently *noisy*, and the original polynomial p is not designed to handle that noise. What we need is a *robust* approximating polynomial for f, to use the term introduced by Buhrman et al. [13]. Formally, a robust approximating polynomial for f is a real polynomial $p_{\text{robust}}\colon\mathbb{R}^n \to \mathbb{R}$ such that for every $x\in\{-1,+1\}^n$,

$$|f(x) - p_{\text{robust}}(x+\delta)| < \frac{1}{3}$$

whenever $\delta = (\delta_1,\dots,\delta_n) \in [-1/3,1/3]^n$. Put differently, a robust polynomial is one that approximates f not only on Boolean inputs but also on the much larger domain $[-4/3,-2/3]\cup[2/3,4/3]$. Robust polynomials compose in the natural way: to use the notations of this paragraph, the polynomial $p_{\text{robust}}(q,q,\dots,q)$ is a valid approximating polynomial for $f(g,g,\dots,g)$.

The obvious question is whether robustness comes at a cost. Ideally, one would like to make an approximating polynomial robust with only a constant-factor increase in degree, so that every Boolean function f would have a robust polynomial of degree $\Theta(\widetilde{\deg}(f))$. Similar to the setting of noisy decision trees and broadcast networks, a fairly direct calculation shows that every Boolean function $f\colon\{-1,+1\}^n \to \{-1,+1\}$ has a robust polynomial of degree $O(\widetilde{\deg}(f)\log n)$. Buhrman et al. [13] improved this bound to $\min\{O(n), O(\widetilde{\deg}(f)\log\widetilde{\deg}(f))\}$ using combinatorial arguments and quantum query complexity. In particular, the work of Buhrman et al. shows that parity, majority, and random functions—all of which have approximate degree $\Theta(n)$—also have robust approximating polynomials of degree $\Theta(n)$. It was posed as an open problem in [13] whether an analogous result holds for every Boolean function, i.e., whether every Boolean function has a robust approximating polynomial of degree $\Theta(\widetilde{\deg}(f))$.

Our result. We give a complete solution to the problem of Buhrman et al. [13]. To be precise, we study a more general problem. Buhrman et al. [13] asked whether a polynomial p can be made robust with only a constant-factor increase in degree, *provided* that p approximates a Boolean function. We prove that *every* polynomial $p\colon\{-1,+1\}^n \to [-1,1]$ can be made robust, regardless of whether p approximates a Boolean function.

THEOREM 1. *Let $p\colon\{-1,+1\}^n \to [-1,1]$ be a given polynomial. Then for every $\epsilon > 0$, there is a polynomial p_{robust} of degree $O(\deg p + \log \frac{1}{\epsilon})$ such that for all $x \in \{-1,+1\}^n$ and $\delta \in [-1/3, 1/3]^n$,*

$$|p(x) - p_{\mathrm{robust}}(x+\delta)| < \epsilon.$$

Furthermore, p_{robust} has an explicit, closed-form description.

Theorem 1 shows that real polynomials are robust to noise. In this regard, they behave differently from other computational models such as decision trees [25] and broadcast networks [29], where handling noise provably increases the computational complexity by a superconstant factor. In fact, Theorem 1 reveals a *very* high degree of robustness to noise: the degree of an ϵ-error robust polynomial grows additively rather than multiplicatively with the error parameter ϵ, and the actual dependence on ϵ is only logarithmic. Theorem 1 is easily seen to be tight with respect to all parameters; see Remark 6.2. Theorem 1 has the following consequence, which the reader may find counterintuitive: high-degree polynomials are more easily made robust than low-degree polynomials, in the sense that a degree-d polynomial can be made robust within error $2^{-\Theta(d)}$ with only a constant-factor increase in degree.

A final point of interest is that Theorem 1 gives an explicit, formulaic construction of a robust polynomial p_{robust} in terms of the original polynomial p. Prior to this work, no explicit robust construction was known even for the parity polynomial $p(x) = x_1x_2\cdots x_n$. To quote Buhrman et al. [13], “We are not aware of a direct ‘closed form’ or other natural way to describe a robust degree-$O(n)$ polynomial for the parity of n bits, but can only infer its existence from the existence of a robust quantum algorithm. Given the simplicity of the non-robust representing polynomial for parity, one would hope for a simple closed form for robust polynomials for parity as well.”

As a consequence of Theorem 1, we conclude that the approximate degree behaves nicely under function composition:

COROLLARY. *For all Boolean functions f and g,*

$$\widetilde{\deg}(f(g,g,\dots,g)) = O(\widetilde{\deg}(f)\,\widetilde{\deg}(g)).$$

Prior to this paper, this conclusion was known to hold only for several special functions, e.g., [11, 30, 13], and required quantum query arguments.

Our techniques. We will now overview the techniques of previous work and contrast them with the approach of this paper. Buhrman et al. [13] gave a remarkable quantum algorithm that recovers an n-bit string with constant probability from $O(n)$ noisy queries to the bits of the string. As an immediate corollary, the authors of [13] concluded that every Boolean function has a robust polynomial of degree $O(n)$. Unfortunately, there does not seem to be a way to modify this argument to obtain a degree-$o(n)$ robust polynomial for functions with sublinear approximate degree. With an unrelated, combinatorial argument, the authors of [13] obtained an upper bound of $O(\widetilde{\deg}(f)\log\widetilde{\deg}(f))$ on the degree of a robust polynomial for any given Boolean function f. This combinatorial argument also seems to be of no use in proving Theorem 1. For one thing, it is unclear how to save a logarithmic factor in the combinatorial analysis, and more fundamentally, the combinatorial argument only works for approximating *Boolean* functions rather than *arbitrary* real functions $\{-1,+1\}^n \to [-1,1]$.

We approach the problem of robust approximation differently, with a direct analytic treatment rather than combinatorics or quantum query complexity. Our solution comprises three steps, corresponding to functions of increasing generality:

(i) robust approximation of the parity polynomial, $p(x) = x_1x_2\cdots x_n$;
(ii) robust approximation of homogeneous polynomials, $p(x) = \sum_{|S|=d} a_S \prod_{i\in S} x_i$;
(iii) robust approximation of arbitrary polynomials.

For step (i), we construct an exact representation of the sign function on the domain $[-4/3,-2/3]\cup[2/3,4/3]$ as an analytic series whose coefficients decrease exponentially with degree. Multiplying n such series, we show that the resulting coefficients still decay rapidly enough to allow truncation at degree $O(n)$.

For step (iii), we write a general polynomial $p\colon\{-1,+1\}^n \to [-1,1]$ as the sum of its homogeneous parts $p = p_0 + p_1 + p_2 + \cdots + p_d$, where d is the degree of p. Using approximation theory and a convexity argument, we show that $\|p_i\|_\infty \leqslant 2^{O(d)}$ for all i. For our purposes, all this means is that a robust polynomial for p can be obtained by summing the robust polynomials for all p_i with sufficiently small error, $2^{-\Omega(d)}$. Obtaining such a polynomial for each p_i is the content of step (ii).

