Properties and Applications of Boolean Function Composition

Avishay Tal*

Abstract

For Boolean functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, the function composition of f and g denoted by $f \circ g : \{0,1\}^{nm} \to \{0,1\}$ is the value of f on n inputs, each of them is the calculation of g on a distinct set of m Boolean variables. Motivated by previous works that achieved some of the best separations between complexity measures such as *sensitivity, block-sensitivity, degree, certificate complexity* and *decision tree complexity* we show that most of these complexity measures behave multiplicatively under composition. We use this multiplicative behavior to establish several applications.

First, we give a negative answer for Adam Kalai's question from [MOS04]: *"Is it true that every Boolean function $f : \{0,1\}^n \to \{0,1\}$ with degree as a polynomial over the reals (denoted by $\deg(f)$) at most $n/3$, has a restriction fixing $2n/3$ of its variables under which f becomes a parity function?"* This question was motivated by the problem of learning juntas as it suggests a simple algorithm, faster than that of Mossel et al. We give a counterexample for the question using composition of functions strongly related to the Walsh-Hadamard code. In fact, we show that for every constants $\epsilon, \delta > 0$ there are (infinitely many) Boolean functions $f : \{0,1\}^n \to \{0,1\}$ such that $\deg(f) \leq \epsilon \cdot n$ and under any restriction fixing less than $(1-\delta) \cdot n$ variables, f does not become a parity function.

Second, we show that for composition, the *block sensitivity* (denoted by bs) property has an unusual behavior - namely that $\mathrm{bs}(f \circ g)$ can be larger than $\mathrm{bs}(f) \cdot \mathrm{bs}(g)$. We show that the ratio between these two has a strong connection to the integrality gap of the *Set Packing* problem. In addition, we obtain the best known separation between *block-sensitivity* and *certificate complexity* (denoted by C) giving infinitely many functions f such that $\mathrm{C}(f) \geq \mathrm{bs}(f)^{\log(26)/\log(17)} = \mathrm{bs}(f)^{1.149\ldots}$.

Last, we present a factor 2 improvement of a result by Nisan and Szegedy [NS94], by showing that for all Boolean functions $\mathrm{bs}(f) \leq \deg(f)^2$.

*Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel. Email: avishay.tal@weizmann.ac.il.

1 Introduction

Complexity measures such as *decision tree complexity*, *certificate complexity*, *block sensitivity*, *sensitivity* and *degree as a polynomial* study the low-level complexity of Boolean functions. The decision tree complexity of a Boolean function f, denoted by $\mathrm{D}(f)$, is the minimal number of bits needed to be read from the input in order to be certain of the value of f. Other measures mentioned above are relaxations of this measure. Their study has found many applications in other fields of complexity such as communication complexity, circuit lower bounds, quantum computation and concurrent computation. In this regime complexity refers to a more combinatorial property than computational. We say that two measures M and N are *polynomially related* if there are constants c_1 and c_2 such that for any Boolean function f we have $M(f) = O(N(f)^{c_1})$ and $N(f) = O(M(f)^{c_2})$. In a beautiful line of works, most of these features were proven to be *polynomially related*. In particular [BI87] showed that the deterministic decision tree complexity is at most quadratic in the non-deterministic decision tree complexity (usually called *certificate complexity*). After proving many such relations for different complexity measures, a natural question to ask is whether these inequalities are tight? One way to show tightness is to construct examples for which the gap is tight, for instance provide infinitely many fs such that $M(f) = \Omega(N(f)^{c_1})$.

As in most cases, using the right tool for the job is essential. In this paper we demonstrate the power of quite a simple tool, Boolean function composition (in short BFC), to prove several results on old and new complexity measures of Boolean functions. BFC is defined for any two Boolean functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$ as a function $f \circ g : \{0,1\}^{nm} \to \{0,1\}$ whose value is the value of f on n inputs, each of them is the calculation of g on an independent set of m bits. Figure 1 demonstrates a composition of two Boolean functions. The tool of BFC was used in several previous works to exhibit separations [1] between complexity measures of Boolean functions mentioned above. Generating separations is a very natural application of BFC as many of these complexity measures behave multiplicatively when composing two functions. Namely, for most complexity measures mentioned above $M(f \circ g) \leq M(f) \cdot M(g)$ and under some simple conditions on f and g equality holds. Using this property, one can get a series of functions, $\{f^k\}_{k\in\mathbb{N}}$ (where powering is defined as repeated composition of f to itself) with polynomial gaps between two such complexity measures from any specific function f with the slightest gap between the two complexities. To demonstrate this, say that you are given a function f with $M(f) = 2$ and $N(f) = 4$ where M and N behave multiplicatively under composition. This automatically gives a quadratic separation between the two as $\forall k \in \mathbb{N}\ .\ N(f^k) = 4^k = (2^k)^2 = (M(f^k))^2$.

1.1 Our Results and Techniques

One motivating question for our study was posed by Adam Kalai in [MOS04]. We give an equivalent form of it (the equivalence easily follows from the discussion in [MOS04]):

Question 1.1. *Is it true that for any Boolean function g on n bits which has degree as a polynomial over the reals (denoted $\deg(f)$) at most $n/3$, there is a restriction fixing at most $2/3n$ bits under which g becomes a parity function?*

[1] A separation between two complexities measures M, N is an infinite sequence of functions $\{f_i\}_{i\in\mathbb{N}}$ for which $M(f_i) = \omega(N(f_i))$ as $i \to \infty$. We say a separation is a polynomial separation if $M(f_i) = \Omega(N(f_i)^c)$ for some constant $c > 1$.

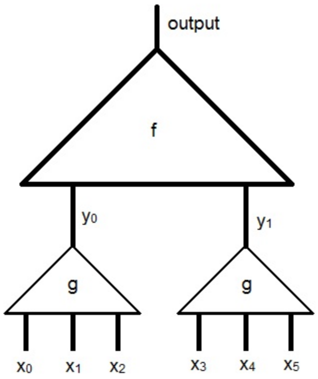

Figure 1: Composition of a 2-variable Boolean function with a 3-variable Boolean function

The question, interesting on its own, gives as an immediate application a learning algorithm for the *junta* problem if true. In short, in the junta learning problem one wants to learn a Boolean function defined over N variables that depends only on an unknown subset of size $n \ll N$ of them. A positive answer for A. Kalai's question gives a roughly $N^{2/3n}$ algorithm for the problem improving Mossel et al. analysis of roughly $N^{n\omega/(\omega+1)}$, where $\omega < 2.374$ is the matrix multiplication exponent. Recent work of [Val12] actually gives a roughly $N^{0.61n}$ learning algorithm for this problem, removing motivation from the possible application.

Let the *minimal restriction size* of g, denoted by $\mathrm{mr}(g)$, be the minimal number of bits needed to be fixed in order for g to become a parity function. We show that the answer to A. Kalai's question is a strong *no* as we demonstrate infinitely many Boolean functions with small degree over the reals, and high minimal restriction size.

Theorem 1. *For any $\epsilon > 0$ and $\delta > 0$ there are infinitely many integers n and Boolean functions $f : \{0,1\}^n \to \{0,1\}$ such that $\deg(f) \leq \epsilon \cdot n$ and $\mathrm{mr}(f) \geq (1-\delta) \cdot n$.*

The proof of this theorem relies on two important ingredients, one of them is a Boolean function derived from the Walsh-Hadamard code with $\mathrm{mr}(f) = n - o(n)$ and degree $n/2$. The second is the behavior of the minimal restriction size and the degree under composition. We show that $\mathrm{mr}(f \circ g) \geq \mathrm{mr}(f) \cdot \mathrm{mr}(g)$, while $\deg(f \circ g) = \deg(f) \cdot \deg(g)$. Combining the two together gives Theorem 1.

We introduce several definitions in order to describe the next problem.

Definition 1.2 (Sensitivity, Block Sensitivity)**.** *For $i \in \{0,\ldots,n-1\}$ denote by e_i the vector whose ith coordinate is 1 and the rest are zeros.*

The sensitivity *of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ on a point x, $\mathrm{sens}(f,x)$, is the number of coordinates $i \in \{0,1\ldots,n-1\}$ such that $f(x) \neq f(x \oplus e_i)$. The* sensitivity *of f, $\mathrm{sens}(f)$, is the maximal $\mathrm{sens}(f,x)$ over all xs.*

The block sensitivity *of f on x, $\mathrm{bs}(f,x)$, is the maximal number of disjoint blocks $B_1,\ldots,B_k \subseteq \{0,1,\ldots,n-1\}$ such that flipping each B_i flips f's value, i.e. $f(x) \neq f(x \oplus \bigoplus_{j \in B_i} e_j)$. The* block sensitivity *of f, $\mathrm{bs}(f)$, is the maximal $\mathrm{bs}(f,x)$ over all xs.*

Definition 1.3 (Certificate Complexity)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $x \in \{0,1\}^n$, a* certificate *for f on x is a set of variables $S \subseteq \{0,1,\dots,n-1\}$ such that any $y \in \{0,1\}^n$ satisfying $x_i = y_i$ for all $i \in S$ has $f(y) = f(x)$.*

The certificate complexity *of f on x is the size of the minimal such certificate. The* certificate complexity *of f, $\mathrm{C}(f)$, is the maximal $\mathrm{C}(f,x)$ over all xs. The* minimal certificate *for f, $\mathrm{C}_{\min}(f)$, is the minimal $\mathrm{C}(f,x)$ over all xs.*

In [Ver10] Elad Verbin noted that (under some condition) $\mathrm{bs}(f^k) \geq \mathrm{bs}(f)^k$, asking whether the converse $\mathrm{bs}(f^k) \leq \mathrm{bs}(f)^k$ is true. This may seem natural, as choosing a complexity measure M to be one of *sensitivity, certificate complexity, decision tree complexity* or *degree* we have $M(f \circ g) \leq M(f) \cdot M(g)$. We answer this question by giving infinitely many examples for which $\mathrm{bs}(f^2) \geq 1.26 \cdot \mathrm{bs}(f)^2$. The main ingredient in the analysis of block sensitivity of BFC is the integer program representation of the problem. We show several interesting results using this perspective. We Introduce two "new" complexity measures: *fractional block sensitivity* (denoted by fbs) which corresponds to the linear relaxation of the problem and *fractional certificate complexity* (denoted by FC) which corresponds to the dual linear program (see Section 5.1 for the exact formulations as an integer and linear program). We show that both measures are the same, and serve as an intermediate measure between block sensitivity and certificate complexity.

Theorem 2. *Let $f : \{0,1\}^n \to \{0,1\}$ and $x \in \{0,1\}^n$ then*

$$\mathrm{bs}(f,x) \leq \mathrm{fbs}(f,x) = \mathrm{FC}(f,x) \leq \mathrm{C}(f,x)$$

Actually, the "new" complexity measures are equal, up to a constant factor, to the randomized certificate complexity defined by Aaronson in [Aar03]. The next theorem shows the important role of fractional block sensitivity in the analysis of block-sensitivity of BFC.

Theorem 3. *Let $f : \{0,1\}^n \to \{0,1\}$, $g : \{0,1\}^m \to \{0,1\}$ be Boolean functions, then $\mathrm{bs}(f \circ g) \leq \mathrm{fbs}(f) \cdot \mathrm{bs}(g)$.*

In fact, we show that this is tight in some weak sense. We then obtain and analyze, using this theorem, the best known separation between block sensitivity and certificate complexity.

Surprisingly, the third application demonstrates that BFC can be used to tighten upper bounds. In particular, we give an improved upper bound to a result by [NS94]:

Theorem 4. *Let f be a Boolean function, then $\mathrm{bs}(f) \leq \deg(f)^2$.*

1.2 Related Work

In [BI87] the decision tree complexity, $\mathrm{D}(f)$, was polynomially related to $\mathrm{C}(f)$ as $\mathrm{C}(f) \leq \mathrm{D}(f) \leq \mathrm{C}(f)^2$. Saks and Wigderson used BFC in [SW86] to give polynomial separations between $\mathrm{C}(f)$, $\mathrm{D}(f)$ and the *Randomized Decision Tree complexity.* In [Nis89], Nisan defined block-sensitivity and showed that $\mathrm{bs}(f) \leq \mathrm{C}(f) \leq \mathrm{bs}(f) \cdot \mathrm{sens}(f)$. In [WZ89] BFC was used to give separations between $\mathrm{sens}(f)$ and $\mathrm{C}_{\min}(f)$, and the behavior of $\mathrm{C}_{\min}$ and sens with respect to composition was analyzed. They gave examples also considered to be the best separations till this work between block-sensitivity and certificate complexity, though the analysis was missing from their work as they didn't even consider block-sensitivity. In [NS94], the degree of a Boolean function as a polynomial over the reals was related to other complexity measures such as bs, C and D. There, it was shown that $\mathrm{bs}(f) \leq 2\deg(f)^2$. In addition, they used BFC to give a separation between sensitivity and degree showing an infinite family $\{f_i\}_{i\in\mathbb{N}}$, such that $\mathrm{sens}(f) = \deg(f)^{\log_2(3)}$. In [NW94], an example by Kushilevitz, using BFC once again, established a better separation between sensitivity and degree. This was used to show a separation between the communication complexity of a Boolean function and the log-rank of the associated matrix. De Wolf and Buhrman gave an excellent survey of the field in [BdW02] with some new unpublished results. Aaronson introduced the notion of *randomized* and *quantum certificate complexity* (denoted RC and QC respectively) in [Aar03] and gave separations between these complexities, block sensitivity and certificate complexity using BFC. Our use of strong duality to show equality between fbs and FC resembles his argument that $\mathrm{QC}(f) = \Theta(\sqrt{\mathrm{RC}(f)})$. Other works in quantum complexity area used BFC as a crucial tool, in particular using Ambainis's *quantum adversary method* (see [Amb03], [HLS07]).

