Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 189 (2015)

Shattered Sets and the Hilbert Function

Shay Moran
Dept. of Computer Science, Technion, Israel,
Microsoft Research, Hertzelia,
Max Planck Institute for Informatics, Saarbrücken
shaymoran1@gmail.com

Cyrus Rashtchian
Dept. of Computer Science & Engineering
University of Washington, Seattle
cyrash@cs.washington.edu

December 6, 2016

Abstract

We study complexity measures on subsets of the boolean hypercube and exhibit connections between algebra (the Hilbert function) and combinatorics (VC theory). These connections yield results in both directions. Our main complexity-theoretic result demonstrates that a large and natural family of linear program feasibility problems cannot be computed by polynomial-sized constant-depth circuits. Moreover, our result applies to a stronger regime in which the hyperplanes are fixed and only the directions of the inequalities are given as input to the circuit. We derive this result by proving that a rich class of extremal functions in VC theory cannot be approximated by low-degree polynomials. We also present applications of algebra to combinatorics. We provide a new algebraic proof of the Sandwich Theorem, which is a generalization of the well-known Sauer-Perles-Shelah Lemma. Finally, we prove a structural result about downward-closed sets, related to the Chvátal conjecture in extremal combinatorics.

1 Introduction

Understanding the properties and structure of subsets of the boolean hypercube is a central theme in theoretical computer science and combinatorics. When studying a family of mathematical objects, endowing the objects with algebraic structure often sheds new light on interesting properties. This phenomena appears classically in areas such as algebraic topology and algebraic geometry. It also provides much utility when studying the boolean hypercube. Let $C \subseteq \{0,1\}^n$ be a subset of the boolean hypercube, and let $\mathbb{F}$ be a field. Consider the linear space of functions from C to $\mathbb{F}$, that is, $\mathbb{F}^C$. This is clearly a $|C|$-dimensional vector space over $\mathbb{F}$. Every function in this space can be represented as a multilinear polynomial with degree at most n. Interestingly, for certain sets C, smaller degree actually suffices. For example, when C is the standard basis, denoted $C = \{\vec{e}_1, \ldots, \vec{e}_m\}$, then any function $f : C \to \mathbb{F}$ can be expressed as the linear function $f(\vec{e}_1)x_1 + \ldots + f(\vec{e}_m)x_m$.

The Hilbert function, denoted $h_d(C, \mathbb{F})$, is the dimension of the space of functions $\{f : C \to \mathbb{F}\}$ that have representations as polynomials with degree at most d. This classical algebraic object will be useful in our study of how the structure of C affects the function space. In complexity theory, Smolensky [Smo93] has used the Hilbert function to unify polynomial approximation lower bounds relating to bounded-depth circuits.

We establish new connections between the Hilbert function and VC theory. Our main technical contributions are bounds $h_d(C, \mathbb{F})$ in terms of basic concepts in VC theory, such as shattering, strong shattering, and down-shifts. Previous results on bounding the Hilbert function utilize a more

ISSN 1433-8092

intricate analysis and focus on symmetric sets, that is, unions of slices of the hypercube [Smo93, BE99]. In addition to giving new bounds on the Hilbert function, our connection between Algebra and Combinatorics allows us to derive results in both directions.

Our main complexity theoretical application is that determining feasibility of a large family of linear programs is hard for the class of bounded-depth circuits. More specifically, let $h_1, \ldots, h_m$ be affine functions. Each sign vector s in $\{\pm\}^m$ defines the following feasibility problem: does there exist $x \in \mathbb{R}^d$ such that $h_i(x) > 0$ when $s_i = +$, and $h_i(x) < 0$ when $s_i = -$, for all $i \in [m]$? This defines a boolean function that takes an input s and outputs one if and only if the problem is feasible. We prove that if $m = 2d + 1$, and the affine functions h_i are in general position, then this function cannot be approximated by low-degree polynomials, over any field. This implies a lower bound on the computability of this function by constant-depth circuits, due to the polynomial approximation technique introduced by Razborov [Raz87] and Smolensky [Smo87]. The above linear programming problem relates to the study of hyperplane arrangements (see the books of Matoušek [Mat02] and Stanley [Sta04] for more details and applications). Our results implicitly provide algebraic proofs of some known facts regarding the combinatorics of hyperplane arrangements.

As a combinatorial application of our bounds on the Hilbert function, we provide a short algebraic proof of the Sandwich Theorem. This theorem comes from VC theory and is a well-studied generalization of the Sauer-Shelah-Perles Lemma [Law83, Paj85, BRI89, BR95, Dre97, ARS02, BCDK06, Mor12, KM13, MR14, MW15]. Similar proofs of related upper bounds have appeared previously, due to Frankl and Pach [BF92], Gurvits [Gur97], and Smolensky [Smo97]. We contribute new lower bounds and applications.

Facts we prove about the function space $\mathbb{F}^C$ also lead to a new result about downward-closed sets. A family D of subsets is *downward-closed* if $b \subseteq a$ and $a \in D$ implies $b \in D$. A theorem of Berge [Ber76] implies that for any downward-closed set D there exists a bijection $\pi : D \to D$ such that $a \cap \pi(a) = \emptyset$ for all $a \in D$. We generalize his result to arbitrary, prescribed intersections. Let $\phi : D \to D$ have the property $\phi(a) \subseteq a$ for all $a \in D$. We show that there always exists a bijection $\pi : D \to D$ such that $a \cap \pi(a) = \phi(a)$. Note that choosing $\phi(a) = \emptyset$ for all a implies Berge's result.

Our algebra-combinatorics connection fits within the framework of the polynomial method. This method has been successful in providing elegant proofs of fundamental results in many areas, such as circuit complexity [Smo87, ABFR94, Raz87, Bei93], discrete geometry [GK15, Dvi09, SSZ15, Tao14], extremal combinatorics [Alo99, Juk11, BF92], and more.

The paper is organized as follows. We state our main theorems in Section 2. In Section 3, we prove our bounds on the Hilbert function. In Section 4, we use our Hilbert function bounds to prove that linear program feasibility is hard for bounded-depth circuits. Finally, in Section 5, we prove results about downward-closed sets. We now review preliminaries.

1.1 Preliminaries

We begin with algebraic preliminaries. Let $C \subseteq \{0, 1\}^n$ and $\mathbb{F}$ be a field. Every $f : C \to \mathbb{F}$ can be expressed as a multilinear polynomial over variables $x_1, \ldots, x_n$ with coefficients in $\mathbb{F}$.

Definition 1.1. *For $d \in [n]$ the* Hilbert function $h_d(C, \mathbb{F})$ *is the dimension of the space of functions $f : C \to \mathbb{F}$ that can be represented as polynomials with degree at most d.*

Notice that $h_d(C, \mathbb{F}) \leqslant \min\{\sum_{j=0}^{d} \binom{n}{j}, |C|\}$. A basic fact about the Hilbert function is that

$$1 = h_0(C, \mathbb{F}) \leqslant h_1(C, \mathbb{F}) \leqslant \ldots \leqslant h_n(C, \mathbb{F}) = |C|.$$

The final equality holds because all $f : C \to \mathbb{F}$ have representations with degree at most n.

It is natural to wonder when the Hilbert function attains its maximum and how the structure of C influences the Hilbert function. We introduce the following measure.

Definition 1.2. *The* **interpolation degree of** C *denoted* $\mathsf{intdeg}(C, \mathbb{F})$ *is the minimum* d *such that any* $f : C \to \mathbb{F}$ *can be expressed as a multilinear polynomial with degree at most* d*. In other words,*

$$\mathsf{intdeg}(C, \mathbb{F}) = \min\{d \in [n] : h_d(C, \mathbb{F}) = |C|\}.$$

Intuitively, a smaller interpolation degree implies a less complex set.

We move on to combinatorial preliminaries. Our bounds on the Hilbert function use basic concepts in VC theory. We define these concepts now.

Definition 1.3. *A subset* $I \subseteq [n]$ *is* **shattered** *by* $C \subseteq \{0,1\}^n$ *if for every pattern* $s : I \to \{0,1\}$ *there is* $c \in C$ *that realizes* s*. In other words, the restriction of* c *to* I *equals* s*. A subset* $I \subseteq [n]$ *is* **strongly shattered** *by* C *if* C *contains all elements of some subcube on* I*. In other words, there exists a pattern* $\bar{s} : ([n] \setminus I) \to \{0,1\}$ *such that all extensions of* $\bar{s}$ *to a vector in* $\{0,1\}^n$ *are in* C*.*

These definitions lead to natural families of sets, which will be important to our work.

