THEORY OF COMPUTING, Volume 9 (18), 2009, pp. 593–615
www.theoryofcomputing.org

SPECIAL ISSUE: ANALYSIS OF BOOLEAN FUNCTIONS

Making Polynomials Robust to Noise*

Alexander A. Sherstov†

Received July 1, 2012; Revised May 14, 2013; Published June 11, 2013

Abstract: A basic question in any model of computation is how to reliably compute a given function when its inputs are subject to noise. Buhrman, Newman, Röhrig, and de Wolf (2003) posed the noisy computation problem for *real polynomials*. We give a complete solution to this problem. For any $\delta > 0$ and any polynomial $p\colon \{0,1\}^n \to [-1,1]$, we construct a corresponding polynomial $p_{\mathrm{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\deg p + \log 1/\delta)$ that is robust to noise in the inputs: $|p(x) - p_{\mathrm{robust}}(x+\varepsilon)| < \delta$ for all $x \in \{0,1\}^n$ and all $\varepsilon \in [-1/3, 1/3]^n$. 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.

ACM Classification: F.0, F.1.3

AMS Classification: 68Q05, 68Q87

Key words and phrases: computation with noise, polynomial approximation, robust polynomials, real polynomials on the Boolean hypercube

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

*An extended abstract of this article appeared in the Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing (STOC'12), pages 747–758, 2012.

†Supported by NSF CAREER award CCF-1149018.

 DOI: 10.4086/toc.2013.v009a018

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 [21, 43, 19, 17, 38], circuits [40, 22, 18, 30, 52, 53], broadcast networks [23, 36, 20, 38, 25, 14, 15], and communication protocols [45, 46, 7, 24]. 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, 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 [45, 46, 7, 24], 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 [42]. 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? This study frequently reveals aspects of the model that would otherwise be overlooked.

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 $|f(x)-p(x)| \le 1/3$ 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 approximate degree of Boolean functions has been studied for over forty years and has enabled spectacular progress in circuit complexity [39, 51, 6, 3, 34, 35, 49, 5], quantum query complexity [4, 9, 2, 1, 29], communication complexity [12, 41, 11, 42, 47], and computational learning theory [54, 31, 28, 32, 33, 48, 50].

The contribution of this paper is to answer a question posed ten years ago by Buhrman et al. [10]. These authors asked whether real polynomials, as a computational model, are robust to noise. Robustness to noise becomes an issue when one wants to do anything nontrivial with approximating polynomials, e.g., compose them. To use a motivating example from [10], 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 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. [10]. Formally, a robust approximating polynomial for f is a real polynomial $p_{\text{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ such that

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

for every $x \in \{-1,+1\}^n$ and every $\varepsilon \in [-1/3, 1/3]^n$. Thus, a robust polynomial approximates f not only on Boolean inputs but also on the 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 $p_{\text{robust}}(q,q,\ldots,q)$ is a valid approximating polynomial for $f(g,g,\ldots,g)$.

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 $\Theta(\widetilde{\deg}(f))$. Analogous to 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. [10] 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 $\Theta(n)$. The authors of [10] asked whether *every* Boolean function f 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. [10]. To be precise, we study a more general question. Buhrman et al. [10] 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 1.1. *Let $p\colon \{-1,+1\}^n \to [-1,1]$ be a given polynomial. Then for every $\delta > 0$, there is a polynomial $p_{\text{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\deg p + \log 1/\delta)$ such that*

$$|p(x) - p_{\text{robust}}(x+\varepsilon)| < \delta$$

for all $x \in \{-1,+1\}^n$ and $\varepsilon \in [-1/3,1/3]^n$. Furthermore, p_{robust} has an explicit, closed-form description.

Theorem 1.1 shows that real polynomials are robust to noise. In this regard, they behave differently from other computational models such as decision trees [21] and broadcast networks [25], where handling noise provably increases the computational complexity by a superconstant factor. In fact, Theorem 1.1 reveals a *very* high degree of robustness: the degree of a δ-error robust polynomial grows additively rather than multiplicatively with the error parameter δ, and the actual dependence on δ is logarithmic. This dependence on δ is best possible; see Remark 6.3 for details. Theorem 1.1 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 1.1 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 $p(x) = x_1 x_2 \cdots x_n$. To quote Buhrman et al. [10], "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 1.1, we conclude that the approximate degree behaves nicely under function composition:

Corollary. For all Boolean functions f and g,

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

Prior to this paper, this conclusion was known to hold only for several special functions, e.g., [8, 27, 10], 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. Using a remarkable quantum query argument, Buhrman et al. [10] gave a robust approximating polynomial of degree $O(\widetilde{\deg}(f))$ for every symmetric Boolean function f as well as every Boolean function f with approximate degree $\Theta(n)$. Unfortunately, the quantum argument does not seem to generalize beyond these two important cases. With an unrelated, combinatorial argument, the authors of [10] 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 1.1. 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, $p(x) = x_1 x_2 \cdots x_n$;

(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 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 $\|p_i\|_\infty \le 2^{O(d)}$ for all i. For our purposes, this means that a robust polynomial for p can be obtained by summing 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 δ, 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 $\delta = 2^{-\Theta(d)}$. Unfortunately, there is no reason to believe that with this δ, 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