2 Preliminaries

We use indices starting from 0 instead of 1, mainly because it makes more sense in section 4. We shall denote the set $\{0, 1, \ldots, n-1\}$ by $[n]$. For a set $S \subseteq [n]$, we shall denote 1_S as the characteristic vector of the set S, i.e. $(1_S)_i = 1$ iff $i \in S$. For an n-dimensional vector space over the field F, the standard basis is denoted by $\{e_i\}_{i\in[n]}$ where e_i denotes the vector with a 1 in the ith coordinate and zeros elsewhere.

f, g will usually denote Boolean functions. n, m will usually denote the number of variables of a Boolean function. ρ, τ will usually denote restrictions to Boolean functions.

Definition 2.1 (Boolean Function Restriction). *For a partial assignment $\rho : [n] \to \{0, 1, *\}$ and a Boolean function $f : \{0,1\}^n \to \{0,1\}$ the restriction of f to ρ, $f|_\rho$ is defined as $f|_\rho(x) = f(y)$ where*

$$y_i = \begin{cases} x_i & \rho(i) = * \\ \rho(i) & \rho(i) \in \{0,1\} \end{cases}$$

for $i \in [n]$. We say that ρ fixes the variables x_i with $\rho(i) \in \{0,1\}$. The size of ρ, denoted by $|\rho|$, is the number of fixed variables.

The function $f|_\rho$ is defined as a function with n variables, but actually only the non-fixed variables are relevant.

2.1 Complexity Measures for Boolean Functions

After defining sensitivity, block sensitivity and certificate complexity in the introduction, we formally define other complexity measures which we will discuss in this paper.

Definition 2.2 (Degree). *The degree of a Boolean function f, $\deg(f)$, is the degree of the unique multivariate multi-linear polynomial over the reals which agrees with f on $\{0,1\}^n$.*

Similarly one can define the degree of a Boolean function over the field $\mathbb{F}_2$.

Definition 2.3 (Decision Tree, [BdW02]). *A* deterministic decision tree *is a rooted ordered binary tree T. Each internal node of T is labeled with a variable x_i and each leaf is labeled with a value 0 or 1. Given an input $x \in \{0,1\}^n$, the tree is evaluated as follows. Start at the root. If this is a leaf then stop. Otherwise, query the variable x_i that labels the root. If $x_i = 0$, then recursively* *evaluate the left subtree, if $x_i = 1$ then recursively evaluate the right subtree. The output of the tree is the value (0 or 1) of the leaf that is reached eventually. Note that an input x deterministically determines the leaf, and thus the output, that the procedure ends up in. We say a decision tree computes f if its output equals $f(x)$, for all $x \in \{0,1\}^n$. The* decision tree complexity *of f, $\mathrm{D}(f)$, is the minimal depth of a decision tree computing f.*

Definition 2.4 (Minimal Restriction Size)**.** *A Boolean function f is a* parity function *(or affine over $\mathbb{F}_2$) if we can write*

$$f(x) = c + \sum_{i \in S} x_i \pmod 2$$

for some $S \subseteq [n]$ and $c \in \{0,1\}$.

The minimal restriction size *of a Boolean function f,* $\mathrm{mr}(f)$*, is the minimal size of a restriction ρ for which $f|_\rho$ is a parity function.*

Note that $\mathrm{C}_{\min}(f) \geq \mathrm{mr}(f)$ as any certificate induces a restriction under which f is constant and in particular a parity function. We introduce the following notation.

Definition 2.5. *For a Boolean function f, $z \in \{0,1\}$ and a measure $M \in \{\mathrm{sens}, \mathrm{bs}, \mathrm{C}\}$ we denote by*

$$M^z(f) \triangleq \max_{x: f(x)=z} M(f,x)$$

2.2 Discrete Fourier Transform

In this section we briefly describe the discrete Fourier transform of Boolean functions - giving notations and stating known facts about it. For a more thorough introduction to the field we suggest the first chapters in O'Donnell's book/blog ([O'D12]).

A discrete Fourier transform is a representation of a Boolean function as a polynomial. Let $f : \{0,1\}^n \to \{0,1\}$. It is convenient to consider f as a function from $\{\pm 1\}^n$ to $\{\pm 1\}$ using the mapping $b \mapsto (-1)^b$, which is equivalent to the linear mapping $b \mapsto 1 - 2b$ restricted to $\{0,1\}$. More formally, we define $F : \{-1,1\}^n \to \{-1,1\}$ by

$$F(x_0, \ldots, x_{n-1}) = 1 - 2 \cdot f\left(\frac{1-x_0}{2}, \frac{1-x_1}{2}, \ldots, \frac{1-x_{n-1}}{2}\right).$$

Using this representation, one can define inner product of Boolean functions as $< f, g >= 1/2^n \cdot \sum_{x \in \{-1,1\}^n} f(x) \cdot g(x)$ where all operations are done in $\mathbb{R}$. For any $S \subseteq [n]$ the S-Fourier character is defined as $\chi_S(x) \triangleq \prod_{i \in S} x_i$. The Fourier characters form an orthonormal basis for the vector space of functions mapping $\{\pm 1\}^n \to \mathbb{R}$. The Fourier transform of F is the representation of F as a linear combination in this basis:

$$F(x) = \sum_{S \subseteq [n]} \chi_S(x) \cdot \hat{F}(S) \ ,$$

where the coefficient $\hat{F}(S)$ is called the S-Fourier coefficient of F. Note that this represents F as a multilinear polynomial. By orthonormality of the characters,

$$\hat{f}(S) \triangleq \hat{F}(S) = \langle \chi_S, F \rangle = 1/2^n \cdot \sum_{x \in \{-1,1\}^n} \chi_S(x) \cdot F(x) = E_{x \in_R \{0,1\}^n}((-1)^{\sum_{i \in S} x_i} \cdot (-1)^{f(x)}) \quad (1)$$

The following fact relates the Fourier coefficients of a Boolean function and its restriction.

Fact 2.6. *Let $f : \{0,1\}^n \to \{0,1\}$, $i \in [n]$, $c \in \{0,1\}$ and $S \subseteq [n] - \{i\}$, then the S-Fourier coefficient of the restricted function $(\widehat{f|_{x_i=c}})(S)$ is equal to $\hat{f}(S) + (-1)^c \cdot \hat{f}(S \cup \{i\})$*

Thus, the largest (in absolute value) Fourier coefficient of $f|_{x_i=c}$ is at most twice the largest Fourier coefficient of f. Induction gives as an immediate corollary the following:

Corollary 2.7. *Let $f : \{0,1\}^n \to \{0,1\}$, and ρ be a restriction fixing k variables, then*

$$\|\widehat{f|_\rho}\|_\infty \leq \|\hat{f}\|_\infty \cdot 2^k$$

The next fact relates the Fourier transform of a Boolean function and the Fourier transform of the function composed with a linear transformation.

Fact 2.8. *Let $f : \{0,1\}^n \to \{0,1\}$ and $A \in (\mathbb{F}_2)^{n \times m}$ then the composition $f \circ A : \{0,1\}^m \to \{0,1\}$ has Fourier spectrum $\widehat{f \circ A}(T) = \sum_{S: A^t \cdot 1_S = 1_T} \hat{f}(S)$,$\forall T \subseteq [m]$ where matrix-vector multiplication is over $\mathbb{F}_2$.*

Proof. Expressing $(f \circ A)$ using the Fourier transform of f

$$\begin{aligned} (f \circ A)(x) &= f(A(x)) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot (-1)^{1_S^t \cdot (A \cdot x)} \\ &= \sum_{S \subseteq [n]} \hat{f}(S) \cdot (-1)^{(A^t \cdot 1_S)^t \cdot x} \\ &= \sum_{T \subseteq [m]} (-1)^{(1_T)^t \cdot x} \cdot \sum_{S: A^t \cdot 1_S = 1_T} \hat{f}(S) \end{aligned}$$

In the last equality we used the fact that when multiplying A by a vector over the field $\mathbb{F}_2$ the result must be a Boolean vector. □

As a special case if $A \in (\mathbb{F}_2)^{n \times n}$ is invertible, then the Fourier spectrum of f and $f \circ A$ are a (linear) permutation of one another.

2.3 Function Composition

Definition 2.9 (Function Composition)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, then the function composition of f and g, $f \circ g : \{0,1\}^{nm} \to \{0,1\}$ is defined as follows:*

$$(f \circ g)\left(x_0^0, x_1^0, \ldots, x_{m-1}^{n-1}\right) = f\left(g\left(x_0^0, x_1^0, \ldots, x_{m-1}^0\right), \ldots, g\left(x_0^{n-1}, x_1^{n-1} \ldots, x_{m-1}^{n-1}\right)\right)$$

Definition 2.10 (Function Powering)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $k \in \mathbb{N}$, then the kth power of f denoted by f^k is defined recursively by $f^1 \triangleq f$ and $f^k \triangleq f \circ (f^{k-1})$ for $k > 1$.*

Definition 2.11 (Good Form)**.** *For $f : \{0,1\}^n \to \{0,1\}$, $M \in \{\mathrm{sens}, \mathrm{bs}, \mathrm{C}\}$ [2] and $z \in \{0,1\}$ we say that f is in (M,z)-good form if $f(z^n) = z$ and $M(f, z^n) = M(f)$.*

Remark 2.12. Any function can be transformed to an (M,z)-good form by applying

$$\tilde{f}(x) = f(b) \oplus z \oplus f(x \oplus b \oplus z^n)$$

for $b \in \{0,1\}^n$ which maximizes the measure M (i.e. $M(f,b) = M(f)$). This transformation does not change the measures $\mathrm{sens}, \mathrm{bs}, \mathrm{C}, \mathrm{C}_{\min}, \mathrm{mr}, \mathrm{D}$ and $\deg$ of the Boolean function. This definition is useful as functions in M-good form have a better lower bound on measure M of their composition.

[2] And also for $M \in \{\mathrm{fbs}, \mathrm{FC}\}$ which we define later

3 Function Composition Properties

The following basic lemma relates complexity measures of functions and their compositions. Properties 5 and 9 were previously proven in [WZ89].

Lemma 3.1 (Function Composition Properties)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$ be Boolean functions, then the following holds:*

1. $\deg(f \circ g) = \deg(f) \cdot \deg(g)$
2. $\mathrm{D}(f \circ g) = \mathrm{D}(f) \cdot \mathrm{D}(g)$
3. $\mathrm{C}(f \circ g) \le \mathrm{C}(f) \cdot \mathrm{C}(g)$
4. *for* $z \in \{0,1\}$*, if* $f(z^n) = g(z^m) = z$ *then* $\mathrm{C}(f \circ g, z^{nm}) \ge \mathrm{C}(f, z^n) \cdot \mathrm{C}(g, z^m)$
5. $\mathrm{sens}(f \circ g) \le \mathrm{sens}(f) \cdot \mathrm{sens}(g)$
6. *for* $z \in \{0,1\}$*, if* $f(z^n) = g(z^m) = z$ *then* $\mathrm{sens}(f \circ g, z^{nm}) \ge \mathrm{sens}(f, z^n) \cdot \mathrm{sens}(g, z^m)$
7. *for* $z \in \{0,1\}$*, if* $f(z^n) = g(z^m) = z$ *then* $\mathrm{bs}(f \circ g, z^{nm}) \ge \mathrm{bs}(f, n^z) \cdot \mathrm{bs}(g, z^m)$
8. $\mathrm{mr}(f \circ g) \ge \mathrm{mr}(f) \cdot \mathrm{C}_{\min}(g)$ *(* $\ge \mathrm{mr}(f) \cdot \mathrm{mr}(g)$*)*
9. $\mathrm{C}_{\min}(f \circ g) \ge \mathrm{C}_{\min}(f) \cdot \mathrm{C}_{\min}(g)$

In particular for $M \in \{\mathrm{bs}, \mathrm{sens}, \mathrm{C}\}$ *if* f, g *are in* (M, z) *good form* $M(f \circ g) \ge M(f) \cdot M(g)$*.*

Proof. We will use in the proofs of these properties the following notations:

$$(f \circ g)\left(x_0^0, x_1^0, \ldots, x_{m-1}^{n-1}\right) = f\left(g\left(x_0^0, x_1^0, \ldots, x_{m-1}^0\right), \ldots, g\left(x_0^{n-1}, x_1^{n-1} \ldots, x_{m-1}^{n-1}\right)\right)$$

x^i will denote the m-bit input to g, $(x_0^i, \ldots, x_{m-1}^i)$. We will sometime abuse notation and denote by x^i as the set of variables $\{x_j^i\}_{j\in[m]}$. $y_0, \ldots, y_{n-1}$ will denote the input variables to f. We will refer to $g(x^i)$ as g_i.

