A PROOF OF THE AARONSON–AMBAINIS CONJECTURE

DAVID TURTUREAN
Physical Superintelligence
david@psi.inc

Abstract. Aaronson and Ambainis conjectured that every bounded polynomial of low degree on the discrete cube has an influential variable. They showed that the conjecture implies that the acceptance probability of every quantum algorithm making T queries can be approximated on most inputs by a classical algorithm making $\mathrm{poly}(T)$ queries. We prove their conjecture: every function $f\colon \{-1,1\}^n \to [-1,1]$ of degree at most $d \geq 1$ has a variable of influence at least $c\,\mathrm{Var}(f)^{5/2}/d^6$, where $c > 0$ is an absolute constant. In the original formulation, for polynomials of degree at most d with values in $[0,1]$ on $\{0,1\}^n$ and variance at least ε, this gives a variable of influence at least $(\varepsilon/d)^7$. The proof is elementary, with steps inspired by works of others. Thus, as Aaronson and Ambainis anticipated, superpolynomial quantum speedups in the query model require structure, such as a promise on the inputs.

Date: November 2026.

2020 *Mathematics Subject Classification.* Primary 06E30; Secondary 05A20, 41A17, 68Q12.

Key words and phrases. Influence, bounded polynomials on the discrete cube, fractional certificates, fractional block sensitivity, quantum query complexity.

AI disclosure

The mathematical statements in this paper were produced by AI models. David Turturean scoped the problem, developed a research harness internally at Physical Superintelligence (PSI), and ran it on PSI's infrastructure. The run that produced this mathematical approach and proof was done with instances of a non-public Anthropic model, serving within the harness to search for existing ideas, to propose new ideas, to combine and test these ideas, and to supervise the work and the communication between other instances. Appendix A gives the timeline of the discovery. The proof of Theorem 1.1 and of Corollary 1.2 has been formalized in the Lean proof assistant and checked by its kernel; the formalization is in the repository `https://github.com/physical-superintelligence/aaronson-ambainis`.

The author prompted an instance of a non-public Anthropic model to package the results into a first draft of this paper, reviewed and edited the draft personally, and is responsible for its correctness.

Acknowledgements

We thank Jim Halverson for discussions, auditing the proof, and comments on a draft, and Phiala Shanahan for comments on the Lean formalization. We are especially grateful to Guy Blanc, Jordan Docter, Carmen Strassle and Li-Yang Tan for their comments and for writing a forthcoming shorter exposition of the proof for experts, and to Scott Aaronson for referring us to them.

Contents

AI disclosure 2
Acknowledgements 2
1. Introduction 4
1.1. The conjecture 5
1.2. Results 5
1.3. Motivation and consequences 8
1.4. Earlier work 9
1.5. Overview of the proof 12
2. Notation and definitions 14
2.1. The cube 14
2.2. Functions and degree 15
2.3. Boolean functions 16
3. A bijection theorem and the determinant inequality 17
3.1. Statements 18
3.2. Lowest monomials 18
3.3. The agreement matrix 19
3.4. Row operations along the wings 21
3.5. Examples and the relation to Reimer's Butterfly Lemma 23
4. The antipodal inequality 25
5. The all-pairs inequality and the two-sided inequality 26
5.1. From antipodal pairs to all pairs 27
5.2. Removing the weights 29
6. The row bound 31
6.1. Chebyshev polynomials and Markov's inequality at an endpoint 32
6.2. Bounded functions and their multilinear extensions 33
6.3. The sensitivity lemma 34
6.4. The block inequality 35
6.5. Fractional packings 35
6.6. Duality 36
6.7. The row bound 38
6.8. The increments at a point 39
7. Real functions: integration over the level 39
7.1. Level functions and their rows 40
7.2. Integration over the level 41
7.3. The third-moment inequality 43
7.4. The mean absolute increments 44
7.5. Relation to the argument of Lovett and Zhang 47
8. Averaging and the proof of the main theorem 47
8.1. Averaging over coordinates 48
8.2. From mean absolute increments to influences 50
8.3. Proof of the main theorem 51
9. Remarks and open questions 53
9.1. Certificates and decision trees 53
9.2. The exponent of the variance 56
9.3. The dependence on the degree 56
9.4. The constant in the row bound 57
9.5. Open questions 57
Appendix A. Timeline of the discovery 58
References 58

1. Introduction

Which problems can a quantum algorithm solve with exponentially fewer queries to its input than any classical algorithm? For decision problems this is possible only if the input is subject to a promise, since for total Boolean functions the quantum and the classical query complexity are polynomially related [BBCMW]. In Simon's problem [Sim] and in Shor's period finding [Sho], for instance, the input is promised to encode a periodic function, a structure that only a minuscule fraction of all inputs have [AA14].

Aaronson and Ambainis made precise the idea that exponential speedups need such structure, through a conjecture which they describe as folklore:

Conjecture 1.5 (folklore) [AA14]**.** *Let Q be a quantum algorithm that makes T queries to a Boolean input $X = (x_1, \dots, x_N)$, and let $\varepsilon > 0$. Then there exists a deterministic classical algorithm that makes* $\mathrm{poly}(T, 1/\varepsilon, 1/\delta)$ *queries to the x_i, and that approximates Q's acceptance probability to within an additive error ε on a $1 - \delta$ fraction of inputs.*

Aaronson and Ambainis show that it is equivalent to the following statement [AA14, Theorem 7.3]: if a problem has a fast quantum algorithm and its promise is satisfied by a constant fraction of all inputs, then a fast classical algorithm solves it on most of these inputs.

Aaronson and Ambainis also reduced this conjecture to a statement about polynomials that does not mention quantum algorithms: bounded polynomials of low degree have influential variables. This is the Aaronson–Ambainis conjecture, which Section 1.1 states in their words. The acceptance probability of a quantum algorithm that makes T queries is such a polynomial, of degree at most $2T$ in the bits of its input and with values in $[0, 1]$ [BBCMW]. If the conjecture holds, a classical algorithm can simulate the quantum algorithm on most inputs by querying a variable of large influence, fixing it, and repeating [AA14, Theorem 3.3].

The conjecture goes back to a lecture of Aaronson in 2008 [Aar08]. O'Donnell's book on the analysis of Boolean functions states it in the notes to Chapter 8 and observes that "if true, this conjecture would have significant consequences for the limitations of efficient quantum computation" [O'D, Section 8.7]. Before the present paper, the known bounds for general bounded polynomials were exponential in the degree [AA14, DFKO, DMP, OZ], and the conjecture had been proved only for special classes of polynomials; Section 1.4 describes this earlier work. An elegant attempt at the conjecture was announced in [KK19], which unfortunately was later withdrawn. We take a different approach.

In this paper we prove the conjecture of Aaronson and Ambainis: every function f on $\{-1, 1\}^n$ with values in $[-1, 1]$ and of degree at most $d \geq 1$ has a variable j with

$$\mathrm{Inf}_j(f) \;\geq\; c\,\frac{\mathrm{Var}(f)^{5/2}}{d^6},$$

where $\mathrm{Inf}_j(f)$ is the influence of the jth variable and $c > 0$ is an absolute constant (Theorem 1.1). In the normalisation of [AA14], a polynomial of degree at most d with values in $[0, 1]$ on $\{0, 1\}^n$ and variance at least ε has a variable of influence at least $(\varepsilon/d)^7$ (Corollary 1.2). The proof is elementary, with steps inspired by work of Tal [Tal]; of Gilmer, Saks and Srinivasan [GSS]; of Agarwal and Ben-David [ABD]; of Reimer [Rei]; and of Lovett and Zhang [LZ].

Consequently the folklore conjecture holds: the acceptance probability of every quantum algorithm that makes T queries can be approximated to within ε on all but a δ fraction of the inputs by a deterministic classical algorithm that makes at most $(4T)^8\varepsilon^{-14}\delta^{-8}$ queries (Section 1.3). The other consequences that Aaronson and Ambainis derive from their conjecture hold as well. For instance, separating BQP from AvgP, the class of problems solvable in polynomial time on most inputs, relative to a random oracle would require a proof that $\mathsf{P} \neq \mathsf{P}^{\#\mathsf{P}}$ [AA14, Theorem 3.5].

We now state the conjecture and our results precisely; Section 1.5 outlines the proof.

1.1. **The conjecture.** For a real polynomial $p\colon \mathbb{R}^N \to \mathbb{R}$, Aaronson and Ambainis [AA14] let

$$\mathrm{Inf}_i[p] \;:=\; \mathbb{E}_{X\in\{0,1\}^N}\big[(p(X)-p(X^i))^2\big]$$

be the influence of the ith variable, where X^i means X with the ith bit flipped, and they state the following conjecture [AA14, Conjecture 1.7], which we quote.

> **Conjecture 1.7** (Bounded Polynomials Have Influential Variables)**.** *Let $p\colon \mathbb{R}^N \to \mathbb{R}$ be a polynomial of degree d. Suppose that $0 \le p(X) \le 1$ for all $X \in \{0,1\}^N$, and*
>
> $$\mathbb{E}_{X\in\{0,1\}^N}\big[(p(X)-\mathbb{E}[p])^2\big] \;\ge\; \varepsilon.$$
>
> *Then there exists an i such that $\mathrm{Inf}_i[p] \ge (\varepsilon/d)^{O(1)}$.*

Corollary 1.2 below proves it with the explicit exponent 7 in place of $O(1)$.

1.2. **Results.** It is convenient to work on the cube $Q_n = \{-1,1\}^n$ with the uniform probability measure; $\mathbb{E}$ and Var denote expectation and variance with respect to it.

For $x \in Q_n$ and $j \in [n] = \{1,\dots,n\}$ let x^j be the point obtained from x by changing the sign of the jth coordinate. For $f\colon Q_n \to \mathbb{R}$ put

$$\mathrm{Inf}_j(f) \;:=\; \mathbb{E}_x\Big[\Big(\frac{f(x)-f(x^j)}{2}\Big)^2\Big].$$

Every function $f\colon Q_n \to \mathbb{R}$ can be written in exactly one way as $f = \sum_{T\subseteq[n]} \hat{f}(T)\chi_T$ with real coefficients $\hat{f}(T)$, where $\chi_T(x) = \prod_{i\in T} x_i$. We say that f has *degree at most d* if $\hat{f}(T) = 0$ whenever $|T| > d$.

Theorem 1.1 (Main theorem)**.** *Let $n \ge 1$ and $d \ge 1$ be integers and let $f\colon \{-1,1\}^n \to [-1,1]$ be a function of degree at most d. Then there is a coordinate $j \in [n]$ with*

$$\mathrm{Inf}_j(f) \;\ge\; \frac{\sqrt{15}}{500\,\pi^4}\cdot\frac{\mathrm{Var}(f)^{5/2}}{d^6}.$$

Remarks on Theorem 1.1. If $\mathrm{Var}(f) = 0$ there is nothing to prove, and a function of degree at most 0 is constant; so the content of the theorem concerns non-constant functions, and for these the assumption $d \ge 1$ is automatic.

The constant $\sqrt{15}/(500\pi^4)$ is approximately $7.95\cdot 10^{-5}$.

It can be multiplied by 4 if the bound of Agarwal and Ben-David [ABD] for the fractional block sensitivity (recalled in Remark 6.9) is used in place of Theorem 6.8 below, and by a

further factor 2 if an inequality of Filmus, Hatami, Keller and Lifshitz [FHKL] is used in place of Corollary 6.10 below; see Remark 8.4.

The original normalisation. The next corollary is the statement in the normalisation of [AA14]. We write $\mathrm{Var}[p]$ for the variance of $p(X)$ when X is uniform on $\{0,1\}^n$.

Corollary 1.2. *Let $d \geq 0$ be an integer, and let $p\colon \mathbb{R}^n \to \mathbb{R}$ be a polynomial of degree at most d such that $0 \leq p(X) \leq 1$ for all $X \in \{0,1\}^n$ and $\mathrm{Var}[p] > 0$. Then $n \geq 1$, $d \geq 1$, and there exists an i such that*

$$\mathrm{Inf}_i[p] \;\geq\; \frac{8\sqrt{15}}{125\,\pi^4} \cdot \frac{\mathrm{Var}[p]^{5/2}}{d^6} \;\geq\; \Big(\frac{\mathrm{Var}[p]}{d}\Big)^7.$$

In particular, Conjecture 1.7 of [AA14] *holds in the following form: if $\mathrm{Var}[p] \geq \varepsilon > 0$, then there exists an i such that*

$$\mathrm{Inf}_i[p] \;\geq\; \frac{8\sqrt{15}}{125\,\pi^4} \cdot \frac{\varepsilon^{5/2}}{d^6} \;\geq\; \Big(\frac{\varepsilon}{d}\Big)^7.$$

Remarks on Corollary 1.2. The corollary follows from Theorem 1.1 by the change of variables $X_i = (1-x_i)/2$ and a numerical estimate; the proof is given at the end of Section 8.

The exponent 7 only absorbs the constant: since $\mathrm{Var}[p] \leq \frac{1}{4}$, we have $\mathrm{Var}[p]^{5/2}/d^6 \geq 128\,(\mathrm{Var}[p]/d)^6$, and $128 \cdot 8\sqrt{15}/(125\pi^4) > \frac{1}{4} \geq \mathrm{Var}[p]/d$.

Rows and mean absolute increments. The proof of Theorem 1.1 uses two properties of the function f, which are expressed with the following notions. For $S \subseteq [n]$ let x^S be the point obtained from x by changing the signs of the coordinates in S.

Definition (Rows)**.** A *row* of a function $f\colon Q_n \to \mathbb{R}$ at a point $x \in Q_n$ is a vector $a \in [0,\infty)^n$ such that

$$|f(x) - f(x^S)| \;\leq\; a(S) := \sum_{i \in S} a_i \qquad \text{for every } S \subseteq [n].$$

The number $\|a\|_1 = \sum_i a_i$ is the *cost* of the row.

Definition (Mean absolute increments)**.** The *mean absolute increment* of f in the coordinate j is

$$L_j(f) \;:=\; \mathbb{E}_x |f(x) - f(x^j)|,$$

and $L(f) = \sum_{j=1}^n L_j(f)$.

The number $\frac{1}{2}L_j(f)$ is known as the L_1 influence of the coordinate j on f [BB, FHKL]. For a polynomial p on $\{0,1\}^N$, Aaronson and Ambainis write $\mathrm{Inf}_i^1[p] = \mathbb{E}|p(X) - p(X^i)|$ for the corresponding quantity [AA14, Section 1.3].

If a row a^x is chosen at every point $x \in Q_n$, the vectors a^x are the rows of a matrix with non-negative entries, with one row for each point of Q_n and one column for each coordinate; hence the name, and the cost of a row is its row sum. Taking $S = \{j\}$ in the definition of a row gives $a^x_j \geq |f(x) - f(x^j)|$, so the mean of the jth column is at least $L_j(f)$.

A function $f\colon Q_n \to [-1,1]$ of degree at most d has a row of cost at most $\pi^2 d^2$ at every point (Theorem 6.8), and it satisfies $L(f) \leq 4d^2$ (Corollary 6.10). This row bound is due to Agarwal and Ben-David [ABD, Theorem 1.7], with the better constant $\pi^2/2$ in place of π^2 (see Remark 6.9). That $L(f)$ is bounded by a polynomial in d was first proved by Bačkurs

and Bavarian [BB], and Filmus, Hatami, Keller and Lifshitz proved $L(f) \le 2d^2$ [FHKL, Theorem 3.3].

The next two theorems concern arbitrary real functions with a row of bounded cost at every point; no bound for the values of f is assumed in them.

Theorem 1.3 (Mean absolute increments)**.** *Let $n \ge 0$, let $f\colon Q_n \to \mathbb{R}$ be a function with $\mathrm{Var}(f) > 0$, and let K be a real number such that for every $x \in Q_n$ there is a row a^x of f at x with $\|a^x\|_1 \le K$. Then $n \ge 1$ and $K > 0$, and there is a coordinate $j \in [n]$ with*

$$\mathbb{E}_x |f(x) - f(x^j)| \;\ge\; \frac{\mathrm{Var}(f)^{3/2}}{6\,K^2}.$$

Theorem 1.4 (Influences)**.** *Let $n \ge 0$, let $f\colon Q_n \to \mathbb{R}$ be a function with $\mathrm{Var}(f) > 0$, and let K be a real number such that for every $x \in Q_n$ there is a row a^x of f at x with $\|a^x\|_1 \le K$. Let $L(f) = \sum_{k=1}^n L_k(f)$. Then $n \ge 1$ and $0 < L(f) \le K$, and there is a coordinate $j \in [n]$ with*

$$\mathrm{Inf}_j(f) \;\ge\; \frac{\sqrt{15}}{125} \cdot \frac{\mathrm{Var}(f)^{5/2}}{L(f)\,K^2}.$$

In particular, $\mathrm{Inf}_j(f) \ge (\sqrt{15}/125)\ \mathrm{Var}(f)^{5/2}/K^3$ for this coordinate j.

Remarks on Theorems 1.3 and 1.4. These are Theorems 7.10 and 8.3 below, where the inequalities are also stated in a form without square roots.

Theorem 1.1 follows from Theorem 1.4 with $K = \pi^2 d^2$ (Theorem 6.8) and $L(f) \le 4d^2$ (Corollary 6.10).

Boolean functions. Theorems 1.3 and 1.4 are derived from an inequality for Boolean functions with a row at every point, the all-pairs inequality (Proposition 5.2), which is described in Section 1.5. For Boolean functions themselves this inequality has a further consequence, Theorem 1.5 below. For a Boolean function $g\colon Q_n \to \{0,1\}$ the difference $|g(x) - g(x^S)|$ is 1 or 0, so a row of g at x is a vector $\lambda \in [0,\infty)^n$ with $\lambda(S) \ge 1$ for every set S that changes the value of g, that is, with $g(x^S) \ne g(x)$.

Rows of Boolean functions are also called fractional certificates [Tal, GSS, KT]: the rows with entries in $\{0,1\}$ are exactly the indicator vectors of the certificates of g at x, that is, of the sets $C \subseteq [n]$ such that g is constant on $\{y : y_i = x_i \text{ for all } i \in C\}$.

We write $\mathrm{I}_j(g) = \Pr_x[g(x) \ne g(x^j)]$ for the influence of the coordinate j on the Boolean function g; thus $\mathrm{I}_j(g) = 4\,\mathrm{Inf}_j(g)$.

Here and below, $1_{\{g(x)=1\}}$ denotes the number which is 1 if $g(x) = 1$ and 0 otherwise, and similarly for other conditions in place of $g(x) = 1$.

Theorem 1.5 (Two-sided inequality)**.** *Let $g\colon Q_n \to \{0,1\}$ be non-constant, and for every $x \in Q_n$ let λ^x be a row of g at x. Then*

$$4\ \mathbb{E}_x\Big[\frac{1_{\{g(x)=1\}}}{\|\lambda^x\|_1}\Big]\ \mathbb{E}_x\Big[\frac{1_{\{g(x)=0\}}}{\|\lambda^x\|_1}\Big] \;\le\; \max_{j\in[n]} \mathrm{I}_j(g).$$

Remarks on Theorem 1.5. This is Theorem 5.4 below.

There is equality if $g(x) = 1$ exactly when $x_1 = 1$ and $\lambda^x = (1, 0, \dots, 0)$ for all x (Example 5.5).

We call the inequality two-sided because the sets $g^{-1}(1)$ and $g^{-1}(0)$ enter in the same way.

For a non-constant Boolean function g with a row of cost at most K at every point it gives a coordinate j with $\mathrm{Inf}_j(g) \geq \mathrm{Var}(g)/K^2$ (Section 9.5). If every point of $g^{-1}(1)$ has a certificate of size at most c_1 and every point of $g^{-1}(0)$ one of size at most c_0, the theorem, applied to the indicator vectors of these certificates, gives $\max_j \mathrm{I}_j(g) \geq 4\Pr[g=1]\Pr[g=0]/(c_0c_1)$; this special case also follows from the inequality of O'Donnell, Saks, Schramm and Servedio [OSSS] for decision trees and the classical bound $D(g) \leq c_0c_1$ for the least depth $D(g)$ of a decision tree that computes g. Section 9.1 compares Theorem 1.5 with the inequality of [OSSS].

1.3. **Motivation and consequences.** The conjecture comes from quantum query complexity. By a lemma of Beals, Buhrman, Cleve, Mosca and de Wolf [BBCMW] (stated as Lemma 3.2 in [AA14]), the acceptance probability of a quantum algorithm that makes T queries to a Boolean input $X \in \{0,1\}^N$ is a real multilinear polynomial in X of degree at most $2T$, and it takes values in $[0,1]$.

Aaronson and Ambainis relate their Conjecture 1.7 to the following conjecture, which they describe as folklore [AA14, Conjecture 1.5]:

> *Let Q be a quantum algorithm that makes T queries to a Boolean input $X = (x_1, \ldots, x_N)$, and let $\varepsilon > 0$. Then there exists a deterministic classical algorithm that makes $\mathrm{poly}(T, 1/\varepsilon, 1/\delta)$ queries to the x_i, and that approximates Q's acceptance probability to within an additive error ε on a $1-\delta$ fraction of inputs.*

Aaronson and Ambainis derive three consequences from Conjecture 1.7 [AA14, Theorem 1.8]. To state them, let $D_\varepsilon(f)$ be the minimum number of queries made by a deterministic algorithm that evaluates a Boolean function $f\colon \{0,1\}^N \to \{0,1\}$ correctly on at least a $1-\varepsilon$ fraction of the inputs. Let $Q_\varepsilon(f)$ be the corresponding minimum for quantum algorithms, which evaluate f with bounded error on such a fraction of the inputs, as in the proof of [AA14, Corollary 3.4].

Finally, let AvgP be the class of languages for which there exists a polynomial-time algorithm that solves a $1-o(1)$ fraction of the instances of each size [AA14, Section 1.1].

Theorem 1.8 of [AA14] reads:

> *Assume Conjecture 1.7. Then*
>
> (i) *Conjecture 1.5 holds.*
>
> (ii) *$D_{\varepsilon+\delta}(f) \leq (Q_\varepsilon(f)/\delta)^{O(1)}$ for all Boolean functions $f\colon \{0,1\}^N \to \{0,1\}$ and all $\varepsilon, \delta > 0$.*
>
> (iii) *If $\mathsf{P} = \mathsf{P}^{\#\mathsf{P}}$, then $\mathsf{BQP}^A \subset \mathsf{AvgP}^A$ with probability 1 for a random oracle A.*

These three implications are proved in [AA14] as Theorem 3.3, Corollary 3.4 and Theorem 3.5.

Because we prove Conjecture 1.7 (Corollary 1.2), the three statements (i), (ii) and (iii) hold.

For (i), the proof of [AA14, Theorem 3.3] with the exponent 7 gives a deterministic classical algorithm that makes at most $(4T)^8\varepsilon^{-14}\delta^{-8}$ queries. Indeed, the acceptance probability has degree at most $2T$, so its total influence is at most $2T$ [AA14, Lemma 3.1]; as long as

the variance of the polynomial obtained by fixing the bits queried so far exceeds $\varepsilon^2\delta/2$, Corollary 1.2 gives a variable of influence greater than $(\varepsilon^2\delta/(4T))^7$, which the algorithm queries next; and by Markov's inequality the algorithm stops within $4T/(\delta(\varepsilon^2\delta/(4T))^7) = (4T)^8\varepsilon^{-14}\delta^{-8}$ queries, except with probability at most $\delta/2$.

As is noted in [AA14, Section 1.1], statement (ii) is a quantum counterpart of a theorem of Smyth, which bounds the approximate decision-tree complexity of a Boolean function by a quantity that is quadratic in its approximate certificate complexity [Smy02, Smy].

Girish, Mittal, Nehoran and Raz [GMNR] replace the uniform distribution on the inputs by other distributions: for a fixed $p \in (0,1)$ and every $n \geq 1$ they let μ_p be the product distribution on $\{0,1\}^n$ under which the coordinates are independent and each is equal to 1 with probability p, and ν_p the uniform distribution on the strings in $\{0,1\}^n$ of Hamming weight $\lfloor pn \rfloor$. Let $\mathcal{D}$ stand for μ_p or for ν_p.

In our notation, the statement $\mathrm{AA}(\mathcal{D})$ of [GMNR, Definition 4.1] reads as follows.

> There are constants $c, C > 0$ and a function $S\colon (0, \frac{1}{4}] \times \{1, 2, \dots\} \to [0,1]$ with $S(v,d) \geq c\,(v/d)^C$ such that for all $n, d \geq 1$ and all $v \in (0, \frac{1}{4}]$, every $f\colon \{0,1\}^n \to [0,1]$ that is given by a multilinear polynomial of degree at most d and whose variance with respect to $\mathcal{D}$ is at least v has a coordinate i with
>
> $$\mathbb{E}_{X\sim\mathcal{D}}\big[(f(X) - f(X^i))^2\big] \;\geq\; S(v,d).$$

It is proved in [GMNR, Theorem 4.3] that the statements $\mathrm{AA}(\mu_p)$ and $\mathrm{AA}(\nu_q)$ are equivalent for all $p, q \in (0,1)$.

It is proved in [GMNR, Theorem 4.14] that each of them implies the following, again in our notation, for every $p \in (0,1)$, for $\mathcal{D} = \mu_p$ as well as for $\mathcal{D} = \nu_p$.

> There is a polynomial P such that for all $n, T \geq 1$, all $\varepsilon \in (0,1)$ and every quantum algorithm that makes at most T queries to an input $X \in \{0,1\}^n$, with acceptance probability $Q(X)$, there is a deterministic classical algorithm that makes at most $P(T, 1/\varepsilon)$ queries and whose output $\widetilde{Q}(X) \in [0,1]$ satisfies
>
> $$\mathbb{E}_{X\sim\mathcal{D}}|\widetilde{Q}(X) - Q(X)| \;\leq\; \varepsilon.$$

The statement $\mathrm{AA}(\mu_{1/2})$ concerns the uniform distribution, and it holds by Corollary 1.2, with $S(v,d) = (v/d)^7$.

Hence, by [GMNR, Theorem 4.3], the statements $\mathrm{AA}(\mu_p)$ and $\mathrm{AA}(\nu_p)$ hold for every $p \in (0,1)$, and so does the conclusion of [GMNR, Theorem 4.14].

In summary, by Corollary 1.2 the statements (i)–(iii) above hold, and for every $p \in (0,1)$ so do $\mathrm{AA}(\mu_p)$, $\mathrm{AA}(\nu_p)$ and the simulation of [GMNR, Theorem 4.14]. Moreover, by [BDST, Claim A.1], Corollary 1.2 implies the conjecture of Blanc, Docter, Strassle and Tan on query weights [BDST, Conjecture 1]; so their Theorem 11 holds without that assumption: if $\mathsf{PromiseBPP} = \mathsf{PromiseBQP}$, then $\mathsf{PromiseBQP}^A \subseteq \mathsf{PromiseHeurBPP}^A$ with probability 1 for a random oracle A [BDST, Theorem 11].

1.4. **Earlier work.** We recall what was known, in the notation of the papers cited.

Boolean functions. O'Donnell, Saks, Schramm and Servedio [OSSS] proved, in the formulation of [AA14, Theorem 1.9], that a Boolean function $f\colon \{0,1\}^N \to \{0,1\}$ with $\Pr[f = 1]$

$\Pr[f = 0] \geq \varepsilon$ has a variable with $\mathrm{Inf}_i[f] \geq 4\varepsilon/D(f)$, where $D(f)$ is the decision-tree complexity of f.

By a theorem of Nisan and Szegedy [NS] the decision-tree complexity of a Boolean function is bounded by a polynomial in its degree, so the Aaronson–Ambainis conjecture was known for polynomials that take only the values 0 and 1 on the cube; see the introduction of [LZ].

