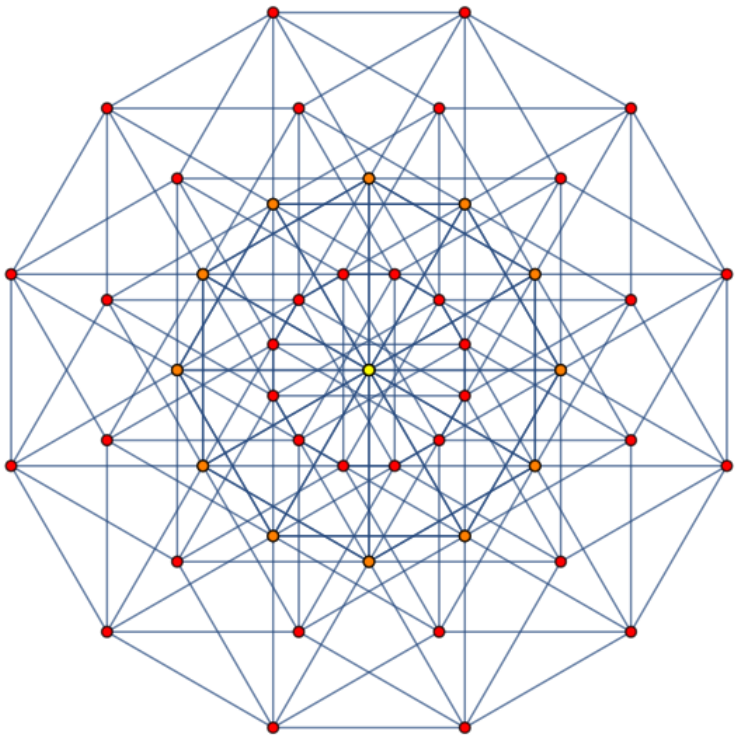
Discrete Analysis
&
Complexity of Quantum Algorithms

Proceedings of the (online) Summer/Fall School

October 11–15 2021



Organizers

Alexandros Eskenazis, University of Cambridge

Paata Ivanisvili, University of California, Irvine

Contents

Cover picture of $\{-1,1\}^6$ courtesy of Wikipedia.

Chapter 1

Every decision tree has an influential variable

after R. O'Donnell, M. Saks, O. Schramm, R.A. Servedio [5]
A summary written by Antonio Ismael Cano Mármol

Abstract. We outline the proof of the OSSS inequality. As an application we give a random decision tree complexity lower bound. Moreover, we introduce an inductive proof [4] of the main inequality.

1.1 Introduction and comparison with previous work

For some $p \in (0,1)$, let $\{-1,1\}^n_{(p)}$ denote the discrete cube endowed with the $p-$biased product measure

$$\mu_{(p)}(x) = p^{|\{i \,:\, x_i=1\}|}\ (1-p)^{|\{i \,:\, x_i=-1\}|}.$$

We will write $\{-1,1\}^n$ instead when referring to the case $p = 1/2$.

Any boolean function $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$ has an associated *influence vector* $(\mathbf{Inf}_1(f), \dots, \mathbf{Inf}_n(f))$, where $\mathbf{Inf}_i(f)$ measures to what extent the value of f depends on variable i, or formally,

$$\begin{aligned}\mathbf{Inf}_i(f) &= \mathop{\mathbf{Pr}}_{x\in\{-1,1\}^n_{(p)}}[f(x) \neq f(x_{[n]\setminus\{i\}} - x_i e_i)] \\ &= 2 \mathop{\mathbf{Pr}}_{\substack{x\in\{-1,1\}^n_{(p)} \\ z_i\in\{-1,1\}}}[f(x) \neq f(x_{[n]\setminus\{i\}} + z_i e_i)].\end{aligned}$$

Here, we denote $x_S = \sum_{i\in S} x_i e_i$ for any $S \subseteq [n] = \{1,\dots,n\}$. The concept of variable influence was introduced by Ben-Or and Linial [1] in a paper in which they made the observation that any balanced function $f : \{-1,1\}^n \to \{-1,1\}$ (i.e. $\mathbf{E}[f] = 0$) satisfies $\mathbf{Inf}_{\max}(f) := \max_{i\in[n]} \mathbf{Inf}_i(f) \geq \frac{1}{n}$. Indeed, this fact follows from the *Efron-Stein inequality*,

$$\mathbf{Var}[f] \leq \sum_{i=1}^{n} \mathbf{Inf}_i(f), \tag{1.1}$$

which relates the influences to the variance $\mathbf{Var}[f] = \mathbf{E}[f^2] - \mathbf{E}[f]^2$. Later, Kahn, Kalai and Linial [3] confirmed a conjecture from [1] by proving that for any *near-balanced* function (i.e. if $|f^{-1}(-1)|/2^n$ and $|f^{-1}(1)|/2^n$ are $\Omega(1)$) it holds

$$\mathbf{Inf}_{\max}(f) \geq \Omega\Big(\frac{\log(n)}{n}\Big). \tag{1.2}$$

This inequality motivates asking about a lower bound in terms of complexity of boolean functions, in particular, *decision tree complexity*.