Definition 1.4. *The* **shattered sets** *with respect to* C *are*

$$\mathsf{str}(C) = \{I \subseteq [n] : I \text{ is shattered by } C\}.$$

The **strongly shattered sets** *with respect to* C *are*

$$\mathsf{sstr}(C) = \{I \subseteq [n] : I \text{ is strongly shattered by } C\}.$$

Definition 1.5. *The* **VC dimension** *of* C *is defined as* $\mathsf{VC}(C) = \max\{|I| : I \in \mathsf{str}(C)\}$.

Note that $\mathsf{sstr}(C) \subseteq \mathsf{str}(C)$ and that both of these families are downward-closed.

We also lower bound the Hilbert function using down-shifts, a standard tool in extremal combinatorics. Let $C \subseteq \{0,1\}^n$ and let $i \in [n]$. We denote as S_i the down-shift operator on the ith coordinate. Obtain the set $S_i(C) \subseteq \{0,1\}^n$ from C as follows. Replace every $c \in C$ such that both (i) $c_i = 1$, and (ii) the ith neighbor[1] of c is not in C with the ith neighbor of c. Authors have referred to this operation as “compression”, “switching”, and “polarization”. Previous works that use down-shifts include [Kle66, Enf70, BL96, GGL$^+$00, GT09, Mor12].

An important property of down-shifts is that they transform an arbitrary subset of $\{0,1\}^n$ into a downward-closed set, without changing cardinality. Specifically, if

$$D = S_n(S_{n-1}(\cdots S_1(C)))$$

is the result of sequentially applying S_i on C for each i, then D is downward-closed. It is also convenient in this context to think of D as a family of subsets of $[n]$ rather than a set of boolean vectors via the natural correspondence between boolean vectors and sets.

2 Our Results

We start with the result about linear program feasibility. We then state the bounds on the Hilbert function in terms of shattered sets and down-shifts. We show this leads to bounded-depth circuit lower bounds. Finally, we state two combinatorial applications.

[1] Vectors $u, v \in \{0,1\}^n$ are ith-neighbors if they differ in coordinate i and are the same elsewhere.

2.1 Linear Program Feasibility

We formalize and prove a strong version of the statement "linear programming feasibility can not be decided by polynomial-sized, constant-depth circuits." Clearly, linear programming being P-complete [DR80] implies a version of this statement for specific linear programs representing functions previously known not to have efficient bounded-depth circuits. We prove a stronger version stating that any linear feasibility problem, in which the number of constraints is roughly twice the number of variables and the constraints are non-degenerate, cannot be decided by an efficient bounded-depth circuit. For a set of hyperplanes $\mathcal{H}$ in $\mathbb{R}^k$ we will define a boolean function $f_{\mathcal{H}}$. It takes orientations as inputs and outputs one if and only if a certain polytope is nonempty. In particular, we establish hardness of this problem even when the hyperplanes are fixed in advance and only the orientations are given as input.

We express linear program feasibility as a boolean function as follows. Specify an arrangement of m hyperplanes $\mathcal{H} = \{h_1, \ldots, h_m\}$ with normal vectors $\vec{n}_i$ and translation scalars b_i as

$$h_i = \{\vec{x} : \langle \vec{n_i}, \vec{x} \rangle = b_i\}.$$

A sign pattern $s \in \{-1,1\}^m$ encodes the following linear programming feasibility problem:

Does there exist $\vec{x} \in \mathbb{R}^k$ satisfying $\mathsf{sign}(\langle x, \vec{n}_i \rangle - b_i) = s_i$ for all $i \in [m]$?

This corresponds to checking the feasibility of a linear program with m constraints and k variables. Define $f_{\mathcal{H}} : \{-1,1\}^m \to \{0,1\}$ as the boolean function such that $f_{\mathcal{H}}(s) = 1$ if and only if the linear program encoded by s is feasible.

As an example, consider the following arrangement in $\mathbb{R}^2$. The three hyperplanes

$$h_1 = \{(x_1,x_2) : 5x_1 + 3x_2 = 3\}, h_2 = \{(x_1,x_2) : 8x_1 - x_2 = 8\}, h_3 = \{(x_1,x_2) : 4x_1 - 5x_2 = 0\}$$

form an arrangement of three lines in the plane. The vector $s = (+1,-1,+1)$ encodes the system

$$\begin{aligned} 5x_1 + 3x_2 &> 3 && (s(1) = +1) \\ 8x_1 - x_2 &< 8 && (s(2) = -1) \\ 4x_1 + 5x_2 &> 0 && (s(3) = +1) \end{aligned}$$

In the example, the system encoded by $(+1,-1,+1)$ is not satisfiable (see Figure 1). For more background material on hyperplane arrangements and related results, see the books by Stanley [Sta04] and Matoušek [Mat02].

We prove the following theorem.

Theorem 2.1. *Let $\mathcal{H}$ be an arrangement of $2k+1$ hyperplanes in $\mathbb{R}^k$ that are in general position. Any $AC^0[p]$ circuit, for a prime p, with depth d computing $f_{\mathcal{H}}$ requires $exp(\Omega(k^{1/2d}))$ gates.*

We prove Theorem 2.1 in Section 4, using the framework of Razborov [Raz87], Smolensky [Smo87].

Explicit Arrangements. The space of oriented hyperplanes is a rich and well-studied object. The books [Mat02, Sta04] provide many facts and examples. The paper [AFR85] and references therein give bounds on how many different boolean functions can be represented as $f_{\mathcal{H}}$ for some hyperplane arrangement $\mathcal{H}$.

General position hyperplane arrangements come from any $2k+1$ vectors in $\mathbb{R}^{k+1}$ such that every $k+1$ of them are linearly independent. For a vector $v \in \mathbb{R}^{k+1}$ the hyperplane has normal $(v_1, \ldots, v_k)$ and translation v_{k+1}. Explicit families of m vectors in $\mathbb{R}^d$ such that every d of them are independent are known for any m, d. For example, take the rows of an $m \times d$ Cauchy or Vandermonde matrix.

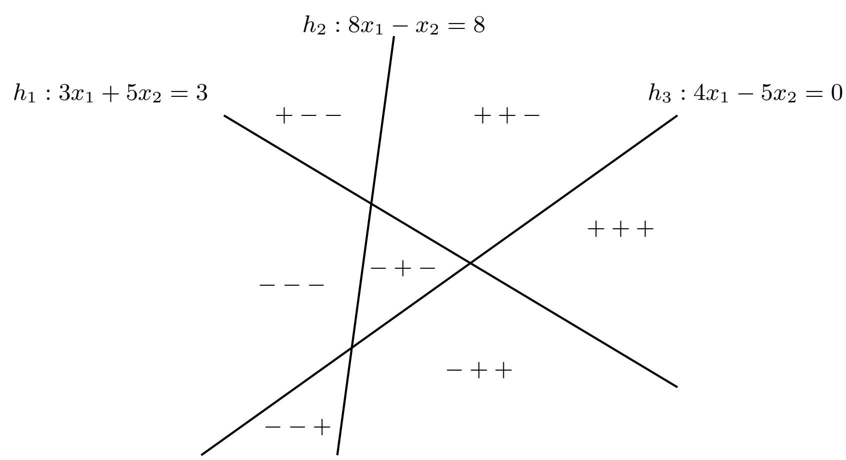

Figure 1: Three lines divide $\mathbb{R}^2$ into seven regions, each labeled by a feasible sign pattern.

2.2 Hilbert Function Bounds

Our results are based on the following theorem.