1. $\deg(f \circ g) = \deg(f) \cdot \deg(g)$. It is obvious that $\deg(f \circ g) \le \deg(f) \cdot \deg(g)$. In the other direction, taking a maximal monomial in f, $c \cdot \prod_{i \in T} y_i$ and a maximal monomial in g, $d \cdot \prod_{j \in S} x_j$, the monomial $c \cdot d^{|T|} \cdot \prod_{i \in T, j \in S} x_j^i$ is in the expansion of $f \circ g$ and there is no way canceling it out. Hence, there is a monomial with degree $|T| \cdot |S| = \deg(f) \cdot \deg(g)$ and coefficient $c \cdot d^{|T|} \neq 0$ in $f \circ g$ expansion.

2. $\mathrm{D}(f \circ g) = \mathrm{D}(f) \cdot \mathrm{D}(g)$. We start by showing $\mathrm{D}(f \circ g) \le \mathrm{D}(f) \cdot \mathrm{D}(g)$. Given optimal decision trees for f and g we construct a decision tree for $f \circ g$. We transform the decision tree for f to a decision tree for $f \circ g$ by replacing each inspection of y_i by a decision tree of depth $\mathrm{D}(g)$ on x^i. This gives a decision tree for $f \circ g$ of depth at most $\mathrm{D}(g) \cdot \mathrm{D}(f)$.

   For the other direction, we use an analogous characterization for the decision tree complexity called the *Adversary Argument* (see [AB09], chapter 12). The decision tree complexity of a Boolean function is the maximal d such that there is a way to answer any less than d adaptively chosen queries in such a way that the answers do not determine f's value. We call this the f*-Adversarial strategy*.

We describe an adversarial strategy for $f \circ g$. When asked about a variable x^i_j such that less than $\mathrm{D}(g)$ variables of x^i were examined up till this point, we will answer according to the g-adversarial strategy, leaving the value of g_i undetermined. Whenever we must determine the value of g_i (i.e. after at least $\mathrm{D}(g)$ queries on variables in x^i), we will pick the value for g_i according to the value we would have given y_i in an f-adversarial strategy. The decision is not certain until at least $\mathrm{D}(f)$ of the g_is were determined. Thus, the decision tree complexity is at least $\mathrm{D}(f) \cdot \mathrm{D}(g)$.

3. $\mathrm{C}(f \circ g) \leq \mathrm{C}(f) \cdot \mathrm{C}(g)$. Let $x \in \{0,1\}^{nm}$ be an input maximizing $\mathrm{C}(f \circ g)$, and let $y \in \{0,1\}^n$ be the intermediate values $y_i = g(x^i)$. f has a certificate S for y of size $\mathrm{C}(f,y) \leq \mathrm{C}(f)$, and g has certificates T_i for each x^i of size $\mathrm{C}(g,x^i) \leq \mathrm{C}(g)$. By definition, the union of certificates $\cup_{i \in S} T_i$ is a certificate for $f \circ g$ at input x: any input $\tilde{x}$ agreeing with x on the certificate has $g(\tilde{x}^i) = g(x^i)\ \forall i \in S$ and since S is a certificate for f we conclude that $(f \circ g)(\tilde{x}) = (f \circ g)(x)$.

4. for $z \in \{0,1\}$, if $f(z^n) = g(z^m) = z$ then $\mathrm{C}(f \circ g, z^{nm}) \geq \mathrm{C}(f, z^n) \cdot \mathrm{C}(g, z^m)$. We show this property for $z = 0$, the case $z = 1$ is similar. Consider the shortest certificate for f on 0^{nm} and assume by contradiction that its size is strictly less than $\mathrm{C}(f, 0^n) \cdot \mathrm{C}(g, 0^m)$. This means that for at most $\mathrm{C}(f, 0^n) - 1$ of $i \in [n]$ the certificate contains $\mathrm{C}(g, 0^m)$ variables or more from x^i. The value of g_i for x^i with less than $\mathrm{C}(g, 0^m)$ variables in the certificate is undetermined under the restriction the certificate induces, since otherwise g would have a shorter certificate on 0^m. This means that the value of g_i is determined for at most $\mathrm{C}(f, 0^n) - 1$ different $i \in [n]$, which shows that the value of $f \circ g$ is undetermined under the restriction the certificate induces, in contradiction to the assumption.

5. $\mathrm{sens}(f \circ g) \leq \mathrm{sens}(f) \cdot \mathrm{sens}(g)$. Let x be an input maximizing $\mathrm{sens}(f \circ g)$ and let $y_i = g(x^i)$ be the intermediate values. Any sensitive coordinate x^i_j must be a sensitive coordinate for g on input x^i, and y_i must be a sensitive coordinate for f on input $(y_0, \ldots, y_{n-1})$. Hence, there are at most $\mathrm{sens}(f) \cdot \mathrm{sens}(g)$ sensitive coordinates for $f \circ g$.

6. for $z \in \{0,1\}$, if $f(z^n) = g(z^m) = z$ then $\mathrm{sens}(f \circ g, z^{nm}) \geq \mathrm{sens}(f, z^n) \cdot \mathrm{sens}(g, z^m)$. We show this property for $z = 0$, the case $z = 1$ is similar. Choosing the input $x = 0^{nm}$, the assumption on g gives that the intermediate values $y_i = g(x^i) = 0$. Since $y = 0$, there are $\mathrm{sens}(f, 0^n)$ many y_is for which if we change their value, the value of $f \circ g$ will change. For each such y_i there are $\mathrm{sens}(g, 0^m)$ variables of x^i such that changing their value from 0 to 1 changes y_i from 0 to 1, and then changing the value of $f \circ g$ from 0 to 1.

7. for $z \in \{0,1\}$, if $f(z^n) = g(z^m) = z$ then $\mathrm{bs}(f \circ g, z^{nm}) \geq \mathrm{bs}(f, n^z) \cdot \mathrm{bs}(g, z^m)$. We show this property for $z = 0$, the case $z = 1$ is similar. Choosing the input $x = 0^{nm}$, the assumption on g gives that the intermediate values $y_i = g(x^i) = 0$. For $i \in [n]$, pick $\mathrm{bs}(g, 0^m)$ disjoint subsets for each set of variables x^i that flips g's value on the 0^m input, denote them by $T^i_1, T^i_2, \ldots, T^i_{\mathrm{bs}(g,0^m)}$. Let $B_1, \ldots, B_{\mathrm{bs}(f,0^n)} \subseteq [n]$ be a maximal set of sensitive blocks for f on input 0^n. For $k = 1, \ldots, \mathrm{bs}(f, 0^n)$, and $\ell = 1, \ldots, \mathrm{bs}(g, 0^m)$ put $A_{k,\ell} \triangleq \cup_{i \in B_k} T^i_\ell$. This defines $\mathrm{bs}(f, 0^n) \cdot \mathrm{bs}(g, 0^m)$ disjoint subsets of $\{0,1\}^{nm}$ that flips the value of $f \circ g$ on 0^{nm}.

8. $\mathrm{mr}(f \circ g) \geq \mathrm{mr}(f) \cdot \mathrm{C}_{\min}(g)$. Let $\rho : [nm] \to \{0,1,*\}$ be a restriction fixing at most $\mathrm{mr}(f) \cdot \mathrm{C}_{\min}(g) - 1$ variables. In at most $\mathrm{mr}(f) - 1$ of the sets x^i, ρ fixes $\mathrm{C}_{\min}(g)$ or more variables. Hence, under the restriction at least $n - \mathrm{mr}(f) + 1$ of the g_is are nonconstants. Assume by

contradiction that the function $(f \circ g)|_\rho$ is affine over $\mathbb{F}_2$. Let $\tau : \{0, 1, \dots, n-1\} \to \{0, 1, *\}$ be the induced restriction of ρ on f, i.e.

$$\tau_i = \begin{cases} * & \rho \text{ does not set the value of } g(x^i) \\ 0 & \rho \text{ sets the value of } g(x^i) \text{ to } 0 \\ 1 & \rho \text{ sets the value of } g(x^i) \text{ to } 1 \end{cases} .$$

For every nonconstant g_i pick two assignments ϕ_i^0, ϕ_i^1 for x^i agreeing with ρ under which g equals 0 and 1 respectively. For g_i that is fixed under ρ pick any two assignments ϕ_i^0, ϕ_i^1 agreeing with ρ. For a bit $b \in \mathbb{F}_2$ and two vectors $u, v \in \mathbb{F}_2^m$ define $\mathrm{MUX}_b(u, v) \triangleq (1-b) \cdot u \oplus b \cdot v$. The bit b chooses between two vectors, if $b = 0$ the output is u and if $b = 1$ the output is v. For constant vectors u, v, $\mathrm{MUX}_b(u, v)$ is affine over $\mathbb{F}_2$ as a function of b. Let

$$h(x_0, x_1, \dots, x_{n-1}) = ((f \circ g)|_\rho)(\mathrm{MUX}_{x_0}(\phi_0^0, \phi_0^1), \mathrm{MUX}_{x_1}(\phi_1^0, \phi_1^1), \dots, \mathrm{MUX}_{x_{n-1}}(\phi_{n-1}^0, \phi_{n-1}^1)) .$$

h is affine over $\mathbb{F}_2$ as a composition of affine transformations. Furthermore, h equals $f|_\tau$ and hence $f|_\tau$ is affine, and we reach a contradiction as τ fixes less than $\mathrm{C}_{\min}(f)$ variables.

9. $\mathrm{C}_{\min}(f \circ g) \geq \mathrm{C}_{\min}(f) \cdot \mathrm{C}_{\min}(g)$. As in 8, fixing less than $\mathrm{C}_{\min}(f) \cdot \mathrm{C}_{\min}(g)$ variables guarantees that at least $n - (\mathrm{C}_{\min}(f) - 1)$ of the g_is are nonconstant and thus $f \circ g$ is nonconstant.

□

Using Lemma 3.1 it is easy to generate polynomial separations between two complexity measures. Let $M \in \{\deg, \mathrm{D}, \mathrm{C}, \mathrm{sens}\}$ and $N \in \{\deg, \mathrm{D}, \mathrm{C}, \mathrm{sens}, \mathrm{bs}, \mathrm{mr}, \mathrm{C}_{\min}\}$ be two complexity measures and suppose we are given a function f, such that $M(f) < N(f)$. If $N \in \{\mathrm{sens}, \mathrm{bs}, \mathrm{C}\}$, we can assume WLOG that f is in $(N, 0)$ good form. Then, taking powers of this function gives infinitely many examples such that $M(f^k) \leq M(f)^k$ and $N(f^k) \geq N(f)^k$ for $k \in \mathbb{N}$. This is a polynomial separation as

$$N(f^k) \geq N(f)^k = M(f)^{k \cdot \log N(f) / \log M(f)} \geq M(f^k)^{\log N(f) / \log M(f)} .$$

Remark 3.2. Note that for $f = \mathrm{OR}_n$ and $g = \mathrm{AND}_n$ the sensitivity, block-sensitivity and certificate complexity of $f \circ g$ is n. This shows that being in M-good form is necessary to achieve $M(f \circ g) \geq M(f) \cdot M(g)$.

4 A. Kalai's Question

In this section we give a negative answer for Question 1.1. First, we construct a function based on the Walsh-Hadamard code. Then, we take powers of this function showing that this is a counterexample for the question.