As is pointed out in [AA14], this does not settle the case of bounded polynomials, because the acceptance probability of a quantum algorithm need not approximate a total Boolean function.

Exponential dependence on the degree. As observed in [AA14], a theorem of Dinur, Friedgut, Kindler and O'Donnell [DFKO] implies the conjecture with the weaker bound $\mathrm{Inf}_i[p] \geq \varepsilon^3/2^{O(d)}$ in place of $(\varepsilon/d)^{O(1)}$.

Defant, Mastyło and Pérez [DMP] derive the bound

$$\max_j \mathrm{Inf}_j(f) \;\geq\; \frac{\mathrm{Var}(f)^2}{d^4 e^{4d}}$$

for functions $f\colon \{-1,1\}^n \to [-1,1]$ of degree d; see also [OZ].

Special classes. Montanaro [Mon] proved the Aaronson–Ambainis conjecture for multilinear forms all of whose coefficients have the same absolute value; here a multilinear form is a polynomial in k blocks of variables in which every monomial is a product of k variables, one from each block (see also [AA14, Section 1.3]).

Ivanisvili [Iva] proved it for polynomials whose Fourier coefficients have the same absolute value on each level, as an application of inequalities of Bohnenblust–Hille type on the cube whose constants grow polynomially in the degree; see also [STV], and [DMP] for the relation between such inequalities and the conjecture.

Bansal, Sinha and de Wolf [BSW] proved that every completely bounded block-multilinear form of degree d with constant variance has a variable with influence at least $1/\mathrm{poly}(d)$. Escudero Gutiérrez [EG] gave another proof of this result and proved the conjecture for homogeneous Fourier completely bounded polynomials.

O'Donnell and Zhao [OZ] showed that the conjecture holds if and only if it holds for one-block decoupled functions.

Bhattacharya [Bha, Theorem 6.4] showed that the conjecture is true for a non-negligible fraction of the random restrictions of a given polynomial, assuming that its variance is not too low.

Restricted quantum algorithms. Two recent papers prove statements of the kind of Conjecture 1.5 of [AA14], which is quoted above, for quantum algorithms with limited adaptivity, without the use of Conjecture 1.7.

Blanc, Docter, Strassle and Tan [BDST] prove that for every quantum algorithm that makes t queries in d rounds, the queries of each round being made in parallel, and for all $\varepsilon, \delta > 0$ there is a classical algorithm that makes $2^{O(d^2)} \cdot (t \log(1/\delta)/\varepsilon)^{O(d)}$ queries and approximates the acceptance probability of the quantum algorithm to within an additive error ε on a $1 - \delta$ fraction of the inputs [BDST, Theorem 1]. For a bounded number d of rounds this number of queries is polynomial in t, $1/\varepsilon$ and $\log(1/\delta)$, whereas the number obtained from Corollary 1.2 through [AA14, Theorem 3.3] is polynomial in t, $1/\varepsilon$ and $1/\delta$, for every quantum algorithm.

Liu and Mutreja [LM] prove Conjecture 1.5 for three classes of quantum algorithms (see [LM, Theorem 1.2], which is stated there informally): algorithms that make a single layer of parallel quantum queries; algorithms in which such a layer is preceded by a bounded, slowly growing number of adaptive quantum queries or by polynomially many adaptive classical queries; and algorithms with a constant number of adaptive layers of parallel queries.

Fractional certificates. Fractional certificates of Boolean functions were introduced independently by Tal [Tal] and by Gilmer, Saks and Srinivasan [GSS]; see the introduction of [KT]. Up to a constant factor, the least cost of a fractional certificate at a point is the randomized certificate complexity which Aaronson had introduced earlier [Aar]; see [GSS, Appendix A].

For a function $f\colon \{0,1\}^n \to [0,1]$, a point x and $\varepsilon > 0$, Lovett and Zhang [LZ] define the ε-fractional certificate complexity of f at x as the least number k for which there is a probability distribution π on $[n]$ such that

$$\pi(B) \;\geq\; \frac{1}{k} \qquad \text{for every set } B \text{ of coordinates with } |f(x \oplus B) - f(x)| \geq \varepsilon,$$

where $x \oplus B$ is x with the bits in B flipped.

They prove that it is at most $\mathrm{poly}(d, 1/\varepsilon, \log n)$ at every point if f is computed by a polynomial of degree d [LZ, Theorem 2].

The fractional certificates of the present paper are the rows of Definition 2.3, which is recalled in Section 1.2. A row a of cost $K > 0$ at x gives, for every $\varepsilon > 0$, an ε-fractional certificate in the sense of Lovett and Zhang, namely $\pi = a/K$ with $k = K/\varepsilon$; the relation of rows to the other notions is described after Definition 2.3.

For rows a bound in terms of the degree alone is known: Agarwal and Ben-David [ABD, Theorem 3.3] proved that the fractional block sensitivity (recalled in Remark 6.9) of every $f\colon \{-1,1\}^n \to [-1,1]$ is at most $\frac{\pi^2}{4}\deg(f)^2$, adapting a method of Anshu, Ben-David and Kundu [ABK] which expands on earlier work of Tal [Tal].

For Boolean functions the bound $\deg(f)^2$ is due to Kulkarni and Tal [KT].

By linear programming duality the bound of Agarwal and Ben-David gives rows of cost at most $\frac{\pi^2}{2}\deg(f)^2$ at every point (Remark 6.9). The statement about rows is [ABD, Theorem 1.7], where it is formulated for functions with values in $[0,1]$, with the constant $\frac{\pi^2}{4}$.

Section 6 contains a complete proof of the bound with a constant that is larger by a factor 2, and discusses the relation with [ABD].

From fractional certificates to the Aaronson–Ambainis conjecture. Lovett and Zhang also show that the Aaronson–Ambainis conjecture would follow from their bound together with a conjectured extension of Talagrand's concentration inequality to the L_∞ norm [LZ, Section 5].

Their conjectured inequality [LZ, Conjecture 37] concerns two subsets X and Y of the cube neither of which has an influential variable, and it asserts that some point of X is close in the L_∞ norm to the convex hull of Y.

Their derivation applies it to a sublevel set and a superlevel set of f; it uses that a fractional certificate at a point bounds from below the L_∞ distance from this point to the convex hull of the set of points at which the value of f differs from its value there by at least ε [LZ, Lemma 28].

Conjecture 37 of Lovett and Zhang is not true as stated: for the tribes function g of width w with 2^w tribes, every point of $X = g^{-1}(1)$ has L_∞ distance $1/w$ from the convex hull of $Y = g^{-1}(0)$, while all the influences of X and of Y in the sense of [LZ, Definition 36] are at most $1/(2^w - 1)$ (Section 9.1). The proof of Theorem 1.1 follows the plan of Lovett and Zhang in passing to the level sets of f, but it replaces their Conjecture 37 by the all-pairs inequality for Boolean functions (Proposition 5.2), in which the rows at the points of value 1 and at the points of value 0 enter in the same way; this inequality is applied at every level and integrated over the levels, so that no single level is chosen (Section 7.5).

1.5. **Overview of the proof.** In this overview, $g\colon Q_n \to \{0,1\}$ is a Boolean function, $\mathcal{O} = g^{-1}(1)$, $\mathcal{N} = g^{-1}(0)$, and

$$b_j(g) \;:=\; \#\{x \in \mathcal{O} :\; x^j \in \mathcal{N}\} \;=\; 2^{n-1}\, \mathrm{I}_j(g) \qquad (j \in [n]).$$

The antipode of x is $\bar{x} = -x$, and $Z(g) = \{u \in \mathcal{O} : \bar{u} \in \mathcal{N}\}$.

The proof of Theorem 1.1 is a chain of six links. The first three concern Boolean functions, the fourth concerns bounded functions of low degree, and the last two concern real functions with a row of bounded cost at every point.

Link 1: a bijection theorem and the determinant inequality (Section 3). A 1-*wing* at $x \in \mathcal{O}$ is a set $A \subseteq [n]$ such that $x^S \in \mathcal{O}$ for every $S \subseteq A$.

Theorem 3.1 states that for every choice of a 1-wing A_x at each $x \in \mathcal{O}$ there is a bijection $\tau\colon \mathcal{O} \to \mathcal{O}$ such that $\tau(x)$ differs from x in every coordinate of A_x, and such that for every j the number of $x \in \mathcal{O}$ with $\tau(x)_j = x_j$ is at most $b_j(g)$.

Without the last condition, the existence of such a bijection is equivalent to the Butterfly Lemma of Reimer [Rei]; see Remark 3.10.

The proof studies the determinant of the matrix with entries $\prod_{i:\, x_i = z_i} s_i$ $(x, z \in \mathcal{O})$, a polynomial in indeterminates $s_1, \dots, s_n$: we show that it is not zero, that every monomial of lowest total degree occurring in it has degree at most $b_j(g)$ in s_j, and that at least one such monomial comes from a permutation of $\mathcal{O}$ compatible with the wings.

If positive lengths $\ell_1, \dots, \ell_n$ are assigned to the coordinates, the bijection yields the *determinant inequality* (Corollary 3.2)

$$\sum_{u \in Z(g)} D_g(u, A_u; \ell) \;\le\; \sum_{j=1}^{n} \ell_j\, b_j(g),$$

where $D_g(u, A; \ell)$ is the least total length of a set G of coordinates, disjoint from A, such that $\bar{u}^{\,G} \in \mathcal{O}$.

Link 2: the antipodal inequality (Section 4). Let a row λ^x of g be given at every point x, and let $w \in (0,\infty)^n$. Theorem 4.3 states that

$$\sum_{u \in Z(g)} \Big(\sum_{j=1}^{n} w_j \lambda_j^u \lambda_j^{\bar{u}} \Big)^{-1} \;\le\; \sum_{j=1}^{n} \frac{b_j(g)}{w_j}.$$

It follows from the determinant inequality with $\ell_j = 1/w_j$: the two rows at u and at $\bar{u}$ single out a 1-wing A at u for which $D_g(u, A; \ell)$ is at least the reciprocal on the left.

Link 3: from antipodal pairs to all pairs (Section 5). Every pair $(x, y) \in \mathcal{O} \times \mathcal{N}$ is a pair of antipodal points of exactly one subcube of Q_n. Applying the antipodal inequality to the restrictions of g to all subcubes and adding up gives the *all-pairs inequality* (Proposition 5.2)

$$\sum_{x\in\mathcal{O}} \sum_{y\in\mathcal{N}} \Big(\sum_{j\colon x_j\neq y_j} w_j \lambda_j^x \lambda_j^y \Big)^{-1} \;\le\; 2^{n-1} \sum_{j=1}^{n} \frac{b_j(g)}{w_j}.$$

Two applications of the Cauchy–Schwarz inequality and an explicit choice of the weights w remove the weights from an inequality of this kind (Lemma 5.3).

For the all-pairs inequality itself the result is Theorem 1.5. That theorem is not used in the proof of Theorem 1.1; the lemma is used again in Link 5.

Link 4: the row bound (Section 6). If $f\colon Q_n \to [-1, 1]$ has degree at most d, then for every x there is a vector $a \in [0, \infty)^n$ with $\|a\|_1 \le \pi^2 d^2$ and

$$|f(x) - f(y)| \;\le \sum_{i\colon x_i\neq y_i} a_i \qquad \text{for all } y$$

(Theorem 6.8).

The proof combines Markov's inequality for the derivative of a polynomial, a composition with Chebyshev polynomials, and linear programming duality.

The bound is known with a better constant, and the proof follows the method of Agarwal and Ben-David [ABD]; see Remark 6.9.

An intermediate step of this proof, the block inequality (Lemma 6.4), also shows that $\sum_j |f(x) - f(x^j)| \le 4d^2$ at every point x (Corollary 6.10).

Link 5: integration over the level (Section 7). Let $f\colon Q_n \to \mathbb{R}$ be a function with a row a^x of cost at most K at every point x. For $w \in (0, \infty)^n$ put

$$s_{xy} \;=\; \sum_{j\colon x_j\neq y_j} w_j \, a_j^x \, a_j^y.$$

For a level t let $g_t = 1_{\{f>t\}}$. If $f(x) \neq t$, the vector a^x, divided by $|f(x) - t|$, is a row of g_t at x. With these rows, and with the row $(1, \dots, 1)$ at the points x with $f(x) = t$, the all-pairs inequality for g_t gives

$$\sum_{x,y\colon\, f(y)<t<f(x)} \frac{(f(x) - t)\,(t - f(y))}{s_{xy}} \;\le\; 2^{n-1} \sum_{j=1}^{n} \frac{b_j(g_t)}{w_j}.$$

We integrate this inequality over t. On the left, the integral of $(f(x) - t)(t - f(y))$ over the levels t between $f(y)$ and $f(x)$ is $(f(x) - f(y))^3/6$. On the right, the integral of $b_j(g_t)$ is $2^{n-1} L_j(f)$.

The result is the *integrated all-pairs inequality* (Proposition 7.4)

$$\sum_{x,y\colon\, f(y)<f(x)} \frac{(f(x) - f(y))^3}{6\, s_{xy}} \;\le\; 4^{n-1} \sum_{j=1}^{n} \frac{L_j(f)}{w_j}.$$

We keep the pairs with $f(y) \le r < f(x)$, for a real number r, and remove the weights with the lemma of Link 3. This gives the *third-moment inequality* (Theorem 7.6)

$$\mathbb{E}\big[(\max(f - r, 0))^3\big] \cdot \Pr[f \le r] \;\le\; \frac{3}{2} K^2 \max_{j\in[n]} L_j(f).$$

Applied at a median m of f, and also to $-f$, it shows that

$$\mathbb{E}|f-m|^3 \;\le\; 6K^2 \max_{j\in[n]} L_j(f).$$

Since $\mathrm{Var}(f)^{3/2} \le \mathbb{E}|f-m|^3$, this is Theorem 1.3.

Link 6: averaging (Section 8). To pass from the numbers $L_j(f)$ to the influences, consider the coordinates k with

$$5\,L(f)\,\mathrm{Inf}_k(f) \;<\; 2\,\mathrm{Var}(f)\,L_k(f).$$

The sum of their influences is at most $\frac{2}{5}\,\mathrm{Var}(f)$. Let h be the function obtained from f by averaging over these coordinates.

Then the function h has a row of cost at most K at every point, and

$$\mathrm{Var}(h) \ge \tfrac{3}{5}\,\mathrm{Var}(f), \qquad L_j(h) \le L_j(f) \;\text{ for every } j,$$
$$L_j(h) = 0 \;\text{ for the coordinates } j \text{ over which we have averaged.}$$

By Theorem 1.3, applied to h, there is a coordinate j with $L_j(h) \ge \mathrm{Var}(h)^{3/2}/(6K^2)$. Since $L_j(h) > 0$, we have not averaged over j, and so

$$5\,L(f)\,\mathrm{Inf}_j(f) \;\ge\; 2\,\mathrm{Var}(f)\,L_j(f) \;\ge\; 2\,\mathrm{Var}(f)\,L_j(h).$$

This gives Theorem 1.4.

With the two results of Link 4, Theorem 1.4 gives Theorem 1.1.

In summary,

bijection theorem (Theorem 3.1) $\Longrightarrow$ determinant inequality (Corollary 3.2)
$\Longrightarrow$ antipodal inequality (Theorem 4.3) $\Longrightarrow$ all-pairs inequality (Proposition 5.2)
$\Longrightarrow$ integrated all-pairs inequality (Proposition 7.4)
$\Longrightarrow$ third-moment inequality (Theorem 7.6) $\Longrightarrow$ Theorem 1.3 (= Theorem 7.10)
$\Longrightarrow$ Theorem 1.4 (= Theorem 8.3),

Theorem 1.4 + row bound (Theorem 6.8) + Corollary 6.10 $\Longrightarrow$ Theorem 1.1.

The arguments are elementary: besides linear algebra over a polynomial ring and the Cauchy–Schwarz inequality, we use Markov's inequality for the derivative of a polynomial at an endpoint of an interval, the separation of a compact convex set from the non-negative orthant, and integrals of polynomials over intervals.

2. Notation and definitions

2.1. **The cube.** Let $n \ge 0$ be an integer, $[n] = \{1,\dots,n\}$, and $Q_n = \{-1,1\}^n$; the cube Q_0 consists of one point.

For $x \in Q_n$ and $S \subseteq [n]$ the point $x^S \in Q_n$ is defined by

$$(x^S)_i = -x_i \quad (i \in S), \qquad (x^S)_i = x_i \quad (i \notin S).$$

We write $x^j = x^{\{j\}}$ and $\bar{x} = x^{[n]} = -x$, the *antipode* of x. Clearly $(x^S)^T = x^{S\triangle T}$, where $\triangle$ is the symmetric difference.

For $x, z \in Q_n$ put

$$\Delta(x,z) = \{i \in [n] : x_i \ne z_i\}, \quad \mathrm{ag}(x,z) = \{i \in [n] : x_i = z_i\} = [n] \setminus \Delta(x,z),$$

the sets of coordinates in which x and z differ and agree. Thus

$$z = x^{\Delta(x,z)} = \bar{x}^{\,\mathrm{ag}(x,z)}, \qquad \Delta(x, x^S) = S, \qquad \mathrm{ag}(\bar{x}, x^S) = S. \tag{2.1}$$

For a vector $a \in \mathbb{R}^n$ and $S \subseteq [n]$ we write $a(S) = \sum_{i\in S} a_i$, and $\|a\|_1 = \sum_{i=1}^n |a_i|$; for $a \in [0,\infty)^n$ this is $a([n])$.

Probabilities, expectations and variances (Pr, $\mathbb{E}$, Var) are taken over a uniformly random point of the cube, each of the 2^n points having probability 2^{-n}; thus $\mathbb{E}f = 2^{-n}\sum_{x\in Q_n} f(x)$ and $\mathrm{Var}(f) = \mathbb{E}[(f-\mathbb{E}f)^2]$ for $f\colon Q_n \to \mathbb{R}$. For every $S \subseteq [n]$ the map $x \mapsto x^S$ is a bijection of Q_n which preserves the measure.

Conventions. An empty sum is 0 and an empty product is 1; the determinant of the matrix with no rows and no columns is 1. We take reciprocals of positive real numbers only; whenever a reciprocal is written, the positivity of the number is proved first. The cardinality of a finite set X is written $|X|$ or $\#X$.

2.2. **Functions and degree.** For $T \subseteq [n]$ let $\chi_T(x) = \prod_{i\in T} x_i$. Since $x_i^2 = 1$ on Q_n, we have $\chi_T\chi_U = \chi_{T\triangle U}$, and $\mathbb{E}\chi_V = 0$ for $V \neq \emptyset$ (the map $x \mapsto x^i$ with $i \in V$ changes the sign of χ_V), while $\chi_\emptyset = 1$. Hence

$$\mathbb{E}[\chi_T\chi_U] \;=\; \begin{cases} 1, & \text{if } T = U, \\ 0, & \text{otherwise.} \end{cases}$$

The 2^n functions χ_T are therefore linearly independent in the space of all functions $Q_n \to \mathbb{R}$, which has dimension 2^n, and so they form a basis: every $f\colon Q_n \to \mathbb{R}$ has a unique expansion

$$f = \sum_{T\subseteq[n]} \hat{f}(T)\,\chi_T, \qquad \hat{f}(T) = \mathbb{E}[f\chi_T] \in \mathbb{R}. \tag{2.2}$$

Let $d \geq 0$ be an integer. The function f has *degree at most* d if $\hat{f}(T) = 0$ for all T with $|T| > d$.

The functions of degree at most d form a linear space. If f has degree at most d and h has degree at most e, then fh has degree at most $d+e$, because $|T\triangle U| \leq |T| + |U|$. A function of degree at most 0 is constant.

The *influence* of the coordinate j on $f\colon Q_n \to \mathbb{R}$ is

$$\mathrm{Inf}_j(f) = \mathbb{E}_x\Big[\Big(\frac{f(x) - f(x^j)}{2}\Big)^2\Big].$$

The *mean absolute increment* of f in the coordinate j is

$$L_j(f) = \mathbb{E}_x|f(x) - f(x^j)|,$$

and we write $L(f) = \sum_{j=1}^n L_j(f)$ for the sum of these numbers.

We shall use the following identity several times. For every function $h\colon Q_n \to \mathbb{R}$ and every real number c,

$$\mathbb{E}\big[(h-c)^2\big] \;=\; \mathrm{Var}(h) + (\mathbb{E}h - c)^2 \;\geq\; \mathrm{Var}(h), \tag{2.3}$$

because $h - c = (h - \mathbb{E}h) + (\mathbb{E}h - c)$ and $\mathbb{E}[h - \mathbb{E}h] = 0$.

The influences and the mean absolute increments are related as follows. For $h(x) = |f(x) - f(x^j)|$ and $c = 0$ the equality in (2.3) and $\mathrm{Var}(h) \geq 0$ give $(\mathbb{E}h)^2 \leq \mathbb{E}[h^2]$, that is,

$$L_j(f)^2 \;\leq\; 4\,\mathrm{Inf}_j(f) \qquad \text{for every } j \in [n]. \tag{2.4}$$

2.3. **Boolean functions.** A *Boolean function* is a function $g\colon Q_n \to \{0,1\}$. For such a function we put

$$\mathcal{O} = \mathcal{O}(g) = \{x \in Q_n : g(x) = 1\}, \qquad \mathcal{N} = \mathcal{N}(g) = \{x \in Q_n : g(x) = 0\},$$

and, for $j \in [n]$,

$$\mathrm{I}_j(g) = \Pr_x\big[g(x) \neq g(x^j)\big], \qquad b_j(g) = \#\{x \in \mathcal{O} :\ x^j \in \mathcal{N}\}.$$

The *zone* of g is the set $Z(g) = \{u \in \mathcal{O} : \bar{u} \in \mathcal{N}\}$ of points of $\mathcal{O}$ whose antipode lies in $\mathcal{N}$.

Lemma 2.1. *For every Boolean function g on Q_n and every $j \in [n]$ we have $b_j(g) = 2^{n-1}\,\mathrm{I}_j(g)$.*

Proof. Let $E_j = \{x \in Q_n : g(x) \neq g(x^j)\}$, so that $\mathrm{I}_j(g) = 2^{-n}|E_j|$. The involution $x \mapsto x^j$ maps E_j onto itself and exchanges $E_j \cap \mathcal{O}$ and $E_j \cap \mathcal{N}$. Hence $|E_j| = 2|E_j \cap \mathcal{O}| = 2\,b_j(g)$. □

Definition 2.2 (Wings)**.** Let g be a Boolean function on Q_n and $x \in \mathcal{O}$. A set $A \subseteq [n]$ is a 1-*wing of g at x* if $x^S \in \mathcal{O}$ for every $S \subseteq A$.

The empty set is a 1-wing at every $x \in \mathcal{O}$, and a subset of a 1-wing at x is a 1-wing at x.

Definition 2.3 (Rows)**.** Let $f\colon Q_n \to \mathbb{R}$ and $x \in Q_n$. A *row of f at x* is a vector $a \in [0,\infty)^n$ such that

$$|f(x) - f(x^S)| \ \le\ a(S) \qquad \text{for every } S \subseteq [n].$$

The number $\|a\|_1$ is the *cost* of the row.

By (2.1), the condition says that $|f(x) - f(y)| \le a(\Delta(x,y))$ for every $y \in Q_n$.

If g is a Boolean function and $x \in Q_n$, a set $S \subseteq [n]$ with $g(x^S) \neq g(x)$ is called a *sensitive block* of g at x. The difference $|g(x) - g(x^S)|$ is 1 if S is a sensitive block at x and 0 otherwise; so a vector $\lambda \in [0,\infty)^n$ is a row of g at x if and only if

$$\lambda(S) \ge 1 \qquad \text{for every sensitive block } S \text{ of } g \text{ at } x, \tag{2.5}$$

in agreement with the definition given in the introduction.

For Boolean functions, rows are the fractional certificates introduced by Tal [Tal] and by Gilmer, Saks and Srinivasan [GSS]. For a function f with values in $[-1,1]$, a vector a is a row of f at x exactly when $a/2$ is a fractional certificate of f at x in the sense of Agarwal and Ben-David [ABD, Section 3.1]. A row bounds every change $|f(x) - f(x^S)|$ at once, whereas the ε-fractional certificates of Lovett and Zhang [LZ] constrain only the sets S with $|f(x) - f(x^S)| \ge \varepsilon$.

We record three simple facts.

Lemma 2.4. *Let g be a Boolean function on Q_n.*

(i) *The vector $(1,\dots,1) \in \mathbb{R}^n$ is a row of g at every point.*
(ii) *If g is not constant, then $n \ge 1$ and every row λ of g at every point satisfies $\|\lambda\|_1 \ge 1$.*
(iii) *If $f\colon Q_n \to \mathbb{R}$, $t \in \mathbb{R}$, $g = 1_{\{f>t\}}$, $x \in Q_n$ with $f(x) \neq t$, and a is a row of f at x, then $a/|f(x) - t|$ is a row of g at x.*

Proof. (i) A sensitive block is not empty, so it contains some i, and the sum in (2.5) is at least 1.

(ii) The cube Q_0 has one point, so $n \geq 1$. Let $x \in Q_n$. Since g is not constant there is y with $g(y) \neq g(x)$, and $S = \Delta(x, y)$ is a sensitive block at x. Hence $\|\lambda\|_1 \geq \lambda(S) \geq 1$.

(iii) Let S be a sensitive block of g at x. Then exactly one of the numbers $f(x)$, $f(x^S)$ is greater than t, and $f(x) \neq t$; so t lies between them and $|f(x) - f(x^S)| \geq |f(x) - t| > 0$. Therefore $a(S) \geq |f(x) - f(x^S)| \geq |f(x) - t|$, which is (2.5) for the vector $a/|f(x) - t|$; its entries are non-negative. □

Example 2.5. Let $n = 2$ and let g be the function with $\mathcal{O} = \{(1, 1)\}$. At $x = (1, 1)$ every non-empty set is a sensitive block, so the rows at x are the vectors λ with $\lambda_1 \geq 1$ and $\lambda_2 \geq 1$; the least cost is 2, and the only 1-wing at x is the empty set.

At $(-1, 1)$ the only sensitive block is $\{1\}$, and $(1, 0)$ is a row of cost 1; similarly $(0, 1)$ is a row at $(1, -1)$. At $(-1, -1)$ the only sensitive block is $\{1, 2\}$, and $(\frac{1}{2}, \frac{1}{2})$ is a row of cost 1.

Further $Z(g) = \{(1, 1)\}$, $b_1(g) = b_2(g) = 1$ and $\mathrm{I}_1(g) = \mathrm{I}_2(g) = \frac{1}{2}$. With these rows the left-hand side of the inequality of Theorem 1.5 is

$$4 \cdot \Big(\frac{1}{4} \cdot \frac{1}{2}\Big) \cdot \Big(\frac{3}{4} \cdot 1\Big) \;=\; \frac{3}{8},$$

and the right-hand side is $\frac{1}{2}$.

3. A bijection theorem and the determinant inequality

This section contains the combinatorial core of the paper: a theorem about permutations of the set $\mathcal{O} = g^{-1}(1)$ of a Boolean function g (Theorem 3.1), and its consequence, the determinant inequality (Corollary 3.2), which is the first link of the proof chain.

At every point x of $\mathcal{O}$ we fix a 1-wing A_x, a set of coordinates that can be changed at x in any combination without leaving $\mathcal{O}$. The bijection theorem permutes the points of $\mathcal{O}$ so that every point changes all the coordinates of its wing (property (a)), while for every coordinate j at most $b_j(g)$ points keep their jth coordinate (property (b)); here $b_j(g)$ is the number of edges in the direction j along which g changes. Summing over the points gives the determinant inequality: from the antipode of a point u whose antipode lies outside $\mathcal{O}$, the coordinates that u keeps lead back into $\mathcal{O}$ without using the wing at u, and property (b) bounds how often each coordinate is used. Section 4 turns this into the antipodal inequality.