A *deterministic decision tree (DDT)* for a boolean function $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$ is a 'deterministic adaptive strategy for reading variables so as to determine the value of f'. Given a function $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$ we define the *DDT complexity of* f as

$$D(f) = \min_{T \text{ DDT for } f} \max_{x \in \{-1,1\}^n_{(p)}} [\#\text{ coordinates queried by } T \text{ on } x].$$

Moreover, we define

$$\delta_i(T) = \mathop{\mathbf{Pr}}_{x \in \{-1,1\}^n_{(p)}} [T \text{ queries } x] \quad \text{and}$$

$$\Delta(T) = \sum_{i=1}^{n} \delta_i(T) = \mathop{\mathbf{E}}_{x \in \{-1,1\}^n_{(p)}} [\#\text{ coords queried by } T \text{ on } x].$$

Also, let $\Delta(f)$ denote the minimum of $\Delta(T)$ over all DDTs T for f. It is worth mentioning that $\Delta(f) \le D(f)$. Any near-balanced function such that $D(f) \le d$ depends on at most 2^d variables, so (1.2) implies $\mathbf{Inf}_{\max}(f) \ge \Omega(d/2^d)$. However, a partial improvement can be obtained.

Theorem 1.1. *Let $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$, and let T be a DDT for f, then*

$$\mathbf{Var}[f] \le \sum_{i=1}^{n} \delta_i(T)\ \mathbf{Inf}_i(f). \tag{1.3}$$

This inequality improves (1.1) and can be generalized to more general contexts. An easy computation yields

$$\Delta(f) \ge \frac{\mathbf{Var}[f]}{\mathbf{Inf}_{\max}(f)},$$

so when $\Delta(f) \le d$ or $D(f) \le d$ and f is near-balanced, we obtain $\mathbf{Inf}_{\max}(f) \ge \Omega(1/d)$, which improves (1.2) if $\Delta(f) = o(n/\log(n))$. The inequality (1.3) seems to be the first quantitatively strong influence lower bound in the literature that takes into account the computational complexity of f.

1.2 Randomized decision tree complexity lower bounds

Given a function $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$, a *randomized decision tree (RDT)* $\mathcal{T}$ for f is a probability distribution over DDTs T for f. For a RDT $\mathcal{T}$ computing f, we define the *RDT complexity of* f as

$$R(f) = \min_{\mathcal{T} \text{ RDTs for} f} \mathop{\mathbf{E}}_{T \sim \mathcal{T}} \max_{x \in \{-1,1\}^n_{(p)}} [\#\text{coords queried by } T \text{ on } x].$$

It is easy to check that $R(f) \le D(f)$, but a reverse inequality is quite subtle. It is well known that $R(f) \ge \Omega(\sqrt{D(f)})$ [2], and the largest known separation is the case $D(f) = n$ and $R(f) \le n^\beta$ with $\beta \simeq 0.753$, which holds for a specific monotone transitive function.

We say that a boolean function f is *monotone* if $f(x) \le f(y)$ whenever $x \le y$, under componentwise partial order. On the other hand, we say f is *transitive* if for each pair $i, j \in [n]$ there exists a permutation σ of $[n]$ satisfying

$$f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \quad \text{for any } x \quad \text{and} \quad \sigma(i) = j.$$

An interesting subclass of transitive boolean functions is that of graph properties: a *property of $v-$vertex graphs* is a set of graphs on a vertex set $V = \{1, \ldots, v\}$ that is invariant under vertex relabelings. Then each graph G can be indentified with a vector $x^G \in \{-1,1\}^{\binom{v}{2}}$, and each property $\mathcal{P}$, with a function $f_{\mathcal{P}}$ such that $f_{\mathcal{P}}(x^G) = 1$ if and only if G satisfies $\mathcal{P}$.

Several lower bounds for the RDT complexity of monotone graph properties have been obtained along the last four decades, but the following lower bound relies on purely probabilistic results and improve them (under some assumptions), and, for monotone transtive functions, is essentially as good as the best unconditional known bound.