Step (ii) is the most difficult part of the proof. A natural approach to the robust approximation of a homogeneous polynomial p is to robustly approximate every monomial in p to within a suitable error ϵ, using the construction from step (i). Since we want the robust polynomial for p to have degree $O(d)$, the smallest setting that we can afford is $\epsilon = 2^{-\Theta(d)}$. Unfortunately, there is no reason to believe that with this ϵ, the proposed robust polynomial will have small error in approximating p. As a matter of fact, a direct calculation even suggests that this approach is doomed: it is straightforward to verify that a homogeneous polynomial $p\colon\{-1,+1\}^n \to [-1,1]$ of degree d can have $\binom{n}{d}$ monomials, each equal to $\pm\bigl(2n\binom{n}{d}\bigr)^{-1/2}$, which suggests that the proposed approximant for p could have error as large as

$$\epsilon\binom{n}{d}\left\{2n\binom{n}{d}\right\}^{-1/2} \gg 1.$$

Surprisingly, we are able to show that the proposed robust approximant for p does work and furthermore has excellent error, $2^{-\Theta(d)}$.

We now describe step (ii) in more detail. The naïve, term-by-term error analysis above ignores key aspects of the problem, such as the convexity of the unit cube $[-1,1]^n$, the metric structure of the hypercube $\{-1,+1\}^n$, and the multilinearity of p. We contribute a novel technique that exploits these considerations. In particular, we are able to express the error in the proposed approximant at any given point $z \in ([-4/3,-2/3]\cup[2/3,4/3])^n$ as an infinite series

$$\sum_{i=1}^{\infty} a_i\, p(z_i),$$

where each $z_i = z_i(z)$ is a suitable point in $[-1,1]^n$, and the coefficients in the series are small and decay rapidly: $\sum_{i=1}^{\infty} |a_i| \leqslant 2^{-\Theta(d)}$. Since p is bounded by 1 in absolute value on the hypercube $\{-1,+1\}^n$, it is also bounded by 1 inside the convex cube $[-1,1]^n$, leading to the desired error estimate. In words, even though the error in the approximation of an individual monomial is relatively large, we show that the errors across the monomials behave in a coordinated way and essentially cancel each other out.

2. Notation and Preliminaries

Throughout this manuscript, we represent the Boolean values "true" and "false" by -1 and $+1$, respectively. In particular, Boolean functions are mappings $f\colon X \to \{-1,+1\}$ for some finite set X such as $X = \{-1,+1\}^n$. The natural numbers are denoted $\mathbb{N} = \{0,1,2,3,\dots\}$. The symbol $\log x$ denotes the logarithm of x to base 2. For a string $x \in \mathbb{R}^n$ and a set $S \subseteq \{1,2,\dots,n\}$, we adopt the shorthand $x|_S = (x_{i_1}, x_{i_2}, \dots, x_{i_{|S|}}) \in \mathbb{R}^{|S|}$, where $i_1 < i_2 < \dots < i_{|S|}$ are the elements of S. The family of all subsets of a given set X is denoted $\mathscr{P}(X)$. The symbol S_n stands for the group of permutations $\sigma\colon \{1,2,\dots,n\} \to \{1,2,\dots,n\}$. A function $\phi\colon \mathbb{R}^n \to \mathbb{R}$ is called *symmetric* if ϕ is invariant under permutations of the variables, i.e., $\phi(x) \equiv \phi(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$ for all $\sigma \in S_n$. We adopt the standard definition of the sign function:

$$\operatorname{sgn} t = \begin{cases} -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases}$$

For a set X, we let $\mathbb{R}^X$ denote the real vector space of functions $X \to \mathbb{R}$. For $\phi \in \mathbb{R}^X$, we write

$$\|\phi\|_\infty = \sup_{x \in X} |\phi(x)|, \qquad \|\phi\|_1 = \sum_{x \in X} |\phi(x)|,$$

where the symbol $\|\phi\|_1$ is reserved for finite X. By the *degree* of a multivariate polynomial p on $\mathbb{R}^n$, denoted $\deg p$, we shall always mean the total degree of p, i.e., the greatest total degree of any monomial of p. The symbol P_d stands for the family of all univariate real polynomials of degree up to d.

Fourier transform. Consider the real vector space of functions $\{-1,+1\}^n \to \mathbb{R}$. For $S \subseteq \{1,2,\dots,n\}$, define $\chi_S\colon \{-1,+1\}^n \to \{-1,+1\}$ by $\chi_S(x) = \prod_{i \in S} x_i$. Then the functions χ_S, $S \subseteq \{1,2,\dots,n\}$, form an orthogonal basis for the vector space in question. In particular, every function $\phi\colon \{-1,+1\}^n \to \mathbb{R}$ has a unique representation as a linear combination of the characters χ_S:

$$\phi = \sum_{S \subseteq \{1,2,\dots,n\}} \hat{\phi}(S)\, \chi_S,$$

where $\hat{\phi}(S) = 2^{-n} \sum_{x \in \{-1,+1\}^n} \phi(x) \chi_S(x)$ is the *Fourier coefficient* of ϕ that corresponds to the character χ_S. Note that

$$\deg \phi = \max\{|S| : \hat{\phi}(S) \neq 0\}.$$

Formally, the Fourier transform is the linear transformation $\phi \mapsto \hat{\phi}$, where $\hat{\phi}$ is viewed as a function $\mathscr{P}(\{1,2,\dots,n\}) \to \mathbb{R}$. In particular, we have the shorthand

$$\|\hat{\phi}\|_1 = \sum_{S \subseteq \{1,2,\dots,n\}} |\hat{\phi}(S)|.$$

Multilinear extensions and convexity. As the previous paragraph shows, associated to every mapping $\phi\colon\{-1,+1\}^n \to \mathbb{R}$ is a unique multilinear polynomial $\tilde{\phi}\colon\mathbb{R}^n \to \mathbb{R}$ such that $\phi \equiv \tilde{\phi}$ on $\{-1,+1\}^n$. In discussing the Fourier transform, we identified ϕ with its multilinear extension $\tilde{\phi}$ to $\mathbb{R}^n$, and will continue to do so throughout this paper. Among other things, this convention allows one to evaluate ϕ everywhere in $[-1,1]^n$ as opposed to just $\{-1,+1\}^n$. It is a simple but important fact that for every $\phi\colon\{-1,+1\}^n \to \mathbb{R}$,

$$\max_{x\in[-1,1]^n} |\phi(x)| = \max_{x\in\{-1,+1\}^n} |\phi(x)| = \|\phi\|_\infty.$$

To see this, fix $\xi \in [-1,1]^n$ arbitrarily and consider the probability distribution on strings $x \in \{-1,+1\}^n$ whereby $x_1,\dots,x_n$ are distributed independently and $\mathbf{E}[x_i] = \xi_i$ for all i. Then $\phi(\xi) = \mathbf{E}[\phi(x)]$ by multilinearity, so that $|\phi(\xi)| \leqslant \max_{x\in\{-1,+1\}^n} |\phi(x)|$.

3. A Robust Polynomial for Parity

The objective of this section is to construct a low-degree robust polynomial for the parity function. In other words, we will construct a polynomial $p\colon\mathbb{R}^n \to \mathbb{R}$ of degree $O(n)$ such that $p(x_1,x_2,\dots,x_n) \approx \prod \operatorname{sgn} x_i$ whenever the input variables are close to Boolean: $x_1,x_2,\dots,x_n \in [-4/3,-2/3] \cup [2/3,4/3]$. Recall that our eventual goal is a robust polynomial for every bounded real function. To this end, the parity approximant p that we are to construct needs to possess a key additional property: the error $p(x) - \prod \operatorname{sgn} x_i$, apart from being small, needs to be expressible as a multivariate series in which the coefficients decay rapidly with monomial order.