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

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 a 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: $\sum_{i=1}^{\infty} |a_i| \le 2^{-\Theta(d)}$. 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 $\mathbb{N} = \{0,1,2,3,\ldots\}$. The symbol $\log x$ denotes the logarithm of x to base 2. For a string $x \in \mathbb{R}^n$ and a set $S \subseteq \{1,2,\ldots,n\}$, we adopt the shorthand $x|_S = (x_{i_1}, x_{i_2}, \ldots, x_{i_{|S|}}) \in \mathbb{R}^{|S|}$, where $i_1 < i_2 < \cdots < i_{|S|}$ are the elements of S. The family of all subsets of a given set X is denoted $\mathcal{P}(X)$. 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. For a natural number i and a real α, the generalized binomial coefficient $\binom{\alpha}{i}$ is given by

$$\binom{\alpha}{i} = \frac{\alpha(\alpha-1)\cdots(\alpha-i+1)}{i!}.$$

The following identity is well-known and straightforward to show [26, p. 186]:

$$\binom{-1/2}{i} = \left(-\frac{1}{4}\right)^i \binom{2i}{i}. \tag{2.1}$$

Fourier transform

Consider the 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 $\mathcal{P}(\{1,2,\dots,n\}) \to \mathbb{R}$. 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 $x_1,\ldots,x_n$ are distributed independently and $\mathbf{E}[x_i] = \xi_i$ for all i. Then $\phi(\xi) = \mathbf{E}[\phi(x)]$ by multilinearity, so that $|\phi(\xi)| \le \max_{x\in\{-1,+1\}^n} |\phi(x)|$.

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 $p(x_1,x_2,\ldots,x_n) \approx \prod \operatorname{sgn} x_i$ whenever the input variables are close to Boolean: $x_1,x_2,\ldots,x_n \in [-4/3,-2/3]\cup[2/3,4/3]$. 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. [10]:

$$B_d(t) = 2^{-d} \sum_{i=\lceil d/2\rceil}^{d} \binom{d}{i} t^i (1-t)^{d-i}\,.$$

In words, $B_d(t)$ is the probability of observing more heads than tails in a sequence of d independent coin flips, each coming up heads with probability t. By the Chernoff bound, B_d sends $[0,1/4] \to [0,2^{-\Omega(d)}]$ and similarly $[3/4,1] \to [1-2^{-\Omega(d)},1]$. As Buhrman et al. [10] point out, this immediately gives a degree-d approximant for the sign function with exponentially small error on $[-4/3,-2/3]\cup[2/3,4/3]$. Unfortunately, the coefficients of B_d do not exhibit the 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 $x_1,x_2,\ldots,x_n \in (-\sqrt{2},0)\cup(0,\sqrt{2})$,*

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

Proof. Recall the binomial series $(1+t)^\alpha = \sum_{i=0}^\infty \binom{\alpha}{i} t^i$, valid for all $-1 < t < 1$ and all real α. In particular, setting $\alpha = -1/2$ gives

$$\begin{aligned} \frac{1}{\sqrt{1+t}} &= \sum_{i=0}^\infty \binom{-1/2}{i} t^i \\ &= \sum_{i=0}^\infty \left(-\frac{1}{4}\right)^i \binom{2i}{i} t^i, \qquad\qquad -1 < t < 1, \end{aligned} \tag{3.2}$$

where the second step uses (2.1). 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

$$\operatorname{sgn} t = \frac{t}{\sqrt{1+(t^2-1)}} = t \sum_{i=0}^\infty \left(-\frac{1}{4}\right)^i \binom{2i}{i} (t^2-1)^i, \tag{3.3}$$

where the second step holds by (3.2). 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 $\varepsilon \in [0,1)$ and let*

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

Then for every integer $N \geq 1$, there is an (explicitly given) polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$ of degree at most $2N + n$ such that

$$\max_{X^n} |\operatorname{sgn}(x_1 x_2 \cdots x_n) - p(x)| \leq \varepsilon^N (1+\varepsilon)^{n/2} \binom{N+n}{N} n. \tag{3.4}$$

Setting $\varepsilon = 7/9$ and $N = 22n$ 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^{-n} everywhere by a polynomial of degree $O(n)$. This is the desired robust polynomial for parity.

Proof of Theorem 3.2. If $\varepsilon > N/(N+n)$, the right-hand side of (3.4) exceeds

$$\left(\frac{N}{N+n}\right)^N \binom{N+n}{N} \geq 1$$

and therefore the theorem holds trivially with $p = 0$. In what follows, we treat the complementary case:

$$0 \leq \varepsilon \leq \frac{N}{N+n}. \tag{3.5}$$

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

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

One can restate (3.1) in the form

$$\operatorname{sgn}(x_1x_2\cdots x_n) = x_1x_2\cdots x_n \sum_{d=0}^{\infty} \xi_d(x_1,x_2,\dots,x_n), \tag{3.6}$$

where

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

On X^n,

$$\|\xi_d\|_\infty \le \varepsilon^d |\mathcal{I}_d| = \varepsilon^d \binom{d+n-1}{d}.$$