Theorem 1.2. *Let $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$ be a nonconstant monotone transitive function, where p is the critical probability of f. Then*

$$R(f) \geq \Delta(f) \geq \frac{n^{2/3}}{(4p(1-p))^{1/3}}. \tag{1.4}$$

If f corresponds to a $v-$vertex graph property, then

$$R(f) \geq \Delta(f) \geq \frac{(v-1)^{4/3}}{(16p(1-p))^{1/3}}. \tag{1.5}$$

Since $\mathbf{E}[f]$ is a strictly increasing continuous function on p, there exists $p \in (0,1)$ such that $\mathbf{E}[f]$. For that p, and since f is transitive, Theorem 1.1 yields $1 \leq (\mathbf{Inf}(f)/n)\ \Delta(f)$, where $\mathbf{Inf}(f) := \sum_{i=1}^n \mathbf{Inf}_i(f)$. Moreover, using that for any $p \in (0,1)$ and any monotone $f : \{-1,1\}^n_{(p)} \to \{-1,1\}$, it holds $\mathbf{Inf}(f) \leq 2\sqrt{p(1-p)\Delta(f)}$, we obtain (1.4). The identity $n = \binom{v}{2}$ yields (1.5).

1.3 The main inequality

Theorem 1.1 can be formulated and proved in a more general context than the $p-$biased boolean cube. Indeed, let $(\Omega,\mu) = (\Omega_1 \times \ldots \times \Omega_n, \mu_1 \times \ldots \times \mu_n)$ be a $n-$wise product probability space, and let (Z,d) be a metric space. We will consider functions $f : \Omega \to Z$. Then a DDT for f is a rooted directed tree that satisfies

- each internal node v is labeled by a coordinate $i_v \in [n]$,
- each leaf is labeled by an element of Z,
- the emanating edges from an internal node v are in one-to-one correspondence with Ω_{i_v},
- the nodes along every root-leaf path are distinct.

Then, replacing $\{-1,1\}^n_{(p)}$ by (Ω,μ), analogous definitions hold for $D(f)$, $R(f)$, $\Delta(f)$ and probabilities $\delta_i(T)$ for a DDT T. On the other hand, variation and influences are defined as follows

$$\mathbf{Vr}[f] = \mathop{\mathbf{E}}_{(x,y)\in(\Omega,\Omega)}[d(f(x),f(y))], \quad \mathbf{Inf}_i(f) = \mathop{\mathbf{E}}_{(x,z_i)\in\Omega\times\Omega_i}[d(f(x),f(x_{[n]\setminus\{i\}}+z_i)].$$

When $\Omega = \{-1,1\}^n_{(p)}$, and $Z = \{-1,1\}$ is equipped with the distance $d(z,z') = |z-z'| = 2\ \mathbf{1}_{z\neq z'}$, the boolean case is involved.

Theorem 1.3 (OSSS inequality)**.** *Let $f : \Omega \to (Z,d)$, and let T be a DDT computing f. Then*

$$\mathbf{Vr}[f] \leq \sum_{i=1}^n \delta_i(T)\ \mathbf{Inf}_i(f). \tag{1.6}$$

Let x and y be random inputs chosen indepently from Ω. Let s be the number of coordinates queried by T on x, and let $i_1,\ldots,i_s,i_{s+1},\ldots,i_n$ be the sequence of those coordinates with $i_t = \emptyset$ whenever $t > s$. For $t \geq 0$, define the set $J[t] := \{s \geq r > t\}$ and the input $u[t] := x_{J[t]} + y_{[n]\setminus J[t]}$. Since $y = u[s]$ and $f(x) = f(u[0])$ (T computes f), and since d is a distance, then

$$\begin{aligned}\mathbf{Vr}[f] &= \mathbf{E}[d(f(x),f(y))] = \mathbf{E}[d(f(u[0]),f(u[t]))]\\ &\leq \mathbf{E}\Big[\sum_{t=1}^n d(f(u[t-1]),f(u[t]))\Big] = \sum_{t=1}^n\sum_{i=1}^n \mathbf{E}\Big[d(f(u[t-1]),f(u[t]))\ \mathbf{1}_{\{i_t=i\}}\Big]\end{aligned}$$

Linearity of expectation and $\mathbf{1}_{\{t\leq s\}} = \sum_{i=1}^n \mathbf{1}_{\{i_t=i\}}$ implies the previous identity. Now, consider the sequence of values $(x_{i_1},x_{i_2},\ldots,x_{\min\{t-1,s\}})$ read by T by time $t-1$ on input x. This sequence determines i_t. Then, it is easy to show that variables y and $(x_j\ :\ j \neq i_1,\ldots,i_{\min\{t-1,s\}})$ are independent with respect to conditional distribution on $(x_1,\ldots,x_{\min\{t-1,s\}})$, and their respective conditional distributions coincide with their original ones. Therefore, for any $i,t \in [n]$,

$$\mathbf{E}[d(f(u[t-1]),f(u[t]))\mathbf{1}_{\{i_t=i\}} \mid (x_{i_1},\ldots,x_{\min\{t-1,s\}})] = \mathbf{1}_{\{i_t=i\}}\mathbf{Inf}_i(f)$$

since $u[t-1]$ and $u[t]$ only differ on their i_tth coordinate, which are x_{i_t} and y_{i_t} respectively. Taking expectation and summing in t gives

$$\sum_{t=1}^{n} \mathbf{E}\Big[d(f(u[t-1]), f(u[t]))\mathbf{1}_{\{i_t=i\}}\Big] = \sum_{t=1}^{n} \mathbf{Pr}[i_t = i]\ \mathbf{Inf}_i(f) = \delta_i(T)\ \mathbf{Inf}_i(f).$$

So summing in i yields the desired result.