Theorem 2.2. *Any $C \subseteq \{0,1\}^n$ and any $d \in [n]$ satisfy the relationships*

$$|\{I \in \mathsf{sstr}(C) \ : \ |I| \leqslant d\}| \ \leqslant \ h_d(C, \mathbb{F}) \ \leqslant \ |\{I \in \mathsf{str}(C) \ : \ |I| \leqslant d\}|$$

and

$$\max\{|I| : I \in \mathsf{sstr}(C)\} \ \leqslant \ \mathsf{intdeg}(C, \mathbb{F}) \ \leqslant \ \max\{|I| : I \in \mathsf{str}(C)\}.$$

The upper bounds on interpolation degree are not new. Smolensky [Smo97] derives the Sauer-Perles-Shelah Lemma using very similar polynomial-based arguments. The upper bounds on interpolation degree in terms of VC dimension also appear implicitly in the work of Frankl and Pach [BF92] and explicitly in Gurvits [Gur97]. Our technical contributions center around the lower bounds and the applications. We prove Theorem 2.2 in Section 3.1.

We strengthen the lower bound on the Hilbert function in Theorem 2.2 using down-shifts.

Theorem 2.3. *Let $C \subseteq \{0,1\}^n$ and let $D = S_n(S_{n-1}(\ldots S_1(C)))$. Then*

$$|\{I \in D \ : \ |I| \leqslant d\}| \ \leqslant \ h_d(C, \mathbb{F}) \quad \text{and} \quad \max\{|I| : I \in D\} \ \leqslant \ \mathsf{intdeg}(C, \mathbb{F}).$$

In Section 3.2 we prove this theorem and show that the parity function provides a tight example over $GF(2)$. We also discuss how Theorem 2.3 implies the lower bound in Theorem 2.2.

2.3 Low-Degree Polynomial Approximations

Classic results in bounded-depth circuit complexity reduce the task of proving circuit lower bounds to showing that a boolean function has no low-degree approximation [Raz87, Smo87, ABFR94]. Smolensky shows in [Smo93] how to express all known degree lower bounds in terms of the Hilbert function. For a boolean function f consider the set $S = f^{-1}(1)$ as a subset of the boolean cube. Smolensky shows that if $h_d(S, \mathbb{F})$ is large, then f is hard to approximate.

Theorem 2.4 ([Smo93]). *Consider $f : \{0,1\}^n \to \{0,1\}$ and $p : \{0,1\}^n \to \mathbb{F}$. Define $S = f^{-1}(1)$ and fix $d = \lfloor (n - \deg_{\mathbb{F}}(p) - 1)/2 \rfloor$. Then,*

$$\Pr_x[p(x) \neq f(x)] \geqslant \frac{2 \cdot h_d(S, \mathbb{F}) - |S|}{2^n},$$

where x is uniform over $\{0,1\}^n$.

We provide a proof of Smolensky's result in Appendix A for completeness. Theorem 2.2 implies the following corollary in terms of strongly shattered sets.

Corollary 2.5. *Assume n is odd. Consider $f : \{0,1\}^n \to \{0,1\}$. If $|f^{-1}(1)| = 2^{n-1}$ and $\mathsf{sstr}(f^{-1}(1)) = \{I \subseteq [n] : |I| \leqslant \frac{n-1}{2}\}$, then for any polynomial $p \in \mathbb{F}[x_1, \ldots, x_n]$ we have*

$$\Pr_x[p(x) \neq f(x)] \geqslant \frac{1}{2} - \frac{10 \deg_{\mathbb{F}}(p)}{\sqrt{n}},$$

where x is uniform over $\{0,1\}^n$.

Proof. Since $\mathsf{sstr}(f^{-1}(1)) = \{I \subseteq [n] : |I| \leqslant \frac{n-1}{2}\}$, we have that

$$|\{I \in \mathsf{sstr}(C) \ : \ |I| \leqslant d\}| = \sum_{j=0}^{d} \binom{n}{j}$$

for all $d = 0, 1, \ldots, (n-1)/2$. Theorem 2.2 implies that $h_d(f^{-1}(1), \mathbb{F}) = \sum_{j=0}^{d} \binom{n}{j}$ as well. Plugging these into Theorem 2.4, along with $|f^{-1}(1)| = 2^{n-1}$, gives the corollary. □

Bernasconi and Egidi [BE99] thoroughly characterize the Hilbert function for symmetric sets and prove that any nearly-balanced, symmetric boolean function is hard to approximate. They leave as an open question deriving bounds for non-symmetric sets. Our connection to VC theory leads to new families of functions satisfying the conditions of Corollary 2.5. Many of these functions, such as the linear programming feasibility functions from Section 2.1, are non-monotone and non-symmetric. As a final remark, recent work shows that Smolensky's lower bound (and thus our result) extends to nonclassical polynomials [BL15].

2.4 The Sandwich Theorem

The following relationship which is a generalization of the Sauer-Perles-Shelah Lemma was discovered several times and independently [BR95, Paj85, Dre97, ARS02].

Theorem 2.6 (Sandwich Theorem). *For any $C \subseteq \{0,1\}^n$ we have $|\mathsf{sstr}(C)| \leqslant |C| \leqslant |\mathsf{str}(C)|$.*

Since $|\mathsf{str}(C)| \leqslant \sum_{i=0}^{\mathsf{VC}(C)} \binom{n}{i}$, this implies the Sauer-Perles-Shelah Lemma.

Theorem 2.2 yields a new algebraic proof of the Sandwich Theorem. Indeed, this follows from examining the case of $d = n$ and observing that $h_n(C, \mathbb{F}) = |C|$.

The Sandwich Theorem is tight in the sense that there are sets that achieve equality in both of its inequalities[2]. These sets are calles *shattering extremal sets*. For example, downward-closed sets are shattering extremal. Shattering extremal sets have been rediscovered and studied in different contexts [Law83, BRI89, BR95, Dre97, BCDK06, Mor12, KM13, MR14, MW15]. In our context, Corollary 2.5 says that shattering extremal sets S of size $|S| = 2^{n-1}$ and VC dimension $\frac{n-1}{2}$ correspond to boolean functions that cannot be approximated by low-degree polynomials.

[2] In fact, it is well known (see for example [Mor12]) that any set achieving equality in one of the inequalities, also achieves equality in the other.

2.5 Downward-closed Sets and Chvátal's Conjecture

Downward-closed sets have a well-studied, rich combinatorial structure. A theorem of Berge [Ber76] implies the following fact. For any downward-closed set D, there is a bijection $\pi : D \to D$ such that $a \cap \pi(a) = \emptyset$, for all $a \in D$. We refer to such a bijection as a *pseudo-complement*. We prove the following generalization of the existence of a pseudo-complement.

Theorem 2.7. *Let D be any downward-closed set. Fix any mapping $\phi : D \to D$ with the property that $\phi(a) \subseteq a$ for all $a \in D$. Then there exists a bijection $\pi : D \to D$ satisfying the condition that $a \cap \pi(a) = \phi(a)$ for all $a \in D$.*

Note that choosing $\phi(a) = \emptyset$ for all a implies the existence of a pseudo-complement.

In topology, downward-closed sets correspond to simplicial complexes. We think of the ϕ as prescribing intersections. For simplicial complexes, this corresponds to prescribing that complexes intersect in certain faces. We prove Theorem 2.7 in Section 5. Our proof proceeds by proving that a certain matrix is invertible. A non-zero determinant implies that the matrix contains a permutation matrix that yields the desired bijection.

We next discuss the result by Berge for the existence of pseudo-complements and its connections with Chvátal's conjecture in extremal combinatorics [CK72]. Berge's result about pseudo-complements follows from the following stronger theorem that he proved.

Theorem 2.8 ([Ber76])**.** *If D is a downward-closed set, then either D or $D \setminus \emptyset$ can be partitioned into pairs of disjoint sets.*

We need two definitions to explain Berge's motivation. A family B of subsets of $[n]$ is called a *star* if there is an element $x \in [n]$ such that $x \in b$ for all $b \in B$. It is called an *intersecting family* if every pair of sets in B intersects. Chvátal's conjecture is the following.

Conjecture 2.9 (Chvátal's conjecture)**.** *If D is a downward closed set, then the cardinality of the largest star in D is equal to the cardinality of the largest intersecting family in D.*

This conjecture remains open, aside from partial results, such as the following corollary of Berge's theorem.

Corollary 2.10. *In a downward-closed set D, any intersecting family has cardinality at most $|D|/2$.*

We contrast Berge's theorem and our Theorem 2.7. Berge's pair decomposition induces a permutation π such that $\pi(\pi(a)) = a$, whereas a permutation decomposes D into disjoint cycles with unspecified lengths. Many people have observed that the above corollary only needs the pseudo-complement result, instead of the stronger statement in Berge's theorem [And88]. Indeed, consider each disjoint cycle in the guaranteed permutation, and note that at most half of the sets in the cycle may mutually intersect. Therefore, our Theorem 2.7 implies the above corollary.