To obtain this coefficient behavior, we use a carefully chosen approximant for the univariate function $\operatorname{sgn} t$. The simplest candidate is the following ingenious construction due to Buhrman et al. [13]:

$$B_n(t) = 2^{-n} \sum_{i=\lceil n/2\rceil}^{n} \binom{n}{i} t^i (1-t)^{n-i}.$$

In words, $B_n(t)$ is the probability of observing more heads than tails in a sequence of n independent coin flips, each coming up heads with probability t. By the Chernoff bound, B_n sends $[0,1/4] \to [0,2^{-\Omega(n)}]$ and similarly $[3/4,1] \to [1-2^{-\Omega(n)},1]$. As Buhrman et al. [13] point out, this immediately gives a degree-n approximant for the sign function with exponentially small error on $[-4/3,-2/3] \cup [2/3,4/3]$. Unfortunately, the coefficients of this approximating polynomial do not exhibit the kind of rapid decay that we require. Instead, in what follows we use a purely analytic construction based on the Maclaurin series for $1/\sqrt{1+t}$.

LEMMA 3.1. *For $x_1,x_2,\dots,x_n \in (-\sqrt{2},0) \cup (0,\sqrt{2})$,*

$$\operatorname{sgn}(x_1x_2\cdots x_n) = x_1x_2\cdots x_n \sum_{i_1,i_2,\dots,i_n\in\mathbb{N}} \prod_{j=1}^{n} \left(-\frac{1}{4}\right)^{i_j} \binom{2i_j}{i_j} (x_j^2-1)^{i_j}. \tag{3.1}$$

Proof. Recall the binomial series

$$(1+t)^\alpha = \sum_{i=0}^{\infty} \binom{\alpha}{i} t^i, \tag{3.2}$$

with the usual notation $\binom{\alpha}{i} = \alpha(\alpha-1)\cdots(\alpha-i+1)/i!$ for the generalized binomial coefficient. The series (3.2) is valid for all $-1 < t < 1$ and all real α. In particular, setting

$\alpha = -1/2$ gives

$$\begin{aligned} \frac{1}{\sqrt{1+t}} &= \sum_{i=0}^{\infty} \binom{-1/2}{i} t^i \\ &= \sum_{i=0}^{\infty} \left(-\frac{1}{4}\right)^i \binom{2i}{i} t^i, \qquad -1 < t < 1. \end{aligned} \tag{3.3}$$

One easily verifies that this absolutely convergent series is the Maclaurin expansion for $1/\sqrt{1+t}$. For all real t with $0 < |t| < \sqrt{2}$, we have

$$\begin{aligned} \operatorname{sgn} t &= \frac{t}{\sqrt{1+(t^2-1)}} \\ &= t \sum_{i=0}^{\infty} \left(-\frac{1}{4}\right)^i \binom{2i}{i} (t^2-1)^i, \end{aligned} \tag{3.4}$$

where the second step holds by (3.3). For $x_1, x_2, \ldots, x_n \in (-\sqrt{2}, 0) \cup (0, \sqrt{2})$, it follows that

$$\begin{aligned} \operatorname{sgn}(x_1 x_2 \ldots x_n) &= x_1 x_2 \cdots x_n \prod_{j=1}^{n} \left\{ \sum_{i=0}^{\infty} \left(-\frac{1}{4}\right)^i \binom{2i}{i} (x_j^2 - 1)^i \right\} \\ &= x_1 x_2 \cdots x_n \sum_{i_1, i_2, \ldots, i_n \in \mathbb{N}} \prod_{j=1}^{n} \left(-\frac{1}{4}\right)^{i_j} \binom{2i_j}{i_j} (x_j^2 - 1)^{i_j}. \qquad \square \end{aligned}$$

We have reached the main result of this section.

THEOREM 3.2 (Robust polynomial for the parity function). *Fix* $\epsilon \in [0, 1)$ *and let*

$$X = [-\sqrt{1+\epsilon}, -\sqrt{1-\epsilon}] \cup [\sqrt{1-\epsilon}, \sqrt{1+\epsilon}].$$

Then for every natural number N, *there is an* (*explicitly given*) *polynomial* $p \colon \mathbb{R}^n \to \mathbb{R}$ *of degree at most* $2N + n$ *such that*

$$\max_{X^n} |\operatorname{sgn}(x_1 x_2 \cdots x_n) - p(x)| \leqslant \epsilon^N (1+\epsilon)^{n/2} \binom{N+n}{N} N. \tag{3.5}$$

Setting $\epsilon = 7/9$ in this result, we infer that the function $\operatorname{sgn}(x_1 x_2 \cdots x_n)$ with inputs $x_1, x_2, \ldots, x_n \in [-4/3, -2/3] \cup [2/3, 4/3]$ can be approximated to within $2^{-\Omega(n)}$ everywhere by a polynomial of degree $O(n)$. This is the desired robust polynomial for parity.

Proof of Theorem 3.2. For a natural number d, let $\mathscr{I}_d$ stand for the family of n-tuples $(i_1, \ldots, i_n)$ of nonnegative integers such that $i_1 + \cdots + i_n = d$. Clearly,

$$|\mathscr{I}_d| = \binom{d+n-1}{d}.$$

One can restate (3.1) in the form

$$\operatorname{sgn}(x_1 x_2 \cdots x_n) = x_1 x_2 \cdots x_n \sum_{d=0}^{\infty} \xi_d(x_1, x_2, \ldots, x_n), \tag{3.6}$$

where

$$\xi_d(x_1,x_2,\dots,x_n)=\sum_{(i_1,\dots,i_n)\in\mathscr{I}_d}\prod_{j=1}^{n}\left(-\frac{1}{4}\right)^{i_j}\binom{2i_j}{i_j}(x_j^2-1)^{i_j}.$$

On X^n,

$$\|\xi_d\|_\infty\leqslant\epsilon^d|\mathscr{I}_d|=\epsilon^d\binom{d+n-1}{d}.$$

As a result, dropping the terms $\xi_{N+1},\xi_{N+2},\dots$ from the infinite series (3.6) results in a uniform approximant of degree $2N+n$ with pointwise error at most

$$\begin{aligned}(1+\epsilon)^{n/2}\sum_{d=N+1}^{\infty}\epsilon^d\binom{d+n-1}{d}&\\ &\leqslant(1+\epsilon)^{n/2}\,\epsilon^{N+1}\binom{N+n}{N+1}\sum_{d=0}^{\infty}\left(\epsilon\cdot\frac{N+n+1}{N+2}\right)^d.\end{aligned}$$

This gives (3.5) provided that $\epsilon\leqslant N/(N+n)$. For larger ϵ, the bound (3.5) exceeds 1 and thus holds trivially with $p=0$. ∎

4. Reduction to Homogeneous Polynomials

We now turn to the construction of a robust polynomial for any real function on the Boolean cube. Real functions given by homogeneous polynomials on $\{-1,+1\}^n$ are particularly convenient to work with, and the proof is greatly simplified by first reducing the problem to the homogeneous case.

To obtain this reduction, we need to bound the coefficients of a *univariate* polynomial in terms of its degree d and maximum value on [0, 1]. In general, the coefficients can grow quite rapidly with degree. For example, the Chebyshev polynomial of degree d is bounded by 1 in absolute value throughout $[-1,1]$ and nevertheless has leading coefficient 2^{d-1}; see Cheney [16] and Rivlin [50] for an exposition. The following first-principles calculation shows that this rate of growth is the highest possible, up to an asymptotic constant in the exponent. In fact, the proof below works even if the polynomial is known to bounded by 1 on a small finite set of equispaced points in $[-1,1]$, as opposed to all of $[-1,1]$.