4.1 Boolean Functions from The Walsh-Hadamard Code

Example 4.1. *Let $n \in \mathbb{N}$, we define* $\mathrm{WHIP_n} : \{0,1\}^{2^{2n}} \to \{0,1\}$ *as follows. Denote by* $\mathrm{IP_n} : \{0,1\}^{2n} \to \{0,1\}$ *the Boolean function defined as:*

$$\mathrm{IP_n}(y_0, y_1, \dots, y_{2n-1}) = y_0 \cdot y_1 + \ldots + y_{2n-2} \cdot y_{2n-1} \pmod 2 .$$

For a non-negative integer k, denote by k_i the i'th least significant bit of k in its binary representation. Let $\mathrm{WH} \in (\mathbb{F}_2)^{2n \times 2^{2n}}$ *be the matrix defined as* $\mathrm{WH}_{i,k} = k_i$ *for* $i \in [2n], k \in [2^{2n}]$*. Finally, we define* $\mathrm{WHIP_n}$ *as:*

$$\mathrm{WHIP_n}(x) = \mathrm{IP_n}(\mathrm{WH} \cdot x)$$

where the matrix vector multiplication $\mathrm{WH} \cdot x$ *is done over* $\mathbb{F}_2$*.*

This example is a degree 2 polynomial over $\mathbb{F}_2$. Yet, as the next theorem states, f requires a lot of variables to be fixed in order to become a parity function i.e. a degree 1 function over $\mathbb{F}_2$.

Theorem 4.2. $\deg(\mathrm{WHIP_n}) = 2^{2n-1}$ *and* $\mathrm{mr}(\mathrm{WHIP_n}) = 2^{2n} - 2^n$*.*

In order to prove Theorem 4.2 we prove several lemmas first, showing the role of WH and $\mathrm{IP_n}$ in the construction. In the following proofs it is convenient to associate a restriction with an affine vector subspace.

Definition 4.3. *Let* $\rho : [m] \to \{0, 1, *\}$ *be a restriction, then the inputs that agrees with* ρ *form an affine subspace of* $(\mathbb{F}_2)^m$ *defined by* $v_\rho + V_\rho$ *where* $V_\rho = \mathrm{span}\{e_i : \rho(i) = *\}$ *and* $(v_\rho)_i = 1$ *iff* $\rho(i) = 1$ *and* 0 *elsewhere.*

Lemma 4.4 (The "Role" of WH)**.** *For* $S \subseteq [2^{2n}]$*, let* $V = \mathrm{span}(\{e_k\}_{k \in S})$ *be a linear subspace of* $(\mathbb{F}_2)^{2^{2n}}$ *then* $|S| = \dim(V) \le 2^{\dim(\mathrm{WH}(V))}$*.*

Proof. Let $d = \dim(\mathrm{WH}(V))$, and let $U \subseteq (\mathbb{F}_2)^{2n}$ be the orthogonal complement of $\mathrm{WH}(V)$ (with respect to the natural bilinear form $B(x, y) = x^t \cdot y$ over $\mathbb{F}_2$). Fix some basis $\{u_j\}_{j \in [2n-d]}$ for U. For any $k \in S$ we have that $\mathrm{WH}(e_k) \perp U$, thus

$$\forall\ j \in [2n - d]\ ,\ 0 = \sum_{i=0}^{2n-1} (u_j)_i \cdot (\mathrm{WH} \cdot e_k)_i = \sum_{i=0}^{2n-1} (u_j)_i \cdot k_i\ .$$

So each $k \in S$ is a solution to a system of $2n - d$ linearly independent linear equations in its binary representation as an integer. Thus, the number of such indices is at most 2^d. □

We state some slight variant of a known property of $\mathrm{IP_n}$ whose proof is given in Appendix A.

Lemma 4.5. *Let* $A \subseteq (\mathbb{F}_2)^{2n}$ *be an affine subspace of dimension* d *defined by* $v_A + \mathrm{span}\{v_0, \ldots, v_{d-1}\}$*. If* $\mathrm{IP_n}|_A(x_0, \ldots, x_{d-1}) \triangleq \mathrm{IP_n}(v_A + \sum_{i \in [d]} x_i \cdot v_i)$ *is a parity function, then* $d \le n$*.*

This property is essential to achieve the next lemma.

Lemma 4.6 (The "Role" of $\mathrm{IP_n}$)**.** *Let* $f = \mathrm{IP_n} \circ A$ *where* $A \in (F_2)^{2n \times m}$ *is a linear transformation. Let* ρ *be a restriction, under which* $\mathrm{IP_n} \circ A$ *is a parity function, then* $\dim(A(V_\rho)) \le n$*.*

Proof. Denote by $V' = v_\rho + V_\rho$, $U = A(V_\rho)$, $u_\rho = A(v_\rho)$ and let $d = \dim(A(V_\rho))$. We write $U = A(V_\rho)$ as $\mathrm{span}\{u_0, u_1, \ldots, u_{d-1}\}$ where $\{u_i\}_{i \in [d]}$ are linearly independent vectors in $(\mathbb{F}_2)^{2n}$. For $i \in [d]$, fix $v_i \in V_\rho$ to be some arbitrary preimage of u_i by the mapping A. We have

$$\begin{aligned}
\mathrm{IP_n}|_U(x_0, \ldots, x_{d-1}) &= \mathrm{IP_n}(u_\rho + \sum_{i \in [d]} u_i \cdot x_i) \\
&= (\mathrm{IP_n} \circ A)(v_\rho + \sum_{i \in [d]} v_i \cdot x_i) \\
&= f|_\rho(v_\rho + \sum_{i \in [d]} v_i \cdot x_i)\ .
\end{aligned}$$

Using our assumption on $f|_\rho$, this shows that $\mathrm{IP_n}|_U$ is a parity function and by Lemma 4.5 this gives $d \leq n$. □

We are ready to prove Theorem 4.2.

Proof. First, we show that $\deg(\mathrm{WHIP_n}) = 2^{2n-1}$. By Fact 2.8 the nonzero Fourier coefficients of $\mathrm{WHIP_n}$ are $\mathrm{WH}^t \cdot 1_S$ for $S \subseteq [2n]$. Opening this up gives that the nonzero Fourier coefficients are the sets

$$\left\{ k \in [2^{2n}] \;\middle|\; \bigoplus_{i \in S} k_i = 1 \right\}$$

for $S \subseteq [2n]$. The size of each such set is either 0 for $S = \emptyset$ or 2^{2n-1} otherwise, hence the degree of $\mathrm{WHIP_n}$ over the reals is 2^{2n-1}.

Next, we show $\mathrm{mr}(\mathrm{WHIP_n}) \geq 2^{2n} - 2^n$. Let ρ be a restriction under which $\mathrm{WHIP_n}$ becomes a parity function. By Lemma 4.6, $\dim(\mathrm{WH}(V_\rho)) \leq n$. Using Lemma 4.4 we have $\dim(V_\rho) \leq 2^n$, hence ρ fixes at least $2^{2n} - 2^n$ variables.

On the other hand, there's a restriction fixing $2^{2n} - 2^n$ variables that makes $\mathrm{WHIP_n}$ a constant, by fixing $x_k = 0$ for any k such that $(k_0, k_2, \ldots, k_{2n-2}) \neq 0^{\mathrm{n}}$. □

Remark 4.7. The matrix used in the construction of Example 4.1 is the transpose of the encoding matrix of the Walsh-Hadamard error correcting code (alternatively, the parity-check matrix for the Hamming code). The above construction can be generalized to use any ϵ-biased sample space instead of the Walsh-Hadamard code, resulting in similar guarantees on the minimal restriction size and degree. Actually, as the Walsh-Hadamard code is an ϵ-biased sample space with $\epsilon = 0$ it can viewed as a special case of this generalization. We discuss this in Appendix B

4.2 Counterexample for A. Kalai's Question

We combine Theorem 4.2 and Property 8 from Lemma 3.1 to get a counterexample for A. Kalai's question.

Theorem (Theorem 1, restated)**.** *For any $\epsilon > 0$ and $\delta > 0$ there exists infinitely many integers N and Boolean functions $f : \{0,1\}^N \to \{0,1\}$ such that $\deg(f) \leq \epsilon \cdot N$ and $\mathrm{mr}(f) \geq (1-\delta) \cdot N$. In particular, for $\epsilon = \delta < 1/3$ this answers A. Kalai's question.*

Proof. Put $m = \lceil \log_2(\frac{1}{\epsilon}) \rceil$ and choose any integer $n \geq \log_2(\frac{m}{\delta})$ (equivalently $\delta \geq m/2^n$). Take the function g to be $(\mathrm{WHIP_n})^m$. g is defined over $N \triangleq 2^{2nm}$ variables. Applying Property 1 from Lemma 3.1 gives $\deg(g) = (2^{2n-1})^m = 2^{2nm} \cdot 1/2^m \leq \epsilon \cdot N$. On the other hand, applying Property 8 from the same lemma gives

$$\mathrm{mr}(g) \geq \mathrm{mr}(\mathrm{WHIP_n})^m = \left(2^{2n} - 2^n\right)^m = 2^{2nm} \cdot \left(1 - \frac{1}{2^n}\right)^m \underset{\text{Bernoulli}}{\geq} N \cdot \left(1 - \frac{m}{2^n}\right) \geq N \cdot (1-\delta) \ .$$

□

4.3 Polynomial Separations between Degree and Minimal Restriction Size

We demonstrate an infinite family of functions with a polynomial separation between deg and mr

Claim 4.8. *There are infinitely many Boolean functions f such that* $\mathrm{mr}(f) \geq 2/3 \cdot \deg(f)^{log_2 3} = 2/3 \cdot \deg(f)^{1.58\ldots}$.

Proof. The function

$$f(x_0, x_1, x_2, x_3) = (x_0 + x_1) \cdot (x_2 + x_3) + x_0 + x_2 \pmod 2$$

defined over $\{0,1\}^4$ has $\deg(f) = 2$, $\mathrm{C}_{\min}(f) = 3$, $\mathrm{mr}(f) = 2$. Applying Lemma 3.1, $g = f^k$ has $\deg(g) = 2^k$ and

$$\mathrm{mr}(g) \geq \mathrm{mr}(f) \cdot \mathrm{C}_{\min}(f^{k-1}) \geq \mathrm{mr}(f) \cdot \mathrm{C}_{\min}(f)^{k-1} = 2 \cdot 3^{k-1} = 2/3 \cdot \deg(g)^{log_2 3} .$$

□

Note that these functions are not counterexamples for A. Kalai's question as their minimal restriction size is less than $2/3n$.

5 Block-Sensitivity of Function Composition

One may note that a natural bound is missing from Lemma 3.1: "$\mathrm{bs}(f \circ g) \leq \mathrm{bs}(f) \cdot \mathrm{bs}(g)$". While this bound may seem reasonable at first glance, it turns out to be false, as we demonstrate next.

We first show that block sensitivity is actually a subcase of the NP-Complete problem *Set Packing*. The maximal gap between $\mathrm{bs}(f \circ g)$ and $\mathrm{bs}(f) \cdot \mathrm{bs}(g)$ is closely related to the integrality gap between the value of the integer program (IP) for Set Packing and the value of its standard linear relaxation. The linear program (LP) gives tight upper bounds on $\mathrm{bs}(f \circ g)$. The integer/linear programming approach shed interesting light on the problem of block-sensitivity, relating it to a relaxed version which we call *Fractional Block Sensitivity*. In fact, $\mathrm{bs}(f) \leq \mathrm{C}(f)$ is just a corollary of LP duality.

We use the IP/LP formulation of block sensitivity to analyze the powers of a previously known example, demonstrating the gaps between $\mathrm{bs}(f)^2$ and $\mathrm{bs}(f^2)$.

5.1 Block Sensitivity as a Special Case of Set Packing

The problem of (unweighted) *Set Packing* is the following. Given sets $B_0, B_1, \ldots, B_{k-1} \subseteq [n]$ find the largest collection of sets not intersecting. This seems familiar, as the block sensitivity of $f : \{0,1\}^n \to \{0,1\}$ on a given $x \in \{0,1\}^n$ is exactly that problem, where the sets are $B \subseteq [n]$ such that $f(x \oplus 1_B) \neq f(x)$. WLOG, we can consider only subsets minimal to inclusion. Consider the following integer program for Set Packing:

$$\begin{array}{lll} \max & \sum_{j=0}^{k-1} w_j & \\ \text{s.t.} & \sum_{j: i \in B_j} w_j \leq 1 & \forall i \in [n] \\ & w_j \in \{0,1\} & \forall j \in [k] \end{array} \qquad \text{(IP(bs(f)))}$$

The variable w_j equals 1 iff we choose B_j to our collection. Note that the first constraint states that each coordinate is covered at most once. The linear relaxation of this problem is:

$$\begin{array}{lll} \max & \sum_{j=0}^{k-1} w_j & \\ & & \\ \text{s.t.} & \sum_{j:i\in B_j} w_j \leq 1 & \forall i \in [n] \\ & w_j \geq 0 & \forall j \in [k] \end{array} \tag{LP(bs(f))}$$

Next, we define the *Fractional Block Sensitivity* which captures the value of this linear program.

