THE UNIVERSITY OF CHICAGO

POLYNOMIAL REPRESENTATION, FOURIER TRANSFORM AND COMPLEXITY
OF BOOLEAN FUNCTIONS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
MASTER

DEPARTMENT OF COMPUTER SCIENCE

BY
HING YIN TSANG

CHICAGO, ILLINOIS
NOVEMBER 9, 2017
Polynomial representation, Fourier transform and complexity of Boolean functions
(Master Thesis)

Hing Yin Tsang

November 9, 2017

Abstract

In the analysis of Boolean functions, expressing a function as a polynomial over a finite field or taking its Fourier transform (a representation as a real polynomial) of it are two of the most successful techniques and lead to many important results in various areas in theoretical computer science.

In this thesis, we study the Log-rank Conjecture for XOR functions and the Sensitivity Conjecture by exploiting structure of the Fourier transform and polynomial representations of Boolean functions.

For the Log-rank Conjecture for XOR functions, we show that it is true for functions with small spectral norm or low $F_2$-degree. More precisely, we prove the following two bounds.

1. $CC(f(x \oplus y)) = O(\|\hat{f}\|_1 \cdot \log \|\hat{f}\|_0)$, and
2. $CC(f(x \oplus y)) = O(2^{d/2} \log^{\max\{d-2,1\}} \|\hat{f}\|_0)$, where $d = \deg_2(f) \geq 1$.

For the Sensitivity Conjecture, we obtain a pair of conjectures with the property that (i). each of them is a consequence of the Sensitivity Conjecture, (ii). neither one of them is known to imply the Sensitivity Conjecture and (iii) both of them together imply the Sensitivity Conjecture. We also obtain a new sufficient condition for the Sensitivity Conjecture based on a graph that we called the monomial graph associated with the Boolean function. Specifically, we show that in order to prove the Sensitivity Conjecture, it suffices to show that for all Boolean functions $f$, the number of monotone paths from 0$^n$ to some level $\ell = \Omega(n)$ in its monomial graph is at most $s(f)^{O(\ell)}$. On the other hand, we show that for all Boolean functions $f$,

1. the number of monotone paths from 0$^n$ to level $\ell$ in its monomial graph is at most $\min\{2^{\ell^2/2}s(f)^\ell, (s(f)^4s(f))^{\ell}\}$;
2. the number of degree-$\ell$ monomials in its $\mathbb{R}$-representation is at most $(4\ell)^{\ell}s(f)^{\ell}\min\{s(f), \ell\}$.

Contents

1 Introduction 2
2 Acknowledgements 5

*Department of Computer Science, University of Chicago, Chicago, IL. Email: hytsang@cs.uchicago.edu
1 Introduction

One of the most powerful methods in analysis of Boolean function is to represent the Boolean functions as some other objects which allow us to exploit additional structure that is difficult to see when we look at the functions directly.

Minsky and Papert initiated in their book *Perceptrons* [MP68] the study of Boolean functions using their polynomial representations. Since then polynomial representations take part in many important results in lower bounds or pseudorandomness on various computation models.

The other important representation is the Fourier transform. The seminal work of Kahn, Kalai and Linial [KKL88] is probably one of the first paper that use Fourier analysis with deep results in harmonic analysis to prove combinatorial results on Boolean functions. Specifically, they used the hypercontractivity inequality to show that balance Boolean functions must have a coordinate with significant influence. After that the Fourier transform has been widely used in computer science and found applications in learning theory, hardness of approximation, property testing, communication complexity, circuit complexity, and more. See the textbook of O’Donnell [O’D14] on the analysis of Boolean function using Fourier transform.

In this thesis we exploit the structure in the polynomial representation over various fields and the Fourier transform, and use them to study two major open problems in concrete complexity — the Log-rank Conjecture for XOR functions and the Sensitivity Conjecture.

**The Log-rank Conjecture for XOR functions.** The *Log-rank Conjecture* of Lovász and Saks [LS88] asserts that the (deterministic) communication complexity of a Boolean function $F(x, y)$ can be upper bound by a function polylogarithmic in the rank (over $\mathbb{R}$) of its communication matrix $(M_F)_{x,y} = F(x, y)$. The best known upper bound on the communication complexity in terms of rank of the communication matrix was merely linear until the work of Lovett [Lov14], who showed the bound $CC(F) \leq \tilde{O}(\sqrt{\text{rank}_{\mathbb{R}}(M_F)})$ (which is exponential in the log of the rank).
On the other hand, the best known separation between $CC(F)$ and $\log \text{rank}_R(M_F)$ is quadratic [GPW15]. Precisely, Göös, Pitassi and Watson [GPW15] showed that there exists an explicit function $F(x, y)$ with $CC(F) \geq \Omega(\log^2 \text{rank}_R(M_F))$. The Log-rank Conjecture itself is known to be equivalent to many other conjectures [LS88, Lov90, Val04, ASTS+03].

Shi and Zhang [ZS09] suggested to consider the special case of XOR functions due to its connection to Fourier analysis of Boolean functions. Specifically, it can be shown that the rank of the communication matrix is exactly the Fourier sparsity of the underlying function [BC99]. In the paper [ZS10], Shi and Zhang made a crucial observation that communication protocol can simulate any parity decision tree with amount of communication at most twice of the depth of the parity decision tree being simulated, i.e., $CC(f(x \oplus y)) \leq 2D_\oplus(f)$ for all Boolean functions $f$. It immediately implies a sufficient condition for proving the Log-rank Conjecture for XOR functions — it suffices to design a parity decision tree for $f$ such that the depth is at most polynomial in $\log ||f||_q$. Motivated by known relationships between query complexity and a number of complexity measures including block sensitivity and certificate complexity, Shi and Zhang [ZS10] defined the “parity” version of these combinatorial measures and showed that many known relationships for the “non-parity” version extended to similar relationships for the “parity” version.

The Log-rank Conjecture for XOR functions has been confirmed for a few classes of special functions, including symmetric functions [ZS09], LTFs and monotone functions [MO10], AC0 functions [KS13], functions with small spectral norm and low-degree polynomials [TWXZ13], and functions with constant alternating number [LZ16].

The Sensitivity Conjecture. The sensitivity versus block sensitivity problem of Nisan and Szegedy [NS94] asks what is the correct relationship between the sensitivity and block sensitivity. More precisely, the Sensitivity Conjecture asserts that sensitivity and block sensitivity are actually polynomially related.

Since block sensitivity is polynomially related to a number of measures including the query complexity, we can basically ask instead about the relationship between sensitivity and any of them. The separation between block sensitivity and sensitivity was known to be at least quadratic since the work of Rubinstein in 1995 [Rub95], Virza [Vir11] and then Ambainis and Sun [AS11] improved the separation by a constant factor. But the largest known separation is still quadratic. The first super-quadratic separation between sensitivity and any other relevant measures was achieved by Tal [Tal16], who gave a 2.1 power separation between sensitivity and query complexity. This separation was then improved by Ben-David, P. Hatami and Tal [BHT17] to a cubic separation between quantum query complexity and sensitivity. In the same paper they also showed a 2.22 power separation between sensitivity and certificate complexity.

On the other hand, the best known upper bound on any relevant measures in terms of sensitivity is exponential. The first such bound is due to Simon [Sim83]. His bound was the best known upper bound for over twenty years until Kenyon and Kutin [KK04] improved the exponent by a constant factor using a rather different argument. The next improvement was due to Ambainis et al. [ABG+14] using a combinatorial argument similar to Simon’s proof, in which they showed $bs(f) = O(2^{s(f)}s(f))$. The argument of Ambainis et al. has been improved slightly in [APV16] and [HLS16]. But an $O(2^{s(f)}s(f))$ type upper bound is still the best known upper bound on block sensitivity in terms of sensitivity.

The Sensitivity Conjecture was also studied for special classes of functions. It is known that the conjecture is true for symmetric functions, graph properties [Tur84, Sun11], bipartite graph properties [GMSZ13], minterm-transitive functions [Cha05], monotone functions [Nis91], functions with constant alternating number [LZ16], and various classes of functions defined based on the circuits that compute them [Mor14, BLTV16, KT16]. It is also known to be false for real-valued functions on Boolean cubes with range $[0,1]$ [Tal16].

The Sensitivity Conjecture also has many equivalent formulations. We refer to the survey of P. Hatami, Kulkarni and Pankratov [HKP11] for the details.
Our results. Although the Log-rank Conjecture and the Sensitivity Conjecture concern about very different complexity measures, both of them can be reduced to a problem of proving existence of low cost query algorithms. We show that there is a general way to construct decision trees from polynomial representations over various fields. Depending on the field and the way the polynomial is expressed, we obtain different types of decision trees (which allow different types of queries).

For the Log-rank Conjecture for XOR functions, by combining our query algorithm and a known relationship between $\mathbb{F}_2$-degree $\deg_{\mathbb{F}_2}(f)$ and the Fourier sparsity $\|f\|_0$, the conjecture reduces to a problem of finding a large affine subspace on which the Boolean function is a constant. We then confirm the conjecture for the classes of functions with small spectral norm or low $\mathbb{F}_2$-degree (these results were from the joint work of the author with Wong, Xie and Zhang [TWXZ13]). Both results are proved by iteratively refining the affine subspace so that the function becomes closer and closer to be a constant function on the affine subspace. For functions with small spectral norm, we use the largest magnitude of the Fourier coefficients to measure how close a function is to a constant; and for functions with low $\mathbb{F}_2$-degree, we use the spectral norm as a performance measure and show how to reduce the spectral norm by a constant factor using a small number of linear restrictions.

For the Sensitivity Conjecture, as a direct consequence of our query algorithm, we show that the Sensitivity Conjecture is equivalent to proving both of the following conjectures:

1. $C_{\min}(f) \leq s(f)^c$ for some absolute constant $c$, and
2. $\deg_{\mathbb{F}_2}(f) \leq s(f)^c$ for some absolute constant $c$.

Both conjectures are trivial consequences of the Sensitivity Conjecture; none of them alone is known to imply the Sensitivity Conjecture; but interestingly, both of them together imply the Sensitivity Conjecture. As far as we know, it is the first equivalent formulation of the Sensitivity Conjecture with this kind of property.

We then proceed to find useful structure in the polynomial representation. We show that for low-sensitivity functions, their polynomial representations must be sparse at every level (i.e., there are small number of degree-$\ell$ monomials for every $\ell$). This property could be useful for upper bounding minimum certificate complexity of a function if we are able to reduce the sparsity by fixing not too many variables. We also find that there is a strong combinatorial property in the polynomial representation of a full degree function. More precisely, we consider the vertex-induced subgraph of the Boolean cube induced by the set of monomials in the polynomial representation (we call such a graph a monomial graph). For functions with full degree, there must be at least $n!/2^n$ monotone paths from $0^n$ to $1^n$ in the monomial graph of some translation of the function. On the other hand, we show that if the number of monotone paths from $0^n$ to level $\ell$ in the monomial graph is at most $s(f)^{\ell^c}$ for some constant $c > 0$ and some $\ell$ linear in $n$, then the Sensitivity Conjecture follows. We also prove a weaker bound, that the number of such paths is at most $\min\{2^{\ell/2}s(f)^{\ell}, (s(f)^4s(f)^{\ell})\}$ for all $\ell$.

Previous work and new contributions. Parts of the results in this thesis are from the previous work of the author. Specifically, the materials in Section 5 are mostly from the joint work of the author with Wong, Xie and Zhang in [TWXZ13], and have been published in FOCS’13 already. The main decision tree constructions in Section 4 was first introduced in [TWXZ13] for the case of parity decision trees. It was extended to the case of (standard) decision trees in [Tsa14]. Results in Section 6.1 (including, in particular, Conjecture 6.1 and Conjecture 6.2) are mostly presented in [Tsa14].

Materials in Section 6.2.2 and Section 6.2.3 were not published in any conferences or journals before, and they are the new contributions from this thesis. In Section 6.2.2, the definition of the mapping $\Psi_f$ (Definition 6.16 for the case of $F = \mathbb{R}$) is implicit in the work of [GKS15] and [L TZ17]. We explicitly define this notion for any field $F$ and introduce the monomial graphs (Definition 6.18). Other new contributions in this section include the relationship between the
number of monotone paths in $M_f^F$ and the degree versus sensitivity problem (Theorem 6.20); and
the conjecture on the number of monotone paths in $M_f^F$ and a weaker bound (Conjecture 6.21 and
Proposition 6.22).

Our materials on the polynomial sparsity in Section 6.2.3 are also the new contributions from
this thesis. They include in particular the upper bound on the polynomial sparsity in terms of
sensitivity (Theorem 6.28 and Corollary 6.32) and our conjecture on the polynomial sparsity
(Conjecture 6.35). We also observe that the “switching lemma for low-sensitivity functions”
in [GSTW16] confirms an average-case version of our conjecture on the polynomial sparsity
(Corollary 6.37).

Organization. In Section 3, we introduce the complexity measures that will be used throughout the thesis. The precise statement of the Log-rank Conjecture for XOR functions and the Sensitivity Conjecture will be stated in Section 3.4. In Section 4, we show how to use the polynomial representation to construct (parity) decision tree whose depth is bounded by the product of the minimum (parity) certificate complexity and the $F$-degree. In Section 5 we reproduce the argument in [TWXZ13] to prove the Log-rank Conjecture for XOR functions for the classes of functions with small spectral norm or low $F_2$-degree. Section 6 is devoted to our new formulation and sufficient condition for the Sensitivity Conjecture, and related results on polynomial sparsity.

2 Acknowledgements

I would like to thank my advisor Laci Babai for his encouragement and advice. I have learnt
a great deal from him, not only mathematics, but also many other things such as writing and
language. His creativity and hard work have given me a perfect model of the scientist that I
hope to be.

I also thank my other advisor Andy Drucker, who has been listening to my thoughts on the Sensitivity Conjecture for years, and was always able to give me helpful feedback. His constant encouragement has helped me a lot during the hard times. His kind support during the summers also allowed me to focus on research.

Finally, I want to thank Shengyu Zhang, who invited me to the world of theoretical computer science, when I was about to leave in the Chinese University of Hong Kong and wondered what to do next. The materials on the Log-rank Conjecture for XOR functions are the fruit of our first collaboration and are still some of my proudest works I have ever done. I hope that I can live up to his expectations and do work that he will be proud of.