3 The Hilbert Function for Subsets of the Boolean Cube

We prove upper and lower bounds on the Hilbert function. First, we prove the bounds in Theorem 2.2 involving the shattered and the strongly shattered sets. Then, we prove the bounds in Theorem 2.3 using shifting. Finally we consider an example of applying these bounds to analyze the Hilbert function of the parity function.

3.1 Bounding the Hilbert Function Using Shattered Sets

The high-level idea of the proof of Theorem 2.2 is to define a vector space V with $\mathsf{dim}(V) = |C|$ and prove that $|\mathsf{sstr}(C)| \leqslant \mathsf{dim}(V) \leqslant |\mathsf{str}(C)|$. We sandwich the dimension $\mathsf{dim}(V)$ by finding a linearly independent set of size $|\mathsf{sstr}(C)|$ and a spanning set of size $|\mathsf{str}(C)|$.

We analyze the $|C|$-dimensional vector space $\{f : C \to \mathbb{F}\}$. Evaluation on C induces a natural mapping from $P \in \mathbb{F}[x_1, \dots, x_n]$ to the restriction $P|_C \in \{f : C \to \mathbb{F}\}$. The following lemma provides the desired sets of spanning monomials and linearly independent monomials.

Lemma 3.1. *For all fields $\mathbb{F}$ and sets $C \subseteq \{0,1\}^n$ the following two facts hold.*

1. *The monomials $\prod_{i \in I} x_i$ for $I \in \mathsf{str}(C)$ span $\{f : C \to \mathbb{F}\}$.*
2. *The monomials $\prod_{i \in I} x_i$ for $I \in \mathsf{sstr}(C)$ are linearly independent in $\{f : C \to \mathbb{F}\}$.*

Proof. For $I \subseteq [n]$, let x_Y denote the monomial $x_I = \prod_{i \in I} x_i$. For the first item, we express every $f : C \to \mathbb{F}$ as a linear combination of monomials $(x_I)|_C$ where $I \in \mathsf{str}(C)$. It suffices to express the monomials $(x_I)|_C$ for all $I \subseteq [n]$. We prove this by induction. For the base case, if $I \in \mathsf{str}(C)$, we are done. Otherwise, I is not shattered by C and there exists $s \in \{0,1\}^I$ such that for all $c \in C$, we have $c|_I \neq s$. Consider

$$P = \prod_{i \in I} \left(x_i - (1 - s_i) \right).$$

Note that $P(c) = 0$ for all $c \in C$ and hence $P|_C = 0|_C$. Specifically, by expanding the product $\prod_{i \in I} (x_i - (1 - s_i))$ we see

$$0|_C = P = (x_I)|_C + (Q)|_C,$$

where the degree of Q is smaller than $|I|$. By induction, we can write Q as a combination of $x_{I'}$ for $I' \in \mathsf{str}(C)$. Since $(x_I)|_C = (-Q)|_C$ we get that x_I is in this span as well.

We now prove the second item. Consider a linear combination

$$P = \sum_{I \in \mathsf{sstr}(C)} \alpha_I x_I$$

such that not all α_I are zero. We want to show that there is $c \in C$ such that $P(c) \neq 0$. Let $Z \in \mathsf{sstr}(C)$ be a maximal set such that $\alpha_Z \neq 0$. Since Z is strongly shattered by C, there is some $\bar{s} : ([n] \setminus Z) \to \{0,1\}$ such that all extensions of it in $\{0,1\}^n$ are in C. Let $Q(x_i)_{i \in Z}$ be the polynomial obtained by plugging in the values of $\bar{s}$ in the variables of $([n] \setminus Z)$. By maximality of Z it follows that the coefficient of x_Z in Q is $\alpha_Z \neq 0$, and so Q is not the 0 polynomial. Therefore there is $s \in \{0,1\}^Z$ such such that $Q(s) \neq 0$. Pick $c \in C$ such that

$$c_i = \begin{cases} s_i & i \in Z, \\ \bar{s}_i & i \in ([n] \setminus Z). \end{cases}$$

It follows that $P(c) = Q(s) \neq 0$, which finishes the proof. □

We use this lemma to prove bounds on the Hilbert function and interpolation degree.

Proof of Theorem 2.2. For the upper bound on $h_d(C, \mathbb{F})$, the above proof shows how to express all monomials of degree d using monomials of equal or smaller degree. For the lower bound on $h_d(C, \mathbb{F})$, linear independence still holds after restricting set size.

The upper bound on $\mathsf{intdeg}(C, \mathbb{F})$ is immediate. For the lower bound on $\mathsf{intdeg}(C, \mathbb{F})$, since $\mathsf{sstr}(C)$ is downward-closed, the linear independence of the monomials in $\mathsf{sstr}(C)$ implies any maximal degree monomial in $\{(x_I)|_C : I \in \mathsf{sstr}(C)\}$ cannot be expressed solely by lower degree monomials. □

3.2 Down-shifts, Downward-closed Bases, and the Hilbert Function

We prove Theorem 2.3. We also use the theorem to analyze the Hilbert function for the parity function. Theorem 2.3 is a direct corollary of the following theorem.

Theorem 3.2. *Let $C \subseteq \{0,1\}^n$ and let $D = S_n(S_{n-1}(\dots S_1(C)))$. Then the set of monomials $\{\prod_{i \in I} x_i : I \in D\}$ is a basis for the vector space of functions $\{f : C \to \mathbb{F}\}$.*

A theorem, equivalent in content, but expressed with respect to Gröbner bases, is proved in [Més09]. For completeness we include an elementary proof in Appendix B.

The lower bound given in Theorem 2.3 subsumes the lower bound in Theorem 2.2. This is a direct corollary of the following simple lemma.

Lemma 3.3. *Let $C \subseteq \{0,1\}^n$ and let $D = S_n(S_{n-1}(\dots S_1(C)))$. We have that $\mathsf{sstr}(C) \subseteq D$, where we associate $\{0,1\}^n$ with subsets of $[n]$ in the natural way.*

Proof. Since D is downward-closed, it follows that it is shattering extremal and therefore $\mathsf{sstr}(D) = D$. So, it is enough to show that $\mathsf{sstr}(C) \subseteq \mathsf{sstr}(D)$. To this end, it suffices to show that for every class C', $\mathsf{sstr}(C') \subseteq \mathsf{sstr}(S_i(C'))$. Let $I \in \mathsf{sstr}(C')$. Therefore C' contains a subcube B in coordinates I. During the down-shift, B is either shifted or stays in place, but in any case also $S_i(C')$ contains a subcube in coordinates I and therefore $I \in \mathsf{sstr}(S_i(C'))$. □

The Hilbert Function of Parity. A simple example which demonstrates an application of Theorem 2.3 is the parity function. Let P denote the set of all vectors of even hamming weight. Notice that P does not contain subcubes other than $\emptyset$. Therefore, $\mathsf{sstr}(P) = \{\emptyset\}$. As a consequence, the lower bound on the Hilbert function in Theorem 2.2 reveals no information in this case. In contrast, shifting gives a better bound. If we down-shift P, say on the first coordinate, we get the set $S_1(P) = D = \{v : v_1 = 0\}$. Therefore, as D is downward closed, shifting it on other coordinates does not change it. Thus, $S_n(S_{n-1}(\dots S_1(P))) = D$. By Theorem 2.3 we have that $h_d(P, \mathbb{F}) \geqslant \binom{n-1}{\leqslant d} = \binom{n-1}{d} + \binom{n-1}{d-1} + \ldots + \binom{n-1}{0}$.

This lower bound is tight when the field has characteristic two and $d \leqslant n/2$. It suffices to show every polynomial q of degree at most d can be expressed by a polynomial of degree at most d that does not depend on x_1. Therefore the $\binom{n-1}{\leqslant d}$ multilinear monomials that do not depend on x_1 span the space of degree at most d polynomials with domain P. Note that $(x_1 + \ldots + x_n)|_P = 0$, and therefore every appearance of x_1 can be replaced by $x_2 + \ldots + x_n$. This transforms q to a polynomial that does not depend on x_1 without changing the represented function.