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

$$\begin{aligned}
(1+\varepsilon)^{n/2} \sum_{d=N+1}^{\infty} \varepsilon^d \binom{d+n-1}{d} &\le (1+\varepsilon)^{n/2}\,\varepsilon^{N+1} \binom{N+n}{N+1} \sum_{d=0}^{\infty} \left(\varepsilon \cdot \frac{N+n}{N+1}\right)^d \\
&\le (1+\varepsilon)^{n/2}\,\varepsilon^{N+1} \binom{N+n}{N+1} \sum_{d=0}^{\infty} \left(\frac{N}{N+1}\right)^d \qquad \text{by (3.5)} \\
&= (1+\varepsilon)^{n/2}\,\varepsilon^{N+1} \binom{N+n}{N} n.
\end{aligned}$$

□

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 a bound on the coefficients of a *univariate* polynomial in terms of its degree d and its maximum absolute value on the interval $[-1,1]$. This fundamental problem was solved in the nineteenth century by V. A. Markov [37, p. 81], who proved an upper bound of

$$O\left(\frac{(1+\sqrt{2})^d}{\sqrt{d}}\right) \tag{4.1}$$

on the size of the coefficients of any degree-d polynomial that is bounded on $[-1,1]$ in absolute value by 1. Markov further showed that (4.1) is tight. Rather than appeal to this deep result in approximation theory, we give a first-principles proof of a 4^d bound which suffices for our purposes.

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

$$\sum_{i=0}^{d}|a_i| \leq 4^d \max_{j=0,1,\dots,d}\left|p\left(\frac{d-2j}{d}\right)\right|. \tag{4.2}$$

Proof. We will use a common approximation-theoretic technique [13, 44] whereby one expresses p as a linear combination of more structured polynomials and analyzes the latter objects. For this, define $q_0,q_1,\dots,q_d \in P_d$ by

$$q_j(t)=\frac{(-1)^j d^d}{d!\,2^d}\binom{d}{j}\prod_{\substack{i=0\\ i\neq j}}^{d}\left(t-\frac{d-2i}{d}\right), \qquad j=0,1,\dots,d. \tag{4.3}$$

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

$$q_j\left(\frac{d-2i}{d}\right)=\begin{cases}1 & \text{if } i=j,\\ 0 & \text{otherwise.}\end{cases}$$

Lagrange interpolation gives

$$p=\sum_{j=0}^{d} p\left(\frac{d-2j}{d}\right) q_j. \tag{4.4}$$

By (4.3), the absolute values of the coefficients of q_j sum to at most

$$\begin{aligned}\frac{d^d}{d!\,2^d}\binom{d}{j}\prod_{\substack{i=0\\ i\neq j}}^{d}\left(1+\frac{|d-2i|}{d}\right) &= \frac{1}{d!\,2^d}\binom{d}{j}\cdot\frac{1}{d+|d-2j|}\prod_{i=0}^{d}(d+|d-2i|)\\ &= \frac{1}{d!\,2^d}\binom{d}{j}\cdot\frac{1}{d+|d-2j|}\prod_{i=0}^{\lfloor d/2\rfloor}(2d-2i)\prod_{i=\lfloor d/2\rfloor+1}^{d}(2i)\\ &= \frac{1}{d!\,2^d}\binom{d}{j}\cdot\frac{2^{d+1}}{d+|d-2j|}\cdot\frac{d!}{(\lceil d/2\rceil-1)!}\cdot\frac{d!}{\lfloor d/2\rfloor!}\\ &= \binom{d}{j}\cdot\frac{2d}{d+|d-2j|}\binom{d-1}{\lfloor d/2\rfloor}\\ &\leq 2^d\binom{d}{j}.\end{aligned}$$

In view of (4.4), we conclude that the absolute values of the coefficients of p sum to at most

$$\left(\max_{j=0,1,\dots,d}\left|p\left(\frac{d-2j}{d}\right)\right|\right)\sum_{j=0}^{d}2^d\binom{d}{j}=4^d\max_{j=0,1,\dots,d}\left|p\left(\frac{d-2j}{d}\right)\right|. \qquad \square$$

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*

$$\|\phi_i\|_\infty \le 4^d \|\phi\|_\infty\,, \qquad i = 0,1,\ldots,d\,.$$

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

$$\|\phi\|_\infty \ge \left| \mathop{\mathbf{E}}_{z\sim\mu_t} [\phi(x_1z_1,\ldots,x_nz_n)] \right| = \left| \sum_{|S|\le d} \hat{\phi}(S) \mathop{\mathbf{E}}_{z\sim\mu_t}\left[\prod_{i\in S} x_i z_i\right] \right| = \left| \sum_{|S|\le d} \hat{\phi}(S) t^{|S|} \prod_{i\in S} x_i \right| = |p(t)|\,.$$

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 $4^d\|\phi\|_\infty$:

$$|\phi_i(x)| \le 4^d \|\phi\|_\infty\,.$$

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

5 Error cancellation in homogeneous polynomials

Recall that the goal of this paper is to construct a degree-$O(d)$ robust polynomial for every 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 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

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

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 the function $A_v f\colon \{-1,+1\}^n \to \mathbb{R}$ where