It turns out that inequality (1.6) is tight since identity holds for *separated* trees. Moreover, a two-function version of Theorem 1.3 can be proved and used to obtain a estimation for the complexity of approximations of a given function g. Finally, it admits a version when Z is a semimetric space in which the constant on right hand side of (1.6) is greater than one.

1.4 An inductive proof of OSSS inequality

Another proof can be obtained through a inductive argument [4]. Indeed, a two-function version can be proved: let $f, g : \{-1,1\}^n \to \{-1,1\}$ and T a DDT for f. Then

$$|\mathbf{Cov}[f,g]| \leq \sum_{i=1}^{n} \delta_i(T)\ \mathbf{Inf}_i(g) \tag{1.7}$$

where $\mathbf{Cov}[f,g] = \mathbf{E}[(f - \mathbf{E}f)(g - \mathbf{E}g)]$. The proof is based on *martigale differences*: for $i \in [n]$, define

$$c_i(f) = c_i(f; x_1, \ldots, x_n) = \mathbf{E}[f|(x_1, \ldots, x_i)] - \mathbf{E}[f|(x_1, \ldots, x_{i-1})].$$

It is easy to check that $\mathbf{Cov}[f,g] = \sum_{i=1}^{n} \mathbf{E}[c_i(f)c_i(g)]$ and $\mathbf{E}[c_n(f)c_n(g)] \leq \mathbf{Inf}_n(f),\ \mathbf{Inf}_n(g)$. The base case $n = 1$ supposes a simple verification. Let T be a DDT for f whose root has label x_n. Let T_{-1} and T_1 be the left and right subtree. Then for $i \neq n$, $\delta_i(T) = 1/2\ (\delta_i(T_{-1}) + \delta_i(T_1))$, $\mathbf{Inf}_i(g) = 1/2\ (\mathbf{Inf}_i(g_{-1}) + \mathbf{Inf}_i(g_1))$ and $c_i(f) = 1/2\ (c_i(f_{-1}) + c_i(f_1))$. Therefore,

$$\begin{aligned} |\mathbf{Cov}[f,g]| &\leq \frac{1}{4} \sum_{a,b \in \{-1,1\}} \Big| \sum_{i=1}^{n-1} \mathbf{E}[c_i(f_a)c_i(g_b)] \Big| + |\mathbf{E}[c_n(f)c_n(g)]| \\ &= \frac{1}{4} \sum_{i=1}^{n-1} |\mathbf{Cov}[f_a, g_b]| + |\mathbf{E}[c_n(f)c_n(g)]|. \end{aligned}$$

So, since $\delta_n(T) = 1$ and by induction hypothesis, (1.7) follows.

Bibliography

[1] Ben-Or M. and Linial L., *Collective coin flipping.* In: Proceedings of the 26th Annual Symposium on Foundation of Computer Science (FOCS), 408-416 (1985).

[2] Blum M. and Impagliazzo R., *Generic oracles and oracle classes.* In: Proceedings of the 28th Annual Symposium on Foundations of Computer Science, 118-126 (1987).

[3] Kahn J., Kalai I. and Linial N., *The influence of variables on boolean functions.* In: Proceedings of the 29th Annual Symposium on Foundations of Computer Science, 68-80 (1988).

[4] Lee H.K., *Decision Trees and Influence: an Inductive Proof of the OSSS Inequality.* Theory of Computing, 6(1), 81-84.

[5] O'Donnell R., Saks M, Schramm O and Servedio R.A., *Every decision tree has an influential variable.* In: Proceedings of the 46th Annual Symposium on Foundations of Computer Science (FOCS), 31-39 (2005).

Antonio Ismael Cano Mármol, ICMAT
email: **ismael.cano@icmat.es**

Chapter 2

A proof of the sensitivity conjecture

after H. Huang [1]
A summary written by Jaume de Dios Pont

Abstract. The sensitivity theorem (former sensitivity conjecture) relates ways to quantify the complexity, or lack of smoothness a function $f : \{-1,1\}^n \to \{-1,1\}$, the sensitivity $s(f)$ and the degree of f when thought as a polynomial. We provide a self-contained proof of this result.

2.1 Introduction

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a function on the n-dimensional hypercube. We define the sensitivity of f at x, which we will write $s(f,x)$ as the number of inputs $y \in \{-1,1\}^n$ that differ from x at exactly one co-ordinate such that $f(x) \neq f(y)$. We define the sensitivity of f, or $s(f)$ as the maximum of the sensitivity of f over all of its inputs.