The permutation of the bijection theorem is read off from a term of lowest degree of a determinant. The matrix M has rows and columns indexed by $\mathcal{O}$, and its entry at (x, z) is the product of indeterminates s_i over the coordinates i in which x and z agree; a permutation of $\mathcal{O}$ contributes to $\det M$ a monomial in which the exponent of s_j is the number of points that keep their jth coordinate. Combining the rows of pairs of points that differ in one coordinate shows that $\det M$ has many factors $1 - s_j^2$, so that in its terms of lowest degree the exponent of s_j is at most $b_j(g)$ (Section 3.3). Combining the rows of the points of each wing gives a matrix with zeros at the entries that property (a) forbids, without changing the terms of lowest degree; a permutation that produces such a term in the new determinant has both properties (Section 3.4). Section 3.5 contains examples and the relation to Reimer's Butterfly Lemma.

3.1. **Statements.**

Theorem 3.1 (Bijection theorem)**.** *Let $n \geq 0$, let g be a Boolean function on Q_n, and for every $x \in \mathcal{O} = g^{-1}(1)$ let A_x be a 1-wing of g at x. Then there is a bijection $\tau\colon \mathcal{O} \to \mathcal{O}$ such that*

(a) *$\tau(x)_i \neq x_i$ for every $x \in \mathcal{O}$ and every $i \in A_x$, and*
(b) *$\#\{x \in \mathcal{O} :\ \tau(x)_j = x_j\} \leq b_j(g)$ for every $j \in [n]$.*

Let $\ell \in (0,\infty)^n$; we think of ℓ_j as the length of the coordinate j. For $u \in Z(g)$ and a 1-wing A of g at u put

$$D_g(u, A; \ell) \ :=\ \min\bigl\{\ell(G) :\ G \subseteq [n] \setminus A,\ \bar{u}^{\,G} \in \mathcal{O}\bigr\}.$$

The minimum is taken over a non-empty finite family of sets, because $\bar{u}^{\,[n]\setminus A} = u^A \in \mathcal{O}$.

In words, $D_g(u, A; \ell)$ is the least total length of a set of coordinates outside the wing whose change brings the antipode of u back into $\mathcal{O}$.

Corollary 3.2 (Determinant inequality)**.** *Let g be a Boolean function on Q_n, let $\ell \in (0,\infty)^n$, and for every $u \in Z(g)$ let A_u be a 1-wing of g at u. Then*

$$\sum_{u \in Z(g)} D_g(u, A_u; \ell) \ \leq\ \sum_{j=1}^{n} \ell_j\, b_j(g).$$

Proof of Corollary 3.2 from Theorem 3.1. Put $A_x = \emptyset$ for $x \in \mathcal{O} \setminus Z(g)$; the empty set is a 1-wing. Let τ be the bijection of Theorem 3.1 for the wings A_x ($x \in \mathcal{O}$), and for $y \in \mathcal{O}$ put $G_y = \mathrm{ag}(y, \tau(y))$.

By (a), $G_y \cap A_y = \emptyset$, and by (2.1), $\bar{y}^{\,G_y} = \tau(y) \in \mathcal{O}$. Hence for $y \in Z(g)$ the set G_y belongs to the family over which $D_g(y, A_y; \ell)$ is the minimum, so that $D_g(y, A_y; \ell) \leq \ell(G_y)$. For $y \in \mathcal{O} \setminus Z(g)$ we use only $\ell(G_y) \geq 0$. Therefore

$$\begin{aligned}\sum_{u \in Z(g)} D_g(u, A_u; \ell) \ &\leq\ \sum_{y \in \mathcal{O}} \ell\bigl(\mathrm{ag}(y, \tau(y))\bigr) \\ &=\ \sum_{j=1}^{n} \ell_j \cdot \#\{y \in \mathcal{O} :\ \tau(y)_j = y_j\} \ \leq\ \sum_{j=1}^{n} \ell_j\, b_j(g),\end{aligned}$$

where we exchanged the order of summation and then used (b) and $\ell_j \geq 0$. □

The rest of this section is the proof of Theorem 3.1.

3.2. **Lowest monomials.** Let $s_1, \ldots, s_n$ be indeterminates and let $\mathbb{Z}[s]$ be the ring of polynomials in them with integer coefficients. For a vector m of non-negative integers and for a set $S \subseteq [n]$ we write

$$s^m = \prod_{i=1}^{n} s_i^{m_i}, \qquad |m| = \sum_{i=1}^{n} m_i, \qquad s^S = \prod_{i \in S} s_i.$$

A monomial s^m *occurs* in $q \in \mathbb{Z}[s]$ if its coefficient in q is not zero. A *lowest monomial* of q is a monomial s^m which occurs in q and satisfies $|m| \leq |m'|$ for every monomial $s^{m'}$ occurring

in q. A non-zero polynomial has at least one lowest monomial, and the zero polynomial has none.

We write $q(0)$ for the constant coefficient of q, and $\deg_j q$ for the degree of a non-zero polynomial q in the variable s_j. The ring $\mathbb{Z}[s]$ has no zero divisors, and for non-zero $p, q \in \mathbb{Z}[s]$ we have

$$\deg_j(pq) = \deg_j p + \deg_j q.$$

The matrices of this section have their rows and their columns indexed by the same finite set X, and entries in a commutative ring, which is $\mathbb{Z}[s]$ or $\mathbb{R}$. We use the Leibniz formula

$$\det B \;=\; \sum_{\tau} \operatorname{sgn}(\tau) \prod_{x \in X} B[x, \tau(x)], \tag{3.1}$$

in which the sum is over all permutations τ of X and $\operatorname{sgn}(\tau) \in \{1, -1\}$ is the sign of τ, and the rule $\det(BC) = \det B \cdot \det C$; see [Lang, Chapter XIII, §4].

By (3.1), the determinant of a diagonal matrix is the product of its diagonal entries, and a ring homomorphism, applied to all entries of a matrix, maps the determinant of the matrix to the determinant of the new matrix.

The first lemma says that a factor with constant coefficient 1 does not change the lowest monomials. It will be applied to powers of $1 - s_j^2$ (Lemma 3.6) and to the determinants of the matrices P and U of Lemma 3.8.

Lemma 3.3. *Let $u, q \in \mathbb{Z}[s]$ with $u(0) = 1$. Then uq and q have the same lowest monomials.*

Proof. Write u_a and q_b for the coefficients of s^a in u and of s^b in q. We first show:

> *if $|m| \le |m'|$ for every monomial $s^{m'}$ occurring in q, then s^m has the same coefficient in uq as in q.* (3.2)

Indeed, the coefficient of s^m in uq is $\sum_{a+b=m} u_a\, q_b$. If $a \neq 0$ then $|b| < |m|$, so s^b does not occur in q and $q_b = 0$. Only the term with $a = 0$ and $b = m$ remains, and it equals $u_0\, q_m = q_m$.

Let s^m be a lowest monomial of q. By (3.2) its coefficient in uq is $q_m \neq 0$. Suppose that a monomial $s^{m'}$ with $|m'| < |m|$ occurs in uq. Every monomial occurring in q has total degree at least $|m| > |m'|$, so (3.2) applies to m': the coefficient of $s^{m'}$ in uq is $q_{m'}$, which is 0 because $|m'| < |m|$. This is a contradiction. So s^m is a lowest monomial of uq.

Conversely, let s^m be a lowest monomial of uq. Then $uq \neq 0$, so $q \neq 0$, and q has a lowest monomial s^{m_0}. By the previous paragraph s^{m_0} occurs in uq; hence $|m| \le |m_0|$, and $|m_0| \le |m'|$ for every $s^{m'}$ occurring in q. By (3.2) the coefficient of s^m in q equals its coefficient in uq, which is not zero. So s^m occurs in q and has the least total degree there. □

3.3. **The agreement matrix.** Fix a Boolean function g on Q_n. The *agreement matrix* of g is the matrix M with rows and columns indexed by $\mathcal{O}$ and entries

$$M[x, z] \;=\; s^{\operatorname{ag}(x,z)} \;=\; \prod_{i:\, x_i = z_i} s_i \qquad (x, z \in \mathcal{O}).$$

If $\mathcal{O} = \emptyset$, all matrices in this section are empty and have determinant 1; nothing below requires $\mathcal{O} \neq \emptyset$.

Since $M[x,z] = \prod_i s_i \cdot \prod_{i:\, x_i \neq z_i} s_i^{-1}$, the matrix M is, up to the factor $\prod_i s_i$ and the substitution of $1/s_i$ for the weights, a principal submatrix of the Varchenko matrix of the arrangement of the n coordinate hyperplanes in $\mathbb{R}^n$ [Var]: the regions of this arrangement correspond to the points of Q_n, and the entry of the Varchenko matrix at two regions is the product of the weights of the hyperplanes that separate them.

Varchenko proved that the determinant of his matrix is a product of factors of the form $\left(1 - \prod_{i \in I} w_i^2\right)^b$, where w_i is the weight of the ith hyperplane, I runs over certain sets of hyperplanes, and the exponents b are non-negative integers; see [HKY], where the determinants of such submatrices are studied for the vertex sets of partial cubes.

The divisibility of $\det M$ in Lemma 3.5 below is of this kind. We use only the properties of M that are proved below.

We first check that $\det M$ is not the zero polynomial, so that it has lowest monomials.

Lemma 3.4. *Let g be a Boolean function on Q_n and M its agreement matrix. Then* $\det M \neq 0$.

Proof. Let $c = 2^n$, and substitute c for every s_i. The substitution is a ring homomorphism $\mathbb{Z}[s] \to \mathbb{R}$, so it maps $\det M$ to the determinant of the real matrix B with entries $B[x,z] = c^{|\mathrm{ag}(x,z)|}$.

The diagonal entries of B are c^n, and for $z \neq x$ we have $|\mathrm{ag}(x,z)| \leq n-1$. A row has $|\mathcal{O}| - 1 < 2^n = c$ off-diagonal entries. So for every $x \in \mathcal{O}$

$$\sum_{z \in \mathcal{O},\ z \neq x} B[x,z] \ \leq\ (|\mathcal{O}| - 1)\, c^{\,n-1} \ <\ c^n \ =\ B[x,x].$$

A real square matrix B with non-negative entries in which every diagonal entry exceeds the sum of the other entries of its row has $\det B \neq 0$: if $Bv = 0$ with $v \neq 0$, choose x with $|v_x|$ largest; then

$$B[x,x]|v_x| \ =\ \Big|\sum_{z \neq x} B[x,z] v_z\Big| \ \leq\ \sum_{z \neq x} B[x,z]|v_x| \ <\ B[x,x]|v_x|,$$

which is impossible.

Hence the image of $\det M$ is not zero, and so $\det M \neq 0$. □

The next two lemmas bound the exponent of s_j in a lowest monomial of $\det M$. For $j \in [n]$ let

$$\mathrm{Up}_j = \{x \in \mathcal{O} :\ x_j = 1,\ x^j \in \mathcal{O}\}, \qquad e_j = |\mathrm{Up}_j|.$$

Lemma 3.5. *For every $j \in [n]$ there is $R_j \in \mathbb{Z}[s]$ with* $\det M = (1 - s_j^2)^{e_j} R_j$.

Proof. Step 1: an identity for the entries. For all $x, z \in Q_n$,

$$s^{\mathrm{ag}(x,z)} = s_j\, s^{\mathrm{ag}(x^j,z)} \ \text{ if } x_j = z_j, \qquad s^{\mathrm{ag}(x^j,z)} = s_j\, s^{\mathrm{ag}(x,z)} \ \text{ if } x_j \neq z_j, \tag{3.3}$$

because x and x^j agree with z in the same coordinates $i \neq j$, and exactly one of the two agrees with z in the coordinate j.

Step 2: the matrix E. Let E be the matrix indexed by $\mathcal{O}$ with $E[x,y] = 1$ if $x \in \mathrm{Up}_j$ and $y = x^j$, and $E[x,y] = 0$ otherwise. Then $E^2 = 0$: a non-zero product $E[x,y]E[y,z]$ would

need $y = x^j$ with $x_j = 1$, so $y_j = -1$, and also $y \in \mathrm{Up}_j$, so $y_j = 1$. Hence, with I the identity matrix,

$$(I + s_j E)(I - s_j E) = I. \tag{3.4}$$

Step 3: the row operation. Next, $(EM)[x,z] = M[x^j,z]$ if $x \in \mathrm{Up}_j$, and $(EM)[x,z] = 0$ otherwise. So the row of a point $x \notin \mathrm{Up}_j$ in $(I - s_j E)M$ is its row in M, and for $x \in \mathrm{Up}_j$, by (3.3),

$$\big((I - s_j E)M\big)[x,z] = M[x,z] - s_j M[x^j,z] = \begin{cases} 0, & \text{if } z_j = x_j, \\ (1 - s_j^2)\, M[x,z], & \text{if } z_j \neq x_j. \end{cases}$$

Step 4: determinants. In matrix form, $(I - s_j E)M = \Lambda \widetilde{M}$, where Λ is the diagonal matrix with entry $1 - s_j^2$ at the points of Up_j and 1 at the other points, and where $\widetilde{M}[x,z] = 0$ if $x \in \mathrm{Up}_j$ and $z_j = x_j$, and $\widetilde{M}[x,z] = M[x,z]$ otherwise.

Taking determinants,

$$\det(I - s_j E) \cdot \det M \;=\; (1 - s_j^2)^{e_j} \det \widetilde{M}.$$

Multiplying by $\det(I + s_j E)$ and using (3.4) we get $\det M = (1 - s_j^2)^{e_j} R_j$ with

$$R_j \;=\; \det(I + s_j E) \cdot \det \widetilde{M}. \qquad \square$$

Lemma 3.6. *Every lowest monomial s^m of $\det M$ satisfies $m_j \le b_j(g)$ for every $j \in [n]$.*

Proof. Fix j and let R_j be as in Lemma 3.5.

Step 1: $\deg_j \det M \le |\mathcal{O}|$. Every entry of M has degree at most 1 in s_j, and by the Leibniz formula $\det M$ is a signed sum of products of $|\mathcal{O}|$ entries; so $\deg_j \det M \le |\mathcal{O}|$.

Step 2: $|\mathcal{O}| = 2e_j + b_j(g)$. The set $\mathcal{O}$ is the disjoint union of Up_j, of $\{x \in \mathcal{O} : x_j = -1,\ x^j \in \mathcal{O}\}$, which is the image of Up_j under the injective map $x \mapsto x^j$, and of $\{x \in \mathcal{O} : x^j \in \mathcal{N}\}$, which has $b_j(g)$ elements; so $|\mathcal{O}| = 2e_j + b_j(g)$.

Step 3: $\deg_j R_j \le b_j(g)$. By Lemma 3.4, $\det M \neq 0$ and hence $R_j \neq 0$. Since $\deg_j(1 - s_j^2)^{e_j} = 2e_j$, we get

$$2e_j + \deg_j R_j \;=\; \deg_j \det M \;\le\; |\mathcal{O}| \;=\; 2e_j + b_j(g),$$

that is, $\deg_j R_j \le b_j(g)$.

Step 4: lowest monomials. The polynomial $(1 - s_j^2)^{e_j}$ has constant coefficient 1. By Lemma 3.3, s^m is a lowest monomial of R_j; in particular it occurs in R_j, and therefore $m_j \le \deg_j R_j \le b_j(g)$. $\square$

3.4. **Row operations along the wings.** The following identity describes the effect, on one entry of M, of combining the rows of the points x^S, $S \subseteq A$.

Lemma 3.7 (Wing identity)**.** *For all $x, z \in Q_n$ and every set $A \subseteq [n]$,*

$$\sum_{S \subseteq A} \prod_{i \in S} (-s_i) \cdot s^{\mathrm{ag}(x^S, z)} \;=\; \begin{cases} \prod_{i \in A} (1 - s_i^2) \cdot s^{\mathrm{ag}(x,z)}, & \text{if } \mathrm{ag}(x,z) \cap A = \emptyset, \\ 0, & \text{if } \mathrm{ag}(x,z) \cap A \neq \emptyset. \end{cases}$$

Proof. Put $\varphi_i = s_i$ and $\gamma_i = 1$ if $x_i = z_i$, and put $\varphi_i = 1$ and $\gamma_i = s_i$ if $x_i \neq z_i$. Then $s^{\mathrm{ag}(x,z)} = \prod_{i\in[n]} \varphi_i$. Changing the sign of the ith coordinate of x exchanges agreement and disagreement with z in that coordinate, so

$$s^{\mathrm{ag}(x^S,z)} \;=\; \prod_{i\in S} \gamma_i \cdot \prod_{i\notin S} \varphi_i.$$

For $S \subseteq A$ the term of S on the left-hand side is therefore

$$\prod_{i\in S}(-s_i\gamma_i) \cdot \prod_{i\in A\setminus S} \varphi_i \cdot \prod_{i\notin A} \varphi_i,$$

and the sum over all $S \subseteq A$ is

$$\prod_{i\in A}\big(\varphi_i - s_i\gamma_i\big) \cdot \prod_{i\notin A} \varphi_i.$$

For $i \in A$ the factor $\varphi_i - s_i\gamma_i$ equals $s_i - s_i = 0$ if $x_i = z_i$, and $1 - s_i^2$ if $x_i \neq z_i$. If some $i \in A$ has $x_i = z_i$, the product is 0. Otherwise every factor with $i \in A$ is $1 - s_i^2$, and $\varphi_i = 1$ for all $i \in A$, so that $\prod_{i\notin A} \varphi_i = s^{\mathrm{ag}(x,z)}$. □

From now on a 1-wing A_x of g at x is fixed for every $x \in \mathcal{O}$, as in Theorem 3.1.

Define matrices P, U, Q indexed by $\mathcal{O}$, with entries in $\mathbb{Z}[s]$, by

$$P[x,w] = \begin{cases} \prod_{i\in\Delta(x,w)}(-s_i), & \text{if } \Delta(x,w) \subseteq A_x, \\ 0, & \text{otherwise,} \end{cases}$$

$$U[x,x] = \prod_{i\in A_x}(1-s_i^2), \qquad U[x,w] = 0 \;\text{ for } w \neq x,$$

$$Q[x,z] = \begin{cases} s^{\mathrm{ag}(x,z)}, & \text{if } \mathrm{ag}(x,z) \cap A_x = \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

In matrix form the wing identity reads $PM = UQ$.

Lemma 3.8. *Let g, the wings A_x and the matrices P, U, Q be as above.*

(i) $PM = UQ$.
(ii) $\det P$ *and* $\det U$ *have constant coefficient* 1.
(iii) $\det Q \neq 0$, *and every lowest monomial* s^m *of* $\det Q$ *satisfies* $m_j \leq b_j(g)$ *for every* $j \in [n]$.

Proof. (i) Fix $x, z \in \mathcal{O}$. By definition,

$$(PM)[x,z] \;=\; \sum_{w} \prod_{i\in\Delta(x,w)} (-s_i) \cdot M[w,z],$$

where the sum is over the points $w \in \mathcal{O}$ with $\Delta(x,w) \subseteq A_x$.

The map $w \mapsto \Delta(x,w)$ is a bijection from this set of points onto the set of all subsets of A_x: it is injective because $w = x^{\Delta(x,w)}$, and it is onto because for $S \subseteq A_x$ the point x^S lies in $\mathcal{O}$ and $\Delta(x,x^S) = S$. That $x^S \in \mathcal{O}$ is the wing property of A_x; this is the only place in the proof of Theorem 3.1 where it is used. Hence

$$(PM)[x,z] \;=\; \sum_{S\subseteq A_x} \prod_{i\in S}(-s_i) \cdot s^{\mathrm{ag}(x^S,z)},$$

and by Lemma 3.7 with $A = A_x$ this equals $U[x,x] \cdot Q[x,z] = (UQ)[x,z]$.

(ii) The map $q \mapsto q(0)$ is a ring homomorphism $\mathbb{Z}[s] \to \mathbb{Z}$, so $(\det P)(0)$ is the determinant of the integer matrix with entries $P[x,w](0)$. Now $P[x,x] = 1$, the empty product (note that $\Delta(x,x) = \emptyset \subseteq A_x$), and for $w \neq x$ the entry $P[x,w]$ is 0 or the product of the $-s_i$ over the non-empty set $\Delta(x,w)$, so that $P[x,w](0) = 0$. Thus $(\det P)(0) = \det I = 1$.

Also

$$\det U \;=\; \prod_{x\in\mathcal{O}} \prod_{i\in A_x} (1 - s_i^2)$$

has constant coefficient 1.

(iii) By (i),

$$\det P \cdot \det M \;=\; \det U \cdot \det Q.$$

Here $\det P \neq 0$ by (ii) and $\det M \neq 0$ by Lemma 3.4, so the left-hand side is not zero, and therefore $\det Q \neq 0$.

Let s^m be a lowest monomial of $\det Q$. By Lemma 3.3 with $u = \det U$ it is a lowest monomial of $\det U \cdot \det Q = \det P \cdot \det M$, and by Lemma 3.3 with $u = \det P$ it is a lowest monomial of $\det M$. Now Lemma 3.6 gives $m_j \le b_j(g)$. □

Proof of Theorem 3.1. By the Leibniz formula (3.1),

$$\det Q \;=\; \sum_{\tau} \operatorname{sgn}(\tau) \prod_{x\in\mathcal{O}} Q[x,\tau(x)],$$

where the sum is over all permutations τ of $\mathcal{O}$.

Every entry of Q is 0 or a monomial with coefficient 1. So the product belonging to τ is 0 unless

$$\operatorname{ag}(x,\tau(x)) \cap A_x = \emptyset \qquad \text{for every } x \in \mathcal{O}, \tag{3.5}$$

and if (3.5) holds, it is the single monomial $\prod_{x\in\mathcal{O}} s^{\operatorname{ag}(x,\tau(x))}$, in which s_j has the exponent $\#\{x \in \mathcal{O} : \tau(x)_j = x_j\}$.

By Lemma 3.8(iii), $\det Q$ has a lowest monomial s^m, and $m_j \le b_j(g)$ for all j. The coefficient of s^m in $\det Q$ is not zero, so for at least one permutation τ the product belonging to τ is not zero and is equal to s^m.

This τ satisfies (3.5), which is (a), and

$$\#\{x \in \mathcal{O} :\; \tau(x)_j = x_j\} \;=\; m_j \;\le\; b_j(g) \qquad \text{for every } j \in [n],$$

which is (b). □

3.5. **Examples and the relation to Reimer's Butterfly Lemma.**

Example 3.9. (1) Let g be the constant function 1 on Q_n. Then $b_j(g) = 0$ for all j, and (b) says that $\tau(x)_j \neq x_j$ for all x and j; so τ is the antipodal map $x \mapsto \bar{x}$, whatever the wings are.

For $n = 1$ the agreement matrix has diagonal entries s_1 and off-diagonal entries 1; its determinant is $s_1^2 - 1$, and the only lowest monomial is 1, the monomial of the antipodal map.

(2) Let $n = 2$ and $g(x) = 1$ if and only if $x_1 = 1$. Then $\mathcal{O} = \{p, q\}$ with $p = (1,1)$ and $q = (1,-1)$, the set $\{2\}$ is a 1-wing at p and at q, and $b_1(g) = 2$, $b_2(g) = 0$.

The agreement matrix has diagonal entries $s_1 s_2$ and off-diagonal entries s_1, so that $\det M = s_1^2(s_2^2 - 1)$. The only lowest monomial is s_1^2; it is the monomial of the transposition τ of p and q. This τ satisfies (a) for the wings $A_p = A_q = \{2\}$, and in (b) there is equality for both coordinates.

Moreover $Z(g) = \mathcal{O}$, and $D_g(p, \{2\}; \ell) = D_g(q, \{2\}; \ell) = \ell_1$, since the only admissible set is $G = \{1\}$. So the determinant inequality reads $2\ell_1 \le 2\ell_1$.

(3) For the function g of Example 2.5 the matrix M has the single entry $s_1 s_2$, the bijection is the identity, and in (b) there is equality: $1 \le b_j(g) = 1$ for $j = 1, 2$.

Remark 3.10 (Relation to Reimer's Butterfly Lemma). Part (a) of Theorem 3.1, without part (b), is equivalent to the Butterfly Lemma, which Reimer used to prove the van den Berg–Kesten conjecture [Rei].

We recall the lemma in the formulation of Smyth [Smy, Lemma 2.4], transcribed to Q_n and specialised to the case in which the ambient subcube is the whole cube.

A *butterfly* is a pair $B = (a, b)$ of points of Q_n with *body* a and *tip* b; its *red* and *yellow* sets are

$$\mathrm{Red}(B) = \{b^S : \ S \subseteq \Delta(a, b)\},$$
$$\mathrm{Yellow}(B) = \{y \in Q_n : \ y_i \ne b_i \text{ for all } i \in \Delta(a, b)\},$$

the subcubes spanned by a and b, and by a and $\bar{b}$.

The lemma states that if $\mathcal{B}$ is a family of butterflies with distinct tips, and $\mathrm{Red}(\mathcal{B})$ and $\mathrm{Yellow}(\mathcal{B})$ are the unions of the red and of the yellow sets of its members, then $|\mathcal{B}| \le |\mathrm{Red}(\mathcal{B}) \cap \mathrm{Yellow}(\mathcal{B})|$.

The lemma follows from (a). Let $\mathcal{O} = \mathrm{Red}(\mathcal{B})$ and $g = 1_{\mathcal{O}}$. If b is the tip of $B = (a, b) \in \mathcal{B}$, then $A_b = \Delta(a, b)$ is a 1-wing of g at b, because $b^S \in \mathrm{Red}(B) \subseteq \mathcal{O}$ for $S \subseteq A_b$; this is well defined since the tips are distinct. Put $A_x = \emptyset$ for the other points x of $\mathcal{O}$.

A bijection τ with (a) maps the tip b of B to a point of $\mathcal{O}$ which differs from b in every coordinate of $\Delta(a, b)$, that is, to a point of $\mathrm{Red}(\mathcal{B}) \cap \mathrm{Yellow}(B)$. Since τ is injective, $|\mathcal{B}| \le |\mathrm{Red}(\mathcal{B}) \cap \mathrm{Yellow}(\mathcal{B})|$.

(a) *follows from the lemma.* Let g and the wings A_x be as in Theorem 3.1, and for $x \in \mathcal{O}$ let B_x be the butterfly with tip x and body x^{A_x}. Then $\mathrm{Red}(B_x) = \{x^S : S \subseteq A_x\} \subseteq \mathcal{O}$, and $Y_x := \mathcal{O} \cap \mathrm{Yellow}(B_x)$ is the set of points of $\mathcal{O}$ which differ from x in every coordinate of A_x.

For every $T \subseteq \mathcal{O}$ the lemma, applied to the family $\{B_x : x \in T\}$, gives $|T| \le \left|\bigcup_{x \in T} Y_x\right|$. By Hall's theorem on distinct representatives [Hall] there is an injective map $\tau \colon \mathcal{O} \to \mathcal{O}$ with $\tau(x) \in Y_x$ for all x. It is a bijection because $\mathcal{O}$ is finite, and it satisfies (a).

Reimer's proof of the lemma [Rei, Section 2] is also by linear algebra. In the presentation of Borgs, Chayes and Randall [BCR, Section 5], where the lemma is stated in terms of cylinders [BCR, Lemma 4.1], vectors which are tensor products over the coordinates are attached to the points, and the lemma is deduced from the linear independence of these vectors; the main step is carried out over the field with two elements.

The proof of Theorem 3.1 given above works instead with a matrix over a polynomial ring, and it obtains the bijection from a term of lowest degree of the determinant.

Part (b), the control of the number of points which keep their jth coordinate, is what Theorem 3.1 adds to the Butterfly Lemma. The determinant inequality (Corollary 3.2) uses both parts.