3 Complexity measures

3.1 Representations over $F$ and algebraic measures

Let $f : \{0, 1\}^n \to \{0, 1\}$ and $F$ be a field. By interpolation one can find a polynomial $p \in F[x_1, \ldots, x_n]$ such that $p(x) = f(x)$ for all $x \in \{0, 1\}^n$. We can further assume $p$ to be multilinear since $b^2 = b$ for $b \in \{0, 1\}$. Specifically, $p(x)$ has the form

$$p(x) = \sum_{y \in f^{-1}(1)} \left( \prod_{i : y_i = 1} x_i \right) \left( \prod_{i : y_i = 0} (1 - x_i) \right) = \sum_{I \subseteq [n]} c_I x_I,$$

where $x_I = \prod_{i \in I} x_i$. We say that $x_I$ is a monomial of $p$ if $c_I \neq 0$. Moreover, we say that $p$ is the $F$-polynomial representation (or simply $F$-representation) of $f$. We may also abuse notation and use $f$ to denote its own polynomial representation when the underlying field is clear.

It is not difficult to show that such representation is unique. So we can use it to define complexity measures for a Boolean function.
**Definition 3.1** (F-degree). Let \( f : \{0,1\}^n \rightarrow \{0,1\} \) be a Boolean function and \( F \) be a field. The \( F\)-degree \( \deg_F(f) \) of \( f \) is the degree of the unique \( F\)-representation of \( f \). When \( F = \mathbb{R} \), we use \( \text{deg} \) to denote \( \deg_{\mathbb{R}} \).

We will need the following two definitions in Section 4 when we discuss how to design query algorithms using the polynomial representations.

**Definition 3.2** (Linear rank). The \( F\)-linear rank \( \text{lin-rank}_F(f) \) of \( f \) is the minimum \( r \) such that the \( F\)-representation of \( f \) can be written as
\[
f = \ell_1 f_1 + \ell_2 f_2 + \cdots + \ell_r f_r + f_0
\]
such that each \( \ell_i \) is a degree-1 polynomial and each \( f_j \) is a polynomial of degree less than \( \deg_F(f) \).

**Definition 3.3** (Dictator rank). The \( F\)-dictator rank \( \text{dic-rank}_F(f) \) of \( f \) is the minimum \( r \) such that the \( F\)-representation of \( f \) can be written as
\[
f = \ell_1 f_1 + \ell_2 f_2 + \cdots + \ell_r f_r + f_0
\]
such that each \( \ell_i \) is a dictator, i.e., \( \ell_i(x) = x_{k_i} \), and each \( f_j \) is a polynomial of degree less than \( \deg_F(f) \).

Note that the linear rank can be arbitrarily smaller than the dictator rank. For instance over \( \mathbb{F}_2 \), the polynomial \( f = (x_1 + \cdots + x_n)(x_{n+1} + \cdots + x_{2n}) \) has \( \text{lin-rank}_{\mathbb{F}_2}(f) = 1 \) but \( \text{dic-rank}_{\mathbb{F}_2}(f) = n \). Nevertheless, these two measures are closely related. Indeed for \( F = \mathbb{F}_2 \) (which is the case that we study in Section 4 and Section 5), for any \( f \) we have \( \text{lin-rank}_{\mathbb{F}_2}(f) = \min_{L} \{ \text{dic-rank}_{\mathbb{F}_2}(f \circ L) \} \) where the minimum is over all non-singular linear transformations \( L \) on \( \mathbb{F}_2^n \) and \( f \circ L \) is defined by \( f \circ L(x) = f(Lx) \). In other words, the linear rank is the minimum dictator rank over all non-singular linear transformations on the domain.

The following two measures are related to the study of the Sensitivity Conjecture. They will be the main measures discussed in Section 6.2.3.

**Definition 3.4** (Degree-\( \ell \) sparsity). Let \( p \in \mathbb{F}[x_1, \ldots, x_n] \), the degree-\( \ell \) sparsity \( \text{spar}_\ell(p) \) of \( p \) is the number of degree-\( \ell \) monomials in \( p \). Let \( f : \{0,1\}^n \rightarrow \{0,1\} \) be a Boolean function. The degree-\( \ell \) sparsity \( \text{spar}_\ell^F(f) \) of \( f \) over \( F \) is defined as the degree-\( \ell \) sparsity of its \( F\)-representation.

**Definition 3.5** (Polynomial sparsity). Under the same notation above, the polynomial sparsity \( \text{spar}(p) \) of \( p \) is the number of monomials in \( p \), i.e., \( \text{spar}(p) = \sum_{\ell=1}^n \text{spar}_\ell(p) \). The \( F\)- polynomial sparsity \( \text{spar}_\ell^F(f) \) of \( f \) is defined as the polynomial sparsity of its \( F\)-representation.

### 3.2 Fourier transform and Fourier analytic measures

For any function \( f : \{0,1\}^n \rightarrow \mathbb{R} \), the Fourier coefficients are defined by \( \hat{f}(s) = 2^{-n} \sum_x f(x) \chi_s(x) \), where \( \chi_s(x) = (-1)^{s \cdot x} \). Note that \( \langle \cdot , \cdot \rangle \) denote the usual inner product over \( \mathbb{R} \). The function \( f \) can be written as \( f = \sum_s \hat{f}(s) \chi_s \). The Fourier transform is closely related to polynomial representation over \( \mathbb{R} \). Specifically, for \( f : \{+1, -1\}^n \rightarrow \mathbb{R} \), its Fourier expansion is the same as its \( \mathbb{R}\)-polynomial representation.

The following two measures will be the main objects of study in Section 5.

**Definition 3.6** (Fourier sparsity). The Fourier sparsity \( \|\hat{f}\|_0 \) of \( f \) is the number of nonzero Fourier coefficients of \( f \).

**Definition 3.7** (Spectral norm). The spectral norm \( \|\hat{f}\|_1 \) of \( f \) is defined as \( \|\hat{f}\|_1 = \sum_s |\hat{f}(s)| \).

Note that \( \|\hat{f}\|_1 \) can be arbitrarily smaller than \( \|\hat{f}\|_0 \). For instance, the AND function has \( \|\hat{f}\|_1 = 1 \) but \( \|\hat{f}\|_0 = 2^n \).
For any function \( f : \{0,1\}^n \rightarrow \mathbb{R} \), Parseval’s Identity says that \( \sum_s \hat{f}(s)^2 = \mathbb{E}_x[f(x)^2] \). In particular, when the range of \( f \) is \( \{+1,-1\} \), we have \( \sum_s \hat{f}(s)^2 = 1 \). We sometimes use \( \hat{f} \) to denote the vector of the multiset \( \{\hat{f}(s) : s \in \{0,1\}^n\} \).

In some of the proofs in Section 5, we may need to consider the derivative of a function and deal with functions of the form \( f \cdot g \). The following simple fact tells us how the Fourier coefficients of \( fg \) are related to those of \( f \) and \( g \).

**Proposition 3.8** (Convolution). Let \( f,g : \{0,1\}^n \rightarrow \mathbb{R} \), the Fourier spectrum of \( fg \) is given by the following formula:

\[
\hat{f}g(s) = \sum_t \hat{f}(t)\hat{g}(s+t).
\]

Using this proposition, one can characterize the Fourier spectra of \( \{+1,-1\} \)-valued functions as follows.

**Proposition 3.9.** Let \( f : \{0,1\}^n \rightarrow \mathbb{R} \). Then \( |f(x)| = 1 \) for all \( x \in \{0,1\}^n \) if and only if

\[
\sum_{t \in \{0,1\}^n} \hat{f}^2(t) = 1, \quad \text{and} \quad \sum_{t \in \{0,1\}^n} \hat{f}(t)\hat{f}(s+t) = 0, \quad \forall s \in \{0,1\}^n \setminus \{0^n\}.
\]

Another fact that easily follows from the convolution formula is the following.

**Lemma 3.10.** Let \( f,g : \{0,1\}^n \rightarrow \mathbb{R} \), then \( ||\hat{f}g||_0 \leq ||\hat{f}||_0||\hat{g}||_0 \) and \( ||\hat{f}g||_1 \leq ||\hat{f}||_1||\hat{g}||_1 \).

### 3.3 Computational models and combinatorial measures

#### 3.3.1 Computational models

A **decision tree** is a rooted binary tree \( T \) such that each internal node is labeled by a variable \( x_i \) for some \( i \in [n] \) and each leaf is labeled by either 0 or 1. Given \( x \in \{0,1\}^n \), \( T \) computes \( T(x) \) as follows: start at the root, if it is a leaf then output the value of this leaf, otherwise query the variable labeling the current node. Recurse on the left and right subtree if the outcome of the query is 0 and 1 respectively. We say that \( T \) computes \( f \) if \( T(x) = f(x) \) for all \( x \in \{0,1\}^n \).

A **parity decision tree** is defined in the same way except now each internal node can be labeled by an arbitrary \( \mathbb{F}_2 \)-linear function on \( x \) instead of just a variable. Then query of the node is a black-box operations that evaluates the linear function labeling the current node on the input \( x \).

**Definition 3.11** (Query complexity). The **query complexity** \( D(f) \) of \( f \) is the minimum depth of any decision tree that computes \( f \).

**Definition 3.12** (Parity query complexity). The **parity query complexity** \( D_{\oplus}(f) \) of \( f \) is the minimum depth of any parity decision tree that computes \( f \).

Another important computational model in concrete complexity is the (deterministic) communication complexity, introduced in the seminal paper by Yao [Yao79]. Let \( F : X \times Y \rightarrow \{0,1\} \). A computational protocol for (computing) \( F \) can be described as follows: there are two parties Alice and Bob, each holds part of the input. Specifically, Alice has the input \( x \in X \) and Bob has the input \( y \in Y \). Their goal is to compute the value \( F(x,y) \). The protocol proceeds by sending messages (i.e., bit strings) between the two parties in turns, each time only one party can send a message to the other party and the message he/she sends can only depend on all previous messages and the part of the input they have. At the end of the protocol they output a value that is equal to \( F(x,y) \). The cost of the protocol is the maximum number of bits they sent during the protocol over all inputs \( (x,y) \). We refer to the textbook of Kushilevitz and Nisan [KN97] on communication complexity for a more formal definition and many results on communication complexity.

**Definition 3.13** (Deterministic communication complexity). The **deterministic communication complexity** \( CC(F) \) of \( F \) is defined as the minimum cost over all protocols that computes \( F \).
We will focus on communication complexity of a special class of functions that is called the XOR functions. A function \( F(x, y) \) is said to be a XOR function [ZS09] if there exists a Boolean function \( f \) such that \( F(x, y) = f(x \oplus y) \). We use \( CC_{\oplus}(f) \) to denote \( CC(F) \) for XOR functions.

### 3.3.2 Combinatorial measures

We first introduce some notation. Let \( x \in \{0, 1\}^n \), and \( B \subseteq [n] \). We denote by \( x^B \) the string obtained from \( x \) by flipping all the bits in \( B \), i.e., \( (x^B)_i = x_i \) if \( i \not\in B \) and \( (x^B)_i = 1 - x_i \) otherwise. When \( B = \{i\} \) is a singleton, we use \( x^i \) to denote \( x^{\{i\}} \).

**Definition 3.14** (Sensitivity). We say that \( f \) is sensitive to the \( i \)-th variable (or coordinate or \( x_i \)) on \( x \) if \( f(x) \neq f(x^i) \). We also say that \( i \) is a sensitive coordinate of \( f \) on \( x \). The sensitivity \( s(f, x) \) on \( x \) is the number of sensitive coordinates on \( x \). Let \( b \in \{0, 1\} \), the \( b \)-sensitivity \( s^b(f) \) is defined as \( \max_{x \in f^{-1}(b)} \{s(f, x)\} \) and the sensitivity \( s(f) \) of \( f \) is defined as \( \max \{s^0(f), s^1(f)\} \).

**Definition 3.15** (Block sensitivity). Let \( B \subseteq [n] \). We say that \( f \) is sensitive to the block \( B \) on \( x \) if \( f(x) \neq f(x^B) \). We also say that \( B \) is a sensitive block on \( x \). The block sensitivity \( bs(f, x) \) on \( x \) is the maximum number of disjoint sensitive blocks on \( x \), and the block sensitivity \( bs(f) \) of \( f \) is defined as \( \max_{x \in \{0, 1\}^n} bs(f, x) \).

**Definition 3.16** (\( \ell \)-block sensitivity). Let \( \ell \in [n] \). The \( \ell \)-block sensitivity \( bs_\ell(f, x) \) on \( x \) is the maximum number of disjoint sensitivity block on \( x \) such that each sensitivity block has size at most \( \ell \). The \( \ell \)-block sensitivity \( bs_\ell(f) \) of \( f \) is defined as \( \max_{x \in \{0, 1\}^n} bs_\ell(f, x) \).

Note that 1-block sensitivity is the same as sensitivity.

**Definition 3.17** (Certificate complexity). We say that a subcube \( Q \supseteq x \) is a certificate of \( f \) on \( x \) if \( f \) takes constant value \( f(x) \) on \( Q \), i.e. \( f(y) = f(x) \) for all \( y \in Q \). The certificate complexity \( C(f, x) \) on \( x \) is the minimum codimension of any certificate on \( x \). The certificate complexity \( C^b(f) \) is defined as \( \max_{x \in f^{-1}(b)} C(f, x) \), and the certificate complexity \( C(f) \) of \( f \) is defined as \( \max \{C^0(f), C^1(f)\} \).

**Definition 3.18** (Parity certificate complexity). We say that an affine subspace \( H \supseteq x \) is a parity certificate of \( f \) on \( x \) if \( f \) takes constant value \( f(x) \) on \( H \), i.e. \( f(y) = f(x) \) for all \( y \in H \). The parity certificate complexity \( C_{\oplus}(f, x) \) on \( x \) is the minimum codimension of any parity certificate on \( x \). The parity certificate complexity \( C_{\oplus}(f) \) of \( f \) is defined as \( \max_{x \in f^{-1}(\{0, 1\})} C_{\oplus}(f, x) \), and the parity certificate complexity \( C_{\oplus}(f) \) of \( f \) is defined as \( \max \{C_{\oplus}^0(f), C_{\oplus}^1(f)\} \).