LEMMA 4.1 (Coefficients of bounded polynomials). *Let $p(t)=\sum_{i=0}^{d}a_it^i$ be a given polynomial. Then*

$$|a_i|\leqslant(4\mathrm{e})^d\max_{j=0,1,\dots,d}\left|p\left(\frac{j}{d}\right)\right|,\qquad i=0,1,\dots,d. \tag{4.1}$$

Proof. The first step is to express p as a linear combination of more structured polynomials, by means of Lagrange interpolation with nodes $\{i/d : i=0,1,2,\dots,d\}$. For this, define $q_0,q_1,\dots,q_d\in P_d$ by

$$q_j(t)=\frac{(-1)^{d-j}d^d}{d!}\binom{d}{j}\prod_{\substack{i=0\\i\neq j}}^{d}\left(t-\frac{i}{d}\right),\qquad j=0,1,\dots,d.$$

One easily verifies that these polynomials behave like delta functions, in the sense that for $i, j = 0, 1, 2, \ldots, d$,

$$q_j\left(\frac{i}{d}\right) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$p = \sum_{j=0}^{d} p\left(\frac{j}{d}\right) q_j.$$

By linearity, it suffices to bound the coefficients of the q_j. The closed form for these polynomials reveals the following rough estimate: if $q_j(t) = \sum_{i=0}^{d} b_{ij} t^i$, then

$$|b_{ij}| \leqslant \frac{d^d}{d!} \binom{d}{i} \binom{d}{j} \leqslant (2\mathrm{e})^d \binom{d}{j}.$$

As a result,

$$|a_i| \leqslant \left(\sum_{j=0}^{d} |b_{ij}|\right) \max_{j=0,1,\ldots,d} \left| p\left(\frac{j}{d}\right) \right| \leqslant (4\mathrm{e})^d \max_{j=0,1,\ldots,d} \left| p\left(\frac{j}{d}\right) \right|. \qquad \square$$

We are now prepared to give the desired reduction to the homogeneous case.

THEOREM 4.2. *Let* $\phi\colon \{-1,+1\}^n \to \mathbb{R}$ *be a given function,* $\deg \phi = d$. *Write* $\phi = \phi_0 + \phi_1 + \cdots + \phi_d$, *where* $\phi_i\colon \{-1,+1\}^n \to \mathbb{R}$ *is given by* $\phi_i = \sum_{|S|=i} \hat{\phi}(S) \chi_S$. *Then*

$$\|\phi_i\|_\infty \leqslant (4\mathrm{e})^d \|\phi\|_\infty, \qquad i = 0, 1, \ldots, d.$$

The above result gives an upper bound on the infinity norm of the homogeneous parts of a polynomial ϕ in terms of the infinity norm of ϕ itself. Note that the bound is entirely independent of the number of variables. For our purposes, Theorem 4.2 has the following consequence: a robust polynomial for ϕ can be obtained by constructing robust polynomials with error $2^{-\Omega(d)} \|\phi\|_\infty$ separately for each of the homogeneous parts. The homogeneous problem will be studied in the next section.

Proof of Theorem 4.2. Pick a point $x \in \{-1,+1\}^n$ arbitrarily and fix it for the remainder of the proof. Consider the univariate polynomial $p \in P_d$ given by

$$p(t) = \sum_{i=0}^{d} \phi_i(x) t^i.$$

For $-1 \leqslant t \leqslant 1$, consider the probability distribution μ_t on the Boolean cube $\{-1,+1\}^n$ whereby each bit is independent and has expected value t. Then

$$\begin{aligned}\|\phi\|_\infty &\geqslant \left|\mathop{\mathbf{E}}_{z\sim\mu_t}[\phi(x_1z_1,\dots,x_nz_n)]\right|\\ &= \left|\sum_{|S|\leqslant d}\hat\phi(S)\mathop{\mathbf{E}}_{z\sim\mu_t}\left[\prod_{i\in S}x_iz_i\right]\right|\\ &= \left|\sum_{|S|\leqslant d}\hat\phi(S)t^{|S|}\prod_{i\in S}x_i\right|\\ &= |p(t)|.\end{aligned}$$

Hence, p is bounded on $[-1,1]$ in absolute value by $\|\phi\|_\infty$. By Lemma 4.1, it follows that the coefficients of p do not exceed $(4\mathrm{e})^d\|\phi\|_\infty$:

$$|\phi_i(x)| \leqslant (4\mathrm{e})^d\|\phi\|_\infty.$$

Since the choice of $x \in \{-1,+1\}^n$ was arbitrary, the theorem follows. ☐

5. Error Cancellation in Homogeneous Polynomials

In Section 3, we constructed a robust polynomial for the parity function. Recall that the goal of this paper is to construct a degree-$O(d)$ robust polynomial for any degree-d real function $\phi\colon\{-1,+1\}^n \to [-1,1]$. By the results of Section 4, we may now assume that ϕ is homogeneous:

$$\phi = \sum_{|S|=d}\hat\phi(S)\chi_S.$$

A naïve approach would be to use the construction of Section 3 and robustly approximate each parity χ_S to within $2^{-\Theta(d)}$ by a degree-$O(d)$ polynomial. Unfortunately, it is unclear whether the resulting polynomial would be a good approximant for ϕ. Indeed, as explained in the introduction, the cumulative error in this approximation could conceivably be as large as $n^{\Omega(d)}2^{-\Theta(d)} \gg 1$. The purpose of this section is to prove that, for a careful choice of approximants for the χ_S, the errors do not compound but instead partially cancel, resulting in a cumulative error of $2^{-\Theta(d)}$. The proof is rather technical. To simplify the exposition, we first illustrate our technique in the simpler setting of $\{-1,+1\}^n$ and then adapt it to our setting of interest, $\mathbb{R}^n$.

Error cancellation on the Boolean hypercube. Let $\phi\colon\{-1,+1\}^n \to \mathbb{R}$ be a degree-d homogeneous polynomial. Our goal is to show that perturbing the Fourier characters of ϕ in a suitable, coordinated manner results in partial cancellation of the errors and does not change the value of ϕ by much relative to the norm $\|\phi\|_\infty$. A precise statement follows.

THEOREM 5.1. *Let $\phi\colon\{-1,+1\}^n \to \mathbb{R}$ be given such that $\hat\phi(S) = 0$ whenever $|S| \neq d$. Fix an arbitrary symmetric function $\delta\colon\{-1,+1\}^d \to \mathbb{R}$ and define $\Delta\colon\{-1,+1\}^n \to \mathbb{R}$ by*

$$\Delta(x) = \sum_{|S|=d}\hat\phi(S)\delta(x|_S).$$

Then

$$\|\Delta\|_\infty \leqslant \frac{d^d}{d!}\,\|\phi\|_\infty \|\hat{\delta}\|_1.$$

In the above result, δ should be thought of as the error in approximating individual characters χ_S, whereas Δ is the cumulative error so incurred. The theorem states that the cumulative error exceeds the norm of ϕ and δ by a factor of only e^d, which is substantially smaller than the factor of $n^{\Omega(d)}$ growth that one could expect *a priori*.