This is a notion of *complexity*, or lack of *smoothness* for functions on the hypercube. Multiple other notions of complexity for boolean functions have been studied and related to each other through the years. Amongst those we can highlight the block sensitivity and the degree.

Block Sensitivity $bs(f)$

Given a subset I of $[n] := \{1, \dots, n\}$, and binary string $x = (x_1, \dots x_n) \in Q^n$ we define $T_I x$ as

$$(T_I x)_j := x_j (-1)^{\mathbb{1}_I(j)} = \begin{cases} -x_j & \text{if } j \in I \\ x_j & \text{if } j \in [n] \setminus I \end{cases}$$

For a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ and a point x in $\{0,1\}^n$ the quantity $bs(f,x)$ (block sensitivity at x) counts how many disjoint subsets $I_1, I_2, \dots, I_{bs(f,x)}$ of $[n]$ one can simultaneously find such that $f(x) \neq f(T_{I_k} x)$. In particular, $bs(f,x) \geq s(f,x)$, since $s(f,x)$ adds the further constraint that $|I_k| = 1$.

We define the block sensitivity of f, or $bs(f)$ as the maximum of the sensitivity of f over all of its inputs.

The degree $\deg(f)$

A function $f : \{-1,1\}^n \to \{-1,1\}$ can be thought as the restriction of a polynomial $P_f : \mathbb{R}^n \to \mathbb{R}$. Since $x^k = x^{k-2}$ for $x \in -1, 1$, one can restrict P_f to be in the class of polynomials that are multi-linear, that is, linear in each of their n variables. These polynomials are a basis for the set of functions $f : \{-1,1\}^n \to \mathbb{R}$, ans in particular, P_f is determined uniquely by f.

The degree $\deg(f)$ of P_f is another measure of complexity for boolean functions.

These last two quantities are closely related to each other. Nisan and Szegedy [2] show that $b(s) \leq 2\deg(f)^2$ (which is Topic 12 in this school). This was later improved to $bs(s) \leq \deg(f)^2$ by Tal [3]. Our goal in the rest of these notes will be to show that $\deg(f) \leq s(f)^2$, showing that the three quantities are polynomially related.

2.2 The induced subgraph problem

Gotsman and Linial [4] reduced the problem of relating the sensitivity and the degree to that of understanding the degree of certain induced subgraphs of the hypercube graph Q^n, which has as vertices the elements of $\{-1,1\}^n$, and edges joining vertices that differ only on one bit (coordinate).

Given a graph G, we denote by $\Delta(G)$ the maximun of the degrees of the vertices of G. Gotsman and Linial showed the following:

Theorem 2.1 ([4] , Theorem 2.1)**.** *The following are equivalent for any monotone function $h : \mathbb{N} \to \mathbb{R}$*

($GL1$) *For any induced subgraph H of Q^n with $Q(H) \neq 2^{n-1}$ we have*

$$\max(\Delta(G-H), \Delta(H)) \geq h(n).$$

($GL2$) *For any boolean function f we have $s(f) \geq h(\deg(f))$*

Now, relating the sensitivity and the degree is related to computing the degree of certain induced subgraphs. Huang showed the following holds:

Theorem 2.2 ([1], Theorem 1.1)**.** *If H is an induced subgraph of Q^n with strictly more than 2^{n-1} vertices, then the degree of H is at least $\sqrt{n}$. Therefore, for any boolean function $f : \{-1,1\}^n \to \{-1,1\}$*

$$s(f) \geq \deg(f)$$

Combining this with the inequality $\deg(f) \geq bs(f)^2$, one obtains a polynomial (fourth power) relation between $bs(f)$ and $s(f)$. This is not expected to be sharp: for the best known counterexamples only give a quadratic relation.

2.3 Proof of Theorem 2.2

The proof is a sleek modification the so called "spectral method" for graphs. We will first understand the method in the general setting, and then addapt it to our scenario. The goal is to bound (from below) the maximum degree of a graph using the following two tools:

Lemma 2.3. *Let G be an undirected graph with adjacency matrix A. Let B be a symmetric matrix such that $|B_{ij}| \leq A_{ij}$. Then the degree of G is at least the largest eigenvalue (in absolute value) of B.*

Proof. The largest eigenvalue is the $l^2 \to l^2$ operator norm of B, and the degree of G bounds the $l^1 \to L^\infty$ (and $l^1 \to L^\infty$ by symmetry) norm of B. In particular, the lemma follows from Schur's test.