### 3.3.3 Minimum measures

In the definition of the combinatorial measures above, we first define a local measure on each input \( x \) and then take the maximum over all inputs. Instead of taking the maximum, one can take the minimum over all inputs. We refer to the resulting classes of measures as the “minimum measures”. Here are the precise definitions of them.

**Definition 3.19** (Minimum block sensitivity). The minimum block sensitivity \( bs_{\min}(f) \) of \( f \) is defined as \( bs_{\min}(f) = \min_{x \in \{0, 1\}^n} bs(f, x) \).

**Definition 3.20** (Minimum certificate complexity). The minimum certificate complexity \( C_{\min}(f) \) of \( f \) is defined as \( C_{\min}(f) = \min_{x \in \{0, 1\}^n} C(f, x) \).

**Definition 3.21** (Minimum parity certificate complexity). For \( b \in \{0, 1\} \), the \( b \)-minimum parity certificate complexity of \( f \) is defined as \( C_{\min}^b(f) = \min_{x \in f^{-1}(b)} C_{\oplus}(f) \). The minimum parity certificate complexity \( C_{\min}(f) \) of \( f \) is defined as \( C_{\min}(f) = \min \{C_{\oplus}^0(f), C_{\oplus}^1(f)\} \).

In general these minimum measures can be arbitrarily smaller than their maximum counterparts. For instance, the AND function has \( bs(f) = C(f) = C_{\oplus}(f) = n \), but for the minimum measures we have \( bs_{\min}(f) = C_{\min}(f) = C_{\oplus, \min}(f) = 1 \).
3.4 The main conjectures

It turns out that for many of the combinatorial measures we introduced, they are either polynomially related or one can be arbitrarily smaller than the other (we refer the reader to the survey of Buhrman and de Wolf [BdW02] for many results on the relationships among query complexity, block sensitivity, certificate complexity and more). However, it is not known whether the sensitivity is polynomially related to block sensitivity (and hence all other measures that are polynomially related to block sensitivity). The Sensitivity Conjecture asserts that they are indeed polynomially related.

**Conjecture 3.22 (Sensitivity Conjecture [NS94]).** There exists $c > 0$ such that for all $f : \{0, 1\}^n \to \{0, 1\}$, $bs(f) \leq (s(f))^c$.

On the other hand, for the relationship between the Fourier analytic measures we introduced and the combinatorial measures, it is not difficult to show that both $\log \|f\|_0$ and $\log \|f\|_1$ can be upper bounded by $D_G(f)$, which in turn can be arbitrarily smaller than its non-parity counterpart $D(f)$. From this it also follows that both $\log \|f\|_0$ and $\log \|f\|_1$ can be arbitrarily smaller than $bs(f), C(f), D(f)$ and $\deg(f)$. However, it is not known whether $\log \|f\|_0$ and $D_G(f)$ are polynomially related. It turns out that this problem is closely related to the Log-rank Conjecture for XOR functions, which we formally state below.

**Conjecture 3.23 (Log-rank Conjecture for XOR functions [ZS09]).** There exists $c > 0$ such that for all $F(x, y) = f(x \oplus y)$ where $f : \{0, 1\}^n \to \{0, 1\}$, $CC(F) \leq O(\log^c \text{rank}(M_F))$.

By the definition of XOR functions, it is straightforward to show that $\text{rank}(M_F) = \|f\|_0$. Combining this with the fact that $CC_G(f) \leq 2D_G(f)$ for all $f$. We obtain the following sufficient condition for the Log-rank Conjecture for XOR functions.

**Conjecture 3.24 ([ZS09]).** There exists $c > 0$ such that for all $f : \{0, 1\}^n \to \{0, 1\}$, $D_G(f) \leq O(\log^c \|f\|_0)$.

In fact more is true. This conjecture is not only a sufficient condition for Conjecture 3.23, but is also necessary, i.e., Conjecture 3.24 and Conjecture 3.23 are equivalent. We will discuss it in more detail in Section 5.4.

4 Query algorithms from polynomial representations: degree reduction

We have seen that both the Log-rank Conjecture and the Sensitivity Conjecture can be reduced to the problem of constructing low-depth (parity) decision trees. In this section we show how one can use the structure of the polynomial representations to construct decision trees with depth upper bounded by the product of the minimum certificate complexity and $F$-degree.

The main idea of our query algorithms is to reduce the degree of the function. Let $f$ be a Boolean function. We think it as a polynomial over some field $F$ and write it in the form

$$f = \ell_1f_1 + \cdots + \ell_rf_r + f_0,$$

where each of the $f_i$ has degree strictly less than $\deg_{F}(f)$ and each $\ell_i$ is a linear function over $F$. It is easy to see that if we know the value of each $\ell_1(x), \ldots, \ell_r(x)$ and fix them, then $f$ reduces to a polynomial of strictly smaller degree. We can then repeat this procedure until the degree of the function is less than 1, in which case we know the value of $f(x)$. If we require each $\ell_i$ to be a variable, then we get a decision tree. When $F = \mathbb{F}_2$ and each $\ell_i$ can be any linear function over $\mathbb{F}_2$, the resulting query algorithm can be represented by a parity decision tree. It follows directly by the construction of the query algorithms and the definitions of dictator rank and linear rank that $D(f) \leq \max_Q \{\text{dic-rank}_F(f|Q)\} \deg_{F}(f)$ and $D_G(f) \leq \max_H \{\text{lin-rank}_F(f|H)\} \deg_{F}(f)$. It is not clear to us how to directly upper bound $\text{dic-rank}_F(f)$ and $\text{lin-rank}_F(f)$ in terms of $s(f)$.
and $\log \|\hat{f}\|_0$ respectively, but fortunately we can upper bound $\text{dic-rank}_F$ and $\text{lin-rank}_F$ by the minimum certificate complexities, which in some sense are easier to deal with.

Remark. Note that we may also consider the case with an arbitrary finite field $\mathbb{F}$ and the functions $\ell_i$ being arbitrary linear functions over $\mathbb{F}$, but in this case we no longer have a parity decision tree. The tree we get can have degree up to $|\mathbb{F}|$ at each internal node and will be able to query a linear function of $x$ over the field $\mathbb{F}$. The case of arbitrary $\mathbb{F}$ also introduces new difficulties in our arguments in Section 5. More precisely, let $b \in \mathbb{F}$ and $\ell$ be a linear function over $\mathbb{F}$, the set of $x \in \{0,1\}^n$ satisfying $\ell(x) = b$ is not necessarily a coset in $\mathbb{F}_2$, in which case we are not able to consider the Fourier transform of the restricted function.

Lemma 4.1. For all $f : \mathbb{F}_2^n \rightarrow \{0,1\}$, we have

1. [Tsa14] $\text{dic-rank}_F(f) \leq C_{\min}(f)$ for any field $\mathbb{F}$.

2. [TWXZ13] $\text{lin-rank}_F(f) \leq C_{\oplus,\min}(f)$.

Proof. Let $k = C_{\min}(f)$ and $Q$ be the subcube of codimension $k$ such that $f|_Q$ is identically constant. Without loss of generality, assume $Q = \{x \in \{0,1\}^n \mid x_i = b_i, \forall i \in [k]\}$ where $b_i \in \{0,1\}$. Now we write $f$ as

$$f = x_1 f_1 + \cdots + x_k f_k + f',$$

where $\deg_G(f_i) < \deg_G(f)$ for all $i \in [k]$ and $f'$ does not depend on variables $x_1, \ldots, x_k$. By definition of $Q$ and $f'$, it follows that if we restrict $f$ on $Q$, then we have

$$f|_Q = b_1(f_1|_Q) + \cdots + b_k(f_k|_Q) + f'.$$

Since $f|_Q$ is constant and each of the $f_i|_Q$ has degree less than $\deg_G(f)$, it follows that $f'$ must have degree less than $\deg_G(f)$ as well. By definition of dictator rank, we have $\text{dic-rank}(f) \leq C_{\min}(f)$.

To prove the second item, recall that when $\mathbb{F} = \mathbb{F}_2$, $\text{lin-rank}_{\mathbb{F}_2}(f) = \min_L \{\text{dic-rank}_{\mathbb{F}_2}(f \circ L)\}$ and $C_{\oplus,\min}(f) = \min_L \{C_{\min}(f \circ L)\}$ where the minima are over all nonsingular linear maps $L$ from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$. From this we can complete the proof by observing $\text{lin-rank}_{\mathbb{F}_2}(f) \leq \text{dic-rank}_{\mathbb{F}_2}(f \circ L) \leq C_{\min}(f \circ L)$ for all $L$. In particular, we have $\text{lin-rank}_{\mathbb{F}_2}(f) \leq \min_L \{C_{\min}(f \circ L)\} = C_{\oplus,\min}(f)$.

By this lemma, we have the following theorem relating (parity) query complexity, ($\mathbb{F}_2$) $\mathbb{F}$-degree and minimum (parity) certificate complexity.

Theorem 4.2. For all non-constant $f : \mathbb{F}_2^n \rightarrow \{0,1\}$, we have

1. [Tsa14] $D(f) \leq \max_Q \{\text{dic-rank}(f|_Q)\} \deg_P(f) \leq \max_Q \{C_{\min}(f|_Q)\} \deg_G(f)$ for any field $\mathbb{F}$, and

2. [TWXZ13] $D_\oplus(f) \leq \max_H \{\text{lin-rank}_{\mathbb{F}_2}(f|_H)\} \deg_{\mathbb{F}_2}(f) \leq \max_H \{C_{\oplus,\min}(f|_H)\} \deg_G(f)$,

where the maximum is over all subcubes $Q$ and affine subspaces $H$ respectively.

5 The Log-rank Conjecture for XOR functions

By the query algorithm we discussed in Section 4, to prove the Log-rank Conjecture for XOR functions it suffices to show that for some small field $\mathbb{F}$, there exists a constant $c$ such that both $\deg_G(f)$ and $\text{lin-rank}_G(f)$ are at most $\log^c \|\hat{f}\|_0$.

Note that $\deg(f)$ can be arbitrarily larger than $\log \|\hat{f}\|_0$. So in order to have any upper bound on $\deg_G(f)$ in terms of $\log \|\hat{f}\|_0$, we must consider fields other than $\mathbb{R}$. In general the degree of the polynomial representations can be very different under different fields. It was shown by Gopalan, Lovett and Shpilka that if $\deg_G(f)$ is small, then $\deg_G(f)$ is large for primes $q \neq p$. Formally, they proved the following theorem.
Theorem 5.1 ([GLS09]). Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function which depends on all $n$ variables. Let $p \neq q$ be distinct primes. Then
\[
\deg_{F_2}(f) \geq \frac{n}{\log_2 p} \deg_{F_2}(f)p^{2 \deg_{F_2}(f)}.
\]

Nevertheless, it can be shown that the $F_2$-degree is indeed upper bounded by $\log_2 \|\hat{f}\|_0$.

Proposition 5.2 ([BC99]). For all $f : \{0,1\}^n \to \{0,1\}$, $\deg_{F_2}(f) \leq \log_2 \|\hat{f}\|_0$.

From now on we will restrict our attention to the case $\mathbb{F} = \mathbb{F}_2$ and use lin-rank to denote lin-ranks in $\mathbb{F}_2$. Under this setting the second item in Theorem 4.2 applies, and the only problem left is to upper bound $C_{\oplus,\min}(f)$ in terms of $\log \|\hat{f}\|_1$. However, there is no known polynomial upper bound on $C_{\oplus,\min}(f)$ in terms of $\log \|\hat{f}\|_0$ for general $f$. In this section we prove that (i) it is possible to upper bound $C_{\oplus,\min}(f)$ by the spectral norm $\|\hat{f}\|_1$ for every Boolean function $f$ (note that the $\|\hat{f}\|_1$ can be arbitrarily smaller than $\|\hat{f}\|_0$; and (ii) for functions with constant $F_2$-degree $d \geq 1$, $C_{\oplus,\min}(f)$ can be upper bounded by $O(\log^c \|\hat{f}\|_0)$ where $c$ is a constant that depends only on $d$.

Theorem 5.3 ([TWXZ13]). $C_{\oplus,\min}(f) = O(\|\hat{f}\|_1)$.

Theorem 5.4 ([TWXZ13]). $C_{\oplus,\min}(f) = O(2^{dF/2} \log^{\max\{d-2,1\}} \|\hat{f}\|_1)$, where $d = \deg_{F_2}(f) \geq 1$.

They imply in particular that the Log-rank Conjecture for XOR functions is true for functions with small spectral norm or low $F_2$-degree.

Both theorems assert that there is a large affine subspace on which $f$ is a constant. We will find such an affine subspace iteratively. In each iteration we start with an affine subspace $H$ and we show that we can further restrict $H$ to a new affine subspace $H'$ such that $f$ on $H'$ is “closer” to a constant function in terms of some Fourier analytic measures. It is hence important, to first understand how the Fourier spectrum changes under taking linear restrictions, before we get into the details of the proofs.

5.1 Restriction to affine subspaces

We first understand how the Fourier spectrum changes under restricting a variable $x_i$. Let $f : \{0,1\}^n \to \{0,1\}$, by the Fourier expansion, we can write $f$ as
\[
f(x) = \sum_s \hat{f}(s) \chi_s(x) = \sum_{s : x_i = 0} \left(\hat{f}(s) + \hat{f}(s + e_i) \chi_i(x)\right) \chi_s(x).
\]
By fixing $x_i = b$, we restrict the function to the subcube $Q = \{x \in \{0,1\}^n \mid x_i = b\}$. The resulting function $f|_Q$ will have Fourier expansion
\[
f|_Q(x) = \sum_{s : x_i = 0} \left(\hat{f}(s) + (-1)^b \hat{f}(s + e_i)\right) \chi_s(x).
\]
It follows that by restricting a variable, the Fourier coefficients $\hat{f}(s)$ and $\hat{f}(s + e_i)$ collapse to a single Fourier coefficient of $f|_Q$ with value $\hat{f}(s) + (-1)^b \hat{f}(s + e_i)$.