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

This definition is meaningful because we identify f with its multilinear extension to $[-1,1]^n$, thus making it possible to evaluate f on inputs with fractional coordinates (see Section 2). 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,

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

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 any individual $A_v\phi$ is small, that 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 $|S| = d$,

$$\begin{aligned}(A_v\chi_S)(x) &= \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1z_2\cdots z_d \prod_{j\in S}\left(\frac{1}{d}\sum_{i=1}^{d} z_i x_j^{v_i}\right)\right]\\ &= \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1z_2\cdots z_d \mathop{\mathbf{E}}_{\tau\colon S\to\{1,\ldots,d\}}\left[\prod_{j\in S} z_{\tau(j)} x_j^{v_{\tau(j)}}\right]\right]\\ &= \mathop{\mathbf{E}}_{\tau\colon S\to\{1,\ldots,d\}}\left[\mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1z_2\cdots z_d\prod_{j\in S} z_{\tau(j)}\right]\prod_{j\in S} x_j^{v_{\tau(j)}}\right], \end{aligned} \tag{5.2}$$

where τ chosen uniformly at random from all possible mappings $S \to \{1,2,\ldots,d\}$. The inner expectation in (5.2) 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,

$$\begin{aligned}(A_v\chi_S)(x) &= \mathop{\mathbf{P}}_{\tau}[\tau \text{ is a bijection}]\ \mathop{\mathbf{E}}_{\tau}\left[\prod_{j\in S} x_j^{v_{\tau(j)}}\ \middle|\ \tau \text{ is a bijection}\right]\\ &= \frac{d!}{d^d} \mathop{\mathbf{E}}_{\substack{T\subseteq S,\\ |T|=v_1+\cdots+v_d}} [\chi_T(x)]. \end{aligned} \tag{5.3}$$

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 (5.3). Taking a weighted sum over S and using the linearity of A_v,

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

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:

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

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 $[-\sqrt{1+\varepsilon},-\sqrt{1-\varepsilon}]\cup[\sqrt{1-\varepsilon},\sqrt{1+\varepsilon}]$ for $0<\varepsilon<1$. 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 $|S|\neq d$. Let $0<\varepsilon<1$ and*

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

Then for every integer $D\geq 1$, there is an (explicitly given) polynomial $p\colon\mathbb{R}^d\to\mathbb{R}$ of degree at most $2D+d$ such that

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

obeys

$$\max_{X^n}|\phi(\operatorname{sgn}x_1,\dots,\operatorname{sgn}x_n)-P(x)|\leq(1+\varepsilon)^{d/2}\,\frac{d^d}{d!}\,\varepsilon^D\binom{D+d}{D}d\,\|\phi\|_\infty\,. \tag{5.4}$$

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 $\varepsilon>D/(D+d)$, then the right-hand side of (5.4) is at least

$$\left(\frac{D}{D+d}\right)^D\binom{D+d}{D}\|\phi\|_\infty\geq\|\phi\|_\infty$$

and hence the theorem holds trivially with $p=0$. In what follows, we treat the complementary case $\varepsilon\leq D/(D+d)$. To start with,

$$\begin{aligned}\sum_{\substack{v\in\mathbb{N}^d:\\ v_1+\cdots+v_d\geq D+1}}\varepsilon^{v_1+\cdots+v_d}\underbrace{\prod_{j=1}^d\left(\frac{1}{4}\right)^{v_j}\binom{2v_j}{v_j}}_{\leq 1}&\leq\sum_{\substack{v\in\mathbb{N}^d:\\ v_1+\cdots+v_d\geq D+1}}\varepsilon^{v_1+\cdots+v_d}\\&=\varepsilon^{D+1}\sum_{i=0}^\infty\binom{D+d+i}{D+1+i}\varepsilon^i\\&\leq\varepsilon^{D+1}\binom{D+d}{D+1}\sum_{i=0}^\infty\left(\frac{D+d}{D+1}\right)^i\varepsilon^i\\&\leq\varepsilon^{D+1}\binom{D+d}{D+1}\sum_{i=0}^\infty\left(\frac{D}{D+1}\right)^i\\&=\varepsilon^{D+1}\binom{D+d}{D}d\,,\end{aligned}\tag{5.5}$$

where the fourth line in the derivation uses $\varepsilon \le D/(D+d)$. Define p by

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

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

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

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

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

$$\Delta(x) = \phi(\operatorname{sgn}x_1,\dots,\operatorname{sgn}x_n) - P(x). \tag{5.6}$$

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 $A_vf\colon X^n\to\mathbb{R}$ where

$$(A_vf)(x) = \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1z_2\cdots z_d\, f\Bigg(\dots,\underbrace{\frac{1}{d}\sum_{i=1}^{d}\frac{z_ix_j(x_j^2-1)^{v_i}}{\varepsilon^{v_i}\sqrt{1+\varepsilon}}}_{j\text{th coordinate}},\dots\Bigg)\right].$$

The jth coordinate in this expression is bounded by 1 in absolute value because $\sqrt{1-\varepsilon}<|x_j|<\sqrt{1+\varepsilon}$ on X^n. 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

$$\|A_v\phi\|_\infty \le \max_{x\in[-1,1]^n}|\phi(x)| = \max_{x\in\{-1,+1\}^n}|\phi(x)| = \|\phi\|_\infty\,, \tag{5.7}$$