Proof of Theorem 5.1. We adopt the convention that $a^0 = 1$ for all real a. For a given vector $v \in \{0,1\}^d$, consider the operator A_v that takes a function $f\colon \{-1,+1\}^n \to \mathbb{R}$ into another function $A_v f\colon \{-1,+1\}^n \to \mathbb{R}$ where

$$(A_v f)(x) = \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1 z_2 \cdots z_d\, f\left(\frac{1}{d}\sum_{i=1}^d z_i x_1^{v_i}, \ldots, \frac{1}{d}\sum_{i=1}^d z_i x_n^{v_i}\right)\right].$$

It is important to note that A_v is a linear transformation in the vector space $\mathbb{R}^{\{-1,+1\}^n}$. This somewhat magical operator is the key to the proof; the remainder of the proof will provide insight into how this definition could have been arrived at in the first place. To start with,

$$\|A_v\phi\|_\infty \leqslant \max_{x\in[-1,1]^n} |\phi(x)| = \max_{x\in\{-1,+1\}^n} |\phi(x)| = \|\phi\|_\infty, \tag{5.1}$$

where the second step holds by convexity. The strategy of the proof is to express Δ as a linear combination of the $A_v\phi$ with small coefficients. Since the infinity norm of each individual $A_v\phi$ is small, this will give the desired bound on the infinity norm of Δ.

To find what suitable coefficients would be, we need to understand the transformation A_v in terms of the Fourier spectrum. Since A_v is linear and the nonzero Fourier coefficients of ϕ have order d, it suffices to determine the action of A_v on the characters of order d. For every $S \subseteq \{1,2,\ldots,n\}$ with $|S| = d$,

$$\begin{aligned}(A_v\chi_S)(x) &= \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1 z_2\cdots z_d \prod_{j\in S}\left(\frac{1}{d}\sum_{i=1}^d z_i x_j^{v_i}\right)\right]\\ &= \mathop{\mathbf{E}}_{\tau\colon S\to\{1,\ldots,d\}}\left[\mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1 z_2\cdots z_d \prod_{j\in S} z_{\tau(j)}\right]\prod_{j\in S} x_j^{v_{\tau(j)}}\right],\end{aligned}$$

where the outer expectation is over a uniformly random mapping $\tau\colon S \to \{1,2,\ldots,d\}$. The inner expectation over z acts like the indicator random variable for the event that τ is a bijection, i.e., it evaluates to 1 when τ is a bijection and vanishes otherwise. As a result,

$$\begin{aligned}(A_v\chi_S)(x) &= \mathop{\mathbf{P}}_{\tau}[\tau \text{ is a bijection}]\ \mathop{\mathbf{E}}_{\tau}\left[\prod_{j\in S} x_j^{v_{\tau(j)}} \mid \tau \text{ is a bijection}\right]\\ &= \frac{d!}{d^d} \mathop{\mathbf{E}}_{\substack{T\subseteq S,\\ |T|=v_1+\cdots+v_d}} [\chi_T(x)]. \end{aligned}\tag{5.2}$$

By the symmetry of δ,

$$\begin{aligned}\delta(x|_S) &= \sum_{k=0}^{d} \hat{\delta}(\{1,2,\ldots,k\}) \sum_{\substack{T\subseteq S,\\ |T|=k}} \chi_T(x)\\ &= \frac{d^d}{d!}\sum_{k=0}^{d} \hat{\delta}(\{1,2,\ldots,k\})\binom{d}{k}(A_{1^k0^{d-k}}\chi_S)(x),\end{aligned}$$

where the second step uses (5.2). Taking a weighted sum over S and using the linearity of A_v,

$$\begin{aligned}&\sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\ |S|=d}} \hat{\phi}(S)\delta(x|_S)\\ &\qquad = \frac{d^d}{d!}\sum_{k=0}^{d} \hat{\delta}(\{1,2,\ldots,k\})\binom{d}{k}\left(A_{1^k0^{d-k}}\sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\ |S|=d}} \hat{\phi}(S)\chi_S\right)(x),\end{aligned}$$

or equivalently

$$\Delta = \frac{d^d}{d!}\sum_{k=0}^{d} \hat{\delta}(\{1,2,\ldots,k\})\binom{d}{k}A_{1^k0^{d-k}}\phi.$$

In light of (5.1), this representation gives the sought upper bound on the norm of Δ:

$$\|\Delta\|_\infty \leqslant \frac{d^d}{d!}\sum_{k=0}^{d} |\hat{\delta}(\{1,2,\ldots,k\})|\binom{d}{k}\|\phi\|_\infty = \frac{d^d}{d!}\,\|\phi\|_\infty\|\hat{\delta}\|_1. \qquad \square$$

Error cancellation with real variables. We now consider the error cancellation problem in its full generality. Again, our goal will be to show that replacing individual characters with suitable approximants results in moderate cumulative error. This time, however, the input variables are no longer restricted to be Boolean, and can take on arbitrary values in $[-1-\epsilon,-1+\epsilon]\cup[1-\epsilon,1+\epsilon]$ for $0<\epsilon<1$. This in turn means that the error term will be given by an infinite series. Another difference is that the coefficients of the error series will not converge to zero rapidly enough, requiring additional ideas to bound the cumulative error.

THEOREM 5.2. *Let $\phi\colon\{-1,+1\}^n\to\mathbb{R}$ be given such that $\hat{\phi}(S)=0$ whenever $|S|\neq d$. Fix $\epsilon\in[0,1)$ and let*

$$X=[-\sqrt{1+\epsilon},-\sqrt{1-\epsilon}]\cup[\sqrt{1-\epsilon},\sqrt{1+\epsilon}].$$

Then for every natural number D, there is an (explicitly given) polynomial $p\colon\mathbb{R}^d\to\mathbb{R}$ of degree at most $2D+d$ such that

$$P(x)=\sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\ |S|=d}} \hat{\phi}(S)p(x|_S)$$

obeys

$$\max_{X^n} |\phi(\operatorname{sgn} x_1, \ldots, \operatorname{sgn} x_n) - P(x)| \leqslant (1+\epsilon)^{d/2}\, \frac{d^d}{d!}\, \epsilon^D \binom{D+d}{D} D\, \|\phi\|_\infty. \tag{5.3}$$

Proof. As before, we adopt the notational convention that $a^0 = 1$ for all real a. We will follow the proof of Theorem 5.1 as closely as possible, pointing out key differences as we go along.