In fact similar phenomenon occurs when we restrict the function to an affine subspace of codimension 1. In this case we consider a character $\chi_\beta$ which maps $x$ to $(-1)^{\langle \beta, x \rangle}$, and let $H = \{x \in \{0,1\}^n \mid \chi_\beta(x) = b\}$ where $b \in \{+1, -1\}$. Then $H$ is an affine subspace in $F_2^n$ (under the natural identification between $\{0,1\}^n$ and $F_2^n$) of codimension 1. It can be shown that (cf. [OD14], Chapter 3, Section 3.3) in the function $f|_H$, the Fourier coefficients $\hat{f}(s)$ and $\hat{f}(s + \beta)$ collapse to a single Fourier coefficient of $f|_H$ with value $\hat{f}(s) + (-1)^b \hat{f}(s + \beta)$. In particular, it follows that both the Fourier sparsity and spectral norm is non-increasing under restriction to affine subspace.
5.2 Functions with small spectral norm

We will prove Theorem 5.3 in this subsection. The high level idea of the proof is to keep increasing the largest Fourier coefficient by choosing an appropriate linear restriction to collapse the largest and second largest Fourier coefficients. For functions with range \{+1, -1\}, once we have a Fourier coefficient with absolute value 1, we can conclude that the function is a parity on some input variables and hence with at most one more linear restriction we can make it constant.

Proof of Theorem 5.3. Our goal is to find an affine subspace \(H\) of codimension \(O(\|\hat{f}\|_1)\) such that \(f\) is constant on \(H\). We will assume \(f\) is from \(\{0,1\}^n\) to \{+1, -1\} (it is without loss of generality since we can replace \(f\) by \(1 - 2f\), and the spectral norm only increases by a factor of two), in which case the Parseval’s Identity asserts that \(\sum_s f(s)^2 = 1\).

Suppose \(f\) is non-constant. Let \(s = \|\hat{f}\|_0\) and let \(\{\alpha_1, \ldots, \alpha_s\}\) be the set of non-zero Fourier coefficients such that \(|\hat{f}(\alpha_i)| \geq |\hat{f}(\alpha_j)|\) for \(i \leq j\). Let \(a_i = |\hat{f}(\alpha_i)|\). Now consider the following greedy restriction process: impose a linear restriction to collapse the two largest Fourier coefficients such that \(f\) becomes non-constant. Let \(b\) be the subfunction has its largest Fourier coefficient being \(a_1 + a_2\) in absolute value.

By Parseval’s Identity, we have

\[
1 - a_1^2 = \sum_{i \geq 2} a_i^2 \leq a_2 \sum_{i \geq 2} a_i = a_2(\|\hat{f}\|_1 - a_1).
\]

So when \(a_1 \leq 1/2\), the greedy restriction increases the largest Fourier coefficient by

\[
a_2 \geq \frac{1 - a_1^2}{\|\hat{f}\|_1 - a_1} > \frac{3}{4\|\hat{f}\|_1}.
\]

Hence the largest coefficients would be larger than \(1/2\) in \(O(\|\hat{f}\|_1)\) steps.

Next we show that the greedy restriction decreases the spectral norm by at least \(2a_1 = 2\max_s |\hat{f}(s)|\). Define

\[
P_+(\beta) = \{(i,j) \in [s]^2 \mid i < j, \alpha_i + \alpha_j = \beta, \hat{f}(\alpha_i) \cdot \hat{f}(\alpha_j) = 1\}
\]

and

\[
P_-(\beta) = \{(i,j) \in [s]^2 \mid i < j, \alpha_i + \alpha_j = \beta, \hat{f}(\alpha_i) \cdot \hat{f}(\alpha_j) = -1\}
\]

Considering the old and new Fourier spectra, it is easy to see that the drop of the spectral norm is

\[
2 \sum_{(i,j) \in P_-(\beta)} \min\{|\hat{f}(\alpha_i)|, |\hat{f}(\alpha_j)|\}.
\]

Note that the restriction is chosen such that the largest two Fourier coefficients have the same sign, so in particular we have \((1,2) \in P_+(\beta)\). By Proposition 3.9, we have \(\sum_{\alpha_i + \alpha_j = \beta} \hat{f}(\alpha_i) \hat{f}(\alpha_j) = 0\), thus \(\sum_{(i,j) \in P_+(\beta)} a_ia_j = \sum_{(i,j) \in P_-(\beta)} a_ia_j\). Now we have

\[
a_1a_2 \leq \sum_{(i,j) \in P_+(\beta)} a_ia_j = \sum_{(i,j) \in P_-(\beta)} a_ia_j \leq a_3 \sum_{(i,j) \in P_-(\beta)} \min\{a_i, a_j\}.
\]

Therefore, the decrease of the spectral norm is at least \(2a_1 \cdot \frac{a_2}{a_3} \geq 2a_1\). Note that by construction, the greedy restriction only increases the largest Fourier coefficients. Thus once \(a_1 > 1/2\), each greedy restriction decreases the spectral norm by at least 1. So it takes at most \(\|\hat{f}\|_1\) further steps to make the spectral norm to be at most 1, in which case at most one more restriction makes the function constant. Note that \(\|\hat{f}\|_1 \geq 1\) for non-constant \(f\), so the total number of restrictions we made is \(O(\|\hat{f}\|_1)\).

\(\square\)
5.3 Functions with low \( \mathbb{F}_2 \)-degree

We will prove Theorem 5.4 in this section. The proof is by induction on \( \deg_{\mathbb{F}_2} \). The case for degree 1 is trivial and the case for degree 2 is due to the following theorem of Dickson [Dic58].

**Theorem 5.5.** Let \( A \in \{0, 1\}^{n \times n} \) be a symmetric matrix whose diagonal entries are all 0, and define a polynomial \( f(x) = x^T Q x + \ell(x) + \varepsilon \), where \( Q \) is the upper triangle part of \( A \). Then \( \text{lin-rank}(f) \) is equal to the rank of matrix \( A \) over \( \mathbb{F}_2 \).

Note that Dickson’s theorem says that, up to an affine (invertible) linear map, the Fourier spectrum of a degree 2 polynomial is identical to a function on \( \{0, 1\}^n \). This means that \( \hat{f}(x) \) is a subspace of co-dimension at most 4. So it follows that \( C_{\ominus, \min}(f) = \log \| \hat{f} \|_1 + 1 \).

5.3.1 Cubic polynomials

The degree 3 case will serve as the base case in our induction proof of Theorem 5.4.

First we need a lemma that relates the linear rank of a cubic polynomial and the linear ranks of its derivatives. For a Boolean function \( f \), its derivative along direction \( t \in \{0, 1\}^n \) is defined as \( \Delta_t f(x) = f(x) + f(x + t) \) where the addition here is over \( \mathbb{F}_2 \). The following statement is slightly more general than Lemma 3.7 in [HS10], but the same proof goes through.

**Lemma 5.6.** [HS10], Lemma 3.7. Let \( M \) be a collection of quadratic functions satisfying \( \text{lin-rank}(f) \leq r \) for all \( M \cup 2M \) (where \( 2M = \{ f_1 + f_2 \mid f_1, f_2 \in M \} \)), then there is a subspace \( V \) of co-dimension at most \( 4r \) such that \( f|_V \) is a linear function for all \( f \in M \).

Now we can prove the degree 3 case of Theorem 5.4.

**Theorem 5.7.** ([TWXZ13]). For all \( f : \{0, 1\}^n \rightarrow \{0, 1\} \) with \( \deg_{\mathbb{F}_2}(f) = 3 \), it holds that \( \text{lin-rank}(f) = O(\log \| \hat{f} \|_1) \).

**Proof.** Note that \( \Delta_t f \) has \( \mathbb{F}_2 \)-degree at most 2 for all \( t \), and that in general \( \Delta_t f + \Delta_s f = \Delta_{t+s} f + \Delta_t \Delta_s f \). Let \( M \) be the collection of derivative with \( \mathbb{F}_2 \)-degree 2, i.e., \( M = \{ \Delta_t f \mid t \in \{0, 1\}^n, \deg_{\mathbb{F}_2}(\Delta_t f) = 2 \} \), then \( M \) satisfies the condition of Lemma 5.6. Furthermore, each \( \Delta_t f \in M \) has

\[
\text{lin-rank}(\Delta_t f) \leq \log \| \Delta_t \hat{f} \|_1 + 1 \leq 2 \log \| \hat{f} \|_1 + 1,
\]

where the last inequality is because of Lemma 3.10. Let \( r = 2 \log \| \hat{f} \|_1 + 1 \). Now by Lemma 5.6, we know that \( 4r \) restrictions can make all \( \Delta_t f \) in \( M \) to become linear functions. Therefore there is a subspace of co-dimension at most \( 4r \) restricted on which \( \Delta_t f \) are linear functions, for all \( t \in \{0, 1\}^n \). This means that \( f|_V \) has degree at most 2. It follows that \( \text{lin-rank}(f) \leq 4r = O(\log \| \hat{f} \|_1) \).

Using this theorem we can complete the proof of the degree-3 case in Theorem 5.4. Specifically, by Theorem 5.7, in \( O(\log \| \hat{f} \|_1) \) restrictions \( f \) becomes a degree at most 2 polynomial \( f' \). If \( f' \) is constant function then we are done. If not then by the degree 2 or 1 case in Theorem 5.4, it takes at most \( \log \| \hat{f} \|_1 + 1 \leq \log \| \hat{f} \|_1 + 1 \) (since \( f' \) is a subfunction of \( f \)) more restrictions to make the function constant.

5.3.2 Constant-degree polynomials

Now we are ready to prove the general case of Theorem 5.4. We prove the following slightly stronger version. Note that sometimes in the proof we change the range of a Boolean function \( f \) from \( \{0, 1\} \) to \( \{+1, -1\} \), and denote the resulting function as \( f^\pm \). It can be done by letting \( f^\pm = 1 - 2f \). From this it is also easily seen that the spectral norm is increased by at most a factor of 2 and the Fourier sparsity is increased by at most 1 additively. Note that we have...
use the notation $f + g$ for addition of $f$ and $g$ over $\mathbb{F}_2$ before (for instance when we define the derivative). In the following proof we will consider both addition over $\mathbb{F}_2$ and over $\mathbb{R}$. For avoid confusion, we use $\oplus$ for addition over $\mathbb{F}_2$ and $+$ for addition over $\mathbb{R}$.

**Theorem 5.8** ([TWXZ13]). For any non-constant function $f : \{0, 1\}^n \to \{0, 1\}$, we have

\[
\text{lin-rank}(f) \leq C_{\oplus, \min}(f) \leq D_\oplus(f) \leq O(2^{d^2/2} \log^{\max(d-2,1)} \|f\|_1).
\]

**Proof.** We will prove by induction on $\mathbb{F}_2$-degree $d$ that

\[
\text{lin-rank}(f) \leq C_{\oplus, \min}(f) \leq \max_{b \in \{0, 1\}} C_{\oplus, \min}^b(f) \leq D_\oplus(f) \leq B_d(\|f^{\pm}\|_1),
\]

where $B_d(m) < 2^{d/2} \log^{d-2} m$ is some non-decreasing (in both $d$ and $m$) function to be determined later.

We have already proved the claim for $d \leq 3$ in the previous subsection. Now suppose that the bound holds for all polynomials of $\mathbb{F}_2$-degree at most $d - 1$, and consider a function $f$ of degree $d \geq 4$. We will first prove a bound for $C_{\oplus, \min}(f)$, which also bounds $\text{lin-rank}(f)$ from above by **Lemma 4.1**.

First, it is not hard to see that there exists a direction $t \in \{0, 1\}^n - \{0^n\}$ such that $\Delta_tf$ is non-constant (unless $f$ is a linear function, in which case the conclusion holds trivially). Fix such $t$. Since $\deg_{\mathbb{F}_2}(\Delta_tf) \leq d - 1$, by induction hypothesis, it holds that for each $b \in \{0, 1\}$,

\[
C_{\oplus, \min}^b(\Delta_tf) \leq B_{d-1}(\|\Delta_tf^{\pm}\|_1).
\]

Define $f_t(x) = f(x \oplus t)$, then by **Lemma 3.10**, we have

\[
\|(\Delta_tf^{\pm})^{\pm}\|_1 = \|f^{\pm} \cdot f^+_{t}^{\pm}\|_1 \leq \|f^{\pm}\|_1 \|f^+_{t}^{\pm}\|_1 = \|f^{\pm}\|_1^2,
\]

which implies that

\[
C_{\oplus, \min}^b(\Delta_tf) \leq B_{d-1}(\|f^{\pm}\|_1^2).
\]

Since $\Delta_tf$ is non-constant, $C_{\oplus, \min}^b(\Delta_tf) \geq 1$ for both $b = 0$ and $1$. For each $b$, by the definition of $C_{\oplus, \min}^b(\Delta_tf)$, there exist affine subspaces $H_b$ with $\text{codim}(H_b) \leq B_{d-1}(\|f^{\pm}\|_1^2)$ such that $(\Delta_tf)|_{H_b} = b$, which is equivalent to $f(x) \oplus f(x \oplus t) = b$ for all $x \in H_b$.

Define

\[
g_0(x) = \frac{1}{2}(f^+(x) + f^+(x \oplus t)), \quad g_1(x) = \frac{1}{2}(f^+(x) - f^+(x \oplus t)).
\]

These two functions have some important properties. First, it is easy to see from the definition of $g_0$ and $g_1$ that $f^\pm = g_0 + g_1$. Second, note that $g_0$ and $g_1$ are not Boolean functions any more (they take values in $\{-1, 0, +1\}$), but they take very special values on the affine space $H_b$: one always takes value 0, and the other always take values in $\pm\{0, 1\}$. In fact it is not hard to verify that

\[
g_b|_{H_b} = f^\pm|_{H_b} \quad \text{and} \quad g_0|_{H_b} = 0.
\]

Third, note that

\[
\hat{f}^\pm_t(s) = \mathbb{E}_x[f^\pm(x \oplus t)\chi_s(x)] = \mathbb{E}_x[f^\pm(x \oplus t)\chi_s(x \oplus t)] = \hat{f}^\pm(s)\chi_s(t),
\]

and thus

\[
\hat{g}_b(s) = \frac{1}{2}(\hat{f}^\pm(s) + (-1)^b\hat{f}^\pm_t(s)) = \frac{1}{2}(\hat{f}^\pm(s) + (-1)^b\chi_s(t)\hat{f}^\pm(s)).
\]

Therefore, we have

\[
\hat{g}_0(s) = \begin{cases} 
\hat{f}^\pm(s) & s \in t^\perp \\
0 & s \in t^\perp
\end{cases}, \quad \text{and} \quad \hat{g}_1(s) = \begin{cases} 
0 & s \in t^\perp \\
\hat{f}^\pm(s) & s \in t^\perp.
\end{cases}
\]
where \( t^\perp = \{ s \in \{0,1\}^n \mid \chi_1(s) = 1 \} \). Namely \( \hat{g}_0 \) and \( \hat{g}_1 \) each takes half-space of the spectrum.