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 $S=\{1,2,\dots,d\}$:

$$\begin{aligned}(A_v\chi_{\{1,\dots,d\}})(x) &= \mathop{\mathbf{E}}_{z\in\{-1,+1\}^d}\left[z_1z_2\cdots z_d\prod_{j=1}^{d}\left(\frac{1}{d}\sum_{i=1}^{d}\frac{z_ix_j(x_j^2-1)^{v_i}}{\varepsilon^{v_i}\sqrt{1+\varepsilon}}\right)\right]\\ &= \frac{1}{(1+\varepsilon)^{d/2}}\mathop{\mathbf{E}}_{\tau}\left[\mathop{\mathbf{E}}_{z}\left[\prod_{j=1}^{d}z_jz_{\tau(j)}\right]\prod_{j=1}^{d}\frac{x_j(x_j^2-1)^{v_{\tau(j)}}}{\varepsilon^{v_{\tau(j)}}}\right],\end{aligned}$$

where the first expectation is taken over a uniformly random mapping $\tau\colon \{1,2,\ldots,d\} \to \{1,2,\ldots,d\}$. The 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. Thus,

$$\begin{aligned}(A_v\chi_{\{1,\ldots,d\}})(x) &= \frac{1}{(1+\varepsilon)^{d/2}} \mathop{\mathbf{P}}_{\tau}[\tau \text{ is a bijection}] \mathop{\mathbf{E}}_{\tau}\left[\prod_{j=1}^{d} \frac{x_j(x_j^2-1)^{v_{\tau(j)}}}{\varepsilon^{v_{\tau(j)}}} \;\middle|\; \tau \text{ is a bijection}\right] \\ &= \frac{1}{(1+\varepsilon)^{d/2}\varepsilon^{v_1+\cdots+v_d}} \cdot \frac{d!}{d^d} \cdot \mathop{\mathbf{E}}_{\sigma\in S_d}\left[\prod_{j=1}^{d} x_j(x_j^2-1)^{v_{\sigma(j)}}\right]. \end{aligned} \tag{5.8}$$

Now, consider the operator

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

This operator is well-defined because by (5.5), the series in question converges absolutely. By (5.8), 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,

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

where the final step follows by (5.5) and (5.7). In light of (5.6), 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*

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

Then

$$\|\Delta\|_\infty \le (4\mathrm{e})^d\,\|\phi\|_\infty \sum_{i=0}^{d} \|\hat{\delta}_i\|_1\,.$$

The functions $\delta_0,\delta_1,\ldots,\delta_d$ in this result are to be thought of as the errors in the approximation of characters of orders $0,1,\ldots,d$, respectively, and Δ is the cumulative error so incurred. 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

$$\|\Delta\|_\infty \le \sum_{i=0}^{d} \|\Delta_i\|_\infty\,, \tag{6.1}$$

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,

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

By Theorem 4.2,

$$\|\phi_i\|_\infty \le 4^d\|\phi\|_\infty\,, \qquad i = 0,1,\ldots,d\,. \tag{6.3}$$

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

We will now apply a similar argument in the setting of real variables.

Theorem 6.2 (Main Theorem)**.** *Let* $0 < \varepsilon < 1$ *and* $\phi\colon \{-1,+1\}^n \to [-1,1]$ *be given,* $\deg\phi = d$. *Then for each* $\delta > 0$, *there is a polynomial P of degree*

$$O\left(\frac{1}{1-\varepsilon}d + \frac{1}{1-\varepsilon}\log\frac{1}{\delta}\right)$$

such that

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

where $X = [-1-\varepsilon,-1+\varepsilon]\cup[1-\varepsilon,1+\varepsilon]$. *Furthermore, P is given explicitly in terms of the Fourier spectrum of* ϕ.

Letting $\varepsilon = 1/3$ in this result settles Theorem 1.1 in the introduction.

Proof. We first consider $0 \le \varepsilon \le 1/10$, in which case

$$X \subset \left[-\sqrt{1+\frac{1}{4}}, -\sqrt{1-\frac{1}{4}}\right] \cup \left[\sqrt{1-\frac{1}{4}}, \sqrt{1+\frac{1}{4}}\right]. \tag{6.5}$$

Let $D = D(d,\delta) \ge 1$ be a parameter to be chosen later. For $i = 0,1,\dots,d$, consider $\phi_i = \sum_{|S|=i} \hat{\phi}(S)\chi_S$, the degree-$i$ homogeneous part of ϕ. By Theorem 4.2,

$$\|\phi_i\|_\infty \le 4^d, \qquad i = 0,1,\dots,d. \tag{6.6}$$

In light of (6.5), Theorem 5.2 gives explicit polynomials $p_i \colon \mathbb{R}^i \to \mathbb{R}$, $i = 0,1,2,\dots,d$, each of degree at most $2D+d$, such that

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

for some absolute constant $K > 1$. Letting

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

we infer that

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

where the last step uses (6.6). Therefore, (6.4) holds with $D = O(d + \log 1/\delta)$.