4 Linear Programming and Low-degree Polynomial Approximations

We now prove Theorem 2.1. By the Razborov-Smolensky framework, it suffices to prove that $f_{\mathcal{H}}$ cannot be approximated by a low-degree polynomial over any field.[3]

Theorem 4.1. *Let $\mathcal{H}$ be an arrangement of $2k+1$ hyperplanes in $\mathbb{R}^k$ that are in general position. For any any polynomial $p \in \mathbb{F}[x_1, \dots, x_{2k+1}]$ we have*

$$\Pr_s[p(s) \neq f_{\mathcal{H}}(s)] \geqslant \frac{1}{2} - \frac{10 \deg_{\mathbb{F}}(p)}{\sqrt{2k+1}},$$

where s is uniform over $\{-1,1\}^{2k+1}$.

[3] We state the following theorem for $\{-1,1\}$ inputs to $f_{\mathcal{H}}$. This only makes sense for fields containing these elements. When $\mathbb{F} = \mathbb{F}_2$ simply replace $\{-1,1\}$ with $\{0,1\}$ in the definition of $f_{\mathcal{H}}$.

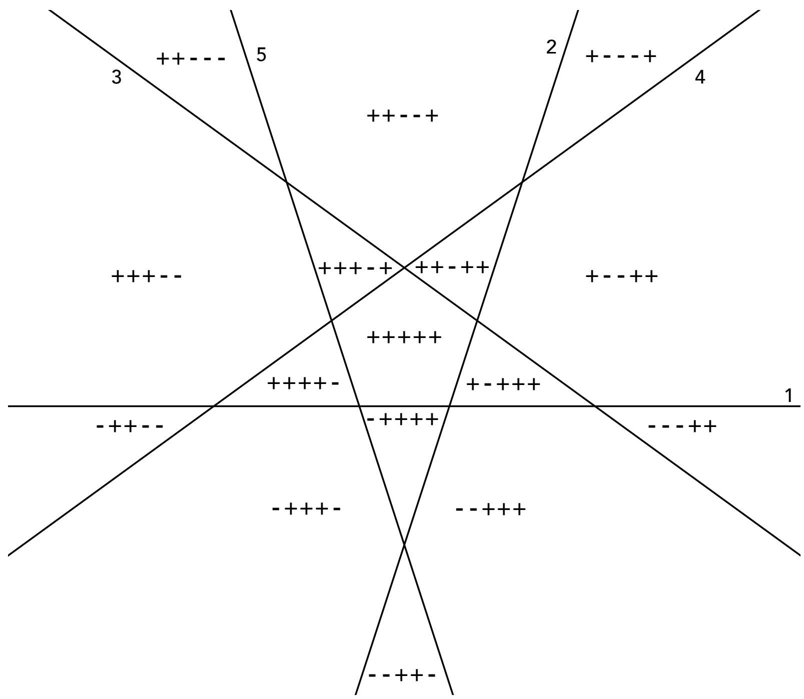

Figure 2: Five hyperplanes divide $\mathbb{R}^2$ into 16 cells. Cell labels in $\{-,+\}^5$ correspond to oriented hyperplane feasibility. Notice that every two coordinates are strongly shattered, but no three coordinates are shattered. This provides a proof-by-picture of Proposition 4.2 for $m = 5$ and $d = 2$.

The proof of Theorem 4.1 proceeds via a reduction to Corollary 2.5. Let

$$S_{\mathcal{H}} = \{s \in \{-1,1\}^n : \ f_{\mathcal{H}}(s) = 1\}.$$

To apply Corollary 2.5 on $f_{\mathcal{H}}$ we will show $|S_{\mathcal{H}}| = 2^{2k}$ and $\mathsf{sstr}(S_{\mathcal{H}}) = \{I \subseteq [2k+1] : |I| \leqslant k\}$. We establish this by the following proposition. The facts we need about hyperplane arrangements follow from standard arguments [GW94, Sta04]. For intuition about the following proposition, see Figure 2 for a pictorial proof in $\mathbb{R}^2$.

Proposition 4.2. *For any m hyperplanes $\mathcal{H}$ in $\mathbb{R}^d$ in general position*

$$\mathsf{sstr}(S_{\mathcal{H}}) = \mathsf{str}(S_{\mathcal{H}}) = \{I \subseteq [m] : |I| \leqslant d\}.$$

Proof. The two lemmas in Appendix C characterize the shattered and strongly shattered sets of $S_{\mathcal{H}}$ when $\mathcal{H}$ is in general position. The first lemma shows $\mathsf{str}(S_{\mathcal{H}}) \subseteq \{I \subseteq [m] : |I| \leqslant d\}$. The second lemma shows $\{I \subseteq [m] : |I| \leqslant d\} \subseteq \mathsf{sstr}(S_{\mathcal{H}})$. Since $\mathsf{sstr}(S_{\mathcal{H}}) \subseteq \mathsf{str}(S_{\mathcal{H}})$ these two lemmas combine to finish the proof. □

Proposition 4.2 implies Theorem 2.1. The equality $\mathsf{sstr}(S_{\mathcal{H}}) = \mathsf{str}(S_{\mathcal{H}})$ along with the Sandwich Theorem implies that $|S_{\mathcal{H}}| = |\mathsf{sstr}(S_{\mathcal{H}})|$. Let k be the ambient dimension in Theorem 2.1. The above proposition for $m = 2k+1$ and $d = k$ gives $|S_{\mathcal{H}}| = 2^{2k}$ and also $\mathsf{sstr}(S_{\mathcal{H}}) = \{I \subseteq [2k+1] : |Y| \leqslant k\}$. Thus $f_{\mathcal{H}}$ satisfies the premises of Corollary 2.5, and Theorem 2.1 follows.

5 Downward-closed Sets and Prescribed Intersections

We prove Theorem 2.7. Let $D \subseteq \{0,1\}^n$ be a downward-closed set. Fix $\phi : D \to D$ with the property that $\phi(a) \subseteq a$ for all $a \in D$. We will show that there exists a bijection $\pi : D \to D$ satisfying the condition that $a \cap \pi(a) = \phi(a)$ for all $a \in D$. We first prove two lemmas about the function space $\{f : D \to GF(2)\}$ and then use these to prove the existence of π. The first lemma holds for all subsets of the boolean cube.

Lemma 5.1. *Let $C \subseteq \{0,1\}^n$ be a subset of the boolean hypercube. The monomials*

$$\prod_{i\in a} x_i \text{ for } a \in C$$

form a basis for $\{f : C \to GF(2)\}$.

Proof. We proceed using induction on $|C|$. When $C = \{a\}$ for $a \in \{0,1\}^n$ the function space has dimension one and the monomial $\prod_{i\in a} x_i$ represent the constant "1" function in this space, which spans it. Let $z \in C$ denote a maximal Hamming weight element in C. Notice $\prod_{i\in z} x_i$ is an indicator function in $\{f : C \to GF(2)\}$ for the input z. By the inductive hypothesis on $(C \setminus \{z\})$, we know the set of monomials $\prod_{i\in a} x_i$ for $a \in (C \setminus \{z\})$ form a basis for $\{f : (C \setminus \{z\}) \to GF(2)\}$. Since $\prod_{i\in z} x_i$ is an indicator function, we may add it to the basis for $\{f : (C \setminus \{z\}) \to GF(2)\}$ and achieve a basis for $\{f : C \to GF(2)\}$. □

We remark that if C is downward-closed, then it is shattering extremal, and the above lemma is a corollary of the Sandwich theorem. We prove the following stronger claim as well.

Lemma 5.2. *Let $D \subseteq \{0,1\}^n$ be a downward-closed set. Fix any mapping $\phi : D \to D$ with the property that $\phi(a) \subseteq a$ for all $a \in D$. The functions*

$$\prod_{i\in\phi(a)} x_i \prod_{i\in a\setminus\phi(a)} (1 + x_i)$$

for $a \in D$ form a basis for $\{f : D \to GF(2)\}$.