This further implies that
\[
\| \hat{f}^\pm \|_1 = \| \hat{g}_0 \|_1 + \| \hat{g}_1 \|_1.
\]

Thus, either \( \| \hat{g}_0 \|_1 \) or \( \| \hat{g}_1 \|_1 \) is at most half of \( \| \hat{f}^\pm \|_1 \). Suppose that \( \| \hat{g}_0 \|_1 \leq \frac{1}{2} \| \hat{f}^\pm \|_1 \). We claim that restricting \( f^\pm \) to \( H_b \) reduces its spectral norm a lot. Indeed, since \( f^\pm|_{H_b} = g_b|_{H_b} \), we have
\[
\| \hat{f}^\pm|_{H_b} \|_1 = \| g_b|_{H_b} \|_1 \leq \| \hat{g}_b \|_1 \leq \frac{1}{2} \| \hat{f} \|_1,
\]

where the first inequality is because of the fact that spectral norm is non-increasing under restriction. To summarize, we have just shown that we can reduce the spectral norm by at least half using at most \( B_{d-1}(\| \hat{f}^\pm \|_1^2) \) linear restrictions.

Now we recursively apply the above steps on the subfunction \( f^\pm|_{H_b} \) until finally we find an affine subspace \( H \) such that \( \| \hat{f}^\pm|_H \|_1 \leq 1 \), at which moment the subfunction is either a constant or a linear function, thus at most one more restriction would give a constant function. In total it takes at most \( B_{d-1}(\| \hat{f}^\pm \|_1^2) \log \| \hat{f}^\pm \|_1 + 1 \) linear restrictions to get a constant function, which implies that
\[
C_{\oplus, \min}(f) \leq B_{d-1}(\| \hat{f}^\pm \|_1^2) \log \| \hat{f}^\pm \|_1 + 1.
\]

Note that by the query algorithm that we used to prove Theorem 4.2, we know that
\[
D_{\oplus}(f) \leq \text{lin-rank}(f) + D_{\oplus}(f')
\]

for some subfunction \( f' \) of \( f \) with \( \text{deg}_{\oplus}(f') < \text{deg}_{\oplus}(f) \). Now by Lemma 4.1, we can use \( C_{\oplus, \min}(f) \) to upper bound \( \text{lin-rank}(f) \). For the second part, since \( \text{deg}_{\oplus}(f') < \text{deg}_{\oplus}(f) \) and \( \| \hat{f}' \|_1 \leq \| \hat{f} \|_1 \), we can apply the induction hypothesis on \( f' \) to upper bound \( D_{\oplus}(f') \). What we get here is
\[
D_{\oplus}(f) \leq B_{d-1}(\| \hat{f}^\pm \|_1^2) \log \| \hat{f}^\pm \|_1 + 1 + B_{d-1}(\| \hat{f}^\pm \|_1^2).
\]

Now define the right-hand side to be \( B_d(\| \hat{f}^\pm \|_1) \), and solve the recursive relation
\[
B_d(m) = B_{d-1}(m^2) \log m + B_{d-1}(m) + 1, \quad B_3(m) = O(\log m + 1),
\]

we get
\[
B_d(m) = (1 + o(1))2^{(d-2)(d-3)/2} \log^{d-1} m,
\]
as desired.

We have just showed that low degree polynomials have small \( C_{\oplus, \min} \) value. We actually conjecture that the bound can be improved to the following.

**Conjecture 5.9 ([TWXZ13]).** There is some absolute constant \( c \) such that for all \( f : \{0,1\}^n \rightarrow \{0,1\} \), \( C_{\oplus, \min}(f) = O(\log^c \| f \|_1) \).

O’Donnell et al. [OST+14] proved that \( C_{\oplus, \min}(f^{\circ k}) \geq \Omega(C_{\min}(f)^k) \) where \( f^{\circ k} \) is the function which results from composing \( f \) with itself \( k \) times. They used this result to construct an explicit \( f \) such that \( C_{\oplus, \min}(f) \geq \Omega((\log \| f \|_1)^{\log 2 + 3}) \). In particular, it shows that \( c \geq \log_2 3 \approx 1.58 \) in Conjecture 5.9. To our knowledge, this is the best known separation between \( C_{\oplus, \min}(f) \) and \( \log \| f \|_1 \).

### 5.4 XOR protocols versus parity decision trees

The conjecture that \( D_{\oplus}(f) \leq \log^c \| f \|_1 \) for some absolute constant \( c \) (Conjecture 3.24) appeared as one possible approach to prove the Log-rank Conjecture for XOR functions (by using the simple but important observation that \( CC_{\oplus}(f) \leq 2D(f) \)). Until recently, it was not known whether this is essentially the only approach, i.e., whether \( CC(F) \) and \( D_{\oplus}(f) \) are actually polynomially related.
It turns out that low cost communication protocol for XOR functions $F(x, y) = f(x \oplus y)$ imply low depth parity decision tree for computing $f$. Precisely we have the following theorem of H. Hatami, Hosseini and Lovett [HHL16].

**Theorem 5.10** ([HHL16]). $D_\oplus(f) \leq O(CC_\oplus(f)^6)$ for all $f$.

As a corollary, this theorem implies the following.

**Corollary 5.11.** Let $\mathcal{M}_1, \mathcal{M}_2$ be any two measures in $\{\text{lin-rank}, C_{\oplus, \text{min}}, D_\oplus, CC_\oplus\}$. Then for all $f$,

$$\mathcal{M}_1(f) \leq (\mathcal{M}_2(f) \log \|\hat{f}\|_0)^c,$$

where $c$ is a constant that depends only on the measures $\mathcal{M}_1$ and $\mathcal{M}_2$.

In particular, it follows that proving the Log-rank Conjecture for XOR functions is equivalent to proving any single one of the linear rank, minimum parity certificate complexity or parity query complexity can be bounded from above by some polynomial in $\log \|\hat{f}\|_0$.

## 6 The Sensitivity Conjecture

In this section, we turn to the study of the Sensitivity Conjecture, which asserts that the block sensitivity can be upper bounded by some polynomial in the sensitivity. We will first discuss an equivalent formulation of the Sensitivity Conjecture using Theorem 4.2 with $F = F_2$. Then we will introduce an approach for the $F$-degree versus sensitivity problem over general field $F$ by considering a refine structure of the $F$-polynomial representations.

### 6.1 A equivalent formulation

The results in this section are from the work of the author in [Tsa14]. By Theorem 4.2 and the fact that sensitivity is non-increasing under taking subfuntions, to prove the Sensitivity Conjecture it suffices to prove the following pair of conjectures.

**Conjecture 6.1** ([Tsa14]). There exists $c > 0$ such that for all $f$, $C_{\text{min}}(f) \leq s(f)^c$.

**Conjecture 6.2** ([BB99, GNS+16]). There exists $c > 0$ such that for all $f$, $\deg_{F_2}(f) \leq s(f)^c$.

**Remark.** Both conjectures are consequences of the Sensitivity Conjecture since $C_{\text{min}}(f) \leq C(f)$ and $\deg_{F_2}(f) \leq \deg(f)$ for all $f$. However, both measures can be arbitrarily smaller than the sensitivity. For instance, let $f$ be the AND function on $n$ variables; then $C_{\text{min}}(f) = 1$ but $s(f) = n$. For $F_2$-degree, let $f$ be the XOR function on $n$ variables, then $\deg_{F_2}(f) = 1$ but $s(f) = n$.

Conjecture 6.2 was first appeared in [BB99] as a related problem to the Sensitivity Conjecture. It was also stated as an open problem in [GNS+16] that may be more approachable than the Sensitivity Conjecture given their results. A more refined relationship of this conjecture to the Sensitivity Conjecture was not known until the work of the author in [Tsa14], in which Conjecture 6.1 was introduced and it was shown that both conjectures together imply the Sensitivity Conjecture.

We would like to rephrase Conjecture 6.2 as a purely combinatorial problem. Let $f$ be a Boolean function. We can always find a subfunction $f'$ on $\deg_{F_2}(f)$ variables such that $\deg_{F_2}(f') = \deg_{F_2}(f)$, i.e., $f'$ has full degree. Moreover, taking restriction does not increase sensitivity. So it follows that $\deg(f) = \deg_{F_2}(f') \leq s(f')^c \leq s(f)^c$. In particular it means that we can always assume $f$ to be of full degree in Conjecture 6.2. One can actually completely remove the reference to $F_2$-degree by using the combinatorial characterization of having full $F_2$-degree. Specifically, a function $f$ has full $F_2$-degree if and only if $|f^{-1}(1)|$ is odd.

On the other hand, since $C(f, x) \leq bs(f, x)s(f)$ for all $x$, we can replace $C_{\text{min}}(f)$ be the even smaller measure $bs_{\text{min}}(f)$ in Conjecture 6.1. To summarize, we turn Conjecture 6.1 and Conjecture 6.2 into the following pair of conjectures.
Conjecture 6.3. There exists $c > 0$ such that for all $f$, $bs_{\min}(f) \leq s(f)^c$.

Conjecture 6.4. There exists $c > 0$ such that if $|f^{-1}(1)|$ is odd, then $s(f) \geq n^c$.

Proposition 6.5 ([Tsa14]). (i). Conjecture 6.3 follows from the Sensitivity Conjecture.

(ii). Conjecture 6.4 follows from the Sensitivity Conjecture.

(iii). Conjecture 6.3 and Conjecture 6.4 together imply the Sensitivity Conjecture.

Proof. (i) and (ii) follows from the fact that $bs_{\min}(f) \leq bs(f)$ and $deg_{\min}(f) \leq deg(f)$. (iii) follows from Theorem 4.2.

To the best of our knowledge, both of the conjectures are open. But there are examples showing that the constants in the conjectures cannot be too small. We note that the $\text{SORT}_4$ function of Ambainis [Amb06] defined by $f(x) = 1$ if either $x_1 \geq x_2 \geq x_3 \geq x_4$ or $x_1 \leq x_2 \leq x_3 \leq x_4$, and $f(x) = 0$ otherwise, has $s(f) = 2$ and $bs_{\min}(f) = 3$, and hence the composition of itself gives separation between $s(f)$ and $bs_{\min}(f)$. For separation between $deg_{\min}(f)$ and $s(f)$, let $f$ be an AND-OR tree with fan-in $\sqrt{n}$ at both levels. It is straightforward to show that $s(f) = \sqrt{n}$ but $deg_{\min}(f) = n$. We summarize such separations as follows.

Proposition 6.6. (i). There is a family of functions with $bs_{\min}(f) = \Omega(s(f)^{bs_3})$.

(ii). There is a family of functions with $deg_{\min}(f) = \Omega(s(f)^2)$.

6.2 $\mathbb{F}$-degree versus sensitivity

We further study the relationship between $\mathbb{F}$-degree and sensitivity, which is one of the subproblems in our equivalent formulation of the Sensitivity Conjecture. We will introduce a general approach for proving polynomial upper bound on the $\mathbb{F}$-degree based on a combinatorial property of the polynomial representation.

6.2.1 Basic structures of polynomial representations

Before we get to the details of our new approach, we collect a set of simple facts about polynomial representations that will be useful later. Let $f : \{0,1\}^n \rightarrow \{0,1\}$; we will denote its $\mathbb{F}$-representation by $p_F = \sum_{I \subseteq [n]} c_I x_I$ where $x_I = \prod_{i \in I} x_i$. When the field $\mathbb{F}$ is clear from the context, we may also abuse notation to use $f$ to denote its own $\mathbb{F}$-representation. We say that $x_I$ is a monomial of $p$ if $c_I \neq 0$. And similarly we say that $x_I$ is a $(\mathbb{F})$-monomial of a Boolean function $f$ if it is a monomial of its $(\mathbb{F})$-polynomial representation. There is a natural identification between elements in $\{0,1\}^n$ and subsets in $[n]$, and we will use the same symbol, say $x$, in both interpretations without specification. For $a, b \in B^n$, we use $a + b$ to denote the entrywise XOR unless otherwise stated.

Proposition 6.7. $f(x) = \sum_{I \subseteq [n]} c_I x_I$. In particular, $f(0^n) = c_\emptyset$.

Proof. If $I \not\subseteq x$, then $x_i = 0$ for some $i \in I$ and hence $\prod_{i \in I} x_i = 0$. Otherwise $x_i = 1$ for all $i \in I$ and $\prod_{i \in I} x_i = 1$.

Definition 6.8 (Minimal monomial). Let $p = \sum_{I \subseteq [n]} c_I x_I$ be a polynomial, we say that $I$ is a minimal monomial if $c_I \neq 0$ and for all nonempty proper subset $J$ of $I$, $c_J = 0$. We say that $x_I$ is a $\mathbb{F}$-minimal monomial of a Boolean function $f$ if it is a minimal monomial of the $\mathbb{F}$-polynomial representation of $f$.

Corollary 6.9 (Minimal monomials are minimal sensitive blocks). Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be a Boolean function. Then over any field $\mathbb{F}$, $x_I$ is a $\mathbb{F}$-minimal monomial of $f$ if and only if $I$ is a minimal sensitive block at $0^n$. Moreover, we have $c_I = 1 - 2c_\emptyset$. 

17
Proof. If \( x_I \) is a minimal monomial, then by definition \( c_J = 0 \) for all nonempty \( J \subseteq I \). So by Proposition 6.7, \( f(1_I) = c_\emptyset + c_I \). Since \( c_I \neq 0 \) and \( f \) is a Boolean function, \( c_\emptyset + c_I = 1 - c_\emptyset \), as desired.