If $\epsilon > D/(D+d)$, then the right member of (5.3) exceeds $\|\phi\|_\infty$ and hence the theorem holds trivially with $p = 0$. We may therefore assume that $\epsilon \leqslant D/(D+d)$, which means in particular that

$$\begin{aligned}\sum_{\substack{v \in \mathbb{N}^d:\\ v_1+\cdots+v_d \geqslant D+1}} \epsilon^{v_1+\cdots+v_d} \prod_{j=1}^{d} \left(\frac{1}{4}\right)^{v_j} \binom{2v_j}{v_j} &\leqslant \sum_{i=D+1}^{\infty} \binom{i+d-1}{i} \epsilon^i \\ &\leqslant \epsilon^D \binom{D+d}{D} D. \end{aligned} \tag{5.4}$$

Define p by

$$p(x_1, \ldots, x_d) = \sum_{\substack{v \in \mathbb{N}^d:\\ v_1+\cdots+v_d \leqslant D}} \prod_{j=1}^{d} \left(-\frac{1}{4}\right)^{v_j} \binom{2v_j}{v_j} x_j (x_j^2 - 1)^{v_j}.$$

Analogous to the Boolean setting, we will define functions to capture the error in approximating an individual character as well as the cumulative error. Let $\delta\colon X^d \to \mathbb{R}$ and $\Delta\colon X^n \to \mathbb{R}$ be given by

$$\delta(x_1, \ldots, x_d) = \sum_{\substack{v \in \mathbb{N}^d:\\ v_1+\cdots+v_d \geqslant D+1}} \prod_{j=1}^{d} \left(-\frac{1}{4}\right)^{v_j} \binom{2v_j}{v_j} x_j (x_j^2 - 1)^{v_j},$$

$$\Delta(x_1, \ldots, x_n) = \sum_{\substack{S \subseteq \{1,2,\ldots,n\}\\ |S| = d}} \hat{\phi}(S)\delta(x|_S).$$

Lemma 3.1 implies that δ is the error incurred in approximating a single character by p, in other words, $\delta(x_1, \ldots, x_d) = \operatorname{sgn}(x_1 \cdots x_d) - p(x_1, \ldots, x_d)$. Hence, Δ captures the cumulative error:

$$\Delta(x) = \phi(\operatorname{sgn} x_1, \ldots, \operatorname{sgn} x_n) - P(x). \tag{5.5}$$

Recall that our goal is to place an upper bound on $\|\Delta\|_\infty$. For $v \in \mathbb{N}^d$, consider the operator A_v that takes a function $f\colon \{-1,+1\}^n \to \mathbb{R}$ into a function $A_v f\colon X^n \to \mathbb{R}$

where

$$(A_v f)(x) = \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d} \left[z_1 z_2 \cdots z_d\, f\left(\ldots,\ \underbrace{\frac{1}{d}\sum_{i=1}^{d} \frac{z_i x_j (x_j^2-1)^{v_i}}{\epsilon^{v_i}\sqrt{1+\epsilon}}}_{j\text{th coordinate}},\ \ldots \right)\right].$$

This definition departs from the earlier one in Theorem 5.1, where v was restricted to $0/1$ entries. Perhaps the most essential difference is the presence of scaling factors in the denominator—it is what ultimately allows one to bound the cumulative error in the setting of an infinite series. Note that A_v is a linear transformation sending $\mathbb{R}^{\{-1,+1\}^n}$ into $\mathbb{R}^{X^n}$. We further have

$$\|A_v\phi\|_\infty \leqslant \max_{x\in[-1,1]^n} |\phi(x)| = \max_{x\in\{-1,+1\}^n} |\phi(x)| = \|\phi\|_\infty, \tag{5.6}$$

where the first step uses the fact that $A_v\phi$ has domain X^n rather than all of $\mathbb{R}^n$, and the second step holds by convexity.

We proceed to examine the action of A_v on the characters of order d. Since the definition of A_v is symmetric with respect to the n coordinates, it suffices to consider $S = \{1,2,\ldots,d\}$:

$$\begin{aligned}(A_v\chi_{\{1,\ldots,d\}})(x) &= \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d} \left[z_1 z_2\cdots z_d \prod_{j=1}^{d}\left(\frac{1}{d}\sum_{i=1}^{d}\frac{z_i x_j(x_j^2-1)^{v_i}}{\epsilon^{v_i}\sqrt{1+\epsilon}}\right)\right] \\ &= \frac{1}{(1+\epsilon)^{d/2}} \mathop{\mathbf{E}}_{\tau}\left[\mathop{\mathbf{E}}_{z}\left[\prod_{j=1}^{d} z_j z_{\tau(j)}\right] \prod_{j=1}^{d}\frac{x_j(x_j^2-1)^{v_{\tau(j)}}}{\epsilon^{v_{\tau(j)}}}\right],\end{aligned}$$

where the first expectation is taken over a uniformly random mapping $\tau\colon\{1,2,\ldots,d\}\to\{1,2,\ldots,d\}$. Let B stand for the event that τ is a bijection. The expectation over z acts like the indicator random variable for B, i.e., it evaluates to 1 when B occurs and vanishes otherwise. Thus,

$$\begin{aligned}(A_v\chi_{\{1,\ldots,d\}})(x) &= \frac{1}{(1+\epsilon)^{d/2}} \mathop{\mathbf{P}}_{\tau}[B] \mathop{\mathbf{E}}_{\tau}\left[\prod_{j=1}^{d}\frac{x_j(x_j^2-1)^{v_{\tau(j)}}}{\epsilon^{v_{\tau(j)}}} \;\middle|\; B\right] \\ &= \frac{1}{(1+\epsilon)^{d/2}\epsilon^{v_1+\cdots+v_d}}\cdot\frac{d!}{d^d}\cdot \mathop{\mathbf{E}}_{\sigma\in S_d}\left[\prod_{j=1}^{d} x_j(x_j^2-1)^{v_{\sigma(j)}}\right]. \end{aligned}\tag{5.7}$$

Now, consider the operator

$$A = (1+\epsilon)^{d/2}\,\frac{d^d}{d!} \sum_{\substack{v\in\mathbb{N}^d:\\ v_1+\cdots+v_d\geqslant D+1}} \left\{\epsilon^{v_1+\cdots+v_d}\prod_{j=1}^{d}\left(-\frac{1}{4}\right)^{v_j}\binom{2v_j}{v_j}\right\} A_v.$$

This operator is well-defined because by (5.4), the infinite series in question converges absolutely. By (5.7), the symmetry of δ, and linearity, $(A\chi_{\{1,\ldots,d\}})(x) = \delta(x_1,\ldots,x_d)$. Since the definition of A is symmetric with respect to the n coordinates, we conclude that

$$(A\chi_S)(x) = \delta(x|_S)$$

for all subsets $S \subseteq \{1, 2, \ldots, n\}$ of cardinality d. As an immediate consequence,

$$\Delta = \sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\|S|=d}} \hat{\phi}(S) \cdot (A\chi_S) = A\left(\sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\|S|=d}} \hat{\phi}(S)\chi_S\right) = A\phi,$$

where the second step uses the linearity of A. In particular,

$$\begin{aligned}\|\Delta\|_\infty &= \|A\phi\|_\infty \\ &\leqslant (1+\epsilon)^{d/2}\,\frac{d^d}{d!} \sum_{\substack{v\in\mathbb{N}^d:\\ v_1+\cdots+v_d\geqslant D+1}} \epsilon^{v_1+\cdots+v_d}\,\|A_v\phi\|_\infty \prod_{j=1}^{d}\left(\frac{1}{4}\right)^{v_j}\binom{2v_j}{v_j} \\ &\leqslant (1+\epsilon)^{d/2}\,\frac{d^d}{d!}\,\epsilon^D\binom{D+d}{D} D\,\|\phi\|_\infty,\end{aligned}$$

where the final step follows by (5.4) and (5.6). In light of (5.5), the proof is complete. □

6. Main Result

We are now in a position to prove the main result of this paper, which states that every bounded real polynomial can be made robust with only a constant-factor increase in degree. Recall that we have already proved this fact for homogeneous polynomials (see Theorems 5.1 and 5.2). It remains to remove the homogeneity assumption, which we will do using the technique of Section 4. For the purposes of exposition, we will first show how to remove the homogeneity assumption in the much simpler context of Theorem 5.1. Essentially the same technique will then allow us to prove the main result.