Definition 5.1 (Fractional Block Sensitivity)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and $x \in \{0,1\}^n$, then the* Fractional Block Sensitivity $\mathrm{fbs}(f,x)$ *of f on x, is the maximal sum of non-negative weights assigned to a collection of blocks, $B_0, \dots, B_{k-1} \subseteq [n]$ such that for each B_j, $f(x) \neq f(x \oplus 1_{B_j})$ and each coordinate $i \in [n]$ is contained by blocks with total weight at most 1 (relaxed disjointness). The fractional block sensitivity of f,* $\mathrm{fbs}(f)$*, is* $\max_x \mathrm{fbs}(f,x)$*.*

The *dual* program for Set Packing is the well known *Set Cover* problem. With our notations, the dual program can be written as

$$\begin{array}{lll} \min & \sum_{i=0}^{n-1} u_i & \\ & & \\ \text{s.t.} & \sum_{i:i\in B_j} u_i \geq 1 & \forall j \in [k] \\ & u_i \geq 0 & \forall i \in [n] \end{array} \tag{Dual LP}$$

Replacing the constraint $u_i \geq 0$ with $u_i \in \{0,1\}$ defines an IP we call the *"dual integer program"*. In fact, this integer program exactly captures the notion of certificate complexity.

Lemma 5.2. *The value of the dual IP is* $\mathrm{C}(f,x)$*.*

Proof. The set of coordinates i having $u_i = 1$ are a certificate for f on x, and vice versa, for any certificate $S \subseteq [n]$ for f on x the values $u_i = (1_S)_i$ form a feasible solution for the integer program whose value equals the certificate size. □

In order to capture the value of Dual LP we define the *Fractional Certificate Complexity.*

Definition 5.3 (Fractional Certificate Complexity)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and $x \in \{0,1\}^n$, then the* Fractional Certificate Complexity $\mathrm{FC}(f,x)$ *of f on x, is the minimal sum of weights assigned to $i \in [n]$, such that for each y such that $f(x) \neq f(y)$, the sum $\sum_{i:x_i \neq y_i} \mathrm{weight}(i)$ is at least 1. The fractional certificate complexity of f,* $\mathrm{FC}(f)$*, is* $\max_x \mathrm{FC}(f,x)$*.*

As both the primal and dual programs are feasible, *strong duality* states that the optimal values of both LPs are equal. This gives the next theorem.

Theorem (Theorem 2, revisited)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $x \in \{0,1\}^n$ then*

$$\mathrm{bs}(f,x) \leq \mathrm{fbs}(f,x) = \mathrm{FC}(f,x) \leq \mathrm{C}(f,x)$$

And, of course, the same inequality holds when taking the maximum over all $x \in \{0,1\}^n$.

5.1.1 Fractional Certificate Complexity is Randomized Certificate Complexity

The "new" measure we defined is actually the same measure up to a constant as the *randomized certificate complexity* defined by Aaronson.

Definition 5.4 (Randomized Certificate Complexity,[Aar03]). *A* randomized verifier *for input x is a randomized algorithm that, on input y to f, (i) accepts with probability* 1 *if $y = x$, and (ii) rejects with probability at least* $1/2$ *if $f(y) \neq f(x)$. (If $y \neq x$ but $f(y) = f(x)$, the acceptance probability can be arbitrary.) Then* $\mathrm{RC}(f, x)$ *is the minimum expected number of queries used by a randomized verifier for x, and* $\mathrm{RC}(f)$ *is the maximum of* $\mathrm{RC}(f, x)$ *over all x.*

Claim 5.5. $\mathrm{RC}(f, x) = \Theta(\mathrm{FC}(f, x))$.

We use in the proof the variant of nonadaptive randomized certificate:

Definition 5.6 (Nonadaptive Randomized Certificate Complexity, [Aar03]). *Call a randomized verifier for x nonadaptive if, on input y , it queries each y_i with independent probability λ_i, and rejects if and only if it encounters a disagreement with x. Let* $\mathrm{RC}_{\mathrm{na}}(f, x)$ *be the minimum of* $\sum_{i\in[n]} \lambda_i$ *over all nonadaptive verifiers for x.*

Aaronson showed adaptiveness does not help much as: $\mathrm{RC}(f, x) = \Theta(\mathrm{RC}_{\mathrm{na}}(f, x))$. Our proof follows some of his lines.

Proof. We show that $\mathrm{RC}_{\mathrm{na}}(f, x) = \Theta(\mathrm{FC}(f, x))$.

Let $u_0, \ldots, u_{n-1}$ be the weights for the minimal fractional certificate for f on x. By minimality, each u_i is at most 1. Then taking $\lambda_i = u_i$ for $i \in [n]$ gives a non-adaptive verifier such that if $f(x) \neq f(y)$ we have

$$\Pr(\text{The verifier accepts } y) = \prod_{i\in[n]:x_i\neq y_i} (1 - \lambda_i) \leq \prod_{i\in[n]:x_i\neq y_i} e^{-\lambda_i} = e^{-\sum_{i\in[n]:x_i\neq y_i} \lambda_i} \leq e^{-1}$$

where the last inequality is by the definition of a fractional certificate. Thus, $\mathrm{RC}_{\mathrm{na}}(f, x) \leq \mathrm{FC}(f, x)$.

For the second direction, given an optimal non-adaptive verifier with probabilities λ_i take $u_i = 2\lambda_i$. For y such that $f(x) \neq f(y)$ the probability of finding a disagreement is at least half, thus

$$1/2 \leq \Pr(\text{The verifier rejects } y) \underset{\text{UB}}{\leq} \sum_{i:x_i\neq y_i} \lambda_i$$

This gives in turn, $\sum_{i:x_i\neq y_i} u_i \geq 1$ which shows that the u_is are a fractional certificate for x. Thus, we have $\mathrm{FC}(f, x) \leq 2\mathrm{RC}_{\mathrm{na}}(f, x)$. □

5.1.2 Previous Results Regarding Set Cover and Set Packing

As Set Cover and Set Packing are fundamental optimization problems, much research was done on them. We survey some of the results and their connections to our discussion.

First, we note that Nisan's result that $\mathrm{C}(f) \leq \mathrm{bs}(f) \cdot \mathrm{sens}(f)$ ([Nis89]) can be viewed as a special (unweighted) case of the primal-dual approximation algorithm for Set Cover in [BYE81].

Second, [CL10] showed that the integrality gap of Set Packing is no bigger than $\ell - 1 + 1/\ell$ for ℓ being the size of the largest set. This bound is tight for infinitely many ℓs using Fano-plane examples. As $\ell \leq \mathrm{sens}(f)$ this gives

$$\mathrm{fbs}(f) \leq \mathrm{bs}(f) \cdot (\mathrm{sens}(f) - 1 + 1/\mathrm{sens}(f)) \ .$$

Fano-planes gives examples for which $\mathrm{fbs}(f,x) = \mathrm{bs}(f,x) \cdot (\ell - 1 + 1/\ell)$ for arbitrary large ℓs and a specific x. However, as fbs and bs are defined as the maximum over all xs, our attempts to find functions with arbitrary large gaps between bs and fbs have failed so far.

Last, [Lov75] showed that the integrality gap of Set Cover (i.e. the ratio between C and FC) is at most $H_t = \sum_{i=1}^{t} 1/i \approx \ln(t)$ where $t = \max_i |\{B_j \mid i \in B_j\}|$ in our notations. Note that for example for MAJ_n on 0^n, $t = \Theta(2^n/\sqrt{n})$, so this does not imply a logarithmic factor between C and FC. Indeed, we will give an example with a polynomial gap between the two.

5.2 Upper Bounding The Block Sensitivity of BFC

With the IP/LP perspective of block-sensitivity we have the proper tools to prove the next theorem, giving an upper bound on the block-sensitivity of the composition of Boolean functions.

Theorem (Theorem 3, restated)**.** *Let $f : \{0,1\}^n \to \{0,1\}$, $g : \{0,1\}^m \to \{0,1\}$ be Boolean functions, then $\mathrm{bs}(f \circ g) \le \mathrm{fbs}(f) \cdot \mathrm{bs}(g)$. Moreover, this inequality is weakly tight in the following sense: for g such that $\mathrm{bs}^0(g) = \mathrm{bs}^1(g)$ we have that $\mathrm{bs}(f \circ g) \ge \mathrm{fbs}(f) \cdot \mathrm{bs}(g) - 2^n$.*

The theorem shows that if f is some fixed function and we take a sequence of functions $\{g_i\}_{i \in \mathbb{N}}$ such that $\mathrm{bs}^0(g_i) = \mathrm{bs}^1(g_i)$ for all $i \in \mathbb{N}$ and $\lim_{i \to \infty} \mathrm{bs}(g_i) = \infty$ then the ratio between $\mathrm{bs}(f \circ g_i)/\mathrm{bs}(g_i)$ tends to $\mathrm{fbs}(f)$ as $i \to \infty$.

Proof. Let $x = (x_j^i)_{i \in [n], j \in [m]}$ be an input to $f \circ g$. We formulate $\mathrm{bs}(f \circ g, x)$ as an integer program. Let $y \in \{0,1\}^n$ be the input to f that x induces, i.e. $y_i = g(x^i)$ for $i \in [n]$. Denote the minimal sensitive blocks for f on y by $B_0, B_1, \ldots, B_{k-1}$. The key observation is this: a set is a minimal sensitive block for $f \circ g$ on x iff it is a collection of minimal sensitive blocks for several g_is, where the g_is that participate in the collection form a minimal sensitive block for f. Thus, in a collection of disjoint minimal sensitive blocks for $f \circ g$, there can be at most $\mathrm{bs}(g, x^i)$ sets intersecting each x^i. The problem of calculating $\mathrm{bs}(f \circ g, x)$ reduces to finding the maximal number of B_js such that each y_i is covered at most $\mathrm{bs}(g, x^i)$ times. Any feasible solution for this problem can be transformed into a collection of disjoint blocks flipping $f \circ g$ value on x and vice versa. The formulation of this problem in terms of integer programming is:

$$\begin{array}{lll} \max & \sum_{j=0}^{k-1} w_j & \\ \text{s.t.} & \sum_{j: i \in B_j} w_j \le \mathrm{bs}(g, x^i) & \forall i \in [n] \\ & w_j \in \{0, 1, \ldots, \mathrm{bs}(g)\} & \forall j \in [k] \end{array} \tag{IP(bs(fg))}$$

Relaxing this to an LP gives

$$\begin{array}{lll} \max & \sum_{j=0}^{k-1} w_j & \\ \text{s.t.} & \sum_{j: i \in B_j} w_j \le \mathrm{bs}(g, x^i) & \forall i \in [n] \\ & w_j \ge 0 & \forall j \in [k] \end{array} \tag{LP(bs(fg))}$$

Observing the constraints, any feasible solution for the linear program LP(bs(fg)) where each w_j is divided by $\mathrm{bs}(g)$ is a feasible solution for the LP for $\mathrm{fbs}(f, y)$. Thus,

$$\mathrm{fbs}(f, y) \ge \frac{\mathrm{OPT}(LP(bs(fg)))}{\mathrm{bs}(g)} \ge \frac{\mathrm{OPT}(IP(bs(fg)))}{\mathrm{bs}(g)} = \frac{\mathrm{bs}(f \circ g, x)}{\mathrm{bs}(g)} .$$

Rearranging this gives $\mathrm{bs}(f \circ g, x) \leq \mathrm{fbs}(f, y) \cdot \mathrm{bs}(g)$. Taking x which maximizes $\mathrm{bs}(f \circ g)$ gives the desired upper bound.

For the tightness part, let $p^0, p^1 \in \{0,1\}^m$ be points achieving the maximal $\mathrm{bs}^0(g), \mathrm{bs}^1(g)$ respectively, and let y be the point achieving the maximal $\mathrm{fbs}(f)$. Consider the block sensitivity of $f \circ g$ on the point $x = (x^i_j)_{i \in [n], j \in [m]}$ where $x^i = p^{y_i}$. Note that the additive gap between IP(bs(fg)) and LP(bs(fg)) is at most k. Indeed, if the optimal values of the LP are w^*_j, then $w_j = \lfloor w^*_j \rfloor$ is a feasible solution for the IP with value at least $\mathrm{OPT}(LP(bs(fg))) - k$. Our choice of x gives $\mathrm{bs}(g, x^i) = \mathrm{bs}(g)$ for all $i \in [n]$, hence the value of LP(bs(fg)) is exactly $\mathrm{bs}(g)$ times the value of LP(bs(f)) as the constraints are equivalent up to the constant factor $\mathrm{bs}(g)$. Hence,

$$\mathrm{fbs}(f) \cdot \mathrm{bs}(g) = \mathrm{OPT}(LP(bs(fg))) \leq \mathrm{OPT}(IP(bs(fg))) + k = \mathrm{bs}(f \circ g, x) + k \leq \mathrm{bs}(f \circ g) + k\ .$$

As $k \leq 2^n$ this completes the proof. □

The behavior of fbs with respect to composition resembles that of sens and C.