Proof. Let $\mathcal{B}$ denote the set of polynomials that we wish to show is a basis. Since the cardinality of $\mathcal{B}$ is $|D|$ it is enough to show that it is a spanning set. By Lemma 5.1, it is enough to show that every monomial of the form $\prod_{i\in a} x_i$ for $a \in D$ can be expressed as a linear combination of polynomials in $\mathcal{B}$. We proceed by induction on the size of a. The case of $a = \emptyset$ is trivial. For the induction step, let $a \in D$ be non-empty. Expand the polynomial

$$\prod_{i\in\phi(a)} x_i \prod_{i\in a\setminus\phi(a)} (1 + x_i) = \left(\prod_{i\in a} x_i\right) + r,$$

where r is a linear combination of monomials $\prod_{i\in b} x_i$ for $b \subseteq a$ and $b \neq a$. Since D is downward-closed, by induction hypothesis r is in the span of $\mathcal{B}$. Thus,

$$\prod_{i\in a} x_i = \left(\prod_{i\in\phi(z)} x_i \prod_{i\in a\setminus\phi(a)} (1 + x_i)\right) + r$$

is also in the span of $\mathcal{B}$, and we are done. □

Proof of Theorem 2.7. We show there exists a bijection $\pi : D \to D$ such that $a \cap \pi(a) = \phi(a)$ for all $a \in D$, for the given map ϕ. Consider the $|D| \times |D|$ boolean matrix M defined as follows. Index the rows and columns both by D, and define the element in location $(a, b) \in D \times D$ to be one if and only if $a \cap b = \phi(a)$. We claim that M is nonsingular. Indeed, the rows of M correspond to the functions in Lemma 5.2. Since they form a basis, the row space of M is $|D|$-dimensional. This implies the determinant of M is nonzero. There must exist a permutation $\pi : [n] \to [n]$ such that $\prod_{i=1}^{|D|} M_{i,\pi(i)} = 1$. By the definition of M, we found the bijection π we were looking for. □

6 Conclusion

We exhibited a connection between algebra and combinatorics. We provided a general way to lower bound the Hilbert function. We showed a new family of functions cannot be approximated by low-degree polynomials. We provided a polynomial method proof of the Sandwich theorem and for a new theorem about prescribed intersections.

6.1 Open Directions

Our work suggests that the interpolation degree is a useful complexity measure on subsets of the boolean hypercube. Therefore, an open direction is to better understand the structure of sets with low interpolation degree. As noted by Remscrim [Rem16], one can equivalently define interpolation degree in terms of the rank of a certain incidence matrix. The matrix corresponds to the monomials in our Lemma 5.1 with a cut-off on the degree. For the case of interpolation degree one, this characterization is particularly simple.

Proposition 6.1. *A set $C \subseteq \{0,1\}^n$ has $\mathsf{intdeg}(C, \mathbb{F}) = 1$ if and only if the boolean vectors corresponding to C are affinely independent in $\mathbb{F}^n$.*

We are curious if other properties of the vectors in C correspond to implications for the interpolation degree. Even for interpolation degree two, the algebraic/matrix description becomes more opaque and less intuitive than the characterization in the above proposition. Since

$$\mathsf{intdeg}(C, \mathbb{F}) \leqslant \mathsf{VC}(C),$$

any combinatorial characterization may also shed new light on the structure of sets with VC dimension two, for which our understanding is lacking [AMY14, MSWY15].

Acknowledgements. We thank Paul Beame, Eli Ben-Sasson, Yuval Filmus, Ariel Gabizon, Sivaramakrishnan Natarajan Ramamoorthy, Anup Rao, Mert Sağlam, and Amir Yehudayoff for helping us improve the presentation of the paper. We also thank the anonymous reviewers for their comments.

References

[ABFR94] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

[AFR85] Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical Realization of Set Systems and Probabilistic communication Complexity. In *FOCS*, pages 277–280. IEEE, 1985.

[Alo99] Noga Alon. Combinatorial Nullstellensatz. *Combinatorics Probability and Computing*, 8(1):7–30, 1999.

[AMY14] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank, VC dimension and spectral gaps. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:135, 2014.

[And88] Ian Anderson. Combinatorics of Finite sets. *Bull. Amer. Math. Soc.*, 1988.

[ARS02] Richard P. Anstee, Lajos Rónyai, and Attila Sali. Shattering news. *Graphs and Combinatorics*, 18(1):59–73, 2002.

[BCDK06] Hans-Jürgen Bandelt, Victor Chepoi, Andreas W. M. Dress, and Jack H. Koolen. Combinatorics of lopsided sets. *Eur. J. Comb.*, 27(5):669–689, 2006.

[BE99] Anna Bernasconi and Lavinia Egidi. Hilbert Function and Complexity Lower Bounds for Symmetric Boolean Functions. *Information and Computation*, 153(1):1–25, 1999.

[Bei93] Richard Beigel. The Polynomial Method in Circuit Complexity. In *In Proceedings of the 8th IEEE Structure in Complexity Theory Conference*. Citeseer, 1993.

[Ber76] C Berge. A Theorem Related to the Chvátal conjecture. In *Proceedings, 5th British Combinatorial Conference*, 1976.

[BF92] L Babai and P Frankl. *Linear Algebra Methods in Combinatorics.* University of Chicago, 1992.

[BL96] Béla Bollobás and Imre Leader. Sums in the grid. *Discrete Mathematics*, 162(1-3):31–48, 1996.

[BL15] Abhishek Bhowmick and Shachar Lovett. Nonclassical Polynomials as a Barrier to Polynomial Lower Bounds. In *30th Conference on Computational Complexity*, page 72, 2015.

[BR95] Béla Bollobás and A. J. Radcliffe. Defect sauer results. *J. Comb. Theory, Ser. A*, 72(2):189–208, 1995.

[BRI89] Béla Bollobás, A. J. Radcliffe, and Leader I. Reverse kleitman inequalities. *Proc. London Math. Soc., Ser. A*, (3) 58:153–168, 1989.

[CK72] Vašek Chvátal and Donald Knuth. *Selected Combinatorial Research Problems.* Stanford University, 1972.

[DR80] David P Dobkin and Steven P Reiss. The Complexity of Linear Programming. *Theoretical Computer Science*, 11:1–18, 1980.

[Dre97] A.W.M. Dress. Towards a theory of holistic clustering. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, 37 Amer. Math. Soc.:271–289, 1997.

[Dvi09] Zeev Dvir. On the size of Kakeya sets in finite fields. *Journal of the American Mathematical Society*, 22(4):1093–1097, 2009.

[Enf70] Per Enflo. On the nonexistence of uniform homeomorphisms betweenl p -spaces. *Arkiv för Matematik*, 8(2):103–105, 1970.

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

[GK15] Larry Guth and Nets Hawk Katz. On the Erdos Distinct Distances Problem in the Plane. *Annals of Mathematics*, 181(1):155–190, 2015.

[GT09] Ben Joseph Green and Terence Tao. Freiman's theorem in finite fields via extremal set theory. *Combinatorics, Probability & Computing*, 18(3):335–355, 2009.

[Gur97] Leonid Gurvits. Linear algebraic proofs of VC-dimension based inequalities. In *Computational Learning Theory*, pages 238–250. Springer, 1997.

[GW94] Bernd Gärtner and Emo Welzl. Vapnik-Chervonenkis dimension and (pseudo-) hyperplane arrangements. *Discrete & Computational Geometry*, 12(1):399–432, 1994.

[Juk11] Stasys Jukna. *Extremal Combinatorics: with Applications in Computer Science*. Springer Science & Business Media, 2011.

[Kle66] Daniel J. Kleitman. Families of non-disjoint subsets. *Journal of Combinatorial Theory*, 1(1):153 – 155, 1966.

[KM13] László Kozma and Shay Moran. Shattering, graph orientations, and connectivity. *Electr. J. Comb.*, 20(3):P44, 2013.

[Law83] J. Lawrence. Lopsided sets and orthant-intersection by convex sets. *Pac. J. Math.*, 104(1):155–173, 1983.

[Mat02] Jiří Matoušek. *Lectures on discrete geometry*, volume 212. Springer New York, 2002.

[Més09] Tamás Mészáros. S-extremal set systems and Gröbner Bases. Master's thesis, Budapest University of Technology and Economics, 2009.

[Mor12] S. Moran. Shattering Extremal Systems. Master's thesis, Saarland University, Saarbrücken, Germany, 2012.