For a more direct proof, let v be an eigenvector associated to the largest eigenvalue of B, and assume v_k is the largest component (in magnitude) of v. Then

$$|\lambda v_k| = |(Av)_k| = \left|\sum_{j=1}^n B_{kj} v_j\right| \leq |v_k| \sum_{j=1}^n A_{kj} \leq |v_k| \deg(G)$$

and the inequality follows by dividing by $|v_k|$ on both sides. □

For the second tool, we define a *principal submatrix* of B as one that is obtained by removing the same set of rows and columns from B.

Lemma 2.4 (Cauchy's Interlace Theorem)**.** *Let B be a symmetric $n \times n$ real matrix with eigenvalues $\beta_1 \geq \lambda_2 \cdots \geq \beta_n$. Let $\tilde{B}$ be an $m \times m$ principal submatrix of B, with eigenvalues $\tilde{\beta}_1 \geq \cdots \geq \tilde{\beta}_m$. Then*

$$\beta_i \geq \tilde{\beta}_i \geq \beta_{n-m+i} \tag{2.1}$$

This follows essentially from the Courant-Fischer characterization of the eigenvalues of a symmetric matrix. The term *Cauchy's Interlace Theorem* is sometimes used for the case $m = n - 1$. The other cases can be deduced from this one by induction removing one row/column after the other.

Combining this two pieces of information we get immediate bounds to the maximum degree $\Delta(H)$ of a subgraph as follows:

Corollary 2.5. *Let G be a graph with n vertices and adjacency matrix A. If B is a $n \times n$ matrix with $|B_{ij}| \leq |A_{ij}|$ and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ then any induced subgraph $H \leq G$ with cardinality $> k$ will have maximum degree $\Delta(H)$ at least λ_{n-k+1}.*

Proof. Apply Lemma 2.3 to H. The adjacency A_H matrix of H is a principal submatrix of A. Construct the matrix B_H by removing/keeping the same rows as columns from B as was done to construct A_H from A. Applying Cauchy's Interlace Theorem (Lemma 2.4) with data H, A_H, B_H gives the proof. □

Application of the spectral method to the hypercube

If the vertices of the hypercube graph Q^n are sorted in lexicographic order, the adjacency matrices A_n on Q^n satisfy the following recurrence relation

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A_n = \begin{pmatrix} A_{n-1} & I_{2^{n-1}} \\ I_{2^{n-1}} & A_{n-1} \end{pmatrix}$$

in other words, the cube Q_n is formed by getting two copies of Q_{n-1} (corresponding to the sub-matrices A_{n-1}) and joining the two copies of each vertex in each cube (giving rise to the $I_{2^{n-1}}$ matrices).

We will build the matrices B_n (in the spirit of Lemma 2.3) similarly, by the recurrence relation

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad B_n = \begin{pmatrix} B_{n-1} & I_{2^{n-1}} \\ I_{2^{n-1}} & -B_{n-1} \end{pmatrix}.$$

This construction already guarantees that $|(B_n)_{ij}| \leq (A_n)_{ij}$, one of the conditions to apply Lemma 2.3. Moreover, it has particularly nice spectral properties:

Proposition 2.6. *The matrices B_n have the following properties:*

1. $B_n^2 = n \cdot I_{2^n}$
2. $tr(B_n) = 0$
3. *Exactly half of the eigenvalues of B_n are $\sqrt{n}$. The other half are $-\sqrt{n}$.*

Proof sketch. The first equality is proven by induction on n, using the recursive definition of B_n. It already implies that all the eigenvalues are $\pm\sqrt{n}$. The second equality follows by direct inspection. Since the trace is the sum of the eigenvalues, half of them must be $\sqrt{n}$ and half $-\sqrt{n}$. □

These are all the tools we need to show Theorem 2.2 (assuming Theorem 2.1, which will be proven in the next section):

Proof of Theorem 2.2. We can apply Corollary 2.5 to the matrices B_n (using that $|(B_n)_{ij}| \leq (A_n)_{ij}$. If $j > 2^{n-1}$ then $2^n - j + 1 \leq 2^{n-1}$. The first 2^{n-1} eigenvalues of B_n are $\sqrt{n}$. Therefore the maximum vertex degree of any vertex-induced subgraph with $j > 2^{n-1}$ verices will be at least the $(2^n - j + 1)$-th largest eigeivalue of B_n, that is, $\sqrt{n}$. □

2.4 The subgraph problem and the degree bounds

The goal of this section is to give a proof of Theorem 2.1, following the origina proof in [4]. The first step is to simplify the statements $(GL1)$ and $(GL2)$ of the theorem to simpler, but equivalent, statements. On one hand $(GL1)$ can be transformed into a statement about sensitivity by studying the *indicator function* of the vertices of H. On the other hand $(GL2)$ can be reduced to the case when the degree of f is maximal. That makes Theorem 2.1 equivalent to the following proposition:

Proposition 2.7. *The following are equivalent for any monotone function $h : \mathbb{N} \to \mathbb{R}$*