To handle the case $1/10 < \varepsilon < 1$, basic approximation theory [44] gives an explicit univariate polynomial r of degree $O(1/(1-\varepsilon))$ that sends $[-1-\varepsilon, -1+\varepsilon] \to [-11/10, -9/10]$ and $[1-\varepsilon, 1+\varepsilon] \to [9/10, 11/10]$. In particular, we have $|\phi(\operatorname{sgn} x_1,\dots,\operatorname{sgn} x_n) - P(r(x_1),\dots,r(x_n))| < \delta$ everywhere on X^n, where P is the approximant constructed in the previous paragraph. □

Remark 6.3. The degree of the robust polynomial in Theorem 1.1 grows additively with $O(\log 1/\delta)$, where δ is the error parameter. As stated in the introduction, this dependence on δ is best possible. 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, $p(1,1,\dots,1) = 1$ and $p(-1,-1,\dots,-1) = -1$. As a result, the univariate polynomial $p_{\text{robust}}(t,t,\dots,t)$ would need to approximate the sign function on $[-4/3,-2/3] \cup [2/3,4/3]$ to within δ, which forces $\deg(p_{\text{robust}}) \ge \Omega(\log 1/\delta)$ by basic approximation theory [16].

Acknowledgments

The author would like to thank Mark Braverman, Harry Buhrman, Vitaly Feldman, Dmitry Gavinsky, Ankur Moitra, Ryan O'Donnell, Ronald de Wolf, and the anonymous reviewers for their useful feedback. In particular, the author is thankful to Ryan for the hard-to-find reference [37] on the growth rate of coefficients of bounded polynomials.

References

[1] SCOTT AARONSON: Limitations of quantum advice and one-way communication. *Theory of Computing*, 1(1):1–28, 2005. Preliminary version in CCC'04. [doi:10.4086/toc.2005.v001a001] 594

[2] SCOTT AARONSON AND YAOYUN SHI: Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004. Preliminary version in STOC'02. [doi:10.1145/1008731.1008735] 594

[3] JAMES ASPNES, RICHARD BEIGEL, MERRICK L. FURST, AND STEVEN RUDICH: The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994. Preliminary version in STOC'91. [doi:10.1007/BF01215346] 594

[4] ROBERT BEALS, HARRY BUHRMAN, RICHARD CLEVE, MICHELE MOSCA, AND RONALD DE WOLF: Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001. Preliminary version in FOCS'98. [doi:10.1145/502090.502097] 594

[5] PAUL BEAME AND DANG-TRINH HUYNH-NGOC: Multiparty communication complexity and threshold circuit size of AC^0. *SIAM Journal on Computing*, 41(3):484–518, 2012. Preliminary version in FOCS'09. [doi:10.1137/100792779] 594

[6] RICHARD BEIGEL, NICK REINGOLD, AND DANIEL A. SPIELMAN: PP is closed under intersection. *J. Comput. System Sci.*, 50(2):191–202, 1995. Preliminary version in STOC'91. [doi:10.1006/jcss.1995.1017] 594

[7] MARK BRAVERMAN AND ANUP RAO: Towards coding for maximum errors in interactive communication. In *Proc. 43rd STOC*, pp. 159–166. ACM Press, 2011. [doi:10.1145/1993636.1993659] 594

[8] HARRY BUHRMAN, RICHARD CLEVE, AND AVI WIGDERSON: Quantum vs. classical communication and computation. In *Proc. 30th STOC*, pp. 63–68. ACM Press, 1998. [doi:10.1145/276698.276713] 596

[9] HARRY BUHRMAN, RICHARD CLEVE, RONALD DE WOLF, AND CHRISTOF ZALKA: Bounds for small-error and zero-error quantum algorithms. In *Proc. 40th FOCS*, pp. 358–368. IEEE Comp. Soc. Press, 1999. [doi:10.1109/SFFCS.1999.814607] 594

[10] HARRY BUHRMAN, ILAN NEWMAN, HEIN RÖHRIG, AND RONALD DE WOLF: Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007. Preliminary version in STACS'05. [doi:10.1007/s00224-006-1313-z] 594, 595, 596, 599

[11] HARRY BUHRMAN, NIKOLAI K. VERESHCHAGIN, AND RONALD DE WOLF: On computation and communication with small bias. In *Proc. 22nd IEEE Conf. on Computational Complexity (CCC'07)*, pp. 24–32. IEEE Comp. Soc. Press, 2007. (See also SAGA'07.). [doi:10.1109/CCC.2007.18] 594

[12] HARRY BUHRMAN AND RONALD DE WOLF: Communication complexity lower bounds by polynomials. In *Proc. 16th IEEE Conf. on Computational Complexity (CCC'01)*, pp. 120–130. IEEE Comp. Soc. Press, 2001. [doi:10.1109/CCC.2001.933879] 594

[13] ELLIOTT W. CHENEY: *Introduction to Approximation Theory*. Chelsea Publishing, New York, 2nd edition, 1982. 602

[14] CHINMOY DUTTA, YASHODHAN KANORIA, D. MANJUNATH, AND JAIKUMAR RADHAKRISHNAN: A tight lower bound for parity in noisy communication networks. In *Proc. 19th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'08)*, pp. 1056–1065. ACM Press, 2008. [ACM:1347198] 594