[MR14] Tamás Mészáros and Lajos Rónyai. Shattering-extremal set systems of VC dimension at most 2. *Electr. J. Comb.*, 21(4):P4.30, 2014.

[MSWY15] Shay Moran, Amir Shpilka, Avi Wigderson, and Amir Yehudayoff. Teaching and compressing for low vc-dimension. In *FOCS*, 2015.

[MW15] Shay Moran and Manfred K. Warmuth. Labeled compression schemes for extremal classes. *CoRR*, abs/1506.00165, 2015.

[Paj85] A. Pajor. Sous-espaces l_1^n des espaces de banach. *Travaux en Cours. Hermann, Paris*, 1985.

[Raz87] Alexander A Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes*, 41(4):333–338, 1987.

[Rem16] Zachary Remscrim. The Hilbert Function, Algebraic Extractors, and Recursive Fourier Sampling. *ECCC Technical Report TR 16-020*, 2016.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82. ACM, 1987.

[Smo93] R. Smolensky. On Representations by Low-degree Polynomials. In *FOCS*, 1993.

[Smo97] Roman Smolensky. Well-Known Bound for the VC-Dimension Made Easy. *Computational Complexity*, 6(4):299–300, 1997.

[SSZ15] Micha Sharir, Adam Sheffer, and Joshua Zahl. Improved Bounds for Incidences Between Points and Circles. *Combinatorics, Probability and Computing*, 24(03):490–520, 2015.

[Sta04] Richard P Stanley. Introduction to Hyperplane Arrangements. *Lecture notes, IAS/Park City Mathematics Institute*, 2004.

[Tao14] Terence Tao. Algebraic Combinatorial Geometry: the Polynomial Method in Arithmetic Combinatorics, Incidence Combinatorics, and Number Theory. *EMS Surveys in Mathematical Sciences*, 1(1):1–46, 2014.

A Proof of Smolensky's Lower bound

For completeness, we provide a proof of Smolensky's Theorem 2.4. We introduce the following notation that will be convenient. For any set $C \subseteq \{0,1\}^n$ and degree $d \in [n]$ let $\mathsf{E}^{d,C}$ denote the evaluation (or restriction) mapping

$$\mathsf{E}^{d,C} : \mathbb{F}[x_1, \ldots, x_n]^{\leqslant d} \to \{f : C \to \mathbb{F}\},$$

where $\mathbb{F}[x_1, \ldots, x_n]^{\leqslant d}$ denotes polynomials with degree at most d over $\mathbb{F}$. Note that $\mathsf{E}^{d,C}$ is a linear map and $h_d(C, \mathbb{F}) = \mathrm{rk}_{\mathbb{F}}(\mathsf{E}^{d,C})$. Associate the map $\mathsf{E}^{d,C}$ with a matrix such that the rows correspond to elements of C, the columns correspond to degree d monomials, and entries are evaluations.

A.1 Proof of Theorem 2.4

Smolensky's Theorem follows from two lemmas. We assume all ranks are over $\mathbb{F}$.

Lemma A.1. *For any $S, T \subseteq \{0,1\}^n$ and any $d \in [n]$ we have $|S \setminus T| \geqslant h_d(S, \mathbb{F}) - h_d(T, \mathbb{F})$.*

Proof. Notice that $\mathsf{E}^{d,S}$ and $\mathsf{E}^{d,T}$ have identical submatrices induced by the $S \cap T$ rows. Denote the submatrices as $A = \mathsf{E}^{d,S\cap T}$ and $B = \mathsf{E}^{d,S\setminus T}$ and $C = \mathsf{E}^{d,T\setminus S}$. We want to prove

$$\mathrm{rk}(B) \geqslant \mathrm{rk}\begin{pmatrix} A \\ B \end{pmatrix} - \mathrm{rk}\begin{pmatrix} A \\ C \end{pmatrix}. \tag{1}$$

Equation 1 actually holds for any three matrices, following from the bounds

$$\mathrm{rk}\begin{pmatrix} A \\ B \end{pmatrix} \leqslant \mathrm{rk}(A) + \mathrm{rk}(B) \qquad \text{and} \qquad \mathrm{rk}\begin{pmatrix} A \\ C \end{pmatrix} \geqslant \mathrm{rk}(A).$$

The fact that $\mathrm{rk}(B) = \mathrm{rk}(\mathsf{E}^{d,S\setminus T}) \leqslant |S \setminus T|$ concludes the proof. □

Lemma A.2 ([Smo93])**.** *Let $p : \{0,1\}^n \to \mathbb{F}$ and define $P = \{x : p(x) \neq 0\}$. If $d < (n - \deg(p))/2$ then*

$$\mathrm{rk}(\mathsf{E}^{d,P}) \leqslant |P|/2.$$

Proof. Assume for contradiction that $\mathrm{rk}(\mathsf{E}^{d,P}) = r > |P|/2$. Let M_1 be an $r \times r$ full-rank sub-matrix. Our goal is to find two degree d polynomials q_1 and q_2 such that the product $q_1 q_2$ is the indicator function for some row in M_1. Assume without loss of generality the matrix $\mathsf{E}^{d,P}$ looks like

$$\left(\begin{array}{c|c} M_1 & \cdots \\ \hline M_2 & \cdots \end{array}\right).$$

Let Q_1 and Q_2 denote the rows corresponding to M_1 and M_2, respectively. We start by constructing q_2 to be zero on Q_2. Since the rank satisfies $r > |P|/2$, we know M_2 has more columns than rows, and thus we can find a vector v_2 such that $\mathsf{E}^{d,P} v_2$ is zero on Q_2. Since M_1 has full-rank, we know that $\mathsf{E}^{d,P} v_2$ is nonzero on some point x^* corresponding to a row in Q_1. Notice that for any $v \in \mathbb{F}^{|P|}$ the vector $\mathsf{E}^{d,P} v$ corresponds to a degree d polynomial. Let q_2 be the polynomial for v_2.

Since M_1 has full rank, we can find a vector v_1 and corresponding polynomial q_1 such that $q_1 q_2$ is the indicator function for x^*. To conclude, notice that $p q_1 q_2$ interpreted as a function on $\{0,1\}^n$ is nonzero on only a single point. By inspection this means $\deg(p q_1 q_2) \geqslant n$, contradicting the assumption that its degree is at most $2d + \deg(p) < n$.

□

With the above lemmas in hand, we can prove Smolensky's theorem.

Proof of Theorem 2.4. Let $F = f^{-1}(0)$ and $P = \{x : p(x) \neq 0\}$ and $d = (n - \deg(p) - 1)/2$. Notice

$$\Pr_x[p(x) \neq f(x)] \geqslant |F \triangle P|/2^n.$$

We show $|F \triangle P| \geqslant 2h_d(F, \mathbb{F}) - |F|$. Indeed, this follows from the above two lemmas:

$$\begin{aligned}
|F \setminus P| &\geqslant h_d(F, \mathbb{F}) - h_d(P, \mathbb{F}) && \text{(By Lemma A.1)} \\
&\geqslant h_d(F, \mathbb{F}) - \frac{|P|}{2} && \text{(Lemma A.2 implies } h_d(P, \mathbb{F}) \leqslant \tfrac{|P|}{2}) \\
&= h_d(F, \mathbb{F}) - \frac{|F| + |P \setminus F| - |F \setminus P|}{2}. && (|P| = |F| + |P \setminus F| - |F \setminus P|.)
\end{aligned}$$

Therefore, by rearranging the above expressions, we see that

$$\frac{1}{2}|F \triangle P| \geqslant h_d(F, \mathbb{F}) - \frac{|F|}{2}.$$

□

B Downward-closed Bases from Shifting

In this section, we prove Theorem 3.2, which says that $\{x_I : I \in D\}$ is a basis for the vector space of functions $\{f : C \to \mathbb{F}\}$, where $D = S_n(S_{n-1}(\ldots S_1(C)))$. The proof of Theorem 3.2 is by induction on n. For the induction to work we prove a stronger statement. Consider the lexicographical order on the set of all multilinear monomials. That is, $m_1 < m_2$ if the smallest i such that x_i appears in exactly one of m_1, m_2 appears in m_2. For example, $x_1 > x_2 x_3 x_4 \ldots x_n$.

We are now ready to state the stronger statement we will prove.