THEOREM 6.1. *Let* $\phi\colon\{-1,+1\}^n \to \mathbb{R}$ *be given,* $\deg\phi = d$. *Fix symmetric functions* $\delta_i\colon\{-1,+1\}^i \to \mathbb{R}$, $i = 0, 1, 2, \ldots, d$, *and define* $\Delta\colon\{-1,+1\}^n \to \mathbb{R}$ *by*

$$\Delta(x) = \sum_{|S|\leqslant d} \hat{\phi}(S)\delta_{|S|}(x|_S).$$

Then

$$\|\Delta\|_\infty \leqslant 30^d\,\|\phi\|_\infty \sum_{i=0}^{d} \|\hat{\delta}_i\|_1.$$

The functions $\delta_0, \delta_1, \ldots, \delta_d$ in this result are to be thought of as perturbations of characters of orders $0, 1, \ldots, d$, respectively, and Δ is the cumulative error incurred as a result of these perturbations. As the theorem shows, the cumulative error exceeds the norms of the functions involved by a factor of only $2^{O(d)}$, which is independent of the number of variables.

Proof of Theorem 6.1. We have

$$(6.1) \qquad \|\Delta\|_\infty \leqslant \sum_{i=0}^{d} \|\Delta_i\|_\infty,$$

where $\Delta_i\colon\{-1,+1\}^n \to \mathbb{R}$ is given by $\Delta_i(x) = \sum_{|S|=i} \hat{\phi}(S)\delta_i(x|_S)$. For $i = 0,1,\ldots,d$, consider $\phi_i = \sum_{|S|=i} \hat{\phi}(S)\chi_S$, the degree-$i$ homogeneous part of ϕ. By Theorem 5.1,

$$\|\Delta_i\|_\infty \leqslant \mathrm{e}^i \|\phi_i\|_\infty \|\hat{\delta}_i\|_1. \tag{6.2}$$

By Theorem 4.2,

$$\|\phi_i\|_\infty \leqslant (4\mathrm{e})^d \|\phi\|_\infty, \qquad i = 0,1,\ldots,d. \tag{6.3}$$

Combining (6.1)–(6.3) completes the proof. ∎

We will now apply a similar argument in the setting of real variables. For convenience of notation, we will work with the domain $[-1,-\epsilon] \cup [\epsilon,1]$ rather than $[-1-\epsilon,-1+\epsilon] \cup [1-\epsilon,1+\epsilon]$. Since ϵ ranges freely in $(0,1)$ in both cases, these two choices are equivalent (simply scale the input variables by an appropriate absolute constant).

Main Theorem. *Let $X = [-1,-\epsilon] \cup [\epsilon,1]$. Let $\phi\colon\{-1,+1\}^n \to [-1,1]$ be given, $\deg\phi = d$. Then for each $\delta > 0$, there is a polynomial P of degree $O\big(\frac{1}{\epsilon}d + \frac{1}{\epsilon}\log\frac{1}{\delta}\big)$ such that*

$$\max_{X^n} |\phi(\operatorname{sgn} x_1,\ldots,\operatorname{sgn} x_n) - P(x)| < \delta. \tag{6.4}$$

Furthermore, P is given explicitly in terms of the Fourier spectrum of ϕ.

Letting $\epsilon = 1/2$ immediately implies the main result of this paper, stated as Theorem 1 in the introduction.

Proof. We first consider the case $7/8 \leqslant \epsilon \leqslant 1$. Let $D = D(d,\delta)$ be a parameter to be chosen later. For $i = 0,1,\ldots,d$, consider $\phi_i = \sum_{|S|=i} \hat{\phi}(S)\chi_S$, the degree-$i$ homogeneous part of ϕ. By Theorem 4.2,

$$\|\phi_i\|_\infty \leqslant (4\mathrm{e})^d, \qquad i = 0,1,\ldots,d. \tag{6.5}$$

Theorem 5.2 gives explicit polynomials $p_i\colon\mathbb{R}^i \to \mathbb{R}$, $i = 0,1,2,\ldots,d$, each of degree at most $2D + d$, such that

$$\max_{X^n} \left| \phi_i(\operatorname{sgn} x_1,\ldots,\operatorname{sgn} x_n) - \sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\ |S|=i}} \hat{\phi}(S)p_i(x|_S) \right| \leqslant \frac{K^d D}{2^D}\|\phi_i\|_\infty$$

for some absolute constant $K > 1$. Letting

$$P(x) = \sum_{|S|\leqslant d} \hat{\phi}(S)p_{|S|}(x|_S),$$

we infer that

$$\max_{X^n} |\phi(\operatorname{sgn} x_1,\ldots,\operatorname{sgn} x_n) - P(x)| \leqslant \frac{K^d D}{2^D}\sum_{i=0}^{d}\|\phi_i\|_\infty \leqslant \frac{(d+1)(4\mathrm{e}K)^d D}{2^D},$$

where the last step uses (6.5). Therefore, (6.4) holds with $D = O(d + \log\frac{1}{\delta})$.

To handle the case $\epsilon < 7/8$, basic approximation theory [50] gives an explicit univariate polynomial r of degree $O(1/\epsilon)$ that sends $[-1,-\epsilon] \to [-1,-7/8]$ and $[\epsilon,1] \to [7/8,1]$. In particular, we have $|\phi(\operatorname{sgn} x_1,\ldots,\operatorname{sgn} x_n) - P(r(x_1),\ldots,r(x_n))| < \delta$ everywhere on X^n, where P is the approximant constructed in the previous paragraph. ∎

REMARK 6.2. As stated in the introduction, Theorem 1 gives the best possible upper bound on the degree of a robust polynomial p_{robust} in terms of the degree of the original polynomial p and the error parameter ϵ. To see this, we may assume that p takes on -1 and $+1$ on the hypercube $\{-1, +1\}^n$; this can be achieved by appropriately translating and scaling p, without increasing its infinity norm beyond 1. Without loss of generality, $p(1, 1, \ldots, 1) = 1$ and $p(-1, -1, \ldots, -1) = -1$. As a result, the univariate polynomial $p_{\mathrm{robust}}(t, t, \ldots, t)$ would need to approximate the sign function on $[-4/3, -2/3] \cup [2/3, 4/3]$ to within ϵ, which forces $\deg(p_{\mathrm{robust}}) \geqslant \Omega\left(\log \frac{1}{\epsilon}\right)$ by basic approximation theory [19]. Finally, $\deg p$ is a trivial lower bound on the degree of p_{robust}.

ACKNOWLEDGMENTS

The author would like to thank Harry Buhrman, Dmitry Gavinsky, Ankur Moitra, Ronald de Wolf, and the anonymous reviewers for their useful feedback.

REFERENCES

[1] S. Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1(1):1–28, 2005.

[2] S. Aaronson and Y. Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004.

[3] A. Ambainis. Polynomial degree vs. quantum query complexity. *J. Comput. Syst. Sci.*, 72(2):220–238, 2006.

[4] A. Ambainis, A. M. Childs, B. Reichardt, R. Špalek, and S. Zhang. Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer. In *Proceedings of the Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 363–372, 2007.

[5] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

[6] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[7] P. Beame, T. Huynh, and T. Pitassi. Hardness amplification in proof complexity. In *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing* (STOC), pages 87–96, 2010.

[8] P. Beame and D.-T. Huynh-Ngoc. Multiparty communication complexity and threshold circuit complexity of AC^0. In *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 53–62, 2009.