($GL1'$) *For any boolean function g with mean not equal to zero there is x with $s(g, x) \leq n - h(n)$.*

($GL2'$) *For any $n \geq 0$ and any boolean function $f : Q^n \to \{-1, 1\}$ $s(f) < h(n)$ implies $d(f) < n$.*

Proof. We will see the equivalence by relating the functions in the statements by $g(x) = f(x)p(x)$, where $p((x_1, x_2, \ldots, x_n)) = \prod_{i=1}^{n} x_i$. This also implies $f(x) = g(x)p(x)$. There are two key relations between f and g:

(A) The function g has mean zero if and only if f has degree n. This is because for any multi-index $I \subset [n]$ multiplication by p sends x^I to its complement: $x^I p(x) = x^{[n]\setminus I}$. Then g has a non-zero constant coefficient if and only if f has a degree n coefficient (coeficients are degree 1 in each separate x_i).

(B) It holds that $s(g, x) = n - s(f, x)$. It holds that $p(x) - = p(T_i x)$, and therefore $f(x) = f(T_i x)$ if and only if $g(x) - = g(T_i x)$ (and viceversa).

$(1') \implies (2')$ If $d(f) = n$ then g does not have mean zero by (A). In particular, $s(g, x) \leq n - h(n)$ at some point x. This implies that $s(f(x)) \geq h(n)$.

$(2') \implies (1')$ If $s(g, x) > n - h(n)$ for all x, then $s(f) < h(n)$ by (B). By (2') therefore $d(f) < n$. Now, by (A) that shows that g has mean zero. □

Bibliography

[1] Huang, H., *Induced subgraphs of hypercubes and a proof of the sensitivity conjecture.* Annals of Mathematics, 190.3, 949-955 (2019)

[2] Nisan, N., Szegedy, M., *On the degree of Boolean functions as real polynomials.* Computational complexity 4, no. 4, 301-313 (1994)

[3] Tal, A, *Properties and applications of Boolean function composition.* Proceedings of the 4th conference on Innovations in Theoretical Computer Science, 441-454 (2013)

[4] Gotsman, C., Linial, N., *The equivalence of two problems on the cube.* Journal of Combinatorial Theory, Series A 61, no. 1, 142-146 (1992)

[5] Marton, K., *Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration.* The Annals of Probability, 24, 857-866 (1996)

Jaume de Dios Pont, UCLA
email: **jdedios@math.ucla.edu**

Chapter 3

Quantum Mechanics Helps in Searching for a Needle in a Haystack

after L. K. Grover [1]
A summary written by Valeria Fragkiadaki

Abstract. Quantum mechanics can speed up a range of search applications over unsorted data. For example, there is a quantum algorithm that can obtain one's phone number over a phone directory of N names arranged randomly in only $O(\sqrt{N})$ accesses to the database compared to at least $0.5N$ accesses needed by any classical algorithm.

This Letter presents a quantum mechanical algorithm for the following search problem that is polynomially faster than any classical algorithm. Search problem: Suppose there is an unsorted database containing N items and we want to find one out of them that satisfies a given condition. We can check if an item satisfies the condition in one step. The most efficient classical algorithm for this examines the items one by one requiring an average of $0.5N$ items to be examined before finding the desired one.

However, quantum mechanical systems can be in *superpositions* of states and simultaneously examine multiple items allowing a certain probability of examining the desired object. This Letter shows that using the same amount of hardware as in the classical case, but having the input and output in superspositions of states, we can find an object in $O(\sqrt{N})$ *quantum mechanical steps* instead of $O(N)$ classical steps.

3.1 Quantum mechanical algorithms

In a quantum computer the logic circuitry and time steps are essentially classical, the biggest difference are the memory *bits* that hold the variables. A classical bit can have state 0 or 1, while a quantum bit can have a state which could be $\overline{0}$ or $\overline{1}$ (computational basis vectors), but it could also be in a superposition state (linear combination of states): $\overline{\psi} = \alpha\overline{0} + \beta\overline{1}$, where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. We say that α, β are the amplitudes of the states $\overline{0}$ and $\overline{1}$ respectively.

The quantum mechanical operations that can be performed are unitary operations, i.e. unitary matrices, that act on a small number of bits, i.e. vectors, in each step. The quantum search algorithm of this letter is a sequence of the following three unitary operations on a pure state followed by a measurement operation:

The Walsh-Hadamard operation performed on a single bit is represented by the matrix

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

i.e. a bit in the state $\overline{0} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is transformed into a superposition in the two states $(1/\sqrt{2})\overline{0} + (1/\sqrt{2})\overline{1} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. Similarly, a bit in the state $\overline{1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is transformed into $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$, i.e. the magnitude of the

amplitude in each state is $1/\sqrt{2}$, but the *phase* of the amplitude (a constant multiplier of the form $e^{i\theta}$) in the state $\overline{1}$ is inverted. Now for the first operation we need, consider the possible states of the system to be $N := 2^n$ so they can be described by n bits. Then, we can perform the transformation M on each bit independently in sequence, thus changing the state of the system. If we start the system with all n bits in the first state, i.e. $\overline{x} = \overline{0} \otimes \overline{0} \otimes \cdots \otimes \overline{0}$, we get a configuration $\overline{\psi} = 2^{-n/2} \sum_{x=0}^{2^n-1} \overline{x}$, where $\overline{x}$ is the binary representation of x. This way we can create a superposition in which the amplitude of the system being in any of the 2^n basic states is equal.