Now, let \( I \) be a minimal sensitive block at \( 0^n \), i.e. \( f(1_I) \neq f(0^n) \). By Proposition 6.7 again, \( \sum_{J \subseteq I} c_J = f(1_I) \neq f(0^n) = c_\emptyset \). So there exists nonempty \( J \subseteq I \) such that \( c_J \neq 0 \). It is clear that \( J \) contains a set \( K \) such that \( x_K \) is a minimal monomial. By our argument above, we have \( f(1_K) \neq f(0^n) \). It follows that \( J = I \) and \( x_I \) is a minimal monomial. \( \square \)

Remark. Note that the same proof actually shows that even the function is not Boolean, a minimal monomial \( x_I \) still corresponds to a minimal sensitive block \( I \) in the sense that \( f(0^n) \neq f(1_I) \).

Corollary 6.10 (Sensitivity from representation). \( s(f,0^n) = |\{i \in [n] \mid c_{(i)} \neq 0\}| \).

Proof. Since sensitive blocks of size 1 are obviously minimal, by Corollary 6.9 the number of those sensitive blocks is exactly the number of degree-1 minimal monomials. \( \square \)

This fact actually asserts that the coefficients of the degree-1 monomial characterize \( f(x) \) for \(|x| \leq 1 \). In fact it is true in general that coefficients of monomials of degree up to \( \ell \) characterize \( f(x) \) for \(|x| \leq \ell \).

Corollary 6.11. Let \( f,f' \) be Boolean functions and \( p,p' \) their polynomial representations. Then \( c_I = c'_I \) for all \(|I| \leq \ell \) if and only if \( f(x) = f(x) \) for all \(|x| \leq \ell \).

Proof. It is clear by Proposition 6.7 that if \( c_I = c'_I \) for all \(|I| \leq \ell \), then \( f(x) = f'(x) \) for all \(|x| \leq \ell \). Now for the other direction, consider the minimal integer \( k \) such that \( c_I \neq c'_I \) for some \(|I| = k \leq \ell \). By Proposition 6.7, we have \( f(1_I) - f'(1_I) = c_I - c'_I \neq 0 \). \( \square \)

Recall that the polynomial representation is not invariant under translation. However, the coefficients of the maxonomials are indeed invariant up to sign under translation.

Definition 6.12 (Maxonomial). Let \( p = \sum c_I x_I \) be a polynomial, we say that \( x_I \) is a maxonomial of \( p \) if \( c_I \neq 0 \) and for all proper superset \( J \) of \( I \), \( c_J = 0 \). We say that \( x_I \) is a \( \mathbb{F} \)-maxonomial of \( f \) if it is a maxonomial of the \( \mathbb{F} \)-polynomial representation of \( f \).

Definition 6.13 (Translation operator \( T_a \)). Let \( f : \{0,1\}^n \rightarrow \{0,1\} \) and \( a \in \{0,1\}^n \). The translation operator \( T_a \) maps \( f \) to \( T_a f \), which is defined by \( T_a f(x) = f(x + a) \) for all \( x \). Note that the addition is over \( \mathbb{F}_2 \).

Definition 6.14 (Derivative). Let \( p \) be a polynomial over \( \mathbb{F} \), the \( i \)-derivative (or derivative along \( x_i \)) of \( p \) is the polynomial \( \Delta_i p \) defined by

\[
\Delta_i p(x) = p(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - p(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n).
\]

In general, for \( I = \{i_1, i_2, \ldots, i_k\} \), the \( I \)-derivative (or derivative along \( x_I \)) of \( p \) is the \( \Delta_I p = \Delta_{i_1} \Delta_{i_2} \cdots \Delta_{i_k} p \).

Note that the ordering in the sequence of derivatives we take does not matter. When the field \( \mathbb{F} \) being considered is clear from the context, we may sometimes abuse notation and use \( \Delta_i f \) to denote \( \Delta_i p \) where \( p \) is the \( \mathbb{F} \)-representation of \( f \). We can now prove the following.

Proposition 6.15 (Translationally invariant of maxonomials). Let \( f : \{0,1\}^n \rightarrow \{0,1\} \), \( \mathbb{F} \) be a field, and \( \mathcal{I} \) be the set of \( \mathbb{F} \)-maxonomials of \( f \). Then \( \mathcal{I} \) is the set of \( \mathbb{F} \)-maxonomials of \( T_a f \) for all \( a \in \{0,1\}^n \). Moreover, the coefficients of the \( \mathbb{F} \)-maxonomials are invariant up to sign (i.e., additive inverse) under \( T_a \) for all \( a \).
Proof. First we observe that $T_{a+b}f = T_a T_b f$ for all $a, b \in \{0, 1\}^n$. It follows that if $a = e_{i_1} + \cdots + e_{i_k}$, then $T_a f = T_{e_{i_k}} T_{e_{i_{k-1}}} \cdots T_{i_1} f$. Thus, it suffices to prove the claim for $a = e_i$.

Let $p$ be the polynomial representation of $f$ and we write $p$ as
\[
p = x_i \Delta p + g_i = \sum c_I x_I,
\]
where $\Delta p$ and $g_i$ do not depend on $x_i$. Let $p'$ be the polynomial representation of $T_{e_i} f$. Then we have
\[
p' = (1 - x_i) \Delta p + g_i = -x_i \Delta p + (g_i + \Delta p) = \sum c'_I x_I.
\]
From this it is clear that for all $I \ni i$, $c_I = -c'_I$. It follows that for those $I$, if $x_I$ is a maxonomial of $p$ then it is also a maxonomial of $p'$. Now, let $x_I$ be a maxonomial in $p$ such that $I \not\ni i$. Such monomial must be from $g_i$. But the coefficient of $x_j$ for any superset $J$ of $I$ (not necessary proper) in $\Delta p$ is zero, it follows that $x_I$ is still a maxonomial of $p'$. Finally, since $T_{e_i} T_{e_i} f = f$, so if $x_I$ is a maxonomial of $p'$, then it is also a maxonomial of $p$. \[\blacksquare\]

6.2.2 Monomial graphs

Recall that the degree versus sensitivity problem (cf. Conjecture 6.2 and equivalently Conjecture 6.4) asks whether the degree is at most some polynomial in the sensitivity. It is without loss of generality that we can assume the function to be of full degree, and try to prove that it has high sensitivity. One way to prove a claim like this is to identify properties of functions with full degree and then to show that functions with low sensitivity do not have those properties.

The property we consider is natural and simple: if $f$ has degree $n$, then for each $i \in [n]$, at least one of the subfunctions\(^1 \) $f^{i=0}$ or $f^{i=1}$ has full degree (can be both of them if we are considering $\deg_{\otimes}$, and exactly one of them if we are considering $\deg_{\otimes}$). Such property is not new and has already been used in [GKS15] and [LTZ17]. Also note that query complexity has the same property.

In the following, we will show that this property has strong implications to the combinatorial structure of polynomial representations. We start by defining a mapping that is essentially the protocol for the communication game introduced by Gilmer, Koucký and Saks [GKS15] (in fact they used this protocol to show that any low-sensitivity function with full degree can be used to construct a low communication cost protocol for their game). In the following, $S_n$ is the set of all permutations on a $n$-element set and $\mathcal{P}(\{0, 1\}^n)$ is the power set of $\{0, 1\}^n$.

Definition 6.16. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ with $\deg(f) = n$. We define $\Psi_f : S_n \rightarrow \mathcal{P}(\{0, 1\}^n)$ such that $z \in \Psi_f(\sigma)$ if $\deg(f^{(\sigma(1), \ldots, \sigma(i)) = (z_{\sigma(1)}, \ldots, z_{\sigma(i)})}) = n - i$ for all $i$.

It is clear that $\Psi_f(\sigma) \not= \emptyset$ for all $\sigma \in S_n$ since $f$ has full degree. We observe that $\Psi$ behaves nicely under translation of the function that induces it.

Proposition 6.17. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ with $\deg(f) = n$. Then the following holds:

1. $\Psi_{T_{a} f}(\sigma) = \Psi_f(\sigma) + a$ for all $a \in \{0, 1\}^n$, and
2. $\Psi_{f+b}(\sigma) = \Psi_f(\sigma) + b \sigma(\otimes)$ for $b \in \{0, 1\}$,

where addition is over $\mathbb{F}_2$ and $e_i$ is the $i$-th standard basis vector.

Proof. Without loss of generality, we assume $\sigma$ is the identity. For the first item, let $z \in \Psi_f(\sigma)$ and observe that after fixing the first $i$ variables according to $z$, $\deg(f(z_1, \ldots, z_i, x_{i+1}, \ldots, x_n)) = n - i$ if and only if $\deg(f(z_1, \ldots, z_i, x_{i+1} + a_{i+1}, \ldots, x_n + a_n)) = n - i$ since translation does not change the degree. But $f(z_1, \ldots, z_i, x_{i+1} + a_{i+1}, \ldots, x_n + a_n) = T_{e_i} f(z_1 + a_1, \ldots, z_i + a_i, x_{i+1}, \ldots, x_n)$, from this it follows that $z + a \in \Psi_{T_{a} f}(\sigma)$. Now, we can see that if $z + a \in \Psi_{T_{a} f}(\sigma)$, then $z = (z + a) + a \in \Psi_{T_{a} T_{a} f}(\sigma) = \Psi_{f}(\sigma)$.

\(^1\)We use $f^{i=0}$ denote the subfunction of $f$ obtained by fixing the $i$-th bit to $0$. In general, we use $f^{(i_1, \ldots, i_k) = (b_1, \ldots, b_k)}$ to denote the subfunction of $f$ obtained by fixing the $i_j$-th bit to $b_j$ for each $j \in [k]$. 

19
The second item is trivial if \( b = 0 \). Assume \( b = 1 \), since adding a constant 1 to a function does not change degree unless the function itself is a constant, \( \deg(f(z_1, \ldots, z_i, x_{i+1}, \ldots, x_n)) = n - i \) if and only if \( \deg(f(z_1, \ldots, z_i, x_{i+1}, \ldots, x_n) + 1) = n - i \) for \( i < n \). And for \( i = n \), since \( \deg(f(z_1, \ldots, z_{n-1}, x_n)) = 1 \) and \( f(z_1, \ldots, z_n) = 1 \), we have \( f(z_1, \ldots, z_n + 1) = 0 \) and hence \( f(z_1, \ldots, z_n + 1) + 1 = 1 \), which completes the proof. \( \square \)

**Definition 6.18 (Monomial graph).** Let \( f : \{0, 1\}^n \to \{0, 1\} \), and \( \mathbb{F} \) be a field. The \( \mathbb{F} \)-**monomial graph** \( M^\mathbb{F}_f \) of \( f \) is the vertex-induced subgraph of the \( n \)-dimensional Boolean cube induced by the union of \( 0^n \) and the set of monomials in the \( \mathbb{F} \)-representation of \( f \).

When the field is \( \mathbb{R} \), we denote \( M^\mathbb{R}_f \) by \( M_f \) and simply call it the monomial graph of \( f \). Now we are ready to turn the condition of being full degree to a combinatorial structure of the polynomial representation. The connection is based on the simple observation that if \( 0^n \in \Psi_f(\sigma) \), then for each \( i \in [n] \), \( \prod^n_{j=i} x_{\sigma(j)} \) is a monomial of \( f \). By our definition of monomial graph, it follows that this set of monomials corresponds precisely to a monotone path from \( 0^n \) to \( 1^n \) in \( M_f \). Since \( \Psi_f(\sigma) \neq \phi \) for all \( \sigma \), there exists \( z \in \{0, 1\}^n \) such that \( z \) is contained in \( \Psi_f(\sigma) \) for at least \( n!/2^n \) many \( \sigma \). By Proposition 6.17, \( 0^n \) is contained in \( \Psi_{T_z f}(\sigma) \) for the same number of \( \sigma \). So it means that there are at least \( n!/2^n \) monotone paths from \( 0^n \) to \( 1^n \) in the graph \( M_{T_z f} \). To summarize, we have proved the following fact:

**Lemma 6.19.** Let \( f : \{0, 1\}^n \to \{0, 1\} \) with \( \deg_\mathbb{F}(f) = n \). Then there exists \( z \in \{0, 1\}^n \) such that the number of monotone paths from \( 0^n \) to \( 1^n \) in \( M^\mathbb{F}_f \) is at least \( n!/2^n \).

Now we show how the number of monotone paths in the monomial graphs is related to the Sensitivity Conjecture.

**Theorem 6.20.** Let \( h : \mathbb{R}_+ \to \mathbb{R}_+ \) be an increasing function. If for all Boolean functions \( f \) the number of monotone paths from \( 0^n \) to level \( \ell \) in \( M^\mathbb{F}_f \) is at most \( h(s(f))\ell \) for some \( \ell \geq 0.1n \), then \( \deg_\mathbb{F}(f) \leq e2^{10}h(s(f)) \) for all \( f \).

**Proof.** Let \( f : \{0, 1\}^n \to \{0, 1\} \). We can assume without loss of generality that \( \deg_\mathbb{F}(f) = n \). Then by Lemma 6.19, there exists \( z \in \{0, 1\}^n \) such that the number of monotone paths from \( 0^n \) to \( 1^n \) in \( M_{T_z f} \) is at least \( n!/2^n \). On the other hand, we can upper bound the number of those paths by \( h(s(f))\ell \cdot (n - \ell)! \). Thus, we have the inequality

\[
h(s(f))\ell \cdot (n - \ell)! \geq \frac{n!}{2^n},
\]

which implies

\[
h(s(f))\ell \geq \frac{(n/\ell)!}{2^n} \geq \frac{(n/\ell)!}{2^{10\ell}} \geq \frac{1}{2^{10\ell}} \cdot n^\ell \cdot \ell! \geq \frac{1}{2^{10\ell}} \cdot n^\ell \cdot 1 \geq \frac{1}{e^{10}},
\]

where the third inequality follows from the bound \( \binom{n}{\ell} \geq (n/\ell)^\ell \) and the last inequality follows from the fact that \( e^{10} = \sum_{i=0}^{\infty} \frac{1}{i!} \geq \frac{n^\ell}{\ell!} \). The proof is completed by taking \( \ell \)-th roots on both sides and rearranging terms. \( \square \)

**Remark.** The constant 0.1 in the condition \( \ell \geq 0.1n \) is arbitrary. In fact, if we replace it by \( \ell \geq c \cdot n \), then the resulting bound will become \( \deg(f) \leq e2^{10/c}h(s(f)) \). So in order to prove the Sensitivity Conjecture, it suffices to prove the condition for \( \ell \geq n/O(\log s(f)) \).