[9] R. Beigel, N. Reingold, and D. A. Spielman. PP is closed under intersection. *J. Comput. Syst. Sci.*, 50(2):191–202, 1995.

[10] M. Braverman and A. Rao. Towards coding for maximum errors in interactive communication. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (STOC), pages 159–166, 2011.

[11] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing* (STOC), pages 63–68, 1998.

[12] H. Buhrman, R. Cleve, R. de Wolf, and C. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of the Fortieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 358–368, 1999.

[13] H. Buhrman, I. Newman, H. Röhrig, and R. de Wolf. Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007. Preliminary version at quant-ph/0309220, September 2003.

[14] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), pages 24–32, 2007.

[15] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the Sixteenth Annual IEEE Conference on Computational Complexity* (CCC), pages 120–130, 2001.

[16] E. W. Cheney. *Introduction to Approximation Theory*. Chelsea Publishing, New York, 2nd edition, 1982.

[17] C. Dutta, Y. Kanoria, D. Manjunath, and J. Radhakrishnan. A tight lower bound for parity in noisy communication networks. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1056–1065, 2008.

[18] C. Dutta and J. Radhakrishnan. Lower bounds for noisy wireless networks using sampling algorithms. In *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 394–402, 2008.

[19] A. Eremenko and P. Yuditskii. Uniform approximation of $\operatorname{sgn}(x)$ by polynomials and entire functions. *J. d'Analyse Mathématique*, 101:313–324, 2007.

[20] W. S. Evans and N. Pippenger. Average-case lower bounds for noisy Boolean decision trees. *SIAM J. Comput.*, 28(2):433–446, 1998.

[21] W. S. Evans and L. J. Schulman. Signal propagation and noisy circuits. *IEEE Transactions on Information Theory*, 45(7):2367–2373, 1999.

[22] E. Farhi, J. Goldstone, and S. Gutmann. A quantum algorithm for the Hamiltonian NAND tree. *Theory of Computing*, 4(1):169–190, 2008.

[23] U. Feige. On the complexity of finite random functions. *Inf. Process. Lett.*, 44(6):295–296, 1992.

[24] U. Feige and J. Kilian. Finding OR in a noisy broadcast network. *Inf. Process. Lett.*, 73(1-2):69–75, 2000.

[25] U. Feige, P. Raghavan, D. Peleg, and E. Upfal. Computing with noisy information. *SIAM J. Comput.*, 23(5):1001–1018, 1994.

[26] P. Gács and A. Gál. Lower bounds for the complexity of reliable Boolean circuits with noisy gates. *IEEE Transactions on Information Theory*, 40(2):579–583, 1994.

[27] R. G. Gallager. Finding parity in a simple broadcast network. *IEEE Transactions on Information Theory*, 34(2):176–180, 1988.

[28] R. Gelles, A. Moitra, and A. Sahai. Efficient and explicit coding for interactive communication. In *Proceedings of the Fifty-Second Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2011. To appear.

[29] N. Goyal, G. Kindler, and M. E. Saks. Lower bounds for the noisy broadcast problem. *SIAM J. Comput.*, 37(6):1806–1841, 2008.

[30] P. Høyer, M. Mosca, and R. de Wolf. Quantum search on bounded-error inputs. In *Proc. of the 30th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 291–299, 2003.

[31] J. Kahn, N. Linial, and A. Samorodnitsky. Inclusion-exclusion: Exact and approximate. *Combinatorica*, 16(4):465–477, 1996.

[32] A. T. Kalai, A. R. Klivans, Y. Mansour, and R. A. Servedio. Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008.

[33] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007.

[34] D. J. Kleitman, F. T. Leighton, and Y. Ma. On the design of reliable Boolean circuits that contain partially unreliable gates. *J. Comput. Syst. Sci.*, 55(3):385–401, 1997.

[35] A. R. Klivans and R. A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004.

[36] A. R. Klivans and A. A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2–3):97–114, 2007. Preliminary version in *Proceedings of the Nineteenth Annual Conference on Computational Learning Theory* (COLT), 2006.

[37] A. R. Klivans and A. A. Sherstov. Lower bounds for agnostic learning via approximate rank. *Computational Complexity*, 19(4):581–604, 2010. Preliminary version in *Proceedings of the Twentieth Annual Conference on Computational Learning Theory* (COLT), 2007.

[38] M. Krause and P. Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theor. Comput. Sci.*, 174(1–2):137–156, 1997.

[39] M. Krause and P. Pudlák. Computing Boolean functions by polynomials and threshold circuits. *Comput. Complex.*, 7(4):346–370, 1998.

[40] E. Kushilevitz and Y. Mansour. Computation in noisy radio networks. *SIAM J. Discrete Math.*, 19(1):96–108, 2005.

[41] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

[42] M. L. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass., 1969.

[43] I. Newman. Computing in fault tolerance broadcast networks. In *Proceedings of the Nineteenth Annual IEEE Conference on Computational Complexity* (CCC), pages 113–122, 2004.

[44] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[45] R. Paturi and M. E. Saks. Approximating threshold circuits by rational functions. *Inf. Comput.*, 112(2):257–272, 1994.

[46] N. Pippenger. On networks of noisy gates. In *Proceedings of the Twenty-Sixth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 30–38, 1985.

[47] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, Mathematics*, 67:145–159, 2002.

[48] A. A. Razborov and A. A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. Preliminary version in *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2008.

[49] R. Reischuk and B. Schmeltz. Reliable computation with noisy circuits and decision trees—a general $n \log n$ lower bound. In *Proceedings of the Thirty-Second Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 602–611, 1991.

[50] T. J. Rivlin. *An Introduction to the Approximation of Functions*. Dover Publications, New York, 1981.

[51] L. J. Schulman. Communication on noisy channels: A coding theorem for computation. In *Proceedings of the Thirty-Third Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 724–733, 1992.

[52] L. J. Schulman. Coding for interactive communication. *IEEE Transactions on Information Theory*, 42(6):1745–1756, 1996.

[53] A. A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[54] A. A. Sherstov. Approximate inclusion-exclusion for arbitrary symmetric functions. *Computational Complexity*, 18(2):219–247, 2009. Preliminary version in *Proceedings of the Twenty-Third Annual IEEE Conference on Computational Complexity* (CCC), 2008.

[55] A. A. Sherstov. The intersection of two halfspaces has high threshold degree. In *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 343–362, 2009.

[56] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (STOC), 2007.

[57] A. A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. In *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing* (STOC), pages 523–532, 2010.

[58] K.-Y. Siu, V. P. Roychowdhury, and T. Kailath. Rational approximation techniques for analysis of neural networks. *IEEE Transactions on Information Theory*, 40(2):455–466, 1994.

[59] D. A. Spielman. Highly fault-tolerant parallel computation. In *Proceedings of the Thirty-Seventh Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 154–163, 1996.

[60] M. Szegedy and X. Chen. Computing Boolean functions from multiple faulty copies of input bits. *Theor. Comput. Sci.*, 321(1):149–170, 2004.

[61] J. Tarui and T. Tsukiji. Learning DNF by approximating inclusion-exclusion formulae. In *Proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity* (CCC), pages 215–221, 1999.

[62] R. de Wolf. A note on quantum algorithms and the minimal degree of ϵ-error polynomials for symmetric functions. *Quantum Information and Computation*, 8(10):943–950, 2008.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de