Claim 5.7. *Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$ be Boolean functions, then (i) $\mathrm{fbs}(f \circ g) \leq \mathrm{fbs}(f) \cdot \mathrm{fbs}(g)$ and (ii) For $z \in \{0,1\}$ if $f(z^n) = g(z^m) = z$ then $\mathrm{fbs}(f \circ g, z^{nm}) \geq \mathrm{fbs}(f, z^n) \cdot \mathrm{fbs}(g, z^m)$*

We defer the proof to Appendix C.

5.3 Analysis of an Example

In order to demonstrate the analysis of block sensitivity of composed Boolean functions we present the next example.

Example 5.8. *Let f be a function on 6 variables defined by:*

$$f = \mathrm{MAJ}(x_0 \oplus x_1 \oplus x_3, x_0 \oplus x_2 \oplus x_4, x_1 \oplus x_2 \oplus x_5)\ .$$

The example above has $\mathrm{bs}(f, x) = 4\ ,\ \forall x \in \{0,1\}^6$ and $\mathrm{C}(f) = 5$. It was actually presented in [BSW86] and its composition was analyzed in [WZ89] and [Aar03]. It was considered the best separation between bs and C, but the analysis of this function composition was wrong claiming that $\mathrm{bs}(f^k) = \mathrm{bs}(f)^k$. As we shall see, $\mathrm{bs}(f^2) = 18 > 16 = \mathrm{bs}(f)^2$ and in general:

Lemma 5.9. *For $k \geq 1$: $\mathrm{bs}(f^{k+1}) = \lfloor 4.5 \cdot \mathrm{bs}(f^k) \rfloor$ and this value is achieved on any input.*

Proof. Let $\{x^i_j\}_{i \in [6], j \in [6^k]}$ be an input for $f^{k+1} = f \circ f^k$ and let $y_i = f(x^i)$. The minimal sensitive blocks of f at any point are, up to indices name change: $\{0\}, \{1\}, \{2\}, \{3,4\}, \{3,5\}, \{4,5\}$.

The optimal solution for the linear program for $\mathrm{fbs}(f)$ on y is $(1, 1, 1, 0.5, 0.5, 0.5)$ respectively. Consider the IP for $\mathrm{bs}(f \circ f^k, x)$ as in Theorem 3. By induction $\mathrm{bs}(f^k, x^i) = \mathrm{bs}(f^k)$. Put weights $\mathrm{bs}(f^k)$ on each set $\{i\}$ for $i = 0, 1, 2$, $\lfloor \mathrm{bs}(f^k)/2 \rfloor$ on each of the sets $\{3,4\}, \{3,5\}$ and $\lceil \mathrm{bs}(f^k)/2 \rceil$ on the set $\{4,5\}$. This is a feasible solution for the IP, as any $i \in [6]$ was covered at most $\mathrm{bs}(f^k)$ times. The solution size is $\lfloor \mathrm{bs}(f^k) \cdot 4.5 \rfloor$, and as Theorem 3 gives $\mathrm{bs}(f^{k+1}) \leq \mathrm{fbs}(f) \cdot \mathrm{bs}(f^k) = 4.5 \cdot \mathrm{bs}(f^k)$, this is optimal. □

For k large enough, using Lemma 5.9, we derive $\mathrm{bs}(f^k) \approx 0.8877 \cdot 4.5^k$. This shows that for k large enough and $g = f^k$, the ratio $\mathrm{bs}(g^2)/\mathrm{bs}(g)^2 \approx 1/0.8877 > 1.126$. Claim 5.7 gives $\mathrm{fbs}(f^k) = 4.5^k$, so there is a small 1.126 factor separating $\mathrm{bs}(f^k)$ from $\mathrm{fbs}(f^k)$. The next claim, proven in Appendix D, shows that this constant factor is no coincidence.

Claim 5.10. *Let $f : \{0,1\}^n \to \{0,1\}$ then for any integer k the ratio $\mathrm{fbs}(f^k)/\mathrm{bs}(f^k)$ is at most $c(n)$ i.e. independent of k.*

5.4 Block Sensitivity and Certificate Complexity Separation

As we demonstrated above the example considered to have the best separation between block-sensitivity and certificate-complexity has a smaller separation than what was previously claimed: $\mathrm{C}(f^k) = 5^k = \Theta(\mathrm{bs}(f^k)^{\log(5)/\log(4.5)}) = \Theta(\mathrm{bs}(f^k)^{1.07\dots})$. The best separation we could find uses a composition of a symmetric function, and was previously presented in [Aar03]. Our analysis is slightly better than that of Aaronson since we use a $(\mathrm{C},0)$-good form of the function.

Example 5.11. *Consider the symmetric function on 29 variables whose value is 1 iff the weight of the input $\|x\| = \sum_{i=1}^{29} x_i$ is in $\{13, 14, 15, 16\}$.*

This function has certificate complexity 26, achieved on inputs of weight $13, 14, 15, 16$, as we need to expose at least $29-3$ coordinates to convince that $f(x) = 1$. The sensitivity of the function is 17, achieved on inputs of weight 12 or 17. The block sensitivity of the function is also 17, achieved on inputs of weight 12 or 17 using 17 blocks of size 1, or on inputs of weight 13 or 16 using 13 blocks of size 1 in addition to 4 blocks of size 4. One can check that the fractional block sensitivity of f is also 17. Take $\tilde{f}$ to be a $(\mathrm{C},0)$ good form of f, as in Remark 2.12. Lemma 3.1 gives $\forall k \in \mathbb{N} : \mathrm{C}(\tilde{f}^k) = \mathrm{C}(\tilde{f})^k = 26^k$, while Claim 5.7 gives $\forall k \in \mathbb{N} : \mathrm{fbs}(\tilde{f}^k) \le 17^k$. Hence,

$$\mathrm{C}(\tilde{f}^k) \ge \mathrm{fbs}(\tilde{f}^k)^{\log(26)/\log(17)} \ge \mathrm{bs}(\tilde{f}^k)^{\log(26)/\log(17)}$$

6 Improving Nisan-Szegedy Bound

The last application shows that function composition can prove not only separations between complexity measures but even tighten relations between them. Nisan and Szegedy showed that:

Theorem 6.1 ([NS94])**.** *Let f be a Boolean function, then $\mathrm{bs}(f) \le 2\deg(f)^2$.*

We improve their result by a factor 2:

Theorem (Theorem 4, restated)**.** *Let f be a Boolean function, then $\mathrm{bs}(f) \le \deg(f)^2$.*

Proof. Assume by contradiction that there exists a Boolean-function f for which $\mathrm{bs}(f) \ge \deg(f)^2 + 1$. WLOG f is in $(\mathrm{bs}, 0)$ good form. Let $d \triangleq \deg(f) \ge 1$ and take $g = f^{2d^2}$. By Property 1 of Lemma 3.1 $\deg(g) = d^{2d^2}$. By Property 7

$$\mathrm{bs}(g) \ge (d^2+1)^{2d^2} = (d^2)^{2d^2} \cdot \left(1 + \frac{1}{d^2}\right)^{2d^2} \ge^{3} (d^{2d^2})^2 \cdot 2^2 = \deg(g)^2 \cdot 4$$

and this is a contradiction to Theorem 6.1 □

This improves as a corollary another relation between complexity measures:

Corollary 6.2 ([Mid04])**.**

$$\mathrm{C}(f) \le \mathrm{D}(F) \le \mathrm{bs}(f) \cdot \deg(f) \le \deg(f)^3$$

[3] $(1 + 1/x)^x$ is a monotone increasing function for $x \ge 1$, hence $(1 + 1/x)^x \ge 2$ for $x = d^2 \ge 1$

Remark 6.3. The method used in Theorem 4 shows that for any two complexity measures M, N where $M \in \{\mathrm{sens}, \mathrm{bs}, \mathrm{fbs}, \mathrm{C}, \mathrm{C}_{\min}, \mathrm{mr}, \deg, \mathrm{D}\}$ and $N \in \{\mathrm{sens}, \mathrm{fbs}, \mathrm{C}, \deg, \mathrm{D}\}$, a relation of the form $M(f) \leq N(f)^{\alpha+o(1)}$ can be tighten to $M(f) \leq N(f)^{\alpha}$.

7 Discussion

In this paper we saw some applications and interesting questions regarding function composition.

BFC most natural use is to generate asymptotic separations based on a constant (even small) size examples for which two complexity measures differ. Hence, possible further work will be to get better separations by finding better examples. This can be done using computer search. More surprisingly, BFC can help tighten relations between complexity measures as in Theorem 4. Can this technique be used to tighten other relations?

BFC fails to separate complexity measures which behave similarly to composition. In particular, this technique fails to separate bs from fbs as proven in Claim 5.7. An interesting question is whether or not $\mathrm{bs}(f) = \Theta(\mathrm{fbs}(f))$, a negative answer will need to use different techniques than composition. A possible approach may involve the integrality gap shown in Section 5, perhaps using examples closely related to Fano-planes as in [CL10]. Answering this question may shed light on other questions such as bs vs. C.

Another interesting matter raised by Example 4.1 is of the connection between codes and Boolean functions. In particular, it will be interesting to see what more can we learn from this connection.

8 Acknowledgements

I thank Benjamin Eliot Klein for lots of support and advice; my advisor, Ran Raz, for guidance and encouragement; Igor Shinkar for his proof of Lemma B.1 and Gregory Valiant for helpful discussions. I also thank the anonymous referees for helpful comments.

References

[Aar03] S. Aaronson. Quantum certificate complexity. In *IEEE Conference on Computational Complexity*, pages 171–178, 2003.

[AB09] S. Arora and B. Barak. *Computational complexity: a modern approach.* Cambridge University Press, 2009.

[Amb03] Andris Ambainis. Polynomial degree vs. quantum query complexity. In *FOCS*, pages 230–239, 2003.

[BdW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[BI87] M. Blum and R. Impagliazzo. Generic oracles and oracle classes (extended abstract). In *FOCS*, pages 118–126, 1987.

[BKS$^+$10] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: New constructions of condensers, ramsey graphs, dispersers, and extractors. *J. ACM*, 57(4), 2010.

[BSW86] S. Bublitz, U. Schürfeld, and I. Wegener. Properties of complexity measures for prams and wrams. *Theor. Comput. Sci.*, 48(3):53–73, 1986.

[BYE81] R. Bar-Yehuda and S. Even. A linear-time approximation algorithm for the weighted vertex cover problem. *J. Algorithms*, 2(2):198–203, 1981.

[CL10] Y. H. Chan and L. C. Lau. On linear and semidefinite programming relaxations for hypergraph matching. In *SODA*, pages 1500–1511, 2010.

[HLS07] Peter Høyer, Troy Lee, and Robert Spalek. Negative weights make adversaries stronger. In *STOC*, pages 526–535, 2007.

[Lov75] L. Lovasz. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383 – 390, 1975.

[Mid04] G. Midrijanis. Exact quantum query complexity for total boolean functions. *arXiv:quant-ph/0403168v2*, 2004.

[MOS04] E. Mossel, R. O'Donnell, and R. A. Servedio. Learning functions of k relevant variables. *J. Comput. Syst. Sci.*, 69(3):421–434, 2004.

[Nis89] N. Nisan. Crew prams and decision trees. In *STOC*, pages 327–335, 1989.

[NN93] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[NS94] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[NW94] N. Nisan and A. Wigderson. On rank vs. communication complexity. In *FOCS*, pages 831–836, 1994.

[O'D12] R. O'Donnell. Analysis of boolean functions. http://analysisofbooleanfunctions.org/, 2012.

[SW86] Michael E. Saks and Avi Wigderson. Probabilistic boolean decision trees and the complexity of evaluating game trees. In *FOCS*, pages 29–38, 1986.

[Val12] Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and juntas with noise. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:6, 2012.

[Ver10] E. Verbin. comments on my philomath project: Sensitivity versus block-sensitivity. S. Aaronson blog, http://www.scottaaronson.com/blog/?p=453, 2010.

[WZ89] I. Wegener and L. Zádori. A note on the relations between critical and sensitive complexity. *Elektronische Informationsverarbeitung und Kybernetik*, 25(8/9):417–421, 1989.

A Properties of the IP function

The next lemma states a well known property of $\mathrm{IP_n}$ which we later use to prove Lemma 4.5. We include the proof for completeness.