4. The antipodal inequality

The determinant inequality (Corollary 3.2) bounds a sum, over the points u of the zone, of the distances $D_g(u, A_u; \ell)$, for an arbitrary choice of wings. In this section the wings are chosen with the help of rows of g in the sense of Definition 2.3.

A row at u shows that a set of coordinates of small total weight is a 1-wing at u, and a row at the antipode $\bar{u}$ shows that it is expensive to return from $\bar{u}$ to $\mathcal{O}$ without using the coordinates of this wing. The result is a lower bound for $D_g(u, A_u; \ell)$ by the reciprocal of a bilinear expression in the two rows (Lemma 4.2). Summing it over u and applying the determinant inequality gives the antipodal inequality (Theorem 4.3).

This use of the rows at a point and at its antipode has a precedent in the proof of the dual inequality by Kahn, Saks and Smyth [KSS, Section 2], where a certificate at a point and a certificate at its antipode in a subcube give the two wings of a butterfly.

The antipodal inequality (Theorem 4.3) is used in Section 5: applied to the restrictions of g to all subcubes, it gives the all-pairs inequality (Proposition 5.2).

We first check that the expressions whose reciprocals are taken are positive. The lemma which shows this is due to Kulkarni and Tal [KT, Lemma 6.2]; we include its short proof.

Lemma 4.1 (Overlap)**.** *Let g be a Boolean function on Q_n, let $p, q \in Q_n$ with $g(p) \neq g(q)$, let λ be a row of g at p and μ a row of g at q. Then*

$$\sum_{j \in \Delta(p,q)} \min(\lambda_j, \mu_j) \;\geq\; 1.$$

In particular, $\lambda_j \mu_j > 0$ for some $j \in \Delta(p,q)$, and $\sum_{j \in \Delta(p,q)} w_j \lambda_j \mu_j > 0$ for every $w \in (0,\infty)^n$.

Proof. Let $\Delta = \Delta(p,q)$ and $S = \{j \in \Delta : \lambda_j \leq \mu_j\}$. Then $\min(\lambda_j, \mu_j)$ is λ_j for $j \in S$ and μ_j for $j \in \Delta \setminus S$, so

$$\sum_{j \in \Delta} \min(\lambda_j, \mu_j) \;=\; \lambda(S) + \mu(\Delta \setminus S).$$

Since $q = p^{\Delta}$, we have $q^{\Delta \setminus S} = p^S$.

If $g(p^S) \neq g(p)$, then S is a sensitive block at p and $\lambda(S) \geq 1$. Otherwise $g(q^{\Delta\setminus S}) = g(p^S) = g(p) \neq g(q)$, so $\Delta \setminus S$ is a sensitive block at q and $\mu(\Delta \setminus S) \geq 1$. In both cases the sum is at least 1.

In particular $\min(\lambda_j, \mu_j) > 0$ for some $j \in \Delta$, and the remaining statements follow because all $\lambda_i \mu_i$ are non-negative. □

Lemma 4.2 (Pointwise step)**.** *Let g be a Boolean function on Q_n, $u \in Z(g)$, λ a row of g at u, μ a row of g at $\bar{u}$, and $\ell \in (0,\infty)^n$. Then $\sum_j \lambda_j \mu_j / \ell_j > 0$, and with*

$$\omega := \Big(\sum_{j=1}^{n} \frac{\lambda_j \mu_j}{\ell_j} \Big)^{-1}, \qquad A := \{ j \in [n] :\; \ell_j < \omega\, \mu_j \},$$

the set A is a 1-wing of g at u and $D_g(u, A; \ell) \geq \omega$.

Proof. We have $g(u) = 1 \neq 0 = g(\bar{u})$ and $\Delta(u, \bar{u}) = [n]$, so the sum is positive by Lemma 4.1 with $w_j = 1/\ell_j$. By the definition of ω,

$$\sum_{j=1}^{n} \lambda_j \cdot \frac{\omega\mu_j}{\ell_j} = 1, \qquad \text{and} \qquad \frac{\omega\mu_j}{\ell_j} > 1 \ \text{ for } j \in A. \tag{4.1}$$

Step 1: $\lambda(A) < 1$. If $\lambda_j = 0$ for all $j \in A$ this is clear. Otherwise, by (4.1),

$$\lambda(A) \overset{\text{Sec. 2.1}}{=} \sum_{j\in A} \lambda_j \overset{(4.1)}{<} \sum_{j\in A} \lambda_j \frac{\omega\mu_j}{\ell_j} \overset{\text{Def. 2.3}}{\leq} \sum_{j\in [n]} \lambda_j \frac{\omega\mu_j}{\ell_j} \overset{(4.1)}{=} 1.$$

Step 2: A is a 1-wing at u. For $S \subseteq A$ we have $\lambda(S) \leq \lambda(A) < 1$. Since λ is a row at u, the set S is not a sensitive block at u; that is, $g(u^S) = g(u) = 1$.

Step 3: $D_g(u, A; \ell) \geq \omega$. Let $G \subseteq [n] \setminus A$ with $\bar{u}^G \in \mathcal{O}$. Since $g(\bar{u}) = 0$, the set G is a sensitive block at $\bar{u}$, so $\mu(G) \geq 1$. Since G is disjoint from A, we have $\ell_j \geq \omega\mu_j$ for all $j \in G$. Hence $\ell(G) \geq \omega\,\mu(G) \geq \omega$. □

Theorem 4.3 (Antipodal inequality)**.** *Let g be a Boolean function on Q_n, let $w \in (0,\infty)^n$, and for every $x \in Q_n$ let λ^x be a row of g at x. Then $\sum_j w_j \lambda^u_j \lambda^{\bar{u}}_j > 0$ for every $u \in Z(g)$, and*

$$\sum_{u\in Z(g)} \Big(\sum_{j=1}^{n} w_j\, \lambda^u_j\, \lambda^{\bar{u}}_j\Big)^{-1} \ \leq \ \sum_{j=1}^{n} \frac{b_j(g)}{w_j}.$$

Proof. Put $\ell_j = 1/w_j$. For each $u \in Z(g)$, Lemma 4.2 with $\lambda = \lambda^u$ and $\mu = \lambda^{\bar{u}}$ gives the positivity and a 1-wing A_u of g at u with

$$D_g(u, A_u; \ell) \ \geq \ \Big(\sum_{j=1}^{n} \frac{\lambda^u_j\, \lambda^{\bar{u}}_j}{\ell_j}\Big)^{-1} \ = \ \Big(\sum_{j=1}^{n} w_j\, \lambda^u_j\, \lambda^{\bar{u}}_j\Big)^{-1}.$$

We sum over $u \in Z(g)$ and apply Corollary 3.2 to the wings A_u:

$$\sum_{u\in Z(g)} \Big(\sum_{j=1}^{n} w_j\, \lambda^u_j\, \lambda^{\bar{u}}_j\Big)^{-1} \overset{\text{Lem. 4.2}}{\leq} \sum_{u\in Z(g)} D_g(u, A_u; \ell) \overset{\text{Cor. 3.2}}{\leq} \sum_{j=1}^{n} \ell_j\, b_j(g) \ = \ \sum_{j=1}^{n} \frac{b_j(g)}{w_j}. \qquad \square$$

Example 4.4. Let $n \geq 1$ and $g(x) = 1$ if and only if $x_1 = 1$. A set S is a sensitive block, at any point, if and only if $1 \in S$; so the vector $(1, 0, \ldots, 0)$ is a row of g at every point. Take $\lambda^x = (1, 0, \ldots, 0)$ for all x. Then $Z(g) = \mathcal{O}$ has 2^{n-1} points, $\sum_j w_j \lambda^u_j \lambda^{\bar{u}}_j = w_1$, $b_1(g) = 2^{n-1}$ and $b_j(g) = 0$ for $j \geq 2$. Both sides of the antipodal inequality are equal to $2^{n-1}/w_1$.

5. The all-pairs inequality and the two-sided inequality

In the first part of this section, we apply the antipodal inequality (Theorem 4.3) to the restrictions of g to all subcubes; the sum of the resulting inequalities is an inequality for all pairs $(x, y) \in \mathcal{O} \times \mathcal{N}$. This reduction of a statement about all pairs of points to statements about the antipodal pairs of the subcubes is the one by which the van den Berg–Kesten–Reimer inequality is derived from the Butterfly Lemma [Rei, Section 3] (see also [BCR, Section 4]), and by which Kahn, Saks and Smyth derive the dual inequality [KSS, Section 2].

In the second part the weights w are removed: the Cauchy–Schwarz inequality, applied to a sum over the pairs, leads to an expression in w which can be minimised explicitly.

The all-pairs inequality (Proposition 5.2) and the removal of the weights (Lemma 5.3) are used in Section 7.

5.1. **From antipodal pairs to all pairs.** For $F \subseteq [n]$ and $z, u \in Q_n$ let $\gamma_F(z,u) \in Q_n$ be the point with

$$\gamma_F(z,u)_i = u_i \quad (i \in F), \qquad \gamma_F(z,u)_i = z_i \quad (i \notin F).$$

For fixed z the points $\gamma_F(z,u)$, $u \in Q_n$, form the subcube through z in which the coordinates in F are free.

Lemma 5.1 (Gluing)**.** *Let $F \subseteq [n]$.*

(i) *$\gamma_F(z, u^S) = \gamma_F(z,u)^{S \cap F}$ for all $z, u \in Q_n$ and $S \subseteq [n]$. In particular $\gamma_F(z, \bar{u}) = \gamma_F(z,u)^F$, and $\gamma_F(z,u^j)$ is equal to $\gamma_F(z,u)^j$ if $j \in F$ and to $\gamma_F(z,u)$ if $j \notin F$.*

(ii) *For every function $\varphi\colon Q_n \to \mathbb{R}$,*

$$\sum_{z \in Q_n} \sum_{u \in Q_n} \varphi\big(\gamma_F(z,u)\big) \;=\; 2^n \sum_{x \in Q_n} \varphi(x).$$

Proof. (i) For $i \notin F$ both points have the ith coordinate z_i. For $i \in F$ both have the ith coordinate $-u_i$ if $i \in S$, and u_i if $i \notin S$. The special cases are $S = [n]$ and $S = \{j\}$.

(ii) Consider the map $\Phi(z,u) = \big(\gamma_F(z,u), \gamma_F(u,z)\big)$ from $Q_n \times Q_n$ to itself. Let $x = \gamma_F(z,u)$ and $x' = \gamma_F(u,z)$. The point $\gamma_F(x,x')$ agrees with x', hence with z, in the coordinates of F, and with x, hence with z, in the other coordinates; so $\gamma_F(x,x') = z$. In the same way $\gamma_F(x',x) = u$.

Thus $\Phi(\Phi(z,u)) = (z,u)$, and Φ is a bijection. Therefore

$$\sum_{(z,u) \in Q_n \times Q_n} \varphi\big(\gamma_F(z,u)\big) \;=\; \sum_{(x,x') \in Q_n \times Q_n} \varphi(x) \;=\; 2^n \sum_{x \in Q_n} \varphi(x). \qquad \square$$

Proposition 5.2 (All-pairs inequality)**.** *Let g be a Boolean function on Q_n, let $w \in (0,\infty)^n$, and for every $x \in Q_n$ let λ^x be a row of g at x. Then $\sum_{j \in \Delta(x,y)} w_j \lambda^x_j \lambda^y_j > 0$ for all $x \in \mathcal{O}$ and $y \in \mathcal{N}$, and*

$$\sum_{x \in \mathcal{O}} \sum_{y \in \mathcal{N}} \Big(\sum_{j \in \Delta(x,y)} w_j \, \lambda^x_j \, \lambda^y_j \Big)^{-1} \;\le\; 2^{n-1} \sum_{j=1}^{n} \frac{b_j(g)}{w_j}. \tag{5.1}$$

Proof. The positivity is contained in Lemma 4.1.

Step 1: one set of free coordinates. Fix $F \subseteq [n]$. For $x \in Q_n$ put

$$T_F(x) = \Big(\sum_{j \in F} w_j \, \lambda^x_j \, \lambda^{x^F}_j \Big)^{-1} \;\text{ if } g(x) = 1 \text{ and } g(x^F) = 0, \qquad T_F(x) = 0 \;\text{ otherwise.}$$

In the first case the sum is positive by Lemma 4.1, because $\Delta(x, x^F) = F$.

We show that

$$\sum_{x \in Q_n} T_F(x) \;\le\; \sum_{j \in F} \frac{b_j(g)}{w_j}. \tag{5.2}$$

Fix $z \in Q_n$. Let g_z be the Boolean function $g_z(u) = g(\gamma_F(z,u))$ on Q_n (this is the restriction of g to the subcube through z with free coordinates F, written as a function on the whole cube which does not depend on the coordinates outside F; so the antipodal inequality applies without renumbering the coordinates), and for $u \in Q_n$ let $\mu^u \in [0,\infty)^n$ be the vector with

$$\mu^u_j = \lambda_j^{\gamma_F(z,u)} \quad (j \in F), \qquad \mu^u_j = 0 \quad (j \notin F).$$

Then μ^u is a row of g_z at u. Indeed, let S be a sensitive block of g_z at u and $x = \gamma_F(z,u)$. By Lemma 5.1(i), $g(x^{S\cap F}) = g_z(u^S) \neq g_z(u) = g(x)$; so $S \cap F$ is a sensitive block of g at x, and $\mu^u(S) = \lambda^x(S \cap F) \geq 1$.

We apply Theorem 4.3 to g_z with the weights w and the rows μ^u:

$$\sum_{u \in Z(g_z)} \Big(\sum_{j=1}^n w_j \, \mu^u_j \, \mu^{\bar{u}}_j \Big)^{-1} \;\leq\; \sum_{j=1}^n \frac{b_j(g_z)}{w_j}. \tag{5.3}$$

The left-hand side of (5.3). Let $u \in Q_n$ and $x = \gamma_F(z,u)$. By Lemma 5.1(i), $\gamma_F(z,\bar{u}) = x^F$. Hence $u \in Z(g_z)$ if and only if $g(x) = 1$ and $g(x^F) = 0$, and

$$\sum_j w_j \, \mu^u_j \, \mu^{\bar{u}}_j \;=\; \sum_{j \in F} w_j \, \lambda^x_j \, \lambda^{x^F}_j.$$

So the left-hand side is $\sum_{u \in Q_n} T_F(\gamma_F(z,u))$.

The right-hand side of (5.3). For $x \in Q_n$ let $\beta_j(x) = 1$ if $g(x) = 1$ and $g(x^j) = 0$, and $\beta_j(x) = 0$ otherwise; thus $b_j(g) = \sum_{x \in Q_n} \beta_j(x)$.

If $j \notin F$, then $g_z(u^j) = g_z(u)$ for all u by Lemma 5.1(i), and so $b_j(g_z) = 0$. If $j \in F$, then $g_z(u^j) = g(\gamma_F(z,u)^j)$, and so

$$b_j(g_z) \;=\; \sum_{u \in Q_n} \beta_j\big(\gamma_F(z,u)\big).$$

Therefore (5.3) reads

$$\sum_{u \in Q_n} T_F\big(\gamma_F(z,u)\big) \;\leq\; \sum_{j \in F} \frac{1}{w_j} \sum_{u \in Q_n} \beta_j\big(\gamma_F(z,u)\big).$$

This holds for every $z \in Q_n$. Adding over all z, we get

$$\begin{aligned} 2^n \sum_{x \in Q_n} T_F(x) &\overset{\text{Lem. 5.1(ii)}}{=} \sum_{z \in Q_n} \sum_{u \in Q_n} T_F\big(\gamma_F(z,u)\big) \;\leq\; \sum_{j \in F} \frac{1}{w_j} \sum_{z \in Q_n} \sum_{u \in Q_n} \beta_j\big(\gamma_F(z,u)\big) \\ &\overset{\text{Lem. 5.1(ii)}}{=} 2^n \sum_{j \in F} \frac{b_j(g)}{w_j}, \end{aligned}$$

which is (5.2) multiplied by 2^n.

Step 2: all sets of free coordinates. We add (5.2) over all $F \subseteq [n]$. On the left, fix $x \in Q_n$. The map $F \mapsto x^F$ is a bijection from the set of all subsets of $[n]$ onto Q_n, with inverse $y \mapsto \Delta(x,y)$. So $\sum_F T_F(x) = 0$ if $g(x) = 0$, and if $g(x) = 1$, then

$$\sum_{F \subseteq [n]} T_F(x) \;=\; \sum_{y \in \mathcal{N}} \Big(\sum_{j \in \Delta(x,y)} w_j \, \lambda^x_j \, \lambda^y_j \Big)^{-1}.$$

Summing over x gives the left-hand side of (5.1).

On the right, every $j \in [n]$ belongs to exactly 2^{n-1} of the subsets of $[n]$, because $F \mapsto F \triangle \{j\}$ exchanges the subsets which contain j and those which do not. Hence

$$\sum_{F \subseteq [n]} \sum_{j \in F} \frac{b_j(g)}{w_j} \;=\; 2^{n-1} \sum_j \frac{b_j(g)}{w_j}. \qquad \square$$

5.2. **Removing the weights.** We use the Cauchy–Schwarz inequality for finite families of real numbers,

$$\Big(\sum_i a_i b_i\Big)^2 \;\le\; \Big(\sum_i a_i^2\Big)\Big(\sum_i b_i^2\Big).$$

The following lemma contains the step in which the weights are removed. It is formulated for two finite families of non-negative vectors, because it is applied twice: to the rows of a Boolean function in the proof of Theorem 5.4, and to multiples of the rows of a real function in the proof of Theorem 7.6.

Lemma 5.3 (Removing the weights)**.** *Let A and B be non-empty finite sets. For every $x \in A$ let $\lambda^x \in [0,\infty)^n$, for every $y \in B$ let $\mu^y \in [0,\infty)^n$, and for every $(x,y) \in A \times B$ let $J(x,y)$ be a subset of $[n]$. Let $C \ge 0$ be a real number. Suppose that for every $w \in (0,\infty)^n$ the following two conditions are satisfied:*

(i) $\sum_{j \in J(x,y)} w_j \lambda_j^x \mu_j^y > 0$ *for all* $(x,y) \in A \times B$*;*

(ii) $\displaystyle\sum_{x \in A} \sum_{y \in B} \Big(\sum_{j \in J(x,y)} w_j \, \lambda_j^x \, \mu_j^y \Big)^{-1} \;\le\; C \sum_{j=1}^n \frac{1}{w_j}.$

Then $\|\lambda^x\|_1 > 0$ for every $x \in A$, $\|\mu^y\|_1 > 0$ for every $y \in B$, and

$$\Big(\sum_{x \in A} \frac{1}{\|\lambda^x\|_1}\Big)\Big(\sum_{y \in B} \frac{1}{\|\mu^y\|_1}\Big) \;\le\; C.$$

Proof. Let $x \in A$. Since B is not empty, there is a $y \in B$, and by (i) with $w = (1,\dots,1)$ there is a j with $\lambda_j^x \mu_j^y > 0$; so $\|\lambda^x\|_1 \ge \lambda_j^x > 0$. In the same way $\|\mu^y\|_1 > 0$ for every $y \in B$. Put

$$v(x) = \frac{1}{\|\lambda^x\|_1} \quad (x \in A), \qquad u(y) = \frac{1}{\|\mu^y\|_1} \quad (y \in B).$$

Further let

$$V_A = \sum_{x \in A} v(x), \quad V_B = \sum_{y \in B} u(y), \quad X_j = \sum_{x \in A} v(x)^2 \lambda_j^x, \quad Y_j = \sum_{y \in B} u(y)^2 \mu_j^y \quad (j \in [n]).$$

The numbers X_j and Y_j are non-negative, and $V_A > 0$ and $V_B > 0$, because A and B are not empty.

Since $v(x)^2 \|\lambda^x\|_1 = v(x)$ for every $x \in A$, and $u(y)^2 \|\mu^y\|_1 = u(y)$ for every $y \in B$,

$$\sum_{j=1}^n X_j = \sum_{x \in A} v(x)^2 \|\lambda^x\|_1 = \sum_{x \in A} v(x) = V_A, \qquad \sum_{j=1}^n Y_j = V_B. \tag{5.4}$$

Step 1: the Cauchy–Schwarz inequality over the pairs. Fix $w \in (0,\infty)^n$, and for $(x,y) \in A \times B$ let $s_{xy} = \sum_{j \in J(x,y)} w_j \lambda_j^x \mu_j^y$, which is positive by (i). Writing $v(x)u(y) = v(x)u(y)\sqrt{s_{xy}}\cdot$

$(1/\sqrt{s_{xy}})$, we get from the Cauchy–Schwarz inequality

$$(V_A V_B)^2 = \Big(\sum_{x\in A}\sum_{y\in B} v(x)\,u(y)\Big)^2 \;\le\; \Big(\sum_{x\in A}\sum_{y\in B} v(x)^2 u(y)^2\, s_{xy}\Big)\Big(\sum_{x\in A}\sum_{y\in B}\frac{1}{s_{xy}}\Big).$$

By (ii), the second factor is at most $C\sum_j 1/w_j$.

In the first factor we have $s_{xy} \le \sum_{j=1}^n w_j \lambda_j^x \mu_j^y$, because the terms with $j \notin J(x,y)$ are non-negative; hence

$$\begin{aligned}\sum_{x\in A}\sum_{y\in B} v(x)^2 u(y)^2\, s_{xy} \;&\le\; \sum_{x\in A}\sum_{y\in B} v(x)^2 u(y)^2 \sum_{j=1}^n w_j \lambda_j^x \mu_j^y \\ &= \sum_{j=1}^n w_j \Big(\sum_{x\in A} v(x)^2 \lambda_j^x\Big)\Big(\sum_{y\in B} u(y)^2 \mu_j^y\Big) \;=\; \sum_{j=1}^n w_j X_j Y_j.\end{aligned}$$

All the factors are non-negative, so that

$$(V_A V_B)^2 \;\le\; C\Big(\sum_{j=1}^n w_j X_j Y_j\Big)\Big(\sum_{j=1}^n \frac{1}{w_j}\Big) \qquad \text{for every } w \in (0,\infty)^n. \tag{5.5}$$

Step 2: the choice of the weights. Let $r_j = \sqrt{X_j Y_j}$ and $R = \sum_j r_j$. For $\varepsilon > 0$ the weights $w_j = 1/(r_j + \varepsilon)$ are positive, and $w_j X_j Y_j = r_j^2/(r_j+\varepsilon) \le r_j$ and $1/w_j = r_j + \varepsilon$. So (5.5) gives

$$(V_A V_B)^2 \;\le\; C\,R\,(R + n\varepsilon) \qquad \text{for every } \varepsilon > 0,$$

and letting $\varepsilon \to 0$ we obtain

$$(V_A V_B)^2 \;\le\; C\,R^2.$$

(The weights $1/r_j$, which minimise the right-hand side of (5.5) when all r_j are positive, are not available if some r_j is 0.)

Step 3: the Cauchy–Schwarz inequality over the coordinates. By the Cauchy–Schwarz inequality and (5.4),

$$R^2 = \Big(\sum_{j=1}^n \sqrt{X_j}\,\sqrt{Y_j}\Big)^2 \;\le\; \Big(\sum_{j=1}^n X_j\Big)\Big(\sum_{j=1}^n Y_j\Big) = V_A V_B.$$

Conclusion. Steps 2 and 3 give $(V_A V_B)^2 \le C\,V_A V_B$. Since $V_A V_B > 0$ we may divide, and get $V_A V_B \le C$, which is the assertion. □

For the rows of a Boolean function, Lemma 5.3 and the all-pairs inequality (Proposition 5.2) give the two-sided inequality.

Theorem 5.4 (Two-sided inequality)**.** *Let $g\colon Q_n \to \{0,1\}$ be non-constant, and for every $x \in Q_n$ let λ^x be a row of g at x. Then $\|\lambda^x\|_1 \ge 1$ for every x, and*

$$4\,\mathbb{E}_x\Big[\frac{1_{\{g(x)=1\}}}{\|\lambda^x\|_1}\Big]\,\mathbb{E}_x\Big[\frac{1_{\{g(x)=0\}}}{\|\lambda^x\|_1}\Big] \;\le\; \max_{j\in[n]} \mathrm{I}_j(g).$$

Proof. By Lemma 2.4(ii), $n \ge 1$ and $\|\lambda^x\|_1 \ge 1$ for every x. Put

$$\mathrm{I}_{\max} = \max_{j\in[n]} \mathrm{I}_j(g), \qquad C = 4^{\,n-1}\,\mathrm{I}_{\max}.$$

By Lemma 2.1, $b_j(g) = 2^{n-1}\,\mathrm{I}_j(g) \le 2^{n-1}\,\mathrm{I}_{\max}$ for every j.

We apply Lemma 5.3 to the sets $A = \mathcal{O}$ and $B = \mathcal{N}$, which are not empty because g is not constant, to the vectors λ^x ($x \in \mathcal{O}$) and $\mu^y = \lambda^y$ ($y \in \mathcal{N}$), to the sets $J(x,y) = \Delta(x,y)$, and to the constant $C \geq 0$.

Let $w \in (0,\infty)^n$. Condition (i) of the lemma holds by Proposition 5.2. By (5.1) and $b_j(g) \leq 2^{n-1}\,\mathrm{I}_{\max}$,

$$\sum_{x\in\mathcal{O}} \sum_{y\in\mathcal{N}} \Big(\sum_{j\in\Delta(x,y)} w_j\, \lambda_j^x\, \lambda_j^y \Big)^{-1} \;\leq\; 2^{n-1} \sum_{j=1}^{n} \frac{2^{n-1}\,\mathrm{I}_{\max}}{w_j} \;=\; C \sum_{j=1}^{n} \frac{1}{w_j},$$

which is condition (ii). The lemma therefore gives

$$\Big(\sum_{x\in\mathcal{O}} \frac{1}{\|\lambda^x\|_1}\Big)\Big(\sum_{y\in\mathcal{N}} \frac{1}{\|\lambda^y\|_1}\Big) \;\leq\; C \;=\; 4^{n-1}\,\mathrm{I}_{\max}\,.$$

Finally,

$$\begin{aligned} 4\,\mathbb{E}_x\Big[\frac{1_{\{g(x)=1\}}}{\|\lambda^x\|_1}\Big]\,\mathbb{E}_x\Big[\frac{1_{\{g(x)=0\}}}{\|\lambda^x\|_1}\Big] \;&=\; 4\cdot 4^{-n}\Big(\sum_{x\in\mathcal{O}} \frac{1}{\|\lambda^x\|_1}\Big)\Big(\sum_{y\in\mathcal{N}} \frac{1}{\|\lambda^y\|_1}\Big) \\ &\leq\; 4\cdot 4^{-n}\cdot 4^{n-1}\,\mathrm{I}_{\max} \;=\; \mathrm{I}_{\max}\,. \end{aligned}$$

□

Example 5.5 (Equality)**.** Let $n \geq 1$ and $g(x) = 1$ if and only if $x_1 = 1$, with the rows $\lambda^x = (1,0,\ldots,0)$ of Example 4.4. Then $\|\lambda^x\|_1 = 1$ for all x, the two expectations in Theorem 5.4 are equal to $\frac{1}{2}$, and $\mathrm{I}_1(g) = 1$; so both sides of the two-sided inequality are equal to 1. In (5.1) there is equality as well: every pair $(x,y) \in \mathcal{O}\times\mathcal{N}$ has $1 \in \Delta(x,y)$ and $\sum_{j\in\Delta(x,y)} w_j\lambda_j^x\lambda_j^y = w_1$, so the left-hand side is $2^{n-1}\cdot 2^{n-1}/w_1$, and the right-hand side is $2^{n-1}\,b_1(g)/w_1$ with $b_1(g) = 2^{n-1}$.