Lemma B.1. *Let p be a polynomial where x_I is the leading monomial of p. Then*

$$p|_C = 0 \implies I \notin D.$$

The above lemma implies that there is no linear combination of monomials in $\{x_I : I \in D\}$ which represents the zero function on C (because the leading monomial in such a combination will be in D which contradicts the lemma). Since down-shifting C does not change its cardinality, it follows that $|C| = |\{x_I : I \in D\}|$ and therefore $\{x_I : I \in D\}$ is indeed a basis.

In the proof we will use the following "locality" of down-shifts. For $i \in [n]$ define the two sets

$$C_{i=0} = \{c \in C : c_i = 0\}, \qquad \text{and} \qquad C_{i=1} = \{c \in C : c_i = 1\}.$$

For all $i, j \in [n]$ with $i \neq j$, the shifting operator S_j satisfies

$$S_j(C) = S_j(C_{i=0}) \cup S_j(C_{i=1}). \tag{2}$$

Putting it differently: $C_{i=0}, C_{i=1}$ are invariant under S_j when $i \neq j$.

Proof of Lemma B.1. Let $C_0 = \{c \in C : c_n = 0\}$, and $C_1 = \{c \in C : c_n = 1\}$. By the locality of down-shifts (Equation 2):

$$S_{n-1}(S_{n-2}(\ldots S_1(C))) = S_{n-1}(S_{n-2}(\ldots S_1(C_0))) \cup S_{n-1}(S_{n-2}(\ldots S_1(C_1))).$$

For $b \in \{0,1\}$, let

$$D_b = \{w|_{\{1,2,\dots,n-1\}} : w \in S_{n-1}(S_{n-2}(\dots S_1(C_b)))\}.$$

By the induction hypothesis applied on C_b, there are no polynomials that represent the zero function on C_b whose leading monomial is x_I where $I \in D_b$.

Assume towards contradiction some multilinear polynomial p represents the zero function on C and has leading monomial of x_I with $I \in D$. A crucial observation which follows directly from the definition of the down-shift S_n is that if $n \in I$ then $(I \setminus \{n\})$ belongs to D_0 **and** to D_1, and if $n \notin I$ then I belongs to D_0 **or** to D_1. Consider the expansion

$$p = \alpha_I x_I + \sum_{J<I} \alpha_J x_J,$$

where $J < I$ in the lexicographical monomial ordering. We distinguish between two cases.

(i) $n \notin I$. Assume that $I \in D_1$ (the proof for $I \in D_0$ is symmetric). Notice that since x_I is the leading monomial, the monomial $x_{I\cup\{n\}}$ does not appear in p. Therefore, the polynomial that is obtained by setting $x_n = 1$ in p has leading monomial x_I and represents the zero function on C_1, contradicting the induction hypothesis.

(ii) $n \in I$. Here we distinguish between two subcases. If $\alpha_{I\setminus\{n\}} + \alpha_I \neq 0$ then the polynomial that is obtained by setting $x_n = 1$ in p has leading monomial $x_{I\setminus\{n\}}$ and represents the zero function on C_1, contradicting the induction hypothesis. If $\alpha_{I\setminus\{n\}} + \alpha_I = 0$ then $\alpha_{I\setminus\{n\}} \neq 0$ and the polynomial that is obtained by setting $x_n = 0$ in p has leading monomial $x_{I\setminus\{n\}}$ and represents the zero function on C_0, contradicting the induction hypothesis. □

C Hyperplane Arrangement Lemmas

We now prove the two lemmas that we assumed in the proof of Proposition 4.2. Recall $\mathcal{H}$ is in general position if any $k \leqslant d$ hyperplanes intersect in a $(d-k)$-dimensional affine subspace. This means (i) every $d+1$ hyperplanes have an empty intersection, and (ii) every subset of d normal vectors are linearly independent. (if some d normal vectors are linearly dependent then their intersection is either empty or infinite but not 0 dimensional).

Lemma C.1. *Let $\mathcal{H}$ be an arrangement in $\mathbb{R}^d$. If $Y \subseteq [m]$ is shattered by $S_{\mathcal{H}}$ then the hyperplanes h_i for $i \in Y$ have a non-empty intersection. In particular, if $\mathcal{H}$ is in general position then $|Y| \leqslant d$.*

Proof. Let M_Y be the matrix whose rows are the normals $\vec{n}_i, i \in Y$, and let $\vec{b}_Y$ be the vector $(b_i)_{i\in Y}$. Consider the affine transformation $T_Y : \mathbb{R}^d \to \mathbb{R}^Y$ defined by

$$T_Y(\vec{x}) = M_Y \cdot \vec{x} - \vec{b}_Y.$$

We need to show that there is some $\vec{x} \in \cap_{i\in Y} h_i$. Note that $\vec{x} \in \cap_{i\in Y}$ if and only if $T_Y(\vec{x}) = \vec{0}$. Thus, it is enough to show that $\vec{0}$ is in the image of T_Y. Now, the fact that $S_{\mathcal{H}}$ shatters Y amounts to that for every sign vector $\vec{s} \in \{-1,1\}^Y$ there exists $\vec{x} \in \mathbb{R}^m$ such that $\mathsf{sign}(T_Y(\vec{x})) = \vec{s}$. In other words, every orthant contains a vector u in the image of T_Y. Now, the convex hull, C, of these vectors is also contained in the image of T_Y (since the image of an affine transformation is convex). Also, since these vectors contain a vector in every orthant, it follows that $\vec{0} \in C$, and therefore $\vec{0}$ is in the image of T_Y. Indeed, an exercise shows that any set of vectors containing a vector in each orthant also contains the origin in its convex hull. □

Lemma C.2. *Let $\mathcal{H}$ be an arrangement in $\mathbb{R}^d$ that is in general position. If the normals $\vec{n}_i, i \in Y$ are linearly independent then Y is strongly shattered by $S_{\mathcal{H}}$. In particular, any subset of size at most d is strongly shattered by $S_{\mathcal{H}}$.*

Proof. It is enough to prove only for sets Y such that the normals $\vec{n}_i, i \in Y$ form a basis to $\mathbb{R}^d$ (because every independent set can be extended to a basis and every subset of a strongly shattered set is strongly shattered). Let $M = M_{[m]}$ be the matrix whose rows are the normals $\vec{n}_i, i \in [m]$, and let $\vec{b} = \vec{b}_{[m]}$ be the vector $(b_i)_{i\in[m]}$. Consider the affine transformation $T : \mathbb{R}^d \to \mathbb{R}^m$ defined by

$$T(\vec{x}) = M \cdot \vec{x} - \vec{b}.$$

Since $\mathcal{H}$ is in general position, there is a unique $\vec{x}^*$ that lies in the intersection $\cap_{i\in Y} h_i$. In other words, $T(\vec{x}^*)_i = 0$ for every $i \in Y$. Moreover, since every $d+1$ hyperplanes have an empty intersection, it follows that for all $j \notin Y$: $T(\vec{x}^*)_j \neq 0$. We show that Y is strongly shattered by showing that for every sign pattern $s \in \{-1,1\}^Y$ there exists $\vec{u} \in \mathbb{R}^d$ such that for all $i \in Y$

$$\mathsf{sign}(T(\vec{x}^* + \vec{u}))_i = \vec{s}_i,$$

and for all $j \notin Y$

$$\mathsf{sign}(T(\vec{x}^* + \vec{u}))_j = \mathsf{sign}(\vec{x}^*)_j.$$

Indeed, since the normals $\vec{n}_i, i \in Y$ are linearly independent, there is some $\vec{w} \in \mathbb{R}^d$ such that $(M \cdot \vec{w})_i = s_i$. Therefore, for every $\varepsilon > 0$ and $i \in Y$

$$\mathsf{sign}(T(\vec{x}^* + \varepsilon\vec{w}))_i = s_i.$$

Moreover, since for all $j \notin Y$ $T(\vec{x}^*)_j \neq 0$ then the exists some $\varepsilon' > 0$ such that for al $j \notin Y$

$$\mathsf{sign}(T(\vec{x}^* + \varepsilon'\vec{w}))_j = \mathsf{sign}(\vec{x}^*)_j.$$

Thus, picking $\vec{u} = \varepsilon'\vec{w}$ finishes the proof. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de