With this lemma in mind, we make the following conjecture.

**Conjecture 6.21.** For all Boolean functions \( f \) and all positive integers \( \ell \), the number of monotone paths from \( 0^n \) to level \( \ell \) in \( M^\mathbb{F}_f \) is at most \( s(f)^c \ell \), where \( c > 0 \) is an absolute constant.

The constant \( c \) in the conjecture is at least 2, by considering the AND-OR tree of fan-in \( \sqrt{n} \) at both levels. We now prove an upper bound on the number of monotone paths from \( 0^n \) to level \( \ell \) in terms of \( s(f) \) and \( \ell \). In the following we consider \( \mathbb{F} = \mathbb{R} \), which gives the strongest result among all fields.
Proposition 6.22. Let \( f : \{0,1\}^n \to \{0,1\} \) with \( s(f) = s \). Then for all \( \ell \geq 1 \), the number of monotone paths from \( 0^n \) to level \( \ell \) in \( M_f \) is at most \( \min\{2^{\ell/2}s(f)\ell,(s(f)4^{s(f)})\ell\} \).

First we have the following lemma which relates the degree of a node in \( M_f \) with the sensitivity of the function. Note that in the proof we will consider the sensitivity of non-Boolean function on an input \( x \), which is defined in the same way (as the Boolean case) as the number of \( i \) such that \( f(x) \neq f(x^i) \).

Lemma 6.23. Let \( f : \{0,1\}^n \to \{0,1\} \), and \( I \subseteq [n] \). Then the number of monomials \( J \) of \( f \) such that \( |J| = |I| + 1 \) and \( I \subseteq J \) is at most \( 2^{|I|}s(f) \).

Proof. The number of such monomials \( J \) is precisely the number of degree-1 monomials in the polynomial representation of \( \Delta_i f \). By Corollary 6.10, this number is equal to the number of \( i \) such that \( \Delta_i f(0) \neq \Delta_i f(e_i) \) where \( e_i \) is the \( i \)-th standard basis vector. Note that this value is exactly \( s(\Delta_i f,0) \).

Now we claim that \( s(\Delta_i f) \leq 2s(f) \), from this a simple induction will complete the proof. Recall that the derivative \( \Delta_i f \) is defined by

\[
\Delta_i f(x) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = f^{i+1}(x) - f^{i-0}(x).
\]

Thus, we have

\[
\Delta_i f(x) = (f^{i+1}(x) - f^{i-0}(x)) - (f^{i+1}(x') - f^{i-0}(x')).
\]

Now suppose \( \Delta_i f(x) \neq \Delta_i f(x') \), then \( j \neq i \) and we must have \( f^{i+1}(x) \neq f^{i+1}(x') \) or \( f^{i-0}(x) \neq f^{i-0}(x') \), i.e., \( j \) is a sensitive coordinate to \( f^{i+0} \) or \( f^{i-0} \). Our claim follows from the fact that both \( f^{i+0} \) and \( f^{i-0} \) are subfunctions of \( f \) and hence has sensitivity at most \( s(f) \).

Proof of Proposition 6.22. By Lemma 6.23, it is clear that the number of those monotone paths is at most \( 2^n \cdot 2^{1-1} \cdot s(f) \ell \leq 2^{\ell/2}s(f)\ell \).

For the other bound, we use the following theorem of Hans-Ulrich Simon.

Theorem 6.24 ([Sim83]). Let \( f : \{0,1\}^n \to \{0,1\} \) be a Boolean function that depends on all \( n \) variables. Then \( s(f)4^{s(f)} \geq n \).

This theorem asserts that a Boolean function with sensitivity \( s(f) \) depends on at most \( s(f)4^{s(f)} \) many variables. Let \( n' \) be the number of variables that \( f \) depends on, it is clear that \( f \) has a subfunction \( f' \) on \( n' \) variables such that the number of monotone paths from \( 0^{n'} \) to level \( \ell \) in \( M_{f'} \) is the same as that in \( M_f \). Now, it is easy to see that the number of such monotone paths in \( M_{f'} \) is at most \( (n')(n'-1) \cdots (n'-\ell) \leq (n')^\ell \), and the proof is completed by the fact that \( n' \leq s(f')4^{s(f')} \leq s(f)4^{s(f)} \).

One way to try improving the upper bound in Proposition 6.22 is to improve the bound in Lemma 6.23. However, the bound \( s(\Delta_i f) \leq 2^{|I|}s(f) \) in the proof of Lemma 6.23 cannot be improved significantly for \( |I| \leq s(f) \). In particular, the dependence on \( 2^{|I|} \) is essential, even for functions with very low sensitivity.

Proposition 6.25. Let \( f : \{0,1\}^{n-\log n} \to \{0,1\} \) be the address function defined by \( f(x,i) = x_i \). Then for all \( \ell \leq \log n \), there exists \( I \subseteq [n + \log n] \) of size \( \ell \) such that \( s(\Delta_I f,0^{n-\ell}) \geq 2^\ell \).

Proof. By definition, it is easy to see that \( f(x,i) = \sum_{j=1}^n x_j \cdot 1_{\{j\}}(i) \) where \( 1_{\{j\}} \) is the indicator of a singleton. The indicator function \( 1_{\{j\}} \) is essentially a translation of the AND function and has polynomial representation \((\prod_{k:j_k=1}^n k)(\prod_{l:j_l=0}^n (1-l))\). Let \( I \) be any subset of \( \{1,\ldots,\log n\} \) of size \( \ell \). Then \( I \) is a monomial of \( 2^{|I|} \) of those indicators. Since each indicator will be multiplied with a distinct variable \( x_j \), it follows that \( f \) has monomials of the form \( x_ji_I \) for at least \( 2^\ell \) many \( j \in [n] \).
However, the address function is not a counterexample to Conjecture 6.21. Indeed, by looking at its \( \mathbb{R} \)-representation, even though for some vertices there can be many neighbors in the next higher level to go, many of the neighbors have few neighbors to proceed further. In fact, the number of monotone paths from zero to level \( \ell \) in its monomial graph is at most \( O((\ell^2(\log n)^\ell) = s(f)O(\ell) \).

### 6.2.3 Sparsity of polynomial representations

By a result of [GNS+16], functions with sensitivity at most \( k \) are determined by the values on inputs of Hamming weight at most \( 2k \). So it follows from Corollary 6.11 that coefficients of monomials of degree higher than \( 2k \) are completely determined by those of degree at most \( 2k \). It suggests that for low-sensitivity functions, their polynomial representations are highly-structured.

As a related problem to Conjecture 6.21, in this section we will consider the sparsity of the \( \mathbb{R} \)-representations (cf. Definition 3.4 and Definition 6.32), and show that the set of degree-\( \ell \) monomials of a low-sensitivity function is sparse, for each \( \ell \). We first restate the definitions of polynomial sparsity for easy reference.

**Definition 6.26 (Degree-\( \ell \) sparsity).** Let \( p \in \mathbb{F}[x_1, \ldots, x_n] \), the degree-\( \ell \) sparsity \( \text{spar}_\ell(p) \) of \( p \) is the number of degree-\( \ell \) monomials in \( p \). Let \( f : \{0,1\}^n \to \{0,1\} \) be a Boolean function. The degree-\( \ell \) sparsity \( \text{spar}_\ell(f) \) of \( f \) over \( \mathbb{F} \) is defined as the degree-\( \ell \) sparsity of its \( \mathbb{F} \)-representation.

**Definition 6.27 (Polynomial sparsity).** Under the same notations above, the polynomial sparsity \( \text{spar}(p) \) of \( p \) is the number of monomials in \( p \), i.e. \( \text{spar}(p) = \sum_{\ell=1}^{\lfloor n/2 \rfloor} \text{spar}_\ell(p) \). The \( \mathbb{F} \)-polynomial sparsity \( \text{spar}_\ell(f) \) of \( f \) is defined as the polynomial sparsity of its \( \mathbb{F} \)-representation.

Note that the set of monomials in the \( \mathbb{F} \)-representation is always contained in the set of monomials in the \( \mathbb{R} \)-representation regardless what \( \mathbb{F} \) is. So our result implies the same bound for polynomial representations over any other field. We will mostly consider \( \mathbb{R} \)-representation in this section. So we simplify the notation by dropping the superscript, and simply write \( \text{spar} \) and \( \text{spar}_\ell \) when the field is \( \mathbb{R} \).

In general, a Boolean function \( f \) can have \( \text{spar}_\ell(f) \) as large as \( \binom{n}{\ell} \) such as the OR function on \( n \) variables. We show that the degree-\( \ell \) sparsity is much smaller for low-sensitivity functions:

**Theorem 6.28.** Let \( f \) be a Boolean function. Then for all \( \ell \geq 1 \),

\[
\text{spar}_\ell(f) \leq (4e)^\ell s(f)^{\ell - \min(s(f), \ell)}.
\]

It is obvious that \( \text{spar}_\ell(f) \leq \binom{n}{\ell} \), so the bound above is nontrivial only when \( \ell \ll \log n \). To prove this theorem, we use the following result of Kenyon and Kutin on \( \ell \)-block sensitivity.

**Theorem 6.29 ([KK04]).** For \( 2 \leq \ell \leq s(f) \), \( b_{s(f)}(f) \leq \frac{\ell}{\ell - 1} s(f) b_{s(f)-1}(f) \) and \( b_{s(f)}(f) \leq \frac{e}{\ell} s(f)^{\ell} \).

**Definition 6.30.** Let \( f : \{0,1\}^n \to \{0,1\} \). A degree-\( \ell \) cover \( H \) of \( f \) is a subset of \( [n] \) that intersects every monomial of degree \( \ell \) in the \( \mathbb{R} \)-representation of \( f \). Denote \( \tau_\ell(f) \) the minimum size of any degree-\( \ell \) cover of \( f \).

We have the following bound on \( \tau_\ell \).

**Lemma 6.31.** For \( 1 \leq \ell \leq s(f) \),

\[
\tau_\ell(f) \leq b_{s(f)}(f) \leq \frac{e\ell(s(f))^{\ell}}{(\ell - 1)!} \quad \text{and} \quad \tau_\ell(f) \leq s(f) b_{s(f)}(f) \leq \frac{e(s(f))^{s(f)+1}}{(s(f)-1)!} \quad \text{for} \quad \ell > s(f).
\]

**Proof.** Let \( s = s(f) \) and \( m = b_{s(f)}(f, 0) \). Note that for \( \ell > s \), \( b_{s(f)}(f, 0^\ell) = b_{s(f)}(f, 0^n) \). Let \( B = \{B_1, \ldots, B_m\} \) be a family of sensitive \( \ell \)-block of \( f \) at \( 0^n \). We claim that \( H = \bigcup_{i=1}^m B_i \) is a degree-\( \ell \) cover of \( f \). Indeed, if there is a monomial \( x_I \) such that \( I \cap H = \phi \), then there exists a minimal monomial \( x_J \) where \( J \subseteq I \). However, it follows from Corollary 6.9 that \( f \) is also sensitive to block \( J \) on input \( 0^n \), which is a contradiction. \( \square \)
**Proof of Theorem 6.28.** We will prove the claim by induction on \(\ell\). For \(\ell = 1\), by Corollary 6.10, the number of degree-1 monomials is precisely the sensitivity of \(f\) on \(0^n\), which is at most \(s(f)\).

Now, let \(s = s(f)\). We can assume \(s(f) \geq 2\) since the claim is trivial if \(s(f) = 1\). Suppose the number of degree-\(\ell\) monomials is bounded by some function \(B(s, \ell)\) to be determined later. To bound the number of degree-(\(\ell + 1\)) monomials, we assume without loss of generality that \([k]\) is the minimum size degree-(\(\ell + 1\)) cover of \(f\), i.e., \(k = \tau_{\ell+1}(f)\). We can express \(f\) as

\[
f(x_1, \ldots, x_n) = x_1 f_1(x_2, \ldots, x_n) + x_2 f_2(x_3, \ldots, x_n) + \cdots + x_k f_k(x_{k+1}, \ldots, x_n) + f_0(x_{k+1}, \ldots, x_n),
\]

where \(f_0\) has no degree-(\(\ell + 1\)) monomials. Now, observe that \(f_1\) is the difference of two subfunctions of \(f\) (in fact, \(f_1(x_2, \ldots, x_n) = f(1, x_2, \ldots, x_n) - f(0, x_2, \ldots, x_n)\)). Similarly, \(f_2\) and so on are differences of some subfunctions of \(f\). By the expression above, it is clear that the degree-(\(\ell + 1\)) monomials in \(f\) are of the form \(x_i \cdot x_{i'}\) where \(x_i\) is a degree-\(\ell\) monomial in \(f_i\). So by induction hypothesis together with the observation that \(f_i\) is a difference of two subfunctions of \(f\), it follows that the number of degree-\(\ell\) monomials in \(f_i\) is at most \(2B(s, \ell)\). Thus, by Lemma 6.31, we conclude that the number of degree-(\(\ell + 1\)) monomials in \(f\) is at most

\[
B(\ell + 1, s) \leq k \cdot 2B(\ell, s) \leq \frac{2e^{\min\{\ell + 1, s\}}}{\min\{\ell, s - 1\}!} B(\ell, s).
\]

Taking \(B(s, \ell) \leq (4\epsilon)^{\ell \min\{s, \ell\}}\), for \(\ell < s\), we have

\[
B(\ell + 1, s) \leq \frac{2e(\ell + 1)}{\ell!} s^{\ell + 1} \cdot (4\epsilon)^{\ell \cdot s^2}
= (4\epsilon)^{\ell + 1} \cdot s^{\ell + 1} \cdot \frac{(\ell + 1)}{\ell!} \cdot \frac{1}{2}
\leq (4\epsilon)^{\ell + 1} s^{(\ell + 1)^2},
\]

For \(\ell \geq s\), we have

\[
B(\ell + 1, s) \leq (4\epsilon)^{\ell + 1 - s} s^s (\ell + 1 - s) \cdot B(s, s) \leq (4\epsilon)^{\ell + 1} s^{(\ell + 1) - s} s^s = (4\epsilon)^{\ell + 1} s^{(\ell + 1)s},
\]

as desired. \(\square\)

**Theorem 6.28** immediately implies an upper bound on \(\mathbb{F}\)-polynomial sparsity in terms of \(\mathbb{F}\)-degree and sensitivity.