6. The row bound

In this section we prove that a function $f\colon Q_n \to [-1,1]$ of degree at most d has, at every point x, a row of cost at most $\pi^2 d^2$ (Theorem 6.8). The idea, which is that of the proof of Agarwal and Ben-David [ABD] (see Remark 6.9), is as follows.

By linear programming duality, it suffices to bound the total sensitivity $\sum_S \kappa_S |f(x) - f(x^S)|$ of every *fractional packing* κ of blocks, in which the blocks through any one coordinate have total weight at most 1. A fractional packing is approximated by a list of blocks in which every coordinate occurs at most N times.

For such a list we construct a bounded function G of low degree in new variables, one for each block of the list, such that, at the point where all new variables are equal to 1, changing the sign of one new variable has the same effect as flipping the corresponding block in f.

The construction composes f with a Chebyshev polynomial, which takes the value 1 at a coordinate through which no block is flipped and the value -1 at a coordinate through which exactly one block is flipped, and stays in $[-1,1]$ otherwise. Markov's inequality then bounds the sum of the changes of G.

The inequality for lists of blocks (Lemma 6.4) has a second consequence, which is recorded at the end of the section: at every point x, the sum of the numbers $|f(x) - f(x^j)|$ is at most $4d^2$ (Corollary 6.10). It is used in Section 8.

Both results are known, with better constants (Remark 6.9); they are proved here to make the paper self-contained. A reader who accepts Theorem 6.8 and Corollary 6.10 may skip this section: nothing else from it is used in the proof of Theorem 1.1.

6.1. **Chebyshev polynomials and Markov's inequality at an endpoint.** The *Chebyshev polynomials* $T_0, T_1, T_2, \dots$ are defined by

$$T_0(t) = 1, \qquad T_1(t) = t, \qquad T_{k+1}(t) = 2t\,T_k(t) - T_{k-1}(t) \quad (k \geq 1).$$

We use the following standard facts (see [BE]), valid for every integer $k \geq 0$.

(C1) T_k is a polynomial with real coefficients of degree at most k.
(C2) $T_k(\cos\theta) = \cos(k\theta)$ for every real θ.
(C3) $|T_k(t)| \leq 1$ for $t \in [-1,1]$; $T_k(1) = 1$; and $T_k(\cos(\pi/k)) = -1$ if $k \geq 1$.
(C4) $T_k'(1) = k^2$.

Indeed, (C1) is clear from the recurrence.

(C2) holds for $k = 0$ and $k = 1$, and the step from $k-1$ and k to $k+1$ is the identity

$$\cos((k+1)\theta) + \cos((k-1)\theta) \;=\; 2\cos\theta\cos(k\theta).$$

(C3) follows from (C2), because every $t \in [-1,1]$ is of the form $\cos\theta$, $1 = \cos 0$, and $\cos(k \cdot \pi/k) = -1$.

For (C4), differentiate the recurrence and put $t = 1$: by (C3),

$$T_{k+1}'(1) \;=\; 2 + 2\,T_k'(1) - T_{k-1}'(1) \qquad \text{for } k \geq 1.$$

The numbers k^2 satisfy the same recurrence,

$$(k+1)^2 \;=\; 2 + 2k^2 - (k-1)^2,$$

and the initial values agree: $T_0'(1) = 0$ and $T_1'(1) = 1$.

Markov's inequality states that a real polynomial p of degree at most D with $|p| \leq 1$ on $[-1,1]$ satisfies $|p'| \leq D^2$ on $[-1,1]$; see [BE]. We need it only at an endpoint, where it has the following short proof.

Lemma 6.1 (Markov's inequality at an endpoint)**.** *Let $D \geq 0$ be an integer and let p be a polynomial with real coefficients of degree at most D such that $|p(t)| \leq 1$ for all $t \in [-1,1]$. Then $|p'(1)| \leq D^2$. Consequently, if q is a polynomial with real coefficients of degree at most D such that $|q(s)| \leq 1$ for all $s \in [0,2]$, then $|q'(0)| \leq D^2$.*

Proof. Step 1: the nodes. If $D = 0$ then p is constant and $p'(1) = 0$. Let $D \geq 1$, and put $\eta_k = \cos(k\pi/D)$ for $k = 0, 1, \dots, D$. Since the cosine is strictly decreasing on $[0,\pi]$,

$$1 = \eta_0 > \eta_1 > \cdots > \eta_D = -1.$$

Step 2: the interpolation formula. For $k = 0, \dots, D$ let

$$\psi_k(t) = \prod_{i \neq k} \frac{t - \eta_i}{\eta_k - \eta_i},$$

where i runs over $\{0, \dots, D\} \setminus \{k\}$. These are polynomials of degree D with $\psi_k(\eta_k) = 1$ and $\psi_k(\eta_i) = 0$ for $i \neq k$.

If r is a polynomial of degree at most D, then $r - \sum_k r(\eta_k)\psi_k$ has degree at most D and vanishes at the $D+1$ points $\eta_0, \dots, \eta_D$, so it is the zero polynomial. Hence

$$r'(1) = \sum_{k=0}^{D} r(\eta_k)\, \psi_k'(1) \qquad \text{for every polynomial } r \text{ of degree at most } D. \tag{6.1}$$

Step 3: the signs. Next we show that

$$(-1)^k \psi_k'(1) > 0 \qquad \text{for every } k \in \{0, \dots, D\}. \tag{6.2}$$

For $k = 0$, all factors of ψ_0 are equal to 1 at $t = 1$, so

$$\psi_0'(1) \;=\; \sum_{i=1}^{D} \frac{1}{1-\eta_i} \;>\; 0.$$

For $k \geq 1$ the polynomial ψ_k has the factor $t - \eta_0 = t - 1$, and therefore

$$\psi_k'(1) = \frac{\prod_{i \neq 0,k}(1-\eta_i)}{\prod_{i \neq k}(\eta_k - \eta_i)}.$$

The numerator is positive. In the denominator the factor $\eta_k - \eta_i$ is negative for the k indices $i < k$ and positive for the indices $i > k$, so the denominator has the sign $(-1)^k$. This proves (6.2).

Step 4: comparison with T_D. By (C1) and (C2), the polynomial T_D has degree at most D and $T_D(\eta_k) = \cos(k\pi) = (-1)^k$. Applying (6.1) to p and to T_D, and using $|p(\eta_k)| \leq 1$, (6.2) and (C4), we get

$$|p'(1)| \;\leq\; \sum_{k=0}^{D} |\psi_k'(1)| \;=\; \sum_{k=0}^{D} (-1)^k \psi_k'(1) \;=\; T_D'(1) \;=\; D^2.$$

Step 5: the interval $[0,2]$. For the last statement apply the first one to $p(t) = q(1-t)$: its degree is at most D, the number $1 - t$ runs over $[0,2]$ when t runs over $[-1,1]$, and $p'(1) = -q'(0)$. □

6.2. **Bounded functions and their multilinear extensions.** Let $h\colon Q_n \to \mathbb{R}$ have the expansion $h = \sum_{V \subseteq [n]} \hat{h}(V)\chi_V$ of (2.2). Its *multilinear extension* is the function

$$\tilde{h}(u) = \sum_{V \subseteq [n]} \hat{h}(V) \prod_{i \in V} u_i \qquad (u \in \mathbb{R}^n),$$

which agrees with h on Q_n.

The extension is bounded on the solid cube, because its values there are averages of values of h.

Lemma 6.2 (Extension)**.** *If $|h(z)| \leq 1$ for all $z \in Q_n$, then $|\tilde{h}(u)| \leq 1$ for all $u \in [-1,1]^n$.*

Proof. Fix $u \in [-1,1]^n$ and put $W_u(z) = \prod_{i=1}^n \frac{1+z_i u_i}{2}$ for $z \in Q_n$; these numbers are non-negative.

Let $V \subseteq [n]$. The sum over $z \in Q_n$ of the product $W_u(z)\chi_V(z)$ is the product of the sums over the single coordinates $z_i \in \{-1, 1\}$:

$$\sum_{z \in Q_n} W_u(z)\chi_V(z) = \prod_{i \in V}\Big(\frac{1+u_i}{2} - \frac{1-u_i}{2}\Big) \prod_{i \notin V}\Big(\frac{1+u_i}{2} + \frac{1-u_i}{2}\Big) = \prod_{i \in V} u_i.$$

For $V = \emptyset$ this gives $\sum_z W_u(z) = 1$. Multiplying by $\hat{h}(V)$ and summing over V we get

$$\tilde{h}(u) \;=\; \sum_z W_u(z)\, h(z).$$

Hence

$$|\tilde{h}(u)| \;\le\; \sum_z W_u(z)|h(z)| \;\le\; \sum_z W_u(z) \;=\; 1. \qquad \square$$

6.3. **The sensitivity lemma.** In the next two lemmas new variables are indexed by $[k]$, and $\mathbf{1} = (1, \dots, 1) \in Q_k$. Thus $\mathbf{1}^b$ is the point of Q_k whose bth coordinate is -1 and whose other coordinates are 1.

The next lemma is the form in which Markov's inequality is used.

Lemma 6.3 (Sensitivity)**.** *Let $k \ge 0$ and $D \ge 0$ be integers and let $G\colon Q_k \to [-1, 1]$ be a function of degree at most D. Then*

$$\sum_{b=1}^{k} |G(\mathbf{1}) - G(\mathbf{1}^b)| \;\le\; 4D^2.$$

Proof. Let $G = \sum_{U \subseteq [k]} \widehat{G}(U)\chi_U$, with $\widehat{G}(U) = 0$ for $|U| > D$. Since $\chi_U(\mathbf{1}) = 1$, while $\chi_U(\mathbf{1}^b)$ equals -1 if $b \in U$ and 1 if $b \notin U$, we have $G(\mathbf{1}) - G(\mathbf{1}^b) = 2\sum_{U \ni b} \widehat{G}(U)$. Hence, for every $R \subseteq [k]$,

$$\sum_{b \in R}\big(G(\mathbf{1}) - G(\mathbf{1}^b)\big) = 2 \sum_{U \subseteq [k]} \widehat{G}(U)\, |U \cap R|. \tag{6.3}$$

Fix R, and for real s let $u(s) \in \mathbb{R}^k$ be the point with $u(s)_b = 1 - s$ for $b \in R$ and $u(s)_b = 1$ for $b \notin R$. Put

$$\varphi_R(s) = \tilde{G}(u(s)) = \sum_{U \subseteq [k]} \widehat{G}(U)\, (1 - s)^{|U \cap R|}.$$

This is a polynomial in s of degree at most D, because $|U \cap R| \le |U| \le D$ whenever $\widehat{G}(U) \ne 0$. For $s \in [0, 2]$ the point $u(s)$ lies in $[-1, 1]^k$, so $|\varphi_R(s)| \le 1$ by Lemma 6.2.

Lemma 6.1 gives $|\varphi'_R(0)| \le D^2$. Since $\varphi'_R(0) = -\sum_U \widehat{G}(U)|U \cap R|$, we conclude from (6.3) that

$$\Big|\sum_{b \in R}\big(G(\mathbf{1}) - G(\mathbf{1}^b)\big)\Big| \le 2D^2 \qquad \text{for every } R \subseteq [k].$$

Applying this to the set R_+ of all b with $G(\mathbf{1}) - G(\mathbf{1}^b) \ge 0$ and to its complement R_-, we obtain

$$\sum_{b=1}^{k} |G(\mathbf{1}) - G(\mathbf{1}^b)| = \sum_{b \in R_+}\big(G(\mathbf{1}) - G(\mathbf{1}^b)\big) - \sum_{b \in R_-}\big(G(\mathbf{1}) - G(\mathbf{1}^b)\big) \le 4D^2. \qquad \square$$

6.4. **The block inequality.** We now carry out the construction described at the beginning of the section.

Lemma 6.4 (Block inequality)**.** *Let $d \geq 0$ be an integer, let $f\colon Q_n \to [-1,1]$ be a function of degree at most d, and let $x \in Q_n$. Let $S_1, \dots, S_k$ be subsets of $[n]$, not necessarily distinct, and let $N \geq 0$ and $m \geq 1$ be integers such that every $i \in [n]$ lies in at most N of the sets $S_1, \dots, S_k$ and*

$$N\big(1 - \cos(\pi/m)\big) \leq 2.$$

Then

$$\sum_{b=1}^{k} |f(x) - f(x^{S_b})| \;\leq\; 4\,d^2 m^2.$$

Proof. Put $c = 1 - \cos(\pi/m)$, so that $c \geq 0$ and $Nc \leq 2$. For $w \in Q_k$ and $i \in [n]$ let

$$\mathrm{cnt}_i(w) = \#\{b \in [k] :\ i \in S_b,\ w_b = -1\} = \sum_{b:\, i \in S_b} \frac{1 - w_b}{2}, \qquad \theta_i(w) = T_m\big(1 - c\ \mathrm{cnt}_i(w)\big).$$

Let $f = \sum_{V \subseteq [n]} \hat{f}(V)\chi_V$ with $\hat{f}(V) = 0$ for $|V| > d$, and define $G\colon Q_k \to \mathbb{R}$ by

$$G(w) = \tilde{f}\big(x_1\theta_1(w), \dots, x_n\theta_n(w)\big) = \sum_{V \subseteq [n]} \hat{f}(V) \prod_{i \in V} x_i\,\theta_i(w).$$

The function G has the following four properties.

(a) $|G(w)| \leq 1$ *for all* $w \in Q_k$. Since $0 \leq \mathrm{cnt}_i(w) \leq N$, the number $1 - c\ \mathrm{cnt}_i(w)$ lies in $[1 - Nc, 1] \subseteq [-1, 1]$, so $|\theta_i(w)| \leq 1$ by (C3). Thus the point $(x_i\theta_i(w))_{i \in [n]}$ lies in $[-1,1]^n$, and (a) follows from Lemma 6.2 applied to f.

(b) G *has degree at most* dm. As a function of $w \in Q_k$, $1 - c\ \mathrm{cnt}_i$ has degree at most 1, so its jth power has degree at most j, and by (C1) the function θ_i has degree at most m. For $|V| \leq d$ the product $\prod_{i \in V} x_i\theta_i$ is a constant multiple of a product of at most d functions of degree at most m, so it has degree at most dm.

(c) $G(\mathbf{1}) = f(x)$. Indeed $\mathrm{cnt}_i(\mathbf{1}) = 0$ and $\theta_i(\mathbf{1}) = T_m(1) = 1$ for every i, so $G(\mathbf{1}) = \tilde{f}(x) = f(x)$.

(d) $G(\mathbf{1}^b) = f(x^{S_b})$ *for every* $b \in [k]$. The only coordinate of $\mathbf{1}^b$ equal to -1 is the bth one, so $\mathrm{cnt}_i(\mathbf{1}^b)$ is 1 if $i \in S_b$ and 0 otherwise. By (C3), $\theta_i(\mathbf{1}^b) = T_m(\cos(\pi/m)) = -1$ if $i \in S_b$, and $\theta_i(\mathbf{1}^b) = T_m(1) = 1$ otherwise. Hence $x_i\theta_i(\mathbf{1}^b) = (x^{S_b})_i$ for all i, and $G(\mathbf{1}^b) = \tilde{f}(x^{S_b}) = f(x^{S_b})$.

By (a) and (b), Lemma 6.3 applies to G with $D = dm$; by (c) and (d) its conclusion is the assertion. □

6.5. **Fractional packings.** A *fractional packing* on $[n]$ is a family $\kappa = (\kappa_S)_{S \subseteq [n]}$ of real numbers such that

$$\kappa_S \geq 0 \quad \text{for all } S \subseteq [n], \qquad \sum_{S \ni i} \kappa_S \leq 1 \quad \text{for all } i \in [n].$$

A fractional packing is approximated by lists of blocks; in the limit the block inequality gives the following bound.

Lemma 6.5 (Packing bound)**.** *Let $d \geq 0$ be an integer, let $f\colon Q_n \to [-1,1]$ be a function of degree at most d, let $x \in Q_n$, and let κ be a fractional packing on $[n]$. Then*

$$\sum_{S\subseteq[n]} \kappa_S \, |f(x) - f(x^S)| \;\leq\; \pi^2 d^2.$$

Proof. Write $t_S = |f(x) - f(x^S)|$, $\Lambda = \sum_S \kappa_S t_S$ and $C_0 = \sum_S t_S$. Let $M \geq 1$ be an integer, and put

$$N = M^2, \qquad k_S = \lfloor M^2 \kappa_S \rfloor \quad (S \subseteq [n]), \qquad m = \lceil \pi M/2 \rceil.$$

Step 1: a list of blocks. Consider the list of blocks in which every set $S \subseteq [n]$ occurs exactly k_S times. A coordinate i lies in

$$\sum_{S\ni i} k_S \;\leq\; M^2 \sum_{S\ni i} \kappa_S \;\leq\; N$$

blocks of the list.

Further $m \geq 1$ and

$$\pi M/2 \;\leq\; m \;<\; \pi M/2 + 1.$$

Since $1 - \cos\vartheta = 2\sin^2(\vartheta/2) \leq \vartheta^2/2$ for real ϑ,

$$N\big(1 - \cos(\pi/m)\big) \leq \frac{M^2\pi^2}{2m^2} \leq 2.$$

Step 2: the block inequality. Lemma 6.4 therefore gives

$$\sum_{S\subseteq[n]} k_S \, t_S \;\leq\; 4d^2m^2 \;\leq\; d^2(\pi M + 2)^2.$$

As $M^2\kappa_S < k_S + 1$ and $t_S \geq 0$, it follows that

$$M^2\Lambda \leq \sum_S (k_S + 1)\, t_S \leq d^2(\pi M + 2)^2 + C_0, \qquad \text{that is,} \qquad \Lambda \leq d^2\Big(\pi + \frac{2}{M}\Big)^2 + \frac{C_0}{M^2}.$$

Step 3: the limit. This holds for every integer $M \geq 1$, and C_0 does not depend on M. Letting $M \to \infty$ we get $\Lambda \leq \pi^2 d^2$. □

6.6. **Duality.** The passage from packings to rows is an instance of linear programming duality. We derive it from the following separation lemma, whose short proof we include.

For a finite set $\mathcal{S}$ we write $\mathbb{R}^{\mathcal{S}}$ for the space of real vectors $v = (v_s)_{s\in\mathcal{S}}$, with the inner product $\langle v, v'\rangle = \sum_s v_s v'_s$ and the Euclidean norm $|v| = \langle v, v\rangle^{1/2}$.

Lemma 6.6 (Separation from the orthant)**.** *Let $\mathcal{S}$ be a finite set and let $C \subseteq \mathbb{R}^{\mathcal{S}}$ be a non-empty compact convex set which contains no vector with all coordinates non-negative. Then there is a vector $\kappa \in [0,\infty)^{\mathcal{S}}$ such that*

$$\sum_{s\in\mathcal{S}} \kappa_s \, v_s < 0 \qquad \textit{for every } v \in C.$$

Proof. Step 1: the distance to the orthant. For a real number ξ let $\xi^- = \max(-\xi, 0)$, and for $v \in \mathbb{R}^{\mathcal{S}}$ let $\rho(v) = \sum_s (v_s^-)^2$.

If ξ is real and $\zeta \geq 0$, then $(\xi - \zeta)^2 \geq (\xi^-)^2$: this is clear for $\xi \geq 0$, and for $\xi < 0$ we have $\zeta - \xi \geq -\xi > 0$. Hence

$$\rho(v) \leq |v - p|^2 \qquad \text{for all } v \in \mathbb{R}^{\mathcal{S}} \text{ and } p \in [0,\infty)^{\mathcal{S}}. \tag{6.4}$$

Step 2: a nearest point. The function ρ is continuous, and $\rho(v) = 0$ if and only if all coordinates of v are non-negative. Since C is compact and not empty, ρ attains its minimum on C at some point $y \in C$, and $\rho(y) > 0$ by the hypothesis on C.

Let κ be the vector with $\kappa_s = y_s^- \geq 0$. Then $y + \kappa$ has the non-negative coordinates $\max(y_s, 0)$, and

$$|\kappa|^2 = \rho(y), \qquad \langle \kappa, y \rangle = -\rho(y),$$

the latter because $y_s^- \, y_s = -(y_s^-)^2$ for every s.

Step 3: the inequality. Let $v \in C$ and $0 < \varepsilon \leq 1$. The point $y + \varepsilon(v - y)$ belongs to C, because C is convex. By the minimality of $\rho(y)$ and by (6.4) with $p = y + \kappa$,

$$\rho(y) \;\leq\; \rho\big(y + \varepsilon(v-y)\big) \;\leq\; |\varepsilon(v-y) - \kappa|^2 = \varepsilon^2 |v-y|^2 - 2\varepsilon\langle \kappa, v - y\rangle + \rho(y).$$

Hence $\langle \kappa, v - y\rangle \leq \frac{\varepsilon}{2}|v-y|^2$ for all $\varepsilon \in (0,1]$, so that $\langle \kappa, v - y\rangle \leq 0$, and

$$\langle \kappa, v\rangle \;\leq\; \langle \kappa, y\rangle = -\rho(y) < 0. \qquad \square$$

Proposition 6.7 (Duality)**.** *Let $n \geq 0$, let $(t_S)_{S\subseteq[n]}$ be a family of real numbers and let K be a real number. Suppose that*

$$\sum_{S\subseteq[n]} \kappa_S \, t_S \leq K \qquad \textit{for every fractional packing } \kappa \textit{ on } [n].$$

Then there is a vector $a \in [0,\infty)^n$ with $\|a\|_1 \leq K$ and $t_S \leq a(S)$ for every $S \subseteq [n]$.

Proof. The family $\kappa = 0$ is a fractional packing, so $K \geq 0$. Let

$$P = \{a \in [0,\infty)^n : \; \|a\|_1 \leq K\}.$$

This set contains 0, it is convex, and it is compact, being closed and contained in $[0,K]^n$.

Let $\mathcal{S}$ be the set of all subsets of $[n]$, and let $C \subseteq \mathbb{R}^{\mathcal{S}}$ be the image of P under the map $a \mapsto (a(S) - t_S)_{S\in\mathcal{S}}$. This map is the sum of a linear map and a constant, so C is non-empty, compact and convex.

If some vector of C has all coordinates non-negative, then the corresponding $a \in P$ has the required properties. Suppose that this is not the case. By Lemma 6.6 there is a family $\kappa = (\kappa_S)_{S\subseteq[n]}$ of non-negative numbers with

$$\sum_{S\subseteq[n]} \kappa_S \big(a(S) - t_S\big) < 0 \qquad \text{for every } a \in P. \tag{6.5}$$

Put $\vartheta = \sum_S \kappa_S t_S$ and $\sigma_i = \sum_{S\ni i} \kappa_S \geq 0$ for $i \in [n]$. Since

$$\sum_S \kappa_S \, a(S) \;=\; \sum_{i=1}^n a_i \sigma_i,$$

the inequality (6.5) says that

$$\sum_{i=1}^n a_i \sigma_i < \vartheta \qquad \text{for every } a \in P. \tag{6.6}$$

For $a = 0$ this gives $\vartheta > 0$. We distinguish two cases.

Case A: $\sigma_i = 0$ for all $i \in [n]$. Then every non-negative multiple of κ is a fractional packing, in particular $\kappa' = \frac{K+1}{\vartheta}\,\kappa$. But $\sum_S \kappa'_S t_S = K + 1 > K$, contrary to the hypothesis.

Case B: $\sigma_i > 0$ for some $i \in [n]$. Choose $j \in [n]$ with $\sigma_j = \max_i \sigma_i > 0$. Then $\kappa' = \kappa/\sigma_j$ is a fractional packing, because $\sum_{S \ni i} \kappa'_S = \sigma_i/\sigma_j \le 1$ for all i. The hypothesis gives $\vartheta/\sigma_j \le K$, that is, $\vartheta \le K\sigma_j$.

On the other hand, the vector a with $a_j = K$ and $a_i = 0$ for $i \ne j$ belongs to P, and (6.6) gives $K\sigma_j < \vartheta$, a contradiction.

Both cases are impossible, and the proposition is proved. □

6.7. **The row bound.**

Theorem 6.8 (Row bound)**.** *Let $n \ge 0$ and $d \ge 0$ be integers and let $f\colon Q_n \to [-1,1]$ be a function of degree at most d. Then for every $x \in Q_n$ there is a row a of f at x with cost $\|a\|_1 \le \pi^2 d^2$.*

Proof. Fix x and apply Proposition 6.7 to the numbers $t_S = |f(x) - f(x^S)|$ and to $K = \pi^2 d^2$. Its hypothesis is Lemma 6.5. The resulting vector $a \in [0,\infty)^n$ satisfies $\|a\|_1 \le \pi^2 d^2$ and $|f(x) - f(x^S)| \le a(S)$ for all $S \subseteq [n]$, so it is a row of f at x (Definition 2.3). □

Remark 6.9 (The bound of Agarwal and Ben-David). Theorem 6.8 is known, with a better constant.

For $f\colon \{-1,1\}^n \to [-1,1]$, Agarwal and Ben-David [ABD] define the sensitivity of a block B at x as $|f(x) - f(x^B)|/2$, and the fractional block sensitivity $\mathrm{fbs}(f)$ as the largest value of $\sum_B w_B |f(x) - f(x^B)|/2$ over all points x and all fractional packings w.

They prove [ABD, Theorem 3.3] that

$$\mathrm{fbs}(f) \;\le\; \frac{\pi^2}{4}\deg(f)^2.$$

In the notation of Lemma 6.5 this is the bound $\frac{\pi^2}{2}d^2$ in place of $\pi^2 d^2$, and by Proposition 6.7 it gives rows of cost at most $\frac{\pi^2}{2}d^2$.

The statement about rows is [ABD, Theorem 1.7], where it is formulated for functions $f\colon \{0,1\}^n \to [0,1]$ of degree at most d, with the bound $\frac{\pi^2}{4}d^2$.

Their proof follows the strategy of Anshu, Ben-David and Kundu [ABK] for Boolean functions.

It composes f with Chebyshev polynomials of affine functions of new variables, N of them for each coordinate, so that the blocks become disjoint, and then applies, through block sensitivity, the inequality $\sum_i |f(x) - f(x^i)|/2 \le \deg(f)^2$ for bounded functions, which is due to Filmus, Hatami, Keller and Lifshitz [FHKL, Theorem 3.3].

The proof given above uses the same composition with Chebyshev polynomials, with one new variable for each block, and it uses Lemma 6.3 in place of the inequality of [FHKL].

Lemma 6.3 is weaker than that inequality by a factor 2, which accounts for the factor 2 between the two bounds. That the classical inequality of Markov, which is what Lemma 6.3 uses, gives the inequality of [FHKL] with this weaker constant is remarked in [FHKL, Section 3].

6.8. **The increments at a point.** Lemma 6.3 bounds the sum of the numbers $|G(\mathbf{1})-G(\mathbf{1}^b)|$, the increments of G at the point $\mathbf{1}$. The block inequality (Lemma 6.4), applied to blocks with one element, gives the same bound at every point.

This is the second property of bounded functions of low degree, besides the row bound (Theorem 6.8), that enters the proof of Theorem 1.1.

Corollary 6.10 (Sensitivity at every point)**.** *Let $n \geq 0$ and $d \geq 0$ be integers, and let $f\colon Q_n \to [-1,1]$ be a function of degree at most d. Then*

$$\sum_{j=1}^{n} |f(x) - f(x^j)| \;\leq\; 4\,d^2 \qquad \text{for every } x \in Q_n.$$

Consequently $L(f) = \sum_{j=1}^{n} L_j(f) \leq 4\,d^2$.

Proof. Fix $x \in Q_n$. We apply Lemma 6.4 with $k = n$, with the sets $S_b = \{b\}$ for $b \in [n]$, and with $N = 1$ and $m = 1$. Every $i \in [n]$ lies in exactly one of these sets, and $N\big(1-\cos(\pi/m)\big) = 1 - \cos\pi = 2$. Since $x^{S_b} = x^b$, the conclusion of the lemma is the first assertion.