[15] CHINMOY DUTTA AND JAIKUMAR RADHAKRISHNAN: Lower bounds for noisy wireless networks using sampling algorithms. In *Proc. 49th FOCS*, pp. 394–402. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.72] 594

[16] ALEXANDRE EREMENKO AND PETER YUDITSKII: Uniform approximation of sgn x by polynomials and entire functions. *J. d'Analyse Mathématique*, 101(1):313–324, 2007. [doi:10.1007/s11854-007-0011-3] 610

[17] WILLIAM S. EVANS AND NICHOLAS PIPPENGER: Average-case lower bounds for noisy Boolean decision trees. *SIAM J. Comput.*, 28(2):433–446, 1998. Preliminary version in STOC'96. [doi:10.1137/S0097539796310102] 594

[18] WILLIAM S. EVANS AND LEONARD J. SCHULMAN: Signal propagation and noisy circuits. *IEEE Trans. Inform. Theory*, 45(7):2367–2373, 1999. Preliminary version in FOCS'93. [doi:10.1109/18.796377] 594

[19] URIEL FEIGE: On the complexity of finite random functions. *Inform. Process. Lett.*, 44(6):295–296, 1992. [doi:10.1016/0020-0190(92)90102-2] 594

[20] URIEL FEIGE AND JOE KILIAN: Finding OR in a noisy broadcast network. *Inform. Process. Lett.*, 73(1-2):69–75, 2000. [doi:10.1016/S0020-0190(00)00002-8] 594

[21] URIEL FEIGE, PRABHAKAR RAGHAVAN, DAVID PELEG, AND ELI UPFAL: Computing with noisy information. *SIAM J. Comput.*, 23(5):1001–1018, 1994. Preliminary version in STOC'90. [doi:10.1137/S0097539791195877] 594, 595

[22] PÉTER GÁCS AND ANNA GÁL: Lower bounds for the complexity of reliable Boolean circuits with noisy gates. *IEEE Trans. Inform. Theory*, 40(2):579–583, 1994. Preliminary version in FOCS'91. [doi:10.1109/18.312190] 594

[23] ROBERT G. GALLAGER: Finding parity in a simple broadcast network. *IEEE Trans. Inform. Theory*, 34(2):176–180, 1988. [doi:10.1109/18.2626] 594

[24] RAN GELLES, ANKUR MOITRA, AND AMIT SAHAI: Efficient and explicit coding for interactive communication. In *Proc. 52nd FOCS*, pp. 768–777. IEEE Comp. Soc. Press, 2011. [doi:10.1109/FOCS.2011.51] 594

[25] NAVIN GOYAL, GUY KINDLER, AND MICHAEL E. SAKS: Lower bounds for the noisy broadcast problem. *SIAM J. Comput.*, 37(6):1806–1841, 2008. Preliminary version in FOCS'05. [doi:10.1137/060654864] 594, 595

[26] RONALD L. GRAHAM, DONALD E. KNUTH, AND OREN PATASHNIK: *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, Boston, Mass., USA, 2nd edition, 1994. 598

[27] PETER HØYER, MICHELE MOSCA, AND RONALD DE WOLF: Quantum search on bounded-error inputs. In *Proc. 30th Internat. Colloq. on Automata, Languages and Programming (ICALP'03)*, pp. 291–299. Springer, 2003. [doi:10.1007/3-540-45061-0_25] 596

[28] ADAM TAUMAN KALAI, ADAM R. KLIVANS, YISHAY MANSOUR, AND ROCCO A. SERVEDIO: Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008. Preliminary version in FOCS'05. [doi:10.1137/060649057] 594

[29] HARTMUT KLAUCK, ROBERT ŠPALEK, AND RONALD DE WOLF: Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007. Preliminary version in FOCS'04. [doi:10.1137/05063235X] 594

[30] DANIEL J. KLEITMAN, FRANK THOMSON LEIGHTON, AND YUAN MA: On the design of reliable Boolean circuits that contain partially unreliable gates. *J. Comput. System Sci.*, 55(3):385–401, 1997. Preliminary version in FOCS'94. [doi:10.1006/jcss.1997.1531] 594

[31] ADAM R. KLIVANS AND ROCCO A. SERVEDIO: Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. System Sci.*, 68(2):303–318, 2004. Preliminary version in STOC'01. [doi:10.1016/j.jcss.2003.07.007] 594

[32] ADAM R. KLIVANS AND ALEXANDER A. SHERSTOV: Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2-3):97–114, 2007. Preliminary version in COLT'06. [doi:10.1007/s10994-007-5010-1] 594

[33] ADAM R. KLIVANS AND ALEXANDER A. SHERSTOV: Lower bounds for agnostic learning via approximate rank. *Comput. Complexity*, 19(4):581–604, 2010. Preliminary version in COLT'07. [doi:10.1007/s00037-010-0296-y] 594

[34] MATTHIAS KRAUSE AND PAVEL PUDLÁK: On the computational power of depth-2 circuits with threshold and modulo gates. *Theoret. Comput. Sci.*, 174(1-2):137–156, 1997. Preliminary version in STOC'94. [doi:10.1016/S0304-3975(96)00019-9] 594