Lemma A.1. *Denote by* $\mathrm{IP_n} : \{0,1\}^{2n} \to \{0,1\}$ *the Boolean function defined as:*

$$\mathrm{IP_n}(y_0, y_1, \ldots, y_{2n-1}) = y_0 \cdot y_1 + \ldots + y_{2n-2} \cdot y_{2n-1} \pmod 2 .$$

Then, every Fourier coefficient of $\mathrm{IP_n}$ *is of magnitude* 2^{-n}*.*

Proof. Let S be a subset of $\{0,1\}^{2n}$. Using Equation (1) the S-Fourier coefficient of $\mathrm{IP_n}$ is equal to

$$\widehat{\mathrm{IP_n}}(S) = E_{y \in_R \{0,1\}^{2n}} \left((-1)^{\mathrm{IP_n}(y)} \cdot (-1)^{\sum_{i=0}^{2n-1} S_i \cdot y_i} \right)$$

where S_i is the indicator that $i \in S$. Now we may write

$$\widehat{\mathrm{IP_n}}(S) = E_{y \in_R \{0,1\}^{2n}} \left(\prod_{i=0}^{n-1} (-1)^{y_{2i} \cdot y_{2i+1}} \cdot (-1)^{S_{2i} \cdot y_{2i} + S_{2i+1} \cdot y_{2i+1}} \right)$$

and as the different multiplicands are probabilistically independent, we can replace expectation and product. Thus, $\widehat{\mathrm{IP_n}}(S) = \prod_{i=0}^{n-1} E\left((-1)^{y_{2i} \cdot y_{2i+1} + S_{2i} \cdot y_{2i} + S_{2i+1} \cdot y_{2i+1}}\right)$. It can be checked by complete enumeration that the expectancy of each multiplicand is either $1/2$ or $-1/2$, hence $|\widehat{\mathrm{IP_n}}(S)| = 2^{-n}$. □

We are ready to prove Lemma 4.5.

Lemma (Lemma 4.5, restated)**.** *Let* $A \subseteq (\mathbb{F}_2)^{2n}$ *be an affine subspace of dimension* d *defined by* $v_A + \mathrm{span}\{v_0, v_1, \ldots, v_{d-1}\}$*. If* $\mathrm{IP_n}|_A(x_0, x_1, \ldots, x_{d-1}) \triangleq \mathrm{IP_n}(v_A + \sum_{i \in [d]} x_i \cdot v_i)$ *is a parity function, then* $d \leq n$*.*

Proof. Let T be a change of basis matrix mapping $v_i \mapsto e_i$ for $i \in [d]$. If $v_A = \vec{0}$, obviously $T(v_A) = \vec{0}$, else we can assume WLOG that $v_A \notin \mathrm{span}\{v_i : i \in [d]\}$ and since v_A is linearly independent of them we can demand that $T(v_A) = e_d$. Write $\mathrm{IP_n} = (\mathrm{IP_n} \circ T^{-1}) \circ T$, and denote $\mathrm{IP_n} \circ T^{-1}$ by $\tilde{\mathrm{IP_n}}$. As $\tilde{\mathrm{IP_n}}$ is a composition of $\mathrm{IP_n}$ with an invertible linear transformation, Fact 2.8 gives that its Fourier spectrum is a (linear) permutation of the Fourier spectrum of $\mathrm{IP_n}$, hence every coefficient is of magnitude 2^{-n}. Let $\tau : [2n] \to \{0, 1, *\}$ be the restriction defined by:

$$\tau(i) = \begin{cases} * & i \in [d] \\ T(v_A)_i & \textit{otherwise} \end{cases} .$$

If $\mathrm{IP_n}|_A$ is a parity function then so is $\tilde{\mathrm{IP_n}}|_\tau$ as

$$\tilde{\mathrm{IP_n}}|_\tau(x_0, \ldots, x_{2n-1}) = \tilde{\mathrm{IP_n}}(T(v_A) + \sum_{i \in [d]} x_i \cdot e_i) = \mathrm{IP_n}(v_A + \sum_{i \in [d]} x_i \cdot v_i) .$$

According to Corollary 2.7 the maximal Fourier coefficient of $\tilde{\mathrm{IP_n}}|_\tau$ is at most the maximal Fourier coefficient of $\tilde{\mathrm{IP_n}}$ times 2^{2n-d}, hence at most 2^{n-d}. As $\tilde{\mathrm{IP_n}}|_\tau$ becomes parity, there is one Fourier coefficient of magnitude 1, giving $d \leq n$. □

Remark A.2. Lemma 4.5 shows in particular that $\mathrm{IP_n}$ restricted to any $n+1$ dimension affine subspace is nonconstant. This property is called being an *affine disperser for dimension* $n+1$. In fact, any affine disperser for dimension $n+1$ doesn't become a parity function under any restriction to an affine subspace of dimension $n+2$. Thus, our property is almost identical to being an affine disperser for dimension $n+1$. The fact that $\mathrm{IP_n}$ is an affine disperser is regarded to Ben-Sasson, Hoory, Rosenman, Vadhan and Wigderson (in an unpublished manuscript, mentioned in [BKS$^+$10]).

B Generalization to Epsilon-Biased Sets

In the construction of Example 4.1 we can replace the WH matrix by a matrix associated with any ϵ-biased sample space as follows. Let $\mathcal{C} \subseteq {\mathbb{F}_2}^{2n}$, we say $\mathcal{C}$ is a an *ϵ-biased sample space* if for any nonzero linear functional $\Phi : (\mathbb{F}_2)^{2n} \to \mathbb{F}_2$ the expectancy of $E_{v\in_R \mathcal{C}}(\Phi(v))$ is in the range $[1/2-\epsilon, 1/2+\epsilon]$. ϵ-biased sets where first defined and constructed in [NN93], following a sequence of works giving better constructions.

Putting an order on the set $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_m\}$ it is convenient to think of an ϵ-biased set as a $2n \times m$ matrix over $\mathbb{F}_2$ where each XOR of a nonempty subset of rows gives a row with $m \cdot (1/2 \pm \epsilon)$ ones. Let $A \in (\mathbb{F}_2)^{2n\times m}$ be the matrix associated with the sample space $\mathcal{C}$, then A naturally defines a (linear) mapping $A : (\mathbb{F}_2)^m \to (\mathbb{F}_2)^{2n}$. The composition of the mappings $f = \mathrm{IP_n} \circ A$ gives a construction with similar properties to that of Example 4.1.

We begin with the following lemma which is analogous to Lemma 4.4

Lemma B.1. *The bitwise* OR *of* n *rows in an* ϵ*-biased matrix with* m *columns has at most* $m - m \cdot (1-2^{-n}) \cdot (1-2\epsilon)$ *zeros.*

The proof of Lemma B.1 is due to Igor Shinkar.

Proof. We may assume WLOG that A is an ϵ-biased matrix of size $n \times m$. Let

$$\mathrm{NZ} = \{j \in [m] : \exists i \in [n] : A_{i,j} = 1\}$$

be the set of nonzero columns. For a fixed $j \in \mathrm{NZ}$, let I be a set chosen uniformly at random from the nonempty subsets of $[n]$ then

$$\Pr_I\left(\bigoplus_{i\in I} A_{i,j} = 1\right) = \frac{2^{n-1}}{2^n - 1}$$

Let $w_I(j) = \bigoplus_{i\in I} A_{i,j}$ and denote by $\|w_I\|_1 = \sum_{j\in[m]} w_I(j)$ where the sum is over the reals. By the definition of ϵ-biased sets, for all choices of I we have $\|w_I\|_1 \geq m \cdot (1/2-\epsilon)$. This gives

$$m \cdot (1/2-\epsilon) \leq E_I(\|w_I\|_1) = \frac{2^{n-1}}{2^n-1} \cdot |\mathrm{NZ}| \ .$$

Rearranging this we have

$$|\mathrm{NZ}| \geq m \cdot \frac{2^n-1}{2^{n-1}} \cdot (1/2-\epsilon) = m \cdot (1-2^{-n}) \cdot (1-2\epsilon) \ .$$

As the set of zero coordinates is $[m] - \mathrm{NZ}$ its size is at most $m - m \cdot (1-2^{-n}) \cdot (1-2\epsilon)$. □

Claim B.2. *Let $f = \mathrm{IP_n} \circ A$ where A is a $2n \times m$ ϵ-biased matrix then* $\deg(f) = m \cdot (1/2 \pm \epsilon)$ *and* $\mathrm{mr}(f) \geq m \cdot (1 - 2^{-n}) \cdot (1 - 2\epsilon)$

Proof. The proof closely follows the lines of the proof of Theorem 4.2. By Fact 2.8 any nonzero Fourier coefficient subset characteristic vector is the XOR of a nonempty subset of rows in A. By definition the subsets are of size $m \cdot (1/2 \pm \epsilon)$, hence the degree of f over the reals is $m \cdot (1/2 \pm \epsilon)$.

Let ρ be a restriction under which f is parity, and let d be the dimension of $A(V_\rho)$, then by Lemma 4.6 $d \leq n$. Let $U = A(V_\rho)^\perp$, we have that $\dim(U) = 2n - d$. Since performing an invertible linear transformation on the rows of A preserves the property of being a an ϵ-biased matrix, we may assume WLOG that the image of $A(V_\rho)$ is $\mathrm{span}\{e_0, e_1, \ldots, e_{d-1}\}$ and thus $U = \mathrm{span}\{e_d, e_{d+1}, \ldots, e_{2n-1}\}$. Let k be a variable not fixed by ρ, then $A \cdot e_k \perp U$. Equivalently, the kth column in A has $2n - d$ zeros in the last $2n - d$ entries, hence it is a zero of the bitwise OR of the last $2n - d$ rows of A. As $d \leq n$ it is a zero entry in the bitwise OR of the last n rows. By Lemma B.1 the number of such variables is at most $m - m \cdot (1 - 2^{-n}) \cdot (1 - 2\epsilon)$ which completes the proof. □

C Fractional Block Sensitivity of Function Composition

We repeat and prove the next claim.

Claim (Claim 5.7, restated)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$ be Boolean functions, then (i)* $\mathrm{fbs}(f \circ g) \leq \mathrm{fbs}(f) \cdot \mathrm{fbs}(g)$ *and (ii) For $z \in \{0,1\}$ if $f(z^n) = g(z^m) = z$ then* $\mathrm{fbs}(f \circ g, z^{nm}) \geq \mathrm{fbs}(f, z^n) \cdot \mathrm{fbs}(g, z^m)$

Proof. For the first part, we show that $\mathrm{FC}(f \circ g) \leq \mathrm{FC}(f) \cdot \mathrm{FC}(g)$ (which is equivalent to $\mathrm{fbs}(f \circ g) \leq \mathrm{fbs}(f) \cdot \mathrm{fbs}(g)$). Let $x = (x^i_j)_{i \in [n], j \in [m]}$ be an input for $f \circ g$ and let $y \in \{0,1\}^n$ be the intermediate values, $y_i = g(x^i)$. Let $T_i : [m] \to \mathbb{R}^+$ be a minimal fractional certificate for g on x^i i.e. a collection of weights for the variables $\{x^i_j\}_{j \in [m]}$. Let $S : [n] \to \mathbb{R}^+$ be a minimal fractional certificate for f on y. We show that the weights $R : [n] \times [m] \to \mathbb{R}^+$ defined by $R(i,j) = S(i) \cdot T_i(j)$ are a fractional certificate for $f \circ g$. Let $x' = ({x'}^i_j)_{i \in [n], j \in [m]}$ such that $(f \circ g)(x) \neq (f \circ g)(x')$ and let y' be the intermediate values for x' i.e. ${y'}_i = g({x'}^i_j)$. Since $f(y) \neq f(y')$ we have $\sum_{i : y_i \neq {y'}_i} S(i) \geq 1$ and since each T_i is a certificate we have:

$$\forall i : y_i \neq {y'}_i. \sum_{j : x^i_j \neq {x'}^i_j} T_i(j) \geq 1 \,.$$

Thus,

$$\sum_{i,j : x^i_j \neq {x'}^i_j} R(i,j) \geq \sum_{i : y_i \neq {y'}_i} \sum_{j : x^i_j \neq {x'}^i_j} R(i,j) = \sum_{i : y_i \neq {y'}_i} \left(S(i) \cdot \sum_{j : x^i_j \neq {x'}^i_j} T_i(j) \right) \geq \sum_{i : y_i \neq {y'}_i} S(i) \geq 1 \,,$$

showing that R is a fractional certificate for $f \circ g$. The total weight of R is

$$\sum_{i,j} R(i,j) = \sum_i \left(S(i) \cdot \sum_j T_i(j) \right) \leq \mathrm{FC}(f) \cdot \mathrm{FC}(g) \,.$$

For the second part, we show that $\mathrm{fbs}(f \circ g, z^{nm}) \geq \mathrm{fbs}(f, z^n) \cdot \mathrm{fbs}(g, z^m)$. Put $x = z^{nm}$ as the input for $f \circ g$. As $g(z^m) = z$, the intermediate values $y_i = g(x^i) = z$. Let $\{B_k\}_k$ be minimal blocks which are sensitive for f on z^n, let $\{u_k\}_k$ be optimal weights for these blocks achieving $\mathrm{fbs}(f, z^n)$. For $i \in [n]$ let $\{T^i_\ell\}_\ell$ be minimal blocks which are sensitive for g on $x^i = z^m$, and let $\{v^i_\ell\}_\ell$ be optimal weights (independent of i) for these blocks achieving $\mathrm{fbs}(g, z^m)$.