The second assertion follows by taking the expectation over x. □

Corollary 6.10 is the inequality of Filmus, Hatami, Keller and Lifshitz [FHKL, Theorem 3.3], which is recalled in Remark 6.9, with the constant 4 in place of 2; if their inequality is used, the bound $4d^2$ can be replaced by $2d^2$ (see Remark 8.4).

That the sum of the L_1 influences of a bounded function is bounded by a polynomial in its degree was asked by Aaronson and Ambainis [AA14, Section 1.3] and was first proved by Bačkurs and Bavarian [BB].

7. Real functions: integration over the level

The all-pairs inequality of Section 5 concerns Boolean functions. In this section it is transferred to real functions through their level functions, following the plan of Lovett and Zhang [LZ, Section 5], which is recalled in Section 1.4 and compared step by step with the present argument in Section 7.5.

Let $f\colon Q_n \to \mathbb{R}$ be a function with a row at every point. For a real number t, the *level function* $g_t = 1_{\{f>t\}}$ is Boolean, and at every point x with $f(x) \neq t$ the row of f at x, divided by $|f(x) - t|$, is a row of g_t. We apply the all-pairs inequality to every g_t and integrate over t. The result is the integrated all-pairs inequality (Proposition 7.4): it bounds a sum over the pairs of points x, y with $f(y) < f(x)$, in which each pair contributes a multiple of $(f(x) - f(y))^3$, by a sum over the coordinates, in which the coordinate j contributes a multiple of the mean absolute increment $L_j(f)$.

Removing the weights with Lemma 5.3, we obtain an inequality between a third moment of f and the largest of the numbers $L_j(f)$ (Theorem 7.6), and from it a lower bound for the largest $L_j(f)$ in terms of the variance of f (Theorem 7.10). The arguments use only the existence of rows of bounded cost: neither the degree of f nor a bound for its values appears before Corollary 7.11.

7.1. **Level functions and their rows.** Throughout this section, $n \geq 0$ is an integer, $f\colon Q_n \to \mathbb{R}$ is a function, and for every $x \in Q_n$ a row a^x of f at x is given (Definition 2.3).

For $w \in (0,\infty)^n$ and $x, y \in Q_n$ we put

$$s_{xy}(w) \;:=\; \sum_{j \in \Delta(x,y)} w_j\, a_j^x\, a_j^y. \tag{7.1}$$

For a real number ξ we write $\xi^+ = \max(\xi, 0)$. One of the numbers ξ^+ and $(-\xi)^+$ is equal to $|\xi|$, and the other is 0; hence

$$|\xi| = \xi^+ + (-\xi)^+ \qquad \text{and} \qquad |\xi|^3 = (\xi^+)^3 + \big((-\xi)^+\big)^3 \qquad \text{for every real } \xi. \tag{7.2}$$

For a real number t let $g_t = 1_{\{f>t\}}$, that is,

$$g_t(x) = 1 \quad \text{if } f(x) > t, \qquad g_t(x) = 0 \quad \text{if } f(x) \leq t.$$

For $t \in \mathbb{R}$ and $x \in Q_n$ let

$$\lambda^{t,x} = \frac{a^x}{|f(x) - t|} \quad \text{if } f(x) \neq t, \qquad \lambda^{t,x} = (1, \dots, 1) \quad \text{if } f(x) = t.$$

By Lemma 2.4(iii) and (i), the vector $\lambda^{t,x}$ is a row of g_t at x.

At a point x with $f(x) = t$ any other row of g_t would serve as well: the terms in which such a point occurs are discarded in the proof of Lemma 7.2.

The first lemma evaluates the bilinear expression of the all-pairs inequality (Proposition 5.2) for these rows, and it shows that the denominators which occur in this section are positive.

Lemma 7.1. *Let $w \in (0,\infty)^n$.*

(i) *If $t \in \mathbb{R}$ and $x, y \in Q_n$ satisfy $f(y) < t < f(x)$, then*

$$\sum_{j \in \Delta(x,y)} w_j\, \lambda_j^{t,x}\, \lambda_j^{t,y} \;=\; \frac{s_{xy}(w)}{(f(x) - t)\,(t - f(y))}.$$

(ii) *$s_{xy}(w) > 0$ for all $x, y \in Q_n$ with $f(y) < f(x)$.*

Proof. (i) Here $f(x) \neq t$ and $f(y) \neq t$, and $|f(x) - t| = f(x) - t$ and $|f(y) - t| = t - f(y)$. So

$$\lambda_j^{t,x}\, \lambda_j^{t,y} \;=\; \frac{a_j^x\, a_j^y}{(f(x) - t)\,(t - f(y))} \qquad \text{for every } j \in [n],$$

and the identity follows from (7.1).

(ii) Let $f(y) < f(x)$, and put $t = (f(x) + f(y))/2$, so that $f(y) < t < f(x)$. Then $g_t(x) = 1 \neq 0 = g_t(y)$, and $\lambda^{t,x}$ and $\lambda^{t,y}$ are rows of g_t at x and at y.

By Lemma 4.1, the left-hand side of the identity in (i) is positive. The denominator on its right-hand side is positive as well, and therefore $s_{xy}(w) > 0$. □

The next lemma is the all-pairs inequality for the level function g_t, written in terms of f.

Lemma 7.2 (The all-pairs inequality at a level)**.** *Let $w \in (0,\infty)^n$. Then for every real number t*

$$\sum_{\substack{x,y \in Q_n \\ f(y) < f(x)}} \frac{(f(x) - t)^+\,(t - f(y))^+}{s_{xy}(w)} \;\leq\; 2^{n-1} \sum_{j=1}^{n} \frac{b_j(g_t)}{w_j}. \tag{7.3}$$

Proof. The denominators are positive by Lemma 7.1(ii).

Fix t. For the function g_t we have $\mathcal{O} = \{x : f(x) > t\}$ and $\mathcal{N} = \{y : f(y) \le t\}$. Proposition 5.2, applied to g_t with the weights w and the rows $\lambda^{t,x}$, states that

$$\sum_{x:\, f(x)>t} \; \sum_{y:\, f(y)\le t} \Big(\sum_{j\in\Delta(x,y)} w_j\, \lambda_j^{t,x}\, \lambda_j^{t,y} \Big)^{-1} \;\le\; 2^{n-1} \sum_{j=1}^{n} \frac{b_j(g_t)}{w_j},$$

and that all the terms on the left-hand side are positive.

We discard the terms with $f(y) = t$. For the remaining terms we have $f(y) < t < f(x)$, and by Lemma 7.1(i) the term of the pair (x, y) is equal to $(f(x)-t)(t-f(y))/s_{xy}(w)$. Hence

$$\sum_{\substack{x,y\in Q_n \\ f(y)<t<f(x)}} \frac{(f(x)-t)\,(t-f(y))}{s_{xy}(w)} \;\le\; 2^{n-1} \sum_{j=1}^{n} \frac{b_j(g_t)}{w_j}. \tag{7.4}$$

The left-hand sides of (7.3) and (7.4) are equal. Indeed, let $f(y) < f(x)$. If $f(y) < t < f(x)$, then $(f(x)-t)^+ = f(x)-t$ and $(t-f(y))^+ = t-f(y)$. Otherwise $t \le f(y)$ or $t \ge f(x)$; then $(t-f(y))^+ = 0$ or $(f(x)-t)^+ = 0$, and the pair (x, y) contributes 0 to the sum in (7.3). □

7.2. **Integration over the level.** We now integrate (7.3) over t. By definition, $b_j(g_t)$ is the number of points x with $g_t(x) = 1$ and $g_t(x^j) = 0$, that is,

$$b_j(g_t) \;=\; \sum_{x\in Q_n} 1_{\{f(x^j)\le t<f(x)\}} \qquad (j \in [n],\ t \in \mathbb{R}). \tag{7.5}$$

So, as functions of t, the left-hand side of (7.3) is a finite linear combination of functions $t \mapsto (u-t)^+(t-v)^+$ with real numbers u and v, and the right-hand side is a finite linear combination of functions $t \mapsto 1_{\{p\le t<q\}}$ with real numbers p and q.

These functions are piecewise polynomial, and they vanish outside a bounded interval. Their integrals over $\mathbb{R}$, written $\int_{\mathbb{R}} \dots\, dt$, are therefore integrals of polynomials over bounded intervals, and an inequality between two such functions which holds for every t may be integrated.

The next lemma contains the three computations that are needed.

Lemma 7.3. *The following identities hold.*

(i) *For all real numbers u and v,*

$$\int_{\mathbb{R}} (u-t)^+\,(t-v)^+\,dt \;=\; \frac{\big((u-v)^+\big)^3}{6}.$$

(ii) *For all real numbers p and q,*

$$\int_{\mathbb{R}} 1_{\{p\le t<q\}}\,dt \;=\; (q-p)^+.$$

(iii) *For every $j \in [n]$,*

$$\sum_{x\in Q_n} \big(f(x)-f(x^j)\big)^+ \;=\; 2^{n-1}\,L_j(f).$$

Proof. (i) Let $u \le v$. Then the integrand vanishes for every t: its first factor is 0 if $t \ge u$, and its second factor is 0 if $t < u$, because then $t < v$. The right-hand side is 0 as well.

Let $v < u$, and put $\delta = u - v$. The integrand vanishes for $t \le v$ and for $t \ge u$, and it is equal to $(u-t)(t-v)$ for $v \le t \le u$. With the substitution $t = v + \rho$ the integral becomes

$$\int_0^\delta (\delta - \rho)\, \rho \, d\rho \;=\; \frac{\delta^3}{2} - \frac{\delta^3}{3} \;=\; \frac{\delta^3}{6}.$$

(ii) The set of all t with $p \le t < q$ is the interval $[p, q)$ if $p < q$, and it is empty otherwise.

(iii) The map $x \mapsto x^j$ is a bijection of Q_n, and $(x^j)^j = x$. Hence

$$\sum_{x \in Q_n} \big(f(x^j) - f(x)\big)^+ \;=\; \sum_{x \in Q_n} \big(f(x) - f(x^j)\big)^+,$$

and by (7.2)

$$\begin{aligned} 2 \sum_{x \in Q_n} \big(f(x) - f(x^j)\big)^+ &= \sum_{x \in Q_n} \Big(\big(f(x) - f(x^j)\big)^+ + \big(f(x^j) - f(x)\big)^+ \Big) \\ &= \sum_{x \in Q_n} |f(x) - f(x^j)| \;=\; 2^n\, L_j(f). \end{aligned}$$

□

The integration gives the following inequality, which is the analogue of the all-pairs inequality (Proposition 5.2) for real functions. We state it with all its hypotheses.

Proposition 7.4 (Integrated all-pairs inequality)**.** *Let $n \ge 0$, let $f\colon Q_n \to \mathbb{R}$, let $w \in (0,\infty)^n$, and for every $x \in Q_n$ let a^x be a row of f at x. Then the numbers $s_{xy}(w) = \sum_{j \in \Delta(x,y)} w_j a^x_j a^y_j$ are positive for all $x, y \in Q_n$ with $f(y) < f(x)$, and*

$$\sum_{\substack{x,y \in Q_n \\ f(y) < f(x)}} \frac{(f(x) - f(y))^3}{6\, s_{xy}(w)} \;\le\; 4^{\,n-1} \sum_{j=1}^n \frac{L_j(f)}{w_j}. \tag{7.6}$$

Proof. The positivity is Lemma 7.1(ii). The inequality (7.3) holds for every real t, and we integrate it over t.

The left-hand side. Let $f(y) < f(x)$. By Lemma 7.3(i) with $u = f(x)$ and $v = f(y)$, the integral of the term of the pair (x, y) in (7.3) is $(f(x) - f(y))^3/(6\, s_{xy}(w))$. So the integral of the left-hand side of (7.3) is the left-hand side of (7.6).

The right-hand side. For every $j \in [n]$,

$$\int_{\mathbb{R}} b_j(g_t)\, dt \overset{(7.5)}{=} \sum_{x \in Q_n} \int_{\mathbb{R}} 1_{\{f(x^j) \le t < f(x)\}}\, dt \overset{\text{Lem. 7.3(ii)}}{=} \sum_{x \in Q_n} \big(f(x) - f(x^j)\big)^+ \overset{\text{Lem. 7.3(iii)}}{=} 2^{n-1}\, L_j(f).$$

So the integral of the right-hand side of (7.3) is $2^{n-1} \sum_j 2^{n-1} L_j(f)/w_j$, which is the right-hand side of (7.6). □

If f is constant, in particular if $n = 0$, there are no pairs x, y with $f(y) < f(x)$, and the left-hand side of (7.6) is 0.

Example 7.5. Let $n \geq 1$ and $f(x) = x_1$. Then $|f(x) - f(x^S)|$ is 2 if $1 \in S$ and 0 otherwise, so $a^x = (2, 0, \ldots, 0)$ is a row of f at every point x.

We have $f(y) < f(x)$ if and only if $x_1 = 1$ and $y_1 = -1$; there are 4^{n-1} such pairs, and for each of them $(f(x) - f(y))^3 = 8$ and $s_{xy}(w) = 4w_1$. Further $L_1(f) = 2$, and $L_j(f) = 0$ for $j \geq 2$. So the left-hand side of (7.6) is $4^{n-1}/(3w_1)$, and the right-hand side is $2 \cdot 4^{n-1}/w_1$.

7.3. **The third-moment inequality.** In the integrated all-pairs inequality (Proposition 7.4) the weights are removed with Lemma 5.3. We keep only the pairs (x, y) in which $f(x)$ lies above a threshold r and $f(y)$ does not. The result bounds the product of the third moment of the part of f above r and the probability that $f \leq r$.

Theorem 7.6 (Third-moment inequality)**.** *Let $n \geq 0$, let $f\colon Q_n \to \mathbb{R}$, and let K be a real number such that for every $x \in Q_n$ there is a row a^x of f at x with $\|a^x\|_1 \leq K$. Let $M \geq 0$ be a real number such that $L_j(f) \leq M$ for every $j \in [n]$. Then, for every real number r,*

$$\mathbb{E}\big[((f - r)^+)^3\big] \cdot \Pr[f \leq r] \;\leq\; \frac{3}{2} K^2 M, \tag{7.7}$$

$$\mathbb{E}\big[((r - f)^+)^3\big] \cdot \Pr[f \geq r] \;\leq\; \frac{3}{2} K^2 M. \tag{7.8}$$

Proof. The right-hand sides are non-negative, because $M \geq 0$. We prove (7.7) in three steps, and then deduce (7.8) from it.

Fix r, and put

$$A = \{x \in Q_n :\; f(x) > r\}, \qquad B = \{y \in Q_n :\; f(y) \leq r\}.$$

Since $(f(x) - r)^+$ is equal to $f(x) - r$ for $x \in A$ and to 0 for $x \notin A$,

$$\mathbb{E}\big[((f - r)^+)^3\big] = 2^{-n} \sum_{x \in A} (f(x) - r)^3, \qquad \Pr[f \leq r] = 2^{-n}|B|. \tag{7.9}$$

If A or B is empty, then the left-hand side of (7.7) is 0, and the inequality holds. From now on let A and B be non-empty.

Step 1: the pairs in $A \times B$. Let $w \in (0, \infty)^n$, $x \in A$ and $y \in B$. Then $f(y) \leq r < f(x)$. So $s_{xy}(w) > 0$ by Lemma 7.1(ii), and

$$(f(x) - f(y))^3 \;\geq\; (f(x) - r)^3 \;>\; 0.$$

All the terms of the sum in (7.6) are positive. We keep those of the pairs in $A \times B$, multiply by 6, and use $L_j(f) \leq M$. In this way Proposition 7.4 gives, for every $w \in (0, \infty)^n$,

$$\begin{aligned} \sum_{x \in A} \sum_{y \in B} \frac{(f(x) - r)^3}{s_{xy}(w)} \;&\leq\; \sum_{x \in A} \sum_{y \in B} \frac{(f(x) - f(y))^3}{s_{xy}(w)} \;\leq\; \sum_{\substack{x, y \in Q_n \\ f(y) < f(x)}} \frac{(f(x) - f(y))^3}{s_{xy}(w)} \\ &\leq\; 6 \cdot 4^{n-1} \sum_{j=1}^{n} \frac{L_j(f)}{w_j} \;\leq\; 6 \cdot 4^{n-1} M \sum_{j=1}^{n} \frac{1}{w_j}. \end{aligned} \tag{7.10}$$

Step 2: removing the weights. We apply Lemma 5.3 to the sets A and B, to the vectors

$$\lambda^x = \frac{a^x}{(f(x) - r)^3} \quad (x \in A), \qquad \mu^y = a^y \quad (y \in B),$$

to the sets $J(x,y) = \Delta(x,y)$, and to the constant $C = 6 \cdot 4^{n-1} M$. For these data,

$$\sum_{j \in J(x,y)} w_j \, \lambda_j^x \, \mu_j^y \;=\; \frac{s_{xy}(w)}{(f(x)-r)^3} \;>\; 0 \qquad (x \in A,\; y \in B),$$

so condition (i) of the lemma holds, and condition (ii) is (7.10).

The cost of the vector λ^x is $\|a^x\|_1/(f(x)-r)^3$. The lemma therefore shows that the costs $\|a^x\|_1$ for $x \in A$ and $\|a^y\|_1$ for $y \in B$ are positive, and that

$$\Big(\sum_{x \in A} \frac{(f(x)-r)^3}{\|a^x\|_1}\Big)\Big(\sum_{y \in B} \frac{1}{\|a^y\|_1}\Big) \;\le\; 6 \cdot 4^{n-1} M. \tag{7.11}$$

Step 3: the bound for the costs. For every point z of A or of B we have $0 < \|a^z\|_1 \le K$; so $K > 0$ and $1/\|a^z\|_1 \ge 1/K$. All the terms in (7.11) are positive, and therefore

$$\frac{1}{K^2}\Big(\sum_{x \in A} (f(x)-r)^3\Big)|B| \;\le\; 6 \cdot 4^{n-1} M.$$

With (7.9) this gives

$$\mathbb{E}\big[((f-r)^+)^3\big] \cdot \Pr[f \le r] \;=\; 4^{-n}\Big(\sum_{x \in A} (f(x)-r)^3\Big)|B| \;\le\; 4^{-n} \cdot 6 \cdot 4^{n-1} K^2 M \;=\; \frac{3}{2} K^2 M,$$

which is (7.7).

The inequality (7.8). For every x the vector a^x is also a row of the function $-f$ at x, and $L_j(-f) = L_j(f)$ for every j. So (7.7) holds for $-f$ and $-r$ in place of f and r. Since $(-f)-(-r) = r-f$, and since $-f \le -r$ if and only if $f \ge r$, this is (7.8). □

A special case concerns two sets on which the values of f are separated by a gap, that is, subsets of a superlevel set and of a sublevel set of f. These are the sets to which Lovett and Zhang apply their conjectured inequality (Section 1.4).

Corollary 7.7 (Separated sets)**.** *Let f, K and M be as in Theorem 7.6. Let r and $\theta > 0$ be real numbers, and let X and Y be subsets of Q_n such that $f(x) \ge r + \theta$ for all $x \in X$ and $f(y) \le r$ for all $y \in Y$. Then*

$$2\,\theta^3 \cdot \frac{|X|}{2^n} \cdot \frac{|Y|}{2^n} \;\le\; 3\,K^2 M.$$

Proof. We have $((f-r)^+)^3 \ge \theta^3$ on X, and $((f-r)^+)^3 \ge 0$ everywhere; so $\mathbb{E}[((f-r)^+)^3] \ge \theta^3 |X|/2^n$. Further $\Pr[f \le r] \ge |Y|/2^n$. The assertion follows from (7.7), multiplied by 2. □

7.4. **The mean absolute increments.** We apply the third-moment inequality (Theorem 7.6) at a median of f, on both sides, and compare the third absolute moment of f about the median with the variance. Two elementary facts are needed: the existence of a median, and an inequality between the second and the third absolute moment.

Lemma 7.8 (Median)**.** *Let $n \ge 0$ and $f\colon Q_n \to \mathbb{R}$. Then there is a real number m such that*

$$\Pr[f \le m] \ge \frac{1}{2} \qquad \textit{and} \qquad \Pr[f \ge m] \ge \frac{1}{2}.$$

Proof. The function f has finitely many values, and its largest value $v_{\max}$ satisfies $\Pr[f \le v_{\max}] = 1$. Let m be the least value v of f with $\Pr[f \le v] \ge \frac{1}{2}$. Then $\Pr[f \le m] \ge \frac{1}{2}$.

If m is the least value of f, then $\Pr[f \ge m] = 1$. Otherwise let v' be the largest value of f that is less than m. Then $f(x) < m$ if and only if $f(x) \le v'$, and $\Pr[f \le v'] < \frac{1}{2}$ by the choice of m. Hence $\Pr[f \ge m] = 1 - \Pr[f < m] > \frac{1}{2}$. □

The second fact will be applied to the function $z = f - m$.

Lemma 7.9 (Second and third moments)**.** *Let $n \ge 0$. For every function $z\colon Q_n \to \mathbb{R}$,*

$$\big(\mathbb{E}[z^2]\big)^3 \;\le\; \big(\mathbb{E}[|z|^3]\big)^2.$$

Proof. We use the Cauchy–Schwarz inequality of Section 5 twice, for families of numbers indexed by the points $x \in Q_n$: first for the numbers $|z(x)|^{1/2}$ and $|z(x)|^{3/2}$, and then for the numbers 1 and $|z(x)|$. After division by 4^n the two inequalities read

$$\big(\mathbb{E}[z^2]\big)^2 \;\le\; \mathbb{E}[|z|] \cdot \mathbb{E}[|z|^3], \qquad \big(\mathbb{E}[|z|]\big)^2 \;\le\; \mathbb{E}[z^2].$$

We square the first inequality and insert the second:

$$\big(\mathbb{E}[z^2]\big)^4 \;\le\; \big(\mathbb{E}[|z|]\big)^2 \big(\mathbb{E}[|z|^3]\big)^2 \;\le\; \mathbb{E}[z^2] \cdot \big(\mathbb{E}[|z|^3]\big)^2.$$

If $\mathbb{E}[z^2] > 0$, we divide by this number. If $\mathbb{E}[z^2] = 0$, there is nothing to prove. □

We can now prove the main result of this section, which is Theorem 1.3 of the introduction.

Theorem 7.10 (Mean absolute increments)**.** *Let $n \ge 0$, let $f\colon Q_n \to \mathbb{R}$ be a function with $\mathrm{Var}(f) > 0$, and let K be a real number such that for every $x \in Q_n$ there is a row a^x of f at x with $\|a^x\|_1 \le K$. Then $n \ge 1$ and $K > 0$, and there is a coordinate $j \in [n]$ with*

$$\mathrm{Var}(f)^3 \;\le\; 36\,K^4\,L_j(f)^2, \qquad \textit{that is,} \qquad \mathbb{E}_x|f(x) - f(x^j)| \;\ge\; \frac{\mathrm{Var}(f)^{3/2}}{6\,K^2}.$$

Proof. Step 0: preliminaries. Since $\mathrm{Var}(f) > 0$, the function f is not constant. So $n \ge 1$, and for every $x \in Q_n$ there is a point y with $f(y) \ne f(x)$, whence

$$0 < |f(x) - f(y)| \le a^x(\Delta(x,y)) \le \|a^x\|_1 \le K. \tag{7.12}$$

In particular $K > 0$.

Choose $j \in [n]$ such that $L_j(f)$ is as large as possible, and put $M = L_j(f)$. Then $M \ge 0$, and $L_k(f) \le M$ for every $k \in [n]$. Let m be a number with the properties of Lemma 7.8.

Step 1: the third moment about the median. The expectation in (7.7) is non-negative, and $\Pr[f \le m] \ge \frac{1}{2}$. So (7.7) with $r = m$ gives

$$\frac{1}{2}\,\mathbb{E}\big[((f-m)^+)^3\big] \;\le\; \mathbb{E}\big[((f-m)^+)^3\big] \cdot \Pr[f \le m] \;\le\; \frac{3}{2}\,K^2 M.$$

In the same way, (7.8) with $r = m$ and $\Pr[f \ge m] \ge \frac{1}{2}$ give $\frac{1}{2}\,\mathbb{E}[((m-f)^+)^3] \le \frac{3}{2}K^2M$.

We add the two inequalities, multiply by 2, and use the second relation in (7.2) with $\xi = f(x) - m$:

$$\mathbb{E}\big[|f-m|^3\big] \;=\; \mathbb{E}\big[((f-m)^+)^3\big] + \mathbb{E}\big[((m-f)^+)^3\big] \;\le\; 6\,K^2 M. \tag{7.13}$$

Step 2: the variance. By (2.3) with $c = m$ we have $\mathrm{Var}(f) \le \mathbb{E}[(f-m)^2]$. By Lemma 7.9 with $z = f - m$, and by (7.13),

$$\mathrm{Var}(f)^3 \overset{(2.3)}{\le} \big(\mathbb{E}[(f-m)^2]\big)^3 \overset{\text{Lem. 7.9}}{\le} \big(\mathbb{E}[|f-m|^3]\big)^2 \overset{(7.13)}{\le} 36\,K^4M^2.$$

Since $M = L_j(f)$, this is the first form of the assertion. The second form follows by taking square roots, because $K > 0$ and $L_j(f) \ge 0$. □

With the row bound (Theorem 6.8) we obtain a statement about bounded functions of low degree.

Corollary 7.11. *Let $n \ge 0$ and $d \ge 0$ be integers, and let $f\colon Q_n \to [-1,1]$ be a function of degree at most d with $\mathrm{Var}(f) > 0$. Then $n \ge 1$ and $d \ge 1$, and there is a coordinate $j \in [n]$ with*

$$\mathbb{E}_x|f(x) - f(x^j)| \;\ge\; \frac{\mathrm{Var}(f)^{3/2}}{6\,\pi^4\,d^4}.$$

Proof. A function of degree at most 0 is constant, so $d \ge 1$. By Theorem 6.8, the function f has a row of cost at most $\pi^2 d^2$ at every point. Theorem 7.10 with $K = \pi^2 d^2$ gives $n \ge 1$ and the assertion, because $K^2 = \pi^4 d^4$. □

Corollary 7.11 is a quantitative form of the L_1 version of the conjecture, the version in which Aaronson and Ambainis first stated it; see [AA14, Conjecture 6.1], where $\mathbb{E}|p - \mathbb{E}p|$ takes the place of the variance (note that $\mathrm{Var}(f) \ge (\mathbb{E}|f - \mathbb{E}f|)^2$).

They prove that the two versions are equivalent [AA14, Proposition 6.2]. One half of their argument is the inequality (2.4), which we use now.

Together with (2.4), Corollary 7.11 gives a coordinate j with

$$\mathrm{Inf}_j(f) \;\ge\; \frac{L_j(f)^2}{4} \;\ge\; \frac{\mathrm{Var}(f)^3}{144\,\pi^8\,d^8}.$$

Here $\mathrm{Var}(f) \le \mathbb{E}[f^2] \le 1$ by (2.3) with $c = 0$, and therefore

$$\mathrm{Inf}_j(f) \;\ge\; \frac{1}{144\,\pi^8}\Big(\frac{\mathrm{Var}(f)}{d}\Big)^8.$$

So a bound of the kind required in the conjecture of Section 1.1 already follows at this point. Section 8 improves the exponents 3 and 8 of the first of the two bounds to the exponents $\frac{5}{2}$ and 6 of Theorem 1.1.

Remark 7.12. (1) Every function $f\colon Q_n \to \mathbb{R}$ has a row at every point x. Indeed, let $c = \max_y |f(x) - f(y)|$. The vector $(c, \dots, c)$ is a row of f at x, because $|f(x) - f(x^S)| \le c \le c\,|S|$ for every non-empty set S, and $|f(x) - f(x^S)| = 0$ for $S = \emptyset$.