**Corollary 6.32.** Let \(f\) be a Boolean function and \(\mathbb{F}\) be a field. Let \(d = \deg_{\mathbb{F}}(f)\) and \(s = s(f)\). Then

\[
\text{spar}^\mathbb{F}(f) \leq d(4\epsilon)^d s^{d \min\{s, d\}}.
\]

We are not aware of any similar results in the literature. However, when the field is \(\mathbb{R}\), one can get a similar but weaker bound \(\text{spar}(f) \leq 2^{2d^2}\) by considering the decision tree that computes \(f\). Indeed, if \(f\) has query complexity \(D(f)\), then the polynomial sparsity of \(f\) is at most \(2^{2D(f)}\). It can be shown that \(D(f) \leq \deg(f)^3\) [Mid04] and hence we have a bound \(\text{spar}(f) \leq 2^{2d^3}\). In fact, one can improve the bound to \(\text{spar}(f) \leq d^{d \cdot 2d^2}\) using the fact that \(d^{2d} \geq n\) [NS94] for functions depending on all \(n\) variables. The latter bound is comparable with our result when \(s \geq d\) but weaker in the remaining cases. But both of these arguments fail when the field is not \(\mathbb{R}\), and we do not know any comparable results in that case.

**Corollary 6.32** also suggests a problem that may help resolving the Sensitivity Conjecture in combination with the formulation we obtained in the Section 6.1 (cf. Conjecture 6.3 and Conjecture 6.4). More precisely, we have the following conjecture:

**Conjecture 6.33.** Let \(f : \{0, 1\}^n \to \{0, 1\}\) and \(s = s(f)\). Then there exists a subfunction \(f'\) of \(f\) on \(n - O(s)\) variables such that \(\text{spar}^{\mathbb{F}}(f') \leq \text{spar}^{\mathbb{F}}(f)/2\).
Assuming this conjecture, we can find a sequence of subfunctions \( f_1, f_2, \ldots, f_t \) such that \( \text{spar}(f_i) \leq \text{spar}^2(f)/2^i \). Since each step we fix \( O(s) \) many variables, and the \( F_2 \)-polynomial sparsity will drop to constant in at most \( \log \text{spar}^2(f) = O(\deg\_{F_2}(f)^2(\log \deg\_{F_2}(f) + \log(s(f))) \) many steps. It follows that \( f \) is identically constant on some subcube of codimension \( \tilde{O}(\log^{2}(\deg\_{F_2}(f)) (\text{where } \tilde{O} \text{ hides polylogarithmic factor of } \deg\_{F_2}(f) \text{ and } s(f)). \) So if this conjecture is true, it implies in particular that \( \text{Conjecture 6.3} \) follows from \( \text{Conjecture 6.4} \), and hence the Sensitivity Conjecture is equivalent to showing \( \deg\_{F_2}(f) \leq s(f)^c \) for some absolute constant \( c > 0 \). To summarize, we have

**Proposition 6.34.** If \( \text{Conjecture 6.33} \) is true, then the Sensitivity Conjecture is equivalent to prove \( \deg\_{F_2}(f) \leq s(f)^c \) for all \( f \), where \( c \) is an absolute constant.

In fact \( \text{Theorem 6.28} \) is not tight. For \( \ell = 2 \), one can actually show that \( \text{spar}(f) \leq O(s(f)^3) \).

We believe the bound is far from tight in general. And we conjecture that the correct exponent in \( \text{Theorem 6.28} \) is linear in \( \ell \).

**Conjecture 6.35.** Let \( f \) be a Boolean function. Then for all \( \ell \geq 1 \), \( \text{spar}_\ell(f) \leq s(f)^{c-\ell} \), where \( c > 0 \) is an absolute constant.

This conjecture is relevant to the result of Kenyon and Kutin on \( \ell \)-block sensitivity [KK04] since minimal monomials correspond precisely to minimal monomials by Corollary 6.9. Also note that we cannot hope for an upper bound \( s(f)^{O(\ell)} \) for any \( c < 1 \), both for \( s \) large and very small. For instance, both \( \text{AND}_n \) and \( \text{Addr}_{n,i} \) have number of degree-\( \ell \) monomials \( s(f)^{O(\ell)} \) for \( \ell = \Omega(s(f)) \).

Assuming the Sensitivity Conjecture, we have a bound \( bs_\ell(f) \leq s(f)^c \) for all \( \ell \geq 2 \). Then \( \text{Conjecture 6.35} \) will follow by the same argument we used to prove \( \text{Theorem 6.28} \). On the other direction, we do not know whether \( \text{Conjecture 6.35} \) itself is enough to imply the Sensitivity Conjecture. However, it implies in particular that the number of monotone paths in the monomial graph \( M_f \) of \( f \) is at most \( \ell s(f)^{O(\ell)} \), which is close to the bound we conjecture in \( \text{Conjecture 6.21} \).

It is worth mentioning the connection between \( \text{Conjecture 6.35} \) and the recent work of Gopalan, Servedio, Tal and Wigderson [GSTW16]. Gopalan et al. studied the moments of the distribution defined by the Fourier coefficients of a low-sensitivity function. Among many results they proved, they showed a version of switching lemma for low-sensitivity functions, which we state below:

**Lemma 6.36 ([GSTW16]).** Let \( f : \{0, 1\}^n \rightarrow \{0, 1\} \). Then for all \( \ell \geq 1 \),

\[
\Pr_{\rho \sim \mathcal{R}_{\ell,n}} \left[ D(f_\rho) = \ell \right] \leq \frac{(32s(f))^\ell \ell!}{n^\ell}.
\]

In the lemma above, \( \mathcal{R}_{\ell,n} \) is the set of all restrictions from \( n \) variables to \( \ell \) variables, and the probability is over an uniform random \( \rho \) in \( \mathcal{R}_{\ell,n} \). We can think \( f_\rho \) as chosen by first picking a uniform random \( k \)-subset \( I \) of \([n]\), and then fixing each variable outside \( I \) to an uniformly random bit.

We now show that \( \text{Lemma 6.36} \) directly implies an “average-case” of \( \text{Conjecture 6.35} \). Recall that \( T_a \) is the translation operator that takes \( f(x) \) to its translation \( f(x + a) \). We have the following fact.

**Corollary 6.37.** Let \( f : \{0, 1\}^n \rightarrow \{0, 1\} \). Then for all \( \ell \geq 1 \),

\[
\mathbb{E}_a \left[ \text{spar}_\ell(T_a(f)) \right] \leq (32s(f))^\ell.
\]

**Proof.** Since \( \deg(g) \leq D(g) \) for any Boolean function \( g \), \( \text{Lemma 6.36} \) implies

\[
\Pr_{\rho \sim \mathcal{R}_{\ell,n}} \left[ \deg(f_\rho) = \ell \right] \leq \Pr_{\rho \sim \mathcal{R}_{\ell,n}} \left[ D(f_\rho) = \ell \right] \leq \frac{(32s(f))^\ell \ell!}{n^\ell}.
\]
We will count the number of pairs \((I, a) \in \binom{[n]}{r} \times \{0, 1\}^n\) such that \(x_I \) is a degree-\(\ell\) monomial in \(T_a f\). Let \(\mathcal{I}\) be the family of all such pairs. By summing up \(I\) first and then \(a\), it is clear that \(|\mathcal{I}| = 2^n \mathbb{E}_a \{ \text{spar}_x (T_a f) \} \).

Let \(\rho \in \mathcal{R}_{\ell, a} \) and suppose \(\rho\) leaves variables in \(I\) free and restricts every variable outside \(I\) to some value. Then \(f_\rho\) is the restriction of \(f\) on some coset of \(H \) where \(H = \{ x \in \{0, 1\}^n \mid x_i = 0 \ \forall i \notin I \}\) is a \(\ell\)-dimensional subcube. We denote that coset as \(a + H\). Now, if \(\deg(f_\rho) = \ell\), it follows that the restriction of \(T_a f\) on \(H\) also has full degree, i.e. \(\deg(T_a f|_H) = \ell\). Moreover, since translation does not affect the degree, it follows that \(\deg(T_{a+h} f|_H) = \ell\) for all \(h \in H\). But \(\deg(T_{a+h} f|_H) = \ell\) implies that \(x_I\) is a monomial in \(T_{a+h} f\). So each \(\rho\) such that \(\deg(f_\rho) = \ell\) contributes exactly \(2^\ell\) pairs in \(\mathcal{I}\). It is not difficult to see that we do not miss any pair and hence \(|\mathcal{I}| = 2^\ell \cdot 2^{n-\ell} \cdot \binom{n}{r} \cdot \Pr_{\rho \in \mathcal{R}_{\ell, a}}[\deg(f_\rho) = \ell] \leq 2^n (32 s(f))^\ell \). From this the claim follows.

\[\square\]

7 Concluding remarks and future work

In this thesis, we study two major open problems in concrete complexity, namely, the Log-rank Conjecture for XOR functions and the Sensitivity Conjecture. By using the polynomial representations over various fields, we reduced each of them to a problem that relates the key complexity measures to some other measures that was not considered before.

Specifically, for the Log-rank Conjecture for XOR functions, the key measure is the logarithm of the Fourier sparsity \(\log \| \hat{f} \|_0\). Our parity decision tree construction we show that it suffices to prove polynomial upper bounds on minimum parity certificate complexity in terms of \(\log \| \hat{f} \|_0\). On the other hand, the polynomial representation itself also suggests a natural class of functions to consider — functions with low \(\mathbb{F}\)-degree. The combination of these ideas led us to prove the Log-rank Conjecture for special classes of XOR functions, including the functions with small spectral norm and functions with low \(\mathbb{F}_2\)-degree. Both results make use of the structure of the Fourier spectra under taking linear restrictions.

For the Sensitivity Conjecture, the situation is a bit different but the same approach of using polynomial representations to construct low-depth decision trees is still valid. In this case we show that the Sensitivity Conjecture is equivalent to proving polynomial upper bound on both the minimum certificate complexity (or minimum block sensitivity) and the \(\mathbb{F}\)-degree in terms of \(s(f)\). As far as we know it is the first equivalent formulation of the Sensitivity Conjecture that reduces it to a pair of conjectures such that each of them is a trivial consequence of the Sensitivity Conjecture, none of them is known to be sufficient for the Sensitivity Conjecture, but both of them together imply the Sensitivity Conjecture. We then focus on the degree versus sensitivity problem and introduce the monomial graph \(M_f\) associated to the Boolean function \(f\), and show that an upper bound of the form \(s(f)^{O(\ell)}\) on the number of monotone paths in \(M_f\) will imply the Sensitivity Conjecture. On the other hand, we show the number of monotone paths in \(M_f\) is at most \(2^{\ell/2} s(f)^\ell\). We also initiate the study of polynomial sparsity and its connection to the number of monotone paths in monomial graphs.

An obvious next step to consider is to prove or disprove any of the conjectures we make in this thesis. For the Log-rank Conjecture for XOR functions, the key conjecture to address is Conjecture 5.9, which in fact implies the main conjecture. While it may be too difficult to prove or disprove it, one may try to improve the separation between \(C_{\mathbb{F}, \text{min}}(f)\) and \(\log \| \hat{f} \|_1\) achieved in [OST+14]. The recent quadratic separation between communication complexity and log-rank by G"{o}"os et al. [GPW15] may provide useful techniques for improving separation in the XOR functions case.

The other interesting direction would be to study whether the bounds we proved also hold for \textit{one-way} communication complexity (or even non-adaptive parity query complexity). Note that Theorem 5.3 implies \(D_{\mathbb{F}}(f) \leq O(\sqrt{\| \hat{f} \|_0 \log \| \hat{f} \|_0})\). It was shown by Sanyal [San15] that one can achieve the same bound for \textit{non-adaptive} parity query complexity, which directly implies the same bound for one-way communication complexity for XOR functions.

Finally, one may try to improve the bound in Theorem 5.4 by finding a \(t \in \mathbb{F}_2^n\) such that
\[ \| \Delta f \|_0 \leq \| f \|_0^{2-\varepsilon} \] for some constant \( \varepsilon \), proving it will immediately improve the bound in Theorem 5.4 using exactly the same proof.

For the Sensitivity Conjecture, we believe either one of Conjecture 6.3 and Conjecture 6.4 would be easier to prove then the Sensitivity Conjecture itself. Note that proving any one of them implies the other is equivalent to the Sensitivity Conjecture, which is already a very interesting consequence since both of them are plausibly “easier” problems.

For the minimum certificate complexity (or minimum block sensitivity) versus sensitivity problem (cf. Conjecture 6.1 and Conjecture 6.3), the goal is to find a large subcube on which the function is constant. The key steps in the result of H. Hatami, Hosseini and Lovett [HHL16] is of this kind. In their paper [HHL16], they used results in Additive Combinatorics and an information theoretic argument to show that if an XOR function \( F(x, y) = f(x \oplus y) \) has low-cost communication protocol, then \( f \) must be constant on a large affine subspace. It would be interesting to see whether their ideas can be used to address the minimum certificate complexity versus sensitivity problem.

For the degree versus sensitivity problem (cf. Conjecture 6.2 and Conjecture 6.4), we believe that our approach using monomial graph can shed some light on it. We show that the study of polynomial sparsity is related to our monomial graph approach. As far as we know, the only papers that are relevant to our polynomial sparsity problem (cf. Conjecture 6.35) is the paper by Gopalan, Servedio, Tal and Wigderson [GSTW16] and Lovett, Tal and Zhang [LTZ17]. As we have shown, the average-case of Conjecture 6.35 follows directly from the switching lemma in the former paper. The latter paper improved the main result in [GSTW16], but unfortunately it does not imply a better average-case bound for Conjecture 6.35.

Finally, we believe that the study of the structure (say, combinatorial structure or other more refine structure than the degree) of the polynomial representation itself is of independent interest and may find applications other than proving the Sensitivity Conjecture.

References


[Tsa14] Hing Yin Tsang. On boolean functions with low sensitivity. manuscript, 2014. 4, 10, 16, 17


28