We show that the sets $A_{k,\ell} = \cup_{i \in B_k} T^i_\ell$ with weights $w_{k,\ell} = u_k \cdot v_\ell$ are "relaxed disjoint". For any coordinate (i, j) the sum of weights for blocks containing this coordinate is at most 1 as

$$\sum_{k,\ell:(i,j)\in A_{k,\ell}} w_{k,\ell} = \sum_{k:i\in B_k} \sum_{\ell:j\in T^i_\ell} w_{k,\ell} = \sum_{k:i\in B_k} \left(u_k \cdot \sum_{\ell:j\in T^i_\ell} v_\ell \right) \leq \sum_{k:i\in B_k} u_k \leq 1 \ .$$

The total weight assigned to the sets $A_{k,\ell}$ is $\sum_{k,\ell} w_{k,\ell} = \mathrm{fbs}(f, z^n) \cdot \mathrm{fbs}(g, z^m)$. As this is a feasible solution for fractional block sensitivity linear program we conclude that $\mathrm{fbs}(f \circ g, z^{nm}) \geq \mathrm{fbs}(f, z^n) \circ \mathrm{fbs}(g, z^m)$ □

D Gaps between Block Sensitivity and Fractional Block Sensitivity of Composed Functions

In this section we prove Claim 5.10. In the proof we use the next lemmas and the following definition.

Definition D.1. *Let $f : \{0,1\}^n \to \{0,1\}$, we say f is monotone increasing (decreasing) if $x \geq y$ pointwise implies $f(x) \geq f(y)$ ($f(x) \leq f(y)$). We say that f is monotone if either cases hold.*

Lemma D.2 ([Nis89])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone function then* $\mathrm{sens}(f) = \mathrm{bs}(f) = \mathrm{C}(f)$.

Lemma D.3 ([Nis89])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and $z \in \{0,1\}$ then* $\mathrm{C}^z(f) \leq \mathrm{bs}^z(f) \cdot \mathrm{sens}^{1-z}(f)$.

Lemma D.4. *Let f and g be Boolean functions, then* $\mathrm{bs}^z(f \circ g) \geq \mathrm{bs}^z(f) \circ \min(\mathrm{bs}^0(g), \mathrm{bs}^1(g))$

Proof. Let p^0 and p^1 be the inputs maximizing $\mathrm{bs}^0(g)$ and $\mathrm{bs}^1(g)$ respectively and let y be the input maximizing $\mathrm{bs}^z(f)$. Take the input $x = (x^0||x^1|| \ldots ||x^{n-1})$ where $x^i = p^{y_i}$ and consider the block sensitivity of $f \circ g$ on this input. f has $\mathrm{bs}^z(f)$ many disjoint blocks on y and each block can be expanded into at least $\min(\mathrm{bs}^0(g), \mathrm{bs}^1(g))$ disjoint sensitive blocks for $f \circ g$ on x. □

Lemma D.5. *Let f, g be Boolean functions such that f is not monotone and let $z \in \{0,1\}$, then*

1. $\mathrm{bs}(f) \geq 2$
2. $\mathrm{bs}^z(f \circ g) \geq \mathrm{bs}(g)$
3. $\mathrm{fbs}^z(f \circ g) \geq \mathrm{fbs}(g)$
4. *The sequence $\{\mathrm{bs}^z(f^\ell)\}_{\ell \in \mathbb{N}}$ is monotone increasing and tends to infinity.*

Proof. 1. Any function with $\mathrm{bs}(f) = 1$ is either $f(x) = x_i$ or $f(x) = 1 - x_i$ for some i. Either functions are monotone, hence any non-monotone function has $\mathrm{bs}(f) \geq 2$.

2. Let $z' \in \{0,1\}$ be such that $\mathrm{bs}^{z'}(g) = \mathrm{bs}(g)$. Since f is not monotone, there is an input $y = y(z', z)$ and a bit $i \in [n]$ such that $y_i = z'$ and changing y_i from $z' \to 1 - z'$ makes a $z \to 1 - z$ change in the output of f. Let p^0 and p^1 be inputs maximizing $\mathrm{bs}^0(g)$ and $\mathrm{bs}^1(g)$ respectively, and take the input $x = (x^0 || x^1 || \dots || x^{n-1})$ where $x^i = p^{y_i}$. First, $(f \circ g)(x) = f(y) = z$. Second, by our choice of i, changing each sensitive block of g on x^i changes the value of $f \circ g$ and we have $\mathrm{bs}(g, x^i) = \mathrm{bs}(g)$ such sensitive blocks.

3. The bound on $\mathrm{fbs}^z(f \circ g)$ can be proven similarly.

4. By Property 2 $\mathrm{bs}^z(f^{\ell+1}) \geq \mathrm{bs}(f^\ell) \geq \mathrm{bs}^z(f^\ell)$. It remains to show that the sequence diverges. Combining Property 2 above and Lemma D.4 gives that for any $k \in \mathbb{N}$:

$$\begin{aligned} \mathrm{bs}^z(f^{2k}) &\underset{\text{Lemma D.4}}{\geq} \mathrm{bs}^z(f^2) \cdot \min(\mathrm{bs}^0(f^{2k-2}), \mathrm{bs}^1(f^{2k-2})) \\ &\underset{\text{Property 2}}{\geq} \mathrm{bs}(f) \cdot \min(\mathrm{bs}^0(f^{2k-2}), \mathrm{bs}^1(f^{2k-2})) \,. \end{aligned}$$

Induction gives $\mathrm{bs}^z(f^{2k}) \geq \mathrm{bs}(f)^k$ which in turn is at least 2^k by Property 1.

□

Claim (Claim 5.10, restated)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ then for integer $\ell \in \mathbb{N}$ the ratio $\mathrm{fbs}(f^\ell)/\mathrm{bs}(f^\ell)$ is at most $c(n) = 25 \cdot n^2 \cdot 2^n$ i.e. independent of ℓ.*

Proof. If f is monotone then f^ℓ is monotone and by D.2 we have that $\mathrm{bs}(f^\ell) = \mathrm{C}(f^\ell)$. Hence, in this case, $\mathrm{fbs}(f^\ell) = \mathrm{bs}(f^\ell)$, and we can assume that f is not monotone for the rest of the proof.

Let $z \in \{0,1\}$. Denote by $r_\ell^z \triangleq \left(\frac{\mathrm{bs}^z(f^\ell)}{\mathrm{fbs}^z(f^\ell)}\right)$ and $r_\ell \triangleq \min(r_\ell^0, r_\ell^1)$. We show that r_ℓ is a lower bound on the ratio $\mathrm{bs}(f^\ell)/\mathrm{fbs}(f^\ell)$. Let $z' \in \{0,1\}$ such that $\mathrm{fbs}^{z'}(f^\ell) = \mathrm{fbs}(f^\ell)$, then we have

$$\frac{\mathrm{bs}(f^\ell)}{\mathrm{fbs}(f^\ell)} = \frac{\mathrm{bs}(f^\ell)}{\mathrm{fbs}^{z'}(f^\ell)} \geq \frac{\mathrm{bs}^{z'}(f^\ell)}{\mathrm{fbs}^{z'}(f^\ell)} = r_\ell^{z'} \geq r_\ell \,.$$

By Lemma D.5,4, there exists a minimal $m \in \mathbb{N}$ such that $\mathrm{bs}(f^m) \geq 2 \cdot 2^n$. Then, by Lemma D.5,2 for any $z \in \{0,1\}$ we have $\mathrm{bs}^z(f^{m+1}) \geq 2 \cdot 2^n$. On the other hand, since m is the minimal such integer, using Theorem 3 gives

$$\mathrm{bs}^z(f^{m+1}) \leq \mathrm{bs}(f^{m+1}) \leq \mathrm{fbs}(f^2) \cdot \mathrm{bs}(f^{m-1}) \leq n^2 \cdot 2 \cdot 2^n \,.$$

By Lemma D.3 and Theorem 2 we have for all $\ell \in \mathbb{N}$:

$$\mathrm{fbs}^z(f^\ell) \underset{\text{Thm 2}}{\leq} \mathrm{C}^z(f^\ell) \underset{\text{Lemma D.3}}{\leq} \mathrm{bs}^z(f^\ell) \cdot \mathrm{sens}^{1-z}(f^\ell) \leq \mathrm{bs}^z(f^\ell) \cdot \mathrm{bs}^{1-z}(f^\ell)$$

In particular, for $\ell \leq m + 1$ we have

$$r_\ell^z = \frac{\mathrm{bs}^z(f^\ell)}{\mathrm{fbs}^z(f^\ell)} \geq \frac{1}{\mathrm{bs}^{1-z}(f^\ell)} \underset{\text{Lemma D.5,4}}{\geq} \frac{1}{\mathrm{bs}^{1-z}(f^{m+1})} \geq \frac{1}{2 \cdot n^2 \cdot 2^n} \,.$$

Next, we show that for $\ell \geq m + 1$:

$$r_{\ell+1} \geq r_\ell \cdot (1 - 2^{-1-\lfloor \frac{l-(m+1)}{2} \rfloor}) \,.$$

This will finish the proof as it gives a global lower bound on all r_ℓs for $\ell \geq m+1$:

$$r_\ell \geq r_{m+1} \cdot \prod_{i=1}^{\infty} (1-2^{-i})^2 \geq r_{m+1} \cdot 0.08 \geq \frac{1}{2^n \cdot 25 \cdot n^2} \ .$$

Let $x = (x^0 || x^1 || \ldots || x^{n-1})$ be the input maximizing $\mathrm{fbs}^z(f \circ f^\ell)$. Let y be the intermediate values in the composition $f \circ f^\ell$ i.e. $y_i = f^\ell(x^i)$. For p^0, p^1 which maximizes $\mathrm{bs}^0(f^\ell)$, $\mathrm{bs}^1(f^\ell)$ respectively, take $\tilde{x} = (\tilde{x}^0 || \tilde{x}^1 || \ldots || \tilde{x}^{n-1})$ to be an input such that $\tilde{x}^i = p^{y_i}$. Let $B_0, \ldots, B_{k-1}$ be the minimal sensitive blocks for f on y, then the value of $\mathrm{fbs}(f^{\ell+1})$ is the optimal value for the following linear program

$$\begin{array}{lll} \max & \sum_{j=0}^{k-1} w_j & \\ & & \\ \text{s.t.} & \sum_{j: i \in B_j} w_j \leq \mathrm{fbs}(f^\ell, x^i) & \text{for all } i \in [n] \\ & w_j \geq 0 & \text{for all } j \in [k] \end{array}$$

Denoting the optimal weights by w_j^*, and taking $w_j = \lfloor w_j^* \cdot r_\ell \rfloor$ gives

$$\forall i \in [n] : \sum_{j: i \in B_j} w_j \leq \mathrm{bs}(f^\ell, \tilde{x}^i) \ .$$

Thus, $\{w_j\}_{j \in [k]}$ a feasible solution for the block sensitivity integer program for $\mathrm{bs}(f \circ f^\ell, \tilde{x})$. As $(f \circ f^\ell)(\tilde{x}) = z$, this shows that

$$\mathrm{bs}^z(f^{\ell+1}) \geq \mathrm{fbs}^z(f^{\ell+1}) \cdot r_\ell - 2^n \ ,$$

hence the ratio

$$r_{\ell+1}^z = \frac{\mathrm{bs}^z(f^{\ell+1})}{\mathrm{fbs}^z(f^{\ell+1})} \geq r_\ell - \frac{2^n}{\mathrm{fbs}^z(f^{\ell+1})} \ .$$

For z' which minimizes $r_\ell^{z'}$ we have

$$r_{\ell+1}^z \geq r_\ell - \frac{2^n}{\mathrm{fbs}^z(f^{\ell+1})} = r_\ell \left(1 - \frac{2^n}{\mathrm{fbs}^z(f^{\ell+1})} \cdot \frac{\mathrm{fbs}^{z'}(f^\ell)}{\mathrm{bs}^{z'}(f^\ell)}\right) \underset{\text{Lemma D.5,3}}{\geq} r_\ell \left(1 - \frac{2^n}{\mathrm{bs}^{z'}(f^\ell)}\right)$$

Previous lemmas gives

$$\mathrm{bs}^{z'}(f^\ell) \underset{\text{Lemma D.4}}{\geq} \mathrm{bs}^{z'}(f^{\ell-(m+1)}) \cdot \min(\mathrm{bs}^0(f^{m+1}), \mathrm{bs}^1(f^{m+1})) \underset{\text{Lemma D.5,4}}{\geq} 2^{\lfloor \frac{\ell-(m+1)}{2} \rfloor} \cdot (2 \cdot 2^n) \ .$$

Thus $r_{\ell+1} \geq r_\ell \cdot (1 - 2^{-1-\lfloor \frac{\ell-(m+1)}{2} \rfloor})$, which completes the proof. □