So the hypothesis on the rows in Proposition 7.4 and in Theorems 7.6 and 7.10 can be met for every f; what matters is the cost of the rows.

(2) The cost cannot be small compared with the oscillation of f: as in (7.12), every row a of f at x satisfies $\|a\|_1 \ge a(\Delta(x,y)) \ge |f(x) - f(y)|$ for every y. Hence a number K as in Theorem 7.10 is at least $\max f - \min f$.

(3) No bound for the values of f is assumed in Theorems 7.6 and 7.10, and their statements are compatible with scaling.

Let $c > 0$. Then $c\,a^x$ is a row of cf at x, and the passage from f, a^x and K to cf, $c\,a^x$ and cK multiplies $\mathrm{Var}(f)$ by c^2 and $L_j(f)$ by c. So both sides of the inequality $\mathrm{Var}(f)^3 \le 36K^4 L_j(f)^2$ are multiplied by c^6.

7.5. **Relation to the argument of Lovett and Zhang.** In the argument by which Lovett and Zhang derive the conjecture from their conjectured inequality [LZ, Claim 38], the function f has values in $[0,1]$, the number t is the mean of f, and the two sets are

$$X = \{x :\ f(x) \le t - \alpha\}, \qquad Y = \{x :\ f(x) \ge t + \alpha\},$$

where the number $\alpha > 0$, which is at most of the order of the variance of f, is chosen at the end. The argument has four steps, and they are related to this section as follows.

The first step is that a fractional certificate at a point of X bounds from below the L_∞ distance from this point to the convex hull of Y [LZ, Lemma 28]. Section 7.1 takes its place: a row of f at x, divided by $|f(x) - t|$, is a row of g_t at x if $f(x) \ne t$. This is used at every level t, and not only near the mean.

The second step is the conjectured inequality [LZ, Conjecture 37], which would give a coordinate that is influential for X or for Y; it is not true as stated (see Section 9.1). Here the all-pairs inequality is applied to g_t instead (Lemma 7.2). The two sets are now $\{f > t\}$ and $\{f \le t\}$, with no gap between them, and the inequality involves all the pairs of points, one from each set.

In [LZ] the sets X and Y are large because the variance of f is not small [LZ, Claim 30]. Here Theorem 7.6 is applied at a median m of f, so that $\Pr[f \le m]$ and $\Pr[f \ge m]$ are at least $\frac{1}{2}$ (Section 7.4). Corollary 7.7 is the form of Theorem 7.6 for two sets which are separated by a gap, as X and Y are.

The last step of [LZ] passes from the set to the function: α is chosen in such a way that few values of f lie just above the level $t - \alpha$, and for such an α the coordinate is influential for f. This step has no counterpart here. The inequality of Lemma 7.2 is integrated over all levels (Section 7.2), and the integral of its right-hand side is expressed by the increments of f itself. No level is chosen.

As in [LZ], the conclusion is a lower bound for the mean of $|f(x) - f(x^j)|$ for some coordinate j (Theorem 7.10).

8. Averaging and the proof of the main theorem

Theorem 7.10 gives a coordinate j for which the mean absolute increment $L_j(f)$ is large. The main theorem (Theorem 1.1) concerns the influences $\mathrm{Inf}_j(f)$, which are means of squares. In this section we pass from the one to the other.

The idea is the following. Call a coordinate k *light* if $\mathrm{Inf}_k(f)$ is small compared with $L_k(f)$, in the sense made precise in the proof of Theorem 8.3.

The influences of the light coordinates have a small sum, so the function h obtained from f by averaging over the light coordinates retains a fixed fraction of the variance of f.

Averaging increases neither the cost of the rows nor the mean absolute increments, and for the coordinates over which one has averaged the mean absolute increments vanish.

So Theorem 7.10, applied to h, produces a coordinate j that is not light and for which $L_j(f)$ is large. For such a coordinate $\mathrm{Inf}_j(f)$ is large (Theorem 8.3).

Together with the row bound (Theorem 6.8) and Corollary 6.10, this gives Theorem 1.1.

8.1. **Averaging over coordinates.** For a function $f\colon Q_n \to \mathbb{R}$ and a coordinate $j \in [n]$, the *average of f over the coordinate j* is the function $\mathrm{Av}_j f\colon Q_n \to \mathbb{R}$ with

$$(\mathrm{Av}_j f)(x) \;=\; \frac{f(x)+f(x^j)}{2} \qquad (x \in Q_n).$$

The following lemma lists the properties of this function that we need.

Lemma 8.1 (Averaging over one coordinate)**.** *Let $n \geq 1$, let $f\colon Q_n \to \mathbb{R}$ and $j \in [n]$, and put $h = \mathrm{Av}_j f$.*

(i) *$\mathbb{E}h = \mathbb{E}f$, and $\mathrm{Var}(f) = \mathrm{Var}(h) + \mathrm{Inf}_j(f)$.*
(ii) *$\mathrm{Inf}_k(h) \leq \mathrm{Inf}_k(f)$ and $L_k(h) \leq L_k(f)$ for every $k \in [n]$.*
(iii) *$L_j(h) = 0$.*
(iv) *Let K be a real number, and for every $x \in Q_n$ let a^x be a row of f at x with $\|a^x\|_1 \leq K$. Then, for every $x \in Q_n$, the vector $b^x = (a^x + a^{x^j})/2$ is a row of h at x, and $\|b^x\|_1 \leq K$.*

Proof. The map $x \mapsto x^j$ is a bijection of Q_n which preserves the measure; so $\mathbb{E}_x[\varphi(x^j)] = \mathbb{E}_x[\varphi(x)]$ for every function $\varphi\colon Q_n \to \mathbb{R}$. Moreover $(x^j)^j = x$, and therefore

$$h(x^j) = h(x) \qquad \text{for every } x \in Q_n. \tag{8.1}$$

(i) The first statement holds because $\mathbb{E}_x[f(x^j)] = \mathbb{E}f$.

For the second, let $\nu = \mathbb{E}f = \mathbb{E}h$. Fix $x \in Q_n$, and put $p = h(x) - \nu$ and $q = (f(x) - f(x^j))/2$. Then $f(x)-\nu = p+q$ and $f(x^j)-\nu = p-q$, and since $(p+q)^2+(p-q)^2 = 2p^2+2q^2$,

$$(f(x)-\nu)^2 + (f(x^j)-\nu)^2 \;=\; 2\,(h(x)-\nu)^2 + 2\Big(\frac{f(x)-f(x^j)}{2}\Big)^2.$$

Taking the expectation over x, we get $2\,\mathrm{Var}(f) = 2\,\mathrm{Var}(h) + 2\,\mathrm{Inf}_j(f)$.

(ii) Let $k \in [n]$. Fix $x \in Q_n$, and put

$$P = f(x) - f(x^k), \qquad Q = f(x^j) - f\big((x^j)^k\big).$$

Since $(x^k)^j = (x^j)^k$, we have $h(x) - h(x^k) = (P+Q)/2$. Hence, using $(P+Q)^2 \leq 2P^2 + 2Q^2$ for the second inequality,

$$|h(x)-h(x^k)| \;\leq\; \frac{|P|+|Q|}{2}, \qquad \Big(\frac{h(x)-h(x^k)}{2}\Big)^2 \;\leq\; \frac{1}{2}\Big(\frac{P}{2}\Big)^2 + \frac{1}{2}\Big(\frac{Q}{2}\Big)^2.$$

We take the expectation over x. The number Q is obtained from P by the substitution of x^j for x; so

$$\mathbb{E}_x|Q| = \mathbb{E}_x|P| = L_k(f) \qquad \text{and} \qquad \mathbb{E}_x\big[(Q/2)^2\big] = \mathbb{E}_x\big[(P/2)^2\big] = \mathrm{Inf}_k(f).$$

This gives $L_k(h) \leq L_k(f)$ and $\mathrm{Inf}_k(h) \leq \mathrm{Inf}_k(f)$.

(iii) This follows from (8.1).

(iv) The entries of b^x are non-negative, and $\|b^x\|_1 = (\|a^x\|_1 + \|a^{x^j}\|_1)/2 \leq K$.

Let $S \subseteq [n]$. Since $(x^S)^j = (x^j)^S$,

$$h(x) - h(x^S) \;=\; \frac{\big(f(x)-f(x^S)\big) + \big(f(x^j)-f((x^j)^S)\big)}{2}.$$

The vector a^x is a row of f at x, and a^{x^j} is a row of f at x^j. Therefore

$$|h(x) - h(x^S)| \;\le\; \frac{a^x(S) + a^{x^j}(S)}{2} \;=\; b^x(S). \qquad \square$$

For a set $R \subseteq [n]$, the *average of f over the coordinates in R* is the function $\mathrm{Av}_R f\colon Q_n \to \mathbb{R}$ with

$$(\mathrm{Av}_R f)(x) \;=\; 2^{-|R|} \sum_{T \subseteq R} f(x^T) \qquad (x \in Q_n).$$

Thus $\mathrm{Av}_\emptyset f = f$.

The averages over sets are obtained by averaging over one coordinate at a time:

$$\mathrm{Av}_{R\cup\{j\}} f \;=\; \mathrm{Av}_j\big(\mathrm{Av}_R f\big) \qquad \text{for } R \subseteq [n] \text{ and } j \in [n] \setminus R. \tag{8.2}$$

Indeed, the subsets of $R \cup \{j\}$ are the sets T and $T \cup \{j\}$ with $T \subseteq R$, and $x^{T\cup\{j\}} = (x^j)^T$ for these T. So

$$(\mathrm{Av}_{R\cup\{j\}} f)(x) \;=\; 2^{-|R|-1} \sum_{T\subseteq R} \Big(f(x^T) + f\big((x^j)^T\big)\Big) \;=\; \frac{(\mathrm{Av}_R f)(x) + (\mathrm{Av}_R f)(x^j)}{2}.$$

By (8.2), the properties of Lemma 8.1 pass to averages over sets. The next lemma states them in the form in which they are used in the proof of Theorem 8.3.

Lemma 8.2 (Averaging over a set of coordinates)**.** *Let $n \ge 0$, let $f\colon Q_n \to \mathbb{R}$, and let K be a real number such that for every $x \in Q_n$ there is a row of f at x of cost at most K. Let $R \subseteq [n]$, and put $h = \mathrm{Av}_R f$. Then:*

(i) *for every $x \in Q_n$ there is a row of h at x of cost at most K;*
(ii) $\mathrm{Var}(f) \le \mathrm{Var}(h) + \sum_{i\in R} \mathrm{Inf}_i(f)$*;*
(iii) $\mathrm{Inf}_k(h) \le \mathrm{Inf}_k(f)$ *and* $L_k(h) \le L_k(f)$ *for every $k \in [n]$;*
(iv) $L_i(h) = 0$ *for every $i \in R$.*

Proof. We use induction on $|R|$. If $R = \emptyset$, then $h = f$, and the four statements are clear.

Let $R = R' \cup \{j\}$ with $j \notin R'$, and suppose that the lemma holds for R' and the function $h' = \mathrm{Av}_{R'} f$. By (8.2), $h = \mathrm{Av}_j h'$. We apply Lemma 8.1 to the function h' and the coordinate j.

(i) By the statement (i) for R', the function h' has a row of cost at most K at every point. By Lemma 8.1(iv), so has h.

(ii) By Lemma 8.1(i) and by the statement (iii) for R',

$$\mathrm{Var}(h') \;=\; \mathrm{Var}(h) + \mathrm{Inf}_j(h') \;\le\; \mathrm{Var}(h) + \mathrm{Inf}_j(f).$$

Together with the statement (ii) for R' this gives

$$\mathrm{Var}(f) \;\le\; \mathrm{Var}(h') + \sum_{i\in R'} \mathrm{Inf}_i(f) \;\le\; \mathrm{Var}(h) + \sum_{i\in R} \mathrm{Inf}_i(f).$$

(iii) By Lemma 8.1(ii) and by the statement (iii) for R', we have $\mathrm{Inf}_k(h) \le \mathrm{Inf}_k(h') \le \mathrm{Inf}_k(f)$ and $L_k(h) \le L_k(h') \le L_k(f)$ for every k.

(iv) By Lemma 8.1(iii), $L_j(h) = 0$. For $i \in R'$, the statement (iv) for R' and Lemma 8.1(ii) give $0 \le L_i(h) \le L_i(h') = 0$. $\square$

8.2. **From mean absolute increments to influences.** The following theorem is Theorem 1.4 of the introduction. Like Theorem 7.10, it assumes no bound for the values of f.

Theorem 8.3 (Influences)**.** *Let $n \geq 0$, let $f\colon Q_n \to \mathbb{R}$ be a function with $\mathrm{Var}(f) > 0$, and let K be a real number such that for every $x \in Q_n$ there is a row a^x of f at x with $\|a^x\|_1 \leq K$. Let $L(f) = \sum_{k=1}^n L_k(f)$. Then $n \geq 1$ and $0 < L(f) \leq K$, and there is a coordinate $j \in [n]$ with*

$$3\,\mathrm{Var}(f)^5 \;\leq\; 3125\,L(f)^2\,K^4\,\mathrm{Inf}_j(f)^2, \qquad \textit{that is,} \qquad \mathrm{Inf}_j(f) \;\geq\; \frac{\sqrt{15}}{125}\cdot\frac{\mathrm{Var}(f)^{5/2}}{L(f)\,K^2}.$$

In particular, $\mathrm{Inf}_j(f) \geq (\sqrt{15}/125)\,\mathrm{Var}(f)^{5/2}/K^3$ for this coordinate j.

Proof. Step 0: the number $L(f)$. By Theorem 7.10, $n \geq 1$ and $K > 0$, and there is a coordinate $j_0 \in [n]$ with $\mathrm{Var}(f)^3 \leq 36K^4 L_{j_0}(f)^2$. Since $\mathrm{Var}(f) > 0$, this shows that $L_{j_0}(f) > 0$, and so $L(f) \geq L_{j_0}(f) > 0$.

For the upper bound let $x \in Q_n$. Definition 2.3 with $S = \{k\}$ gives $|f(x) - f(x^k)| \leq a^x_k$ for every $k \in [n]$, and therefore

$$\sum_{k=1}^n |f(x) - f(x^k)| \;\leq\; \|a^x\|_1 \;\leq\; K \qquad \text{for every } x \in Q_n.$$

Taking the expectation over x, we get $L(f) \leq K$.

Step 1: the light coordinates. We write $V = \mathrm{Var}(f)$ and $L = L(f)$, and we put

$$R \;=\; \big\{k \in [n] :\; 5\,L\,\mathrm{Inf}_k(f) < 2\,V\,L_k(f)\big\}.$$

All the numbers $L_k(f)$ are non-negative, and their sum is L. Hence

$$5\,L\sum_{k\in R}\mathrm{Inf}_k(f) \;\leq\; 2\,V\sum_{k\in R} L_k(f) \;\leq\; 2\,V\,L,$$

and since $L > 0$ it follows that

$$\sum_{k\in R}\mathrm{Inf}_k(f) \;\leq\; \frac{2}{5}\,V. \tag{8.3}$$

Step 2: averaging. Let $h = \mathrm{Av}_R f$. By Lemma 8.2(ii) and (8.3),

$$\mathrm{Var}(h) \;\geq\; V - \sum_{k\in R}\mathrm{Inf}_k(f) \;\geq\; \frac{3}{5}\,V \;>\; 0.$$

By Lemma 8.2(i), the function h has a row of cost at most K at every point. So Theorem 7.10 applies to h, with the same number K: there is a coordinate $j \in [n]$ with

$$\Big(\frac{3}{5}\,V\Big)^3 \;\leq\; \mathrm{Var}(h)^3 \;\leq\; 36\,K^4\,L_j(h)^2. \tag{8.4}$$

Step 3: the coordinate j is not light. By (8.4) we have $L_j(h) > 0$, and so $j \notin R$ by Lemma 8.2(iv). By the definition of R, this means that $2VL_j(f) \leq 5L\,\mathrm{Inf}_j(f)$. Since $L_j(h) \leq L_j(f)$ by Lemma 8.2(iii), we get

$$0 \;<\; 2\,V\,L_j(h) \;\leq\; 2\,V\,L_j(f) \;\leq\; 5\,L\,\mathrm{Inf}_j(f). \tag{8.5}$$

Step 4: conclusion. We multiply (8.4) by $(2V)^2$ and use (8.5):

$$\frac{108}{125}\,V^5 \;=\; \Big(\frac{3}{5}\,V\Big)^3 (2V)^2 \;\le\; 36\,K^4\,\big(2\,V\,L_j(h)\big)^2 \\ \le\; 36\,K^4\,\big(5\,L\,\mathrm{Inf}_j(f)\big)^2 \;=\; 900\,L^2K^4\,\mathrm{Inf}_j(f)^2.$$

Multiplying by $125/36$, we obtain $3V^5 \le 3125\,L^2K^4\,\mathrm{Inf}_j(f)^2$, which is the first form of the assertion.

The second form follows by taking square roots, because $LK^2 > 0$ and $\mathrm{Inf}_j(f) \ge 0$, and because $\sqrt{3/3125} = \sqrt{15}/125$. Since $L \le K$, the last statement of the theorem follows from the second form. □

8.3. **Proof of the main theorem.**

Proof of Theorem 1.1. Let $n \ge 1$ and $d \ge 1$, and let $f\colon Q_n \to [-1,1]$ be a function of degree at most d. If $\mathrm{Var}(f) = 0$, then every $j \in [n]$ has the required property, because $\mathrm{Inf}_j(f) \ge 0$; and $[n]$ is not empty.

Let $\mathrm{Var}(f) > 0$. By Theorem 6.8, for every $x \in Q_n$ there is a row a^x of f at x with $\|a^x\|_1 \le \pi^2 d^2$. By Corollary 6.10, $L(f) \le 4d^2$.

Theorem 8.3 with $K = \pi^2 d^2$ shows that $L(f) > 0$, and it gives a coordinate $j \in [n]$ with

$$\mathrm{Inf}_j(f) \;\ge\; \frac{\sqrt{15}}{125}\cdot\frac{\mathrm{Var}(f)^{5/2}}{L(f)\,(\pi^2d^2)^2} \;\ge\; \frac{\sqrt{15}}{125}\cdot\frac{\mathrm{Var}(f)^{5/2}}{4d^2\cdot\pi^4d^4} \;=\; \frac{\sqrt{15}}{500\,\pi^4}\cdot\frac{\mathrm{Var}(f)^{5/2}}{d^6}. \qquad \square$$

It remains to prove Corollary 1.2, the form of the theorem in the normalisation of [AA14], which was stated in Section 1.

Proof of Corollary 1.2 from Theorem 1.1. Since $\mathrm{Var}[p] > 0$, the polynomial p is not constant on $\{0,1\}^n$; hence $n \ge 1$ and $d \ge 1$. For $x \in Q_n$ put

$$f(x) \;:=\; 2\,p\Big(\frac{1-x_1}{2},\dots,\frac{1-x_n}{2}\Big) - 1.$$

The map $x \mapsto X = ((1-x_1)/2,\dots,(1-x_n)/2)$ is a bijection from Q_n onto $\{0,1\}^n$ which carries x^i to X^i and the uniform measure to the uniform measure. Hence $|f| \le 1$ on Q_n, $\mathrm{Var}(f) = 4\,\mathrm{Var}[p]$, and

$$\mathrm{Inf}_i(f) = \mathbb{E}_x\Big[\Big(\frac{f(x)-f(x^i)}{2}\Big)^2\Big] = \mathbb{E}_X\big[(p(X)-p(X^i))^2\big] = \mathrm{Inf}_i[p].$$

The function f has degree at most d: expanding p gives f as a real linear combination of products $\prod_i x_i^{k_i}$ with $\sum_i k_i \le d$, and on Q_n such a product equals χ_T with $T = \{i : k_i \text{ odd}\}$, a set of size at most d.

Theorem 1.1 therefore gives an i with

$$\mathrm{Inf}_i[p] = \mathrm{Inf}_i(f) \;\ge\; \frac{\sqrt{15}}{500\,\pi^4}\cdot\frac{(4\,\mathrm{Var}[p])^{5/2}}{d^6} = \frac{8\sqrt{15}}{125\,\pi^4}\cdot\frac{\mathrm{Var}[p]^{5/2}}{d^6},$$

because $4^{5/2} = 32$ and $32/500 = 8/125$.

For the second inequality we show that

$$\frac{8\sqrt{15}}{125\,\pi^4}\cdot\frac{v^{5/2}}{d^6} \;\ge\; \Big(\frac{v}{d}\Big)^7 \qquad \text{for every real number } v \text{ with } 0 < v \le \tfrac14. \tag{8.6}$$

Put $t = v/d$; then $0 < t \le \frac{1}{4}$, because $d \ge 1$. We have

$$\frac{v^{5/2}}{d^6} \;=\; \frac{t^6}{v^{7/2}} \;\ge\; 128\, t^6,$$

since $v^{7/2} \le 4^{-7/2} = \frac{1}{128}$.

Moreover $\sqrt{15} > \frac{7}{2}$, and $\pi^4 < 100$ because $\pi^2 < 10$. So the left-hand side of (8.6) is at least

$$\frac{8 \cdot \frac{7}{2}}{125 \cdot 100} \cdot 128\, t^6 \;=\; \frac{3584}{12500}\, t^6 \;\ge\; \frac{1}{4}\, t^6 \;\ge\; t^7.$$

Since $0 \le p \le 1$ on $\{0,1\}^n$ we have $p^2 \le p$ there, so $\mathrm{Var}[p] \le \mathbb{E}[p] - \mathbb{E}[p]^2 \le \frac{1}{4}$. Hence (8.6) applies to $v = \mathrm{Var}[p]$, and this is the second inequality of the corollary.

For the last sentence of the corollary let $\mathrm{Var}[p] \ge \varepsilon > 0$. Then $\mathrm{Var}[p]^{5/2} \ge \varepsilon^{5/2}$, which gives the first inequality, and $\varepsilon \le \frac{1}{4}$, so that (8.6) applies to $v = \varepsilon$ and gives the second. □

Remark 8.4 (The constants). The proof of Theorem 1.1 uses two properties of a function $f\colon Q_n \to [-1,1]$ of degree at most d: it has a row of cost at most $\pi^2 d^2$ at every point, and $\sum_j |f(x) - f(x^j)| \le 4d^2$ at every point x. For both properties better constants are known.

(1) By [ABD, Theorem 1.7], applied to the function $(1+f)/2$ (equivalently, by [ABD, Theorem 3.3] and Proposition 6.7), the function f has, at every point, a row of cost at most $\frac{\pi^2}{2} d^2$ (Remark 6.9).

If these rows are used in the proof of Theorem 1.1, that is, if Theorem 8.3 is applied with $K = \frac{\pi^2}{2} d^2$, the result is

$$\mathrm{Inf}_j(f) \;\ge\; \frac{\sqrt{15}}{125} \cdot \frac{\mathrm{Var}(f)^{5/2}}{4d^2\,(\pi^2 d^2/2)^2} \;=\; \frac{\sqrt{15}}{125\,\pi^4} \cdot \frac{\mathrm{Var}(f)^{5/2}}{d^6},$$

and the first bound of Corollary 1.2 becomes

$$\mathrm{Inf}_i[p] \;\ge\; \frac{32\sqrt{15}}{125\,\pi^4} \cdot \frac{\mathrm{Var}[p]^{5/2}}{d^6}.$$

Thus the constants of Theorem 1.1 and of Corollary 1.2 are multiplied by 4.

(2) The inequality of Filmus, Hatami, Keller and Lifshitz [FHKL, Theorem 3.3], which is recalled in Remark 6.9, gives $\sum_j |f(x) - f(x^j)| \le 2d^2$ at every point x, and hence $L(f) \le 2d^2$. If it is used in place of Corollary 6.10, the constants of Theorem 1.1, of Corollary 1.2 and of (1) are multiplied by a further factor 2.

Remark 8.5 (The exponents). The exponents arise as follows.

(1) In the integrated all-pairs inequality (7.6), the left-hand side is of the third order in the differences $f(x) - f(y)$, and the right-hand side is of the first order in the increments $f(x) - f(x^j)$.

Of the three factors on the left, two come from the rows of g_t at x and at y, which are the rows of f divided by $|f(x) - t|$ and by $|f(y) - t|$, and one comes from the length of the interval of the levels t between $f(y)$ and $f(x)$. The two rows also account for the factor K^2 in (7.7).

This leads to the inequality $\mathbb{E}|f - m|^3 \le 6K^2 \max_k L_k(f)$ of (7.13), and so to the exponents $\frac{3}{2}$ and 2 of Theorem 7.10.

(2) Theorem 7.10 and the inequality $L_j(f)^2 \le 4\,\mathrm{Inf}_j(f)$ of (2.4) give a coordinate j with $\mathrm{Var}(f)^3 \le 144\,K^4\,\mathrm{Inf}_j(f)$.

The proof of Theorem 8.3 does not use (2.4). The coordinate j found there is not light, that is,

$$\mathrm{Inf}_j(f) \;\geq\; \frac{2\,\mathrm{Var}(f)}{5\,L(f)} \cdot L_j(f).$$

The lower bound for $L_j(f)$ is of the order of $\mathrm{Var}(f)^{3/2}/K^2$. In the proof of Theorem 8.3 it is not squared; it is multiplied by a number of the order of $\mathrm{Var}(f)/L(f)$. The result is of the order of $\mathrm{Var}(f)^{5/2}/(L(f)K^2)$.

(3) The resulting bound is the stronger one. Indeed, $K \geq \max f - \min f$ by Remark 7.12(2).

Further $\max f - \min f \geq 2\,\mathrm{Var}(f)^{1/2}$: this follows from (2.3) with $c = (\max f + \min f)/2$, because $|f(x) - c| \leq (\max f - \min f)/2$ for every x. Hence

$$\frac{\sqrt{15}}{125} \cdot \frac{\mathrm{Var}(f)^{5/2}}{K^3} \;\geq\; \frac{2\sqrt{15}}{125} \cdot \frac{\mathrm{Var}(f)^3}{K^4} \;>\; \frac{\mathrm{Var}(f)^3}{144\,K^4}.$$

(4) For a function $f\colon Q_n \to [-1,1]$ of degree at most d, the number K can be taken of the order of d^2, and $L(f)$ is at most of this order. This gives the exponents $\frac{5}{2}$ and 6 of Theorem 1.1.

9. Remarks and open questions

This section contains a comparison of the two-sided inequality (Theorem 5.4) with results on certificates and decision trees (Section 9.1), examples which concern the exponent of the variance, the dependence on the degree and the constant of the row bound (Sections 9.2, 9.3 and 9.4), and open questions (Section 9.5).

9.1. **Certificates and decision trees.** Let $g\colon Q_n \to \{0,1\}$ be non-constant. A *certificate* of g at x is a set $C \subseteq [n]$ such that $g(y) = g(x)$ for every $y \in Q_n$ with $y_i = x_i$ for all $i \in C$.

These points y are the points x^S with $S \cap C = \emptyset$. So C is a certificate at x if and only if every sensitive block of g at x meets C, that is, by (2.5), if and only if the indicator vector of C is a row of g at x; the cost of this row is $|C|$.

Suppose that g has a certificate of size at most c_1 at every point of $\mathcal{O}$, and a certificate of size at most c_0 at every point of $\mathcal{N}$. For the corresponding rows the two averages in Theorem 5.4 are at least $\Pr[g=1]/c_1$ and $\Pr[g=0]/c_0$, and the theorem gives

$$\max_{j\in[n]} \mathrm{I}_j(g) \;\geq\; \frac{4\,\Pr[g=1]\,\Pr[g=0]}{c_0\,c_1}. \tag{9.1}$$

For the function of Example 2.5 one can take $c_1 = 2$ and $c_0 = 1$, and (9.1) is the bound $\frac{3}{8}$ found there.