[35] MATTHIAS KRAUSE AND PAVEL PUDLÁK: Computing Boolean functions by polynomials and threshold circuits. *Comput. Complexity*, 7(4):346–370, 1998. Preliminary version in FOCS'95. [doi:10.1007/s000370050015] 594

[36] EYAL KUSHILEVITZ AND YISHAY MANSOUR: Computation in noisy radio networks. *SIAM J. Discrete Math.*, 19(1):96–108, 2005. Preliminary version in SODA'98. [doi:10.1137/S0895480103434063] 594

[37] VLADIMIR A. MARKOV: On functions of least deviation from zero in a given interval. Russian Academy of Sciences, St. Petersburg, 1892. In Russian. 601, 611

[38] ILAN NEWMAN: Computing in fault tolerant broadcast networks and noisy decision trees. *Random Structures & Algorithms*, 34(4):478–501, 2009. Preliminary version in CCC'04. [doi:10.1002/rsa.20240] 594

[39] RAMAMOHAN PATURI AND MICHAEL E. SAKS: Approximating threshold circuits by rational functions. *Inform. and Comput.*, 112(2):257–272, 1994. Preliminary version in FOCS'90. [doi:10.1006/inco.1994.1059] 594

[40] NICHOLAS PIPPENGER: On networks of noisy gates. In *Proc. 26th FOCS*, pp. 30–38. IEEE Comp. Soc. Press, 1985. [doi:10.1109/SFCS.1985.41] 594

[41] ALEXANDER A. RAZBOROV: Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003. [doi:10.1070/IM2003v067n01ABEH000422] 594

[42] ALEXANDER A. RAZBOROV AND ALEXANDER A. SHERSTOV: The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. Preliminary version in FOCS'08. [doi:10.1137/080744037] 594

[43] RÜDIGER REISCHUK AND BERND SCHMELTZ: Reliable computation with noisy circuits and decision trees–a general $n\log n$ lower bound. In *Proc. 32nd FOCS*, pp. 602–611. IEEE Comp. Soc. Press, 1991. [doi:10.1109/SFCS.1991.185425] 594

[44] THEODORE J. RIVLIN: *An Introduction to the Approximation of Functions*. Dover Publications, New York, 1981. 602, 610

[45] LEONARD J. SCHULMAN: Communication on noisy channels: A coding theorem for computation. In *Proc. 33rd FOCS*, pp. 724–733. IEEE Comp. Soc. Press, 1992. [doi:10.1109/SFCS.1992.267778] 594

[46] LEONARD J. SCHULMAN: Coding for interactive communication. *IEEE Trans. Inform. Theory*, 42(6):1745–1756, 1996. Preliminary versions in FOCS'92 and STOC'93. [doi:10.1109/18.556671] 594

[47] ALEXANDER A. SHERSTOV: Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008. EATCS. 594

[48] ALEXANDER A. SHERSTOV: The intersection of two halfspaces has high threshold degree. In *Proc. 50th FOCS*, pp. 343–362. IEEE Comp. Soc. Press, 2009. [doi:10.1109/FOCS.2009.18] 594

[49] ALEXANDER A. SHERSTOV: Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in STOC'07. [doi:10.1137/08071421X] 594

[50] ALEXANDER A. SHERSTOV: Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. In *Proc. 42nd STOC*, pp. 523–532. ACM Press, 2010. [doi:10.1145/1806689.1806762] 594

[51] KAI-YEUNG SIU, VWANI P. ROYCHOWDHURY, AND THOMAS KAILATH: Rational approximation techniques for analysis of neural networks. *IEEE Trans. Inform. Theory*, 40(2):455–466, 1994. Preliminary version in NIPS'92. [doi:10.1109/18.312168] 594

[52] DANIEL A. SPIELMAN: Highly fault-tolerant parallel computation. In *Proc. 37th FOCS*, pp. 154–163. IEEE Comp. Soc. Press, 1996. [doi:10.1109/SFCS.1996.548474] 594

[53] MARIO SZEGEDY AND XIAOMIN CHEN: Computing Boolean functions from multiple faulty copies of input bits. *Theoret. Comput. Sci.*, 321(1):149–170, 2004. Preliminary version in LATIN'02. [doi:10.1016/j.tcs.2003.07.001] 594

[54] JUN TARUI AND TATSUIE TSUKIJI: Learning DNF by approximating inclusion-exclusion formulae. In *Proc. 14th IEEE Conf. on Computational Complexity (CCC'99)*, pp. 215–221. IEEE Comp. Soc. Press, 1999. [doi:10.1109/CCC.1999.766279] 594

AUTHOR

Alexander A. Sherstov
Assistant Professor
University of California, Los Angeles
sherstov@cs.ucla.edu
`http://www.cs.ucla.edu/~sherstov`

ABOUT THE AUTHOR

ALEXANDER A. SHERSTOV completed his Ph. D. in 2009 at the University of Texas at Austin. After a postdoc at Microsoft Research, Alexander joined UCLA as an assistant professor of computer science. His research interests include computational complexity theory, computational learning theory, and quantum computing.