We compare (9.1) with the theorem of O'Donnell, Saks, Schramm and Servedio [OSSS].

A *decision tree* is a rooted binary tree in which every inner node is labelled by a coordinate $i \in [n]$, its two outgoing edges are labelled by -1 and 1, every leaf is labelled by 0 or 1, and the labels of the inner nodes on each path from the root to a leaf are distinct.

On the input $x \in Q_n$ one starts at the root and, at an inner node labelled i, follows the edge labelled x_i, until a leaf is reached; the labels of the inner nodes on this path are the coordinates *queried* on x. The tree *computes* g if for every x the label of the leaf reached is $g(x)$.

For a decision tree T let $\delta_j(T)$ be the probability that the coordinate j is queried on a uniformly distributed input, and let $\Delta(T) = \sum_j \delta_j(T)$; this is the average number of coordinates queried.

The *depth* of T is the largest number of coordinates queried on an input, and $D(g)$ is the least depth of a decision tree that computes g.

In this notation, Theorem 1.1 of [OSSS], for the uniform measure, is the first of the inequalities

$$4 \Pr[g=1] \Pr[g=0] \;\le\; \sum_{j=1}^{n} \delta_j(T)\, \mathrm{I}_j(g) \;\le\; \Delta(T) \max_{j\in[n]} \mathrm{I}_j(g), \tag{9.2}$$

valid for every decision tree T that computes g.

(In [OSSS] the functions take the values ± 1; the function $2g-1$ has variance $4\Pr[g=1]\Pr[g=0]$, and its influences in the sense of [OSSS] are the numbers $\mathrm{I}_j(g)$.)

Since $\Delta(T)$ is at most the depth of T, it follows that

$$\max_{j\in[n]} \mathrm{I}_j(g) \;\ge\; \frac{4 \Pr[g=1] \Pr[g=0]}{D(g)}, \tag{9.3}$$

which is the statement quoted in Section 1.4.

The coordinates queried on an input x form a certificate of g at x, because every y that agrees with x in these coordinates follows the same path. Hence one can take $c_0 = c_1 = D(g)$ in (9.1); the resulting bound is weaker than (9.3) by the factor $D(g)$.

In the other direction, $D(g) \le c_0 c_1$ by a classical inequality, found independently by Blum and Impagliazzo [BI], by Hartmanis and Hemachandra [HH] and by Tardos [Tar] (see [Smy02, p. 221] and [KT, Section 6]), so that (9.1) follows from (9.3) as well.

The bounds (9.1) and (9.3) involve the largest size of a certificate and the largest number of queried coordinates, whereas Theorem 5.4 and (9.2) involve averages. This can make a large difference.

Let $n \ge 2$, and let $g(x) = 1$ if $x_1 = 1$ or $x_2 = \cdots = x_n = 1$, and $g(x) = 0$ otherwise; then $\Pr[g=0] = \frac{1}{2} - 2^{-n}$.

Let $u = (-1, 1, \dots, 1)$. Every certificate of g at u contains $\{2, \dots, n\}$, because $g(u) = 1$ and $g(u^i) = 0$ for $i \ge 2$. Hence $c_1 \ge n-1$ in (9.1), and $D(g) \ge n-1$, because the coordinates queried on u form a certificate.

As $\Pr[g=1]\Pr[g=0] \le \frac{1}{4}$, and $c_0 \ge 1$ because a certificate of a non-constant function is not empty, the right-hand sides of (9.1) and (9.3) are at most $1/(n-1)$.

On the other hand, at a point x with $x_1 = 1$ the set $\{1\}$ is a certificate, and at a point x with $x_1 = -1$ and $x_i = -1$, where $i \ge 2$, the set $\{1, i\}$ is a certificate. With the corresponding rows, and any row at u, the left-hand side of the inequality of Theorem 5.4 is at least

$$4 \cdot \frac{1}{2} \cdot \frac{1}{2} \Big(\frac{1}{2} - 2^{-n} \Big) \;=\; \frac{1}{2} - 2^{-n}.$$

The inequality (9.2) gives a bound of the same order for g. Indeed, let T be the decision tree which queries the coordinate 1, stops with the value 1 if $x_1 = 1$, and otherwise queries the coordinates $2, 3, \dots, n$ in this order, stopping with the value 0 at the first i with $x_i = -1$, and with the value 1 if there is no such i.

It computes g, and $\delta_1(T) = 1$ and $\delta_j(T) = 2^{1-j}$ for $j \geq 2$. So $\Delta(T) = 2 - 2^{1-n}$, and (9.2) gives

$$\max_j \mathrm{I}_j(g) \;\geq\; \frac{1 - 4^{1-n}}{2 - 2^{1-n}}.$$

(In fact $\max_j \mathrm{I}_j(g) = \mathrm{I}_1(g) = 1 - 2^{1-n}$.)

Averages of reciprocal costs, in place of the largest cost, are also essential in Section 7. There the vectors $a^x/(f(x)-r)^3$ are inserted into Lemma 5.3, from which Theorem 5.4 was deduced.

These vectors have arbitrary non-negative real entries, their costs $\|a^x\|_1/(f(x)-r)^3$ admit no bound in terms of K (they are large at the points x where $f(x)$ is close to r), and the conclusion of the lemma involves only the sum of the reciprocals of these costs.

Neither of the two averages in Theorem 5.4 can be replaced by the corresponding probability.

Let $w \geq 1$ be an integer, let $m = 2^w$ and $n = wm$, let $T_1, \ldots, T_m$ be pairwise disjoint subsets of $[n]$ with w elements each, and let $g(x) = 1$ if there is a k such that $x_i = 1$ for all $i \in T_k$, and $g(x) = 0$ otherwise; functions of this kind are the tribes functions of Ben-Or and Linial [BL] (see also [OSSS]).

Then $\Pr[g = 0] = (1 - 1/m)^m$. By Bernoulli's inequality, $(1 - 1/m)^{m/2} \geq \frac{1}{2}$, and $(1-1/m)^m \leq e^{-1}$ because $1 - s \leq e^{-s}$ for real s; hence

$$\frac{1}{4} \;\leq\; \Pr[g = 0] \;<\; \frac{1}{2}.$$

For a coordinate $j \in T_k$ we have $g(x) \neq g(x^j)$ if and only if $x_i = 1$ for all $i \in T_k \setminus \{j\}$ and there is no $l \neq k$ such that $x_i = 1$ for all $i \in T_l$; so

$$\mathrm{I}_j(g) \;=\; 2^{1-w}(1 - 1/m)^{m-1} \;\leq\; 2^{1-w}.$$

At a point x with $g(x) = 1$, a set T_k such that $x_i = 1$ for all $i \in T_k$ is a certificate, and its indicator vector is a row of cost w. With these rows at the points of $\mathcal{O}$,

$$4\,\mathbb{E}_x\Big[\frac{1_{\{g(x)=1\}}}{\|\lambda^x\|_1}\Big]\,\Pr[g = 0] \;=\; \frac{4}{w}\,\Pr[g = 1]\,\Pr[g = 0] \;\geq\; \frac{1}{2w} \;\geq\; \frac{2^w}{4w}\max_{j\in[n]} \mathrm{I}_j(g),$$

and $2^w/(4w)$ is unbounded.

So the inequality of Theorem 5.4 does not remain true if its second average is replaced by $\Pr[g = 0]$. The function $1 - g$ has the same rows and the same numbers I_j as g, so the same holds for the first average and $\Pr[g = 1]$.

This is consistent with Theorem 5.4: at a point x with $g(x) = 0$ the m sets $\{i \in T_k : \; x_i = -1\}$ are pairwise disjoint sensitive blocks of g at x, so every row of g at x has cost at least $m = 2^w$.

The same function shows that the conjectured inequality of Lovett and Zhang [LZ, Conjecture 37], which is recalled in Section 1.4, is not true as stated. In the coordinates $(1 + x_i)/2$ used in [LZ], let $X = g^{-1}(1)$ and $Y = g^{-1}(0)$. If T_k consists of ones at a point $x \in X$, then every $y \in Y$ has a zero in T_k, so the mean of $1 - z_i$ over $i \in T_k$ is at least $1/w$ for every z in the convex hull of Y; and the average of the w points of Y obtained from x by setting the jth coordinate of every block of ones to 0 ($j = 1, \ldots, w$) is at L_∞ distance $1/w$ from x. So every point of X has L_∞ distance exactly $1/w$ from the convex hull of Y. On the other hand, the influences of [LZ, Definition 36] are $\Pr_{x\in X}[x^i \notin X] = \mathrm{I}_i(g)/(2\Pr[g = 1]) \leq 2^{-w}$ and

$\Pr_{y\in Y}[y^i \notin Y] = \mathrm{I}_i(g)/(2\Pr[g=0]) = 1/(2^w-1)$, which tend to 0 exponentially in w. With X and Y exchanged the distance is 2^{-w}, so for this pair of sets the failure is one-sided. The example concerns the conjectured inequality for arbitrary sets, not the level sets of functions of low degree to which [LZ, Claim 38] applies it.

9.2. **The exponent of the variance.** Let $n \geq 1$ and let $f = \hat{f}(\emptyset) + \sum_{i=1}^n \hat{f}(\{i\})\, x_i$ be a function of degree at most 1 on Q_n with $|f| \leq 1$. Then

$$f(x) - f(x^j) = 2\hat{f}(\{j\})\, x_j, \qquad \text{so that} \qquad \mathrm{Inf}_j(f) = \hat{f}(\{j\})^2,$$

and $\mathrm{Var}(f) = \sum_{i=1}^n \hat{f}(\{i\})^2$ by the orthonormality of the functions χ_T.

Let x be the point with $x_i = 1$ if $\hat{f}(\{i\}) \geq 0$ and $x_i = -1$ otherwise. Then

$$f(x) = \hat{f}(\emptyset) + \sum_i |\hat{f}(\{i\})| \leq 1 \qquad \text{and} \qquad f(\bar{x}) = \hat{f}(\emptyset) - \sum_i |\hat{f}(\{i\})| \geq -1,$$

whence $\sum_i |\hat{f}(\{i\})| \leq 1$.

Therefore

$$\mathrm{Var}(f) = \sum_{i=1}^n \hat{f}(\{i\})^2 \;\leq\; \max_j |\hat{f}(\{j\})| \cdot \sum_{i=1}^n |\hat{f}(\{i\})| \;\leq\; \max_j |\hat{f}(\{j\})|,$$

that is,

$$\max_{j\in[n]} \mathrm{Inf}_j(f) \;\geq\; \mathrm{Var}(f)^2.$$

For $f = (x_1 + \cdots + x_n)/n$ there is equality: $\mathrm{Var}(f) = 1/n$ and $\mathrm{Inf}_j(f) = 1/n^2$ for every j.

Consequently, if $c > 0$, α and β are real numbers such that $\max_j \mathrm{Inf}_j(f) \geq c\,\mathrm{Var}(f)^\alpha / d^\beta$ for all $n, d \geq 1$ and all $f\colon Q_n \to [-1,1]$ of degree at most d, then $\alpha \geq 2$, because the case $d = 1$ gives $n^{-2} \geq c\, n^{-\alpha}$ for all n.

The same function shows what Theorems 7.10 and 8.3 lose. Let $f = (x_1 + \cdots + x_n)/n$. The vector $(2/n, \ldots, 2/n)$ is a row of f of cost 2 at every point, since $|f(x) - f(x^S)| \leq 2|S|/n$.

These rows have the least possible cost, because every row a of f at x satisfies $a_i \geq |f(x) - f(x^i)| = 2/n$ for all i. Moreover $L_j(f) = 2/n$ for every j, so that $L(f) = 2$.

With $K = 2$, Theorem 7.10 yields a coordinate j with $L_j(f) \geq n^{-3/2}/24$, whereas $L_j(f) = 2/n$. Theorem 8.3 yields a coordinate j with $\mathrm{Inf}_j(f) \geq (\sqrt{15}/1000)\, n^{-5/2}$, whereas $\mathrm{Inf}_j(f) = n^{-2}$.

In both cases the bound is smaller than the true value by a factor of the order of $n^{1/2}$. Since the rows have the least possible cost, none of this loss is due to the rows: it occurs in the passage from the rows to the increments and to the influences.

9.3. **The dependence on the degree.** There is no lower bound for $\max_j \mathrm{Inf}_j(f)$ in terms of $\mathrm{Var}(f)$ alone.

Let $n = 2k+1$, and let $f(x) \in \{-1, 1\}$ be the sign of $x_1 + \cdots + x_n$; this sum is an odd integer, hence not zero. Then $f(\bar{x}) = -f(x)$, so that $\mathbb{E}f = 0$ and $\mathrm{Var}(f) = \mathbb{E}[f^2] = 1$.

Fix j. We have $f(x) \neq f(x^j)$ if and only if exactly k of the $2k$ coordinates x_i with $i \neq j$ are equal to 1, and in this case $\big((f(x) - f(x^j))/2\big)^2 = 1$. Hence

$$\mathrm{Inf}_j(f) \;=\; p_k := \binom{2k}{k} 2^{-2k} \;\leq\; \frac{1}{\sqrt{2k+1}} \;=\; \frac{1}{\sqrt{n}} \qquad (j \in [n]).$$

The inequality holds for $k = 0$, and it passes from k to $k+1$, because $p_{k+1} = p_k \cdot \frac{2k+1}{2k+2}$ and $(2k+1)(2k+3) \le (2k+2)^2$.

So the influences tend to 0 as $n \to \infty$, while the variance is equal to 1.

9.4. **The constant in the row bound.** Let $d \ge 1$ and $n \ge 1$, and let $f(x) = T_d\big((x_1 + \cdots + x_n)/n\big)$ for $x \in Q_n$. These are the functions with which Filmus, Hatami, Keller and Lifshitz show that their inequality $\sum_i |f(x) - f(x^i)|/2 \le d^2$, which is recalled in Remark 6.9, is sharp [FHKL, Section 3].

By (C3), $|f| \le 1$. By (C1), f is a linear combination of the powers s^k, $0 \le k \le d$, of the function $s = (x_1 + \cdots + x_n)/n$, which has degree at most 1; so f has degree at most d.

Let a be a row of f at the point $\mathbf{1} = (1, \ldots, 1) \in Q_n$. For every $i \in [n]$,

$$a_i \;\ge\; |f(\mathbf{1}) - f(\mathbf{1}^i)| \;=\; |T_d(1) - T_d(1 - 2/n)| \;=\; 1 - T_d(1 - 2/n),$$

by (C3). Hence $\|a\|_1 \ge n\big(1 - T_d(1 - 2/n)\big)$, and the right-hand side tends to $2\,T_d'(1) = 2d^2$ as $n \to \infty$, by (C4).

Let us call a real number c *admissible* if for all $n \ge 0$ and $d \ge 0$, every function $Q_n \to [-1, 1]$ of degree at most d has, at every point, a row of cost at most $c\,d^2$.

The example shows that every admissible number is at least 2. The number π^2 is admissible by Theorem 6.8, and $\pi^2/2$ is admissible by Remark 6.9.

9.5. **Open questions.** *The exponents.* Theorem 1.1 is a bound of the form $c\,\mathrm{Var}(f)^\alpha/d^\beta$ with $\alpha = \frac{5}{2}$ and $\beta = 6$. By Section 9.2 the exponent α in such a bound is at least 2, and by Section 9.3 the bound has to depend on d.

We do not know the best exponents. In particular, we do not know whether there is a bound of this form with $\alpha = 2$.

The corresponding question for functions with rows is the following. Is there a constant $c > 0$ such that every non-constant function $f\colon Q_n \to \mathbb{R}$ which has a row of cost at most K at every point has a coordinate j with $\mathrm{Inf}_j(f) \ge c\,\mathrm{Var}(f)^2/K^2$?

By Theorem 6.8, a positive answer would give a bound of the above form with $\alpha = 2$ and $\beta = 4$.

For Boolean functions the answer is positive. Indeed, let $g\colon Q_n \to \{0, 1\}$ be non-constant, with a row of cost at most K at every point. The two averages in Theorem 5.4 are at least $\Pr[g = 1]/K$ and $\Pr[g = 0]/K$. Since $\mathrm{I}_j(g) = 4\,\mathrm{Inf}_j(g)$ and $\mathrm{Var}(g) = \Pr[g = 1]\Pr[g = 0]$, the theorem gives a coordinate j with $\mathrm{Inf}_j(g) \ge \mathrm{Var}(g)/K^2$, and $\mathrm{Var}(g) \ge 4\,\mathrm{Var}(g)^2$ because $\mathrm{Var}(g) \le \frac{1}{4}$.

The bijection. The proof of Theorem 3.1 obtains the bijection τ from a term of a determinant that does not vanish. It would be interesting to have a direct combinatorial construction of a bijection with the properties (a) and (b).

The row bound. Let c^* be the infimum of the admissible numbers of Section 9.4, so that $2 \le c^* \le \pi^2/2$. What is the value of c^*?

Appendix A. Timeline of the discovery

In the run that produced the proof, before any statement or method in the table below was first proposed, various instances of the model brought in from the literature Reimer's inequality and his Butterfly Lemma [Rei], and the dual inequality of Kahn, Saks and Smyth [KSS], and set as a goal a version of these inequalities for fractional certificates. The table gives, for the main statements of the paper, the order in which the statement or the method was first proposed and the order in which a proof can best be identified, retrospectively, as first written down; proofs written down together share a number.

	order proposed	order proved
Two-sided inequality (Theorem 1.5)	1	3
Integration over all levels (Section 7)	2	6
Antipodal inequality (Theorem 4.3)	3	1
All-pairs inequality (Proposition 5.2)	4	3
Determinant inequality (Corollary 3.2)	5	1
Bijection theorem (Theorem 3.1)	6	3
Averaging over coordinates (Section 8)	7	7

The first complete proof was written down during the third day of the run, together with the proofs of the two-sided inequality, the all-pairs inequality and the bijection theorem, and then checked in the Lean proof assistant. It gave the bound $\mathrm{Var}(f)^8/(2^{23}\pi^{12}d^{12})$ in place of the bound of Theorem 1.1, that is, the exponent 22 in place of 7 in Corollary 1.2. After the instances of the model were asked to push on open problems using the intermediate results of the proof, they also improved the exponent of 22 by polishing these intermediate results: the integration over all levels gave the exponent 11, checked in Lean, and then the averaging over coordinates, together with the bound $L(f) \leq 4d^2$ of Corollary 6.10, gave the exponent 7, checked in Lean.

References

[AA14] S. Aaronson and A. Ambainis, *The need for structure in quantum speedups*, Theory Comput. **10** (2014), 133–166.

[Aar] S. Aaronson, *Quantum certificate complexity*, J. Comput. System Sci. **74** (2008), no. 3, 313–322.

[Aar08] S. Aaronson, *How to solve longstanding open problems in quantum computing using only Fourier analysis*, lecture at the workshop Analytic Tools in Computational Complexity, Banff International Research Station, 2008, `http://www.scottaaronson.com/talks/openqc.ppt`.

[ABD] A. Agarwal and S. Ben-David, *Oracle separations for the quantum-classical polynomial hierarchy*, preprint, arXiv:2410.19062v2, 2025; conference version in: 17th Innovations in Theoretical Computer Science Conference (ITCS 2026), Leibniz International Proceedings in Informatics (LIPIcs), vol. 362, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026, Art. 2. Statements are cited here with their numbers in the preprint.

[ABK] A. Anshu, S. Ben-David and S. Kundu, *On query-to-communication lifting for adversary bounds*, in: 36th Computational Complexity Conference (CCC 2021), Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, doi:10.4230/LIPIcs.CCC.2021.30.

[BB] A. Bačkurs and M. Bavarian, *On the sum of L_1 influences*, in: Proceedings of the 29th IEEE Conference on Computational Complexity (CCC 2014), IEEE Computer Society, 2014, 132–143, doi:10.1109/CCC.2014.21.

[BBCMW] R. Beals, H. Buhrman, R. Cleve, M. Mosca and R. de Wolf, *Quantum lower bounds by polynomials*, J. ACM **48** (2001), no. 4, 778–797.

[BCR] C. Borgs, J. T. Chayes and D. Randall, *The van den Berg–Kesten–Reimer inequality: a review*, in: Perplexing Problems in Probability, Progress in Probability, vol. 44, Birkhäuser, Boston, 1999, 159–173. Statements and sections are cited here with their numbers in the technical report MSR-TR-98-42, Microsoft Research, 1998.

[BDST] G. Blanc, J. Docter, C. Strassle and L.-Y. Tan, *Quantum speedups require structure or depth*, preprint, arXiv:2608.19158, 2026; conference version to appear in: 67th Annual Symposium on Foundations of Computer Science (FOCS 2026). Statements are cited here with their numbers in the preprint.

[BE] P. Borwein and T. Erdélyi, *Polynomials and Polynomial Inequalities*, Graduate Texts in Mathematics, vol. 161, Springer-Verlag, New York, 1995.

[Bha] S. K. Bhattacharya, *Aaronson–Ambainis conjecture is true for random restrictions*, preprint, arXiv:2402.13952, 2024; conference version: *Random restrictions of bounded low degree polynomials are juntas*, in: 16th Innovations in Theoretical Computer Science Conference (ITCS 2025), Leibniz International Proceedings in Informatics (LIPIcs), vol. 325, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, Art. 17. Statements are cited here with their numbers in the preprint arXiv:2402.13952v2.

[BI] M. Blum and R. Impagliazzo, *Generic oracles and oracle classes*, in: 28th Annual Symposium on Foundations of Computer Science (FOCS 1987), IEEE, 1987, 118–126.

[BL] M. Ben-Or and N. Linial, *Collective coin flipping, robust voting schemes and minima of Banzhaf values*, in: 26th Annual Symposium on Foundations of Computer Science (FOCS 1985), IEEE, 1985, 408–416.

[BSW] N. Bansal, M. Sinha and R. de Wolf, *Influence in completely bounded block-multilinear forms and classical simulation of quantum algorithms*, in: Computational Complexity Conference (CCC 2022), Leibniz International Proceedings in Informatics (LIPIcs), vol. 234, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, Art. 28.

[DFKO] I. Dinur, E. Friedgut, G. Kindler and R. O'Donnell, *On the Fourier tails of bounded functions over the discrete cube*, Israel J. Math. **160** (2007), 389–412.

[DMP] A. Defant, M. Mastyło and A. Pérez, *On the Fourier spectrum of functions on Boolean cubes*, Math. Ann. **374** (2019), no. 1–2, 653–680. The bound is quoted here from Section 4.1 of the preprint arXiv:1706.03670v1.

[EG] F. Escudero Gutiérrez, *Influences of Fourier completely bounded polynomials and classical simulation of quantum algorithms*, Chicago J. Theoret. Comput. Sci. **2024**, Art. 2.

[FHKL] Y. Filmus, H. Hatami, N. Keller and N. Lifshitz, *On the sum of the L_1 influences of bounded functions*, Israel J. Math. **214** (2016), no. 1, 167–192. Statements and sections are cited here with their numbers in the preprint arXiv:1404.3396v3.

[GMNR] U. Girish, K. Mittal, B. Nehoran and R. Raz, *Distributional variants of the Aaronson–Ambainis conjecture*, preprint, arXiv:2609.35327v1, 2026.

[GSS] J. Gilmer, M. Saks and S. Srinivasan, *Composition limits and separating examples for some Boolean function complexity measures*, Combinatorica **36** (2016), no. 3, 265–311. The appendix is cited here as in the preprint arXiv:1306.0630v1.

[Hall] P. Hall, *On representatives of subsets*, J. London Math. Soc. **10** (1935), 26–30.

[HH] J. Hartmanis and L. A. Hemachandra, *One-way functions, robustness, and the non-isomorphism of NP-complete sets*, in: Proceedings of the 2nd Annual Structure in Complexity Theory Conference, IEEE Computer Society, 1987, 160–174.

[HKY] W. Hochstättler, S. Keip and B. Yazici, *Exploring the Varchenko determinant of partial cubes*, preprint, arXiv:2505.14800, 2025.

[Iva] P. Ivanisvili, *Polynomial growth of Bohnenblust–Hille constants on the Hamming cube*, preprint, arXiv:2609.12427, 2026.

[KK19] N. Keller and O. Klein, *Quantum speedups need structure*, preprint, arXiv:1911.03748 (2019); withdrawn.

[KSS] J. Kahn, M. Saks and C. Smyth, *The dual BKR inequality and Rudich's conjecture*, Combin. Probab. Comput. **20** (2011), no. 2, 257–266; preliminary version: *A dual version of Reimer's inequality and a proof of Rudich's conjecture*, in: Proceedings of the 15th Annual IEEE Conference on Computational Complexity, IEEE Computer Society, 2000, 98–103.

[KT] R. Kulkarni and A. Tal, *On fractional block sensitivity*, Chicago J. Theoret. Comput. Sci. **2016**, Art. 8; Electronic Colloquium on Computational Complexity, Report TR13-168, 2013. Statements and sections are cited here with their numbers in the report.

[Lang] S. Lang, *Algebra*, revised third edition, Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002.

[LM] Q. Liu and S. Mutreja, *Parallel quantum advantage with limited adaptivity requires structure*, preprint, arXiv:2608.20297v1, 2026.

[LZ] S. Lovett and J. Zhang, *Fractional certificates for bounded functions*, in: 14th Innovations in Theoretical Computer Science Conference (ITCS 2023), Leibniz International Proceedings in Informatics (LIPIcs), vol. 251, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 84:1–84:13.

[Mon] A. Montanaro, *Some applications of hypercontractive inequalities in quantum information theory*, J. Math. Phys. **53** (2012), no. 12, 122206.

[NS] N. Nisan and M. Szegedy, *On the degree of Boolean functions as real polynomials*, Comput. Complexity **4** (1994), no. 4, 301–313.

[O'D] R. O'Donnell, *Analysis of Boolean Functions*, Cambridge University Press, New York, 2014.

[OSSS] R. O'Donnell, M. Saks, O. Schramm and R. A. Servedio, *Every decision tree has an influential variable*, in: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), IEEE, 2005, 31–39.

[OZ] R. O'Donnell and Y. Zhao, *Polynomial bounds for decoupling, with applications*, in: Conference on Computational Complexity (CCC 2016), Leibniz International Proceedings in Informatics (LIPIcs), vol. 50, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016, Art. 24.

[Rei] D. Reimer, *Proof of the van den Berg–Kesten conjecture*, Combin. Probab. Comput. **9** (2000), no. 1, 27–32.

[Sho] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput. **26** (1997), no. 5, 1484–1509.

[Sim] D. R. Simon, *On the power of quantum computation*, SIAM J. Comput. **26** (1997), no. 5, 1474–1483.

[Smy] C. D. Smyth, *The conjectures of Rudich, Tardos, and Kusner*, Ph.D. dissertation, Rutgers University, New Brunswick, New Jersey, 2001.

[Smy02] C. D. Smyth, *Reimer's inequality and Tardos' conjecture*, in: Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC 2002), ACM, 2002, 218–221.

[STV] J. Slote, C.-K. Tseng and A. Volberg, *An $m^{2.943}$ Bohnenblust–Hille bound on the Boolean cube*, preprint, arXiv:2609.21144, 2026.

[Tal] A. Tal, *Properties and applications of Boolean function composition*, in: Proceedings of the 4th Conference on Innovations in Theoretical Computer Science (ITCS 2013), ACM, 2013, 441–454.

[Tar] G. Tardos, *Query complexity, or why is it difficult to separate* $\mathrm{NP}^A \cap \mathrm{coNP}^A$ *from* P^A *by random oracles A?*, Combinatorica **9** (1989), no. 4, 385–392.

[Var] A. Varchenko, *Bilinear form of real configuration of hyperplanes*, Adv. Math. **97** (1993), no. 1, 110–144.