Next consider the case when the starting state $\overline{x}$ is another one of the 2^n possible states (not the $\overline{0} \otimes \overline{0} \otimes \cdots \otimes \overline{0}$). Performing the transformation M on each bit we get a superposition of states described by all possible n bit binary strings with amplitude of each state having the same magnitude of $2^{-n/2}$ and sign either $+$ or $-$. The sign of each state $\overline{y}$ is determined by the parity of the bitwise dot product of $\overline{x}$ and $\overline{y}$, i.e. $(-1)^{\overline{x} \cdot \overline{y}}$. This describes the Walsh-Hadamard transformation on n bits.

The third transformation that we will need is the selective rotation of the phase of the amplitude in certain states. The transformation describing this for a 2-state system is of the form

$$\begin{pmatrix} e^{i\phi_1} & 0 \\ 0 & e^{i\phi_2} \end{pmatrix}, \ \phi_1, \phi_2 \in \mathbb{R}.$$

3.2 The abstracted problem

Let a system have $N = 2^n$ states which are labelled $S_1, S_2, ..., S_N$ and are represented as n bit strings. Let there be a unique state, say S_ν, that satisfies the condition $C(S_\nu) = 1$, whereas for all other states S, $C(S) = 0$. Assuming that for each S, the condition $C(S)$ can be evaluated in unit time, the problem is to identify the state S_ν.

3.3 Algorithm

(i) Initialize the system to the superposition $(1/\sqrt{N}, 1/\sqrt{N}, ..., 1/\sqrt{N})$ as discussed in subsection 3.1. This superposition can be obtained in $O(\log N)$ steps.

(ii) Repeat the following unitary operations $O(\sqrt{N})$ times:

(a) Let the system be in any state S:
If $C(S) = 1$, rotate the phase by π radians
If $C(S) = 0$, leave the system unaltered

(b) Apply the diffusion transform D which is defined by the matrix D as follows:

$$D_{ij} = \frac{2}{N}, \text{ if } i \neq j, \quad \text{and} \quad D_{ii} = -1 + \frac{2}{N}.$$

(iii) Measure the resulting state. This will be the state S_ν with a probability of at least 0.5.

3.4 Convergence

The loop in step (ii) above is the heart of the algorithm. Each iteration of this loop increases the amplitude in the desired state by $O(1/\sqrt{N})$, as a result in $O(\sqrt{N})$ repetitions of the loop, the amplitude and hence the probability in the desired state reaches $O(1)$. In order to see that the amplitude increases by $O(1/\sqrt{N})$ in each repetition, we first show that the diffusion transform D is equivalent to the *inversion about average* operation which is a unitary operation.

Let α denote the average amplitude over all states S_i, i.e. if α_i is the amplitude in the i-th state, then $\alpha = \frac{1}{N} \sum_{i=1}^{N} \alpha_i$. Now, observe that the diffusion transform, D, defined in (b) can be represented in the form $D = -I + 2P$, where I is the identity matrix and P is a projection matrix with $P_{i,j} = 1/N, \ \forall i, j$. Notice also that $P^2 = P$ and that P acting on any vector $\overline{v}$ gives a vector each of whose components is equal to the average of all components. Thus, when D acts on an arbitrary vector $\overline{v}$ we get

$$D\overline{v} = (-I + 2P)\overline{v} = -\overline{v} + 2P\overline{v}.$$

Since each component of the vector $P\overline{v}$ is A, where A is the average of all components of the vector $\overline{v}$, the i-th component of $D\overline{v}$ is

$$(D\overline{v})_i = -\overline{v}_i + 2A = A + (A - \overline{v}_i) \tag{3.1}$$

which is precisely the *inversion about average*.

Next consider what happens when we apply this operation to a vector with each of the components, except one, having an amplitude equal to $C/\sqrt{N}$ where $1/2 \leq C \leq 1$; the one component that is different has an amplitude of $-\sqrt{1-(N-1)\frac{C^2}{N}}$, which is approximately $-\sqrt{1-C^2}$. Then, the average A of all components is approximately equal to $C/\sqrt{N}$, thus each of the $(N-1)$ components which are approximately equal to the average do not change significantly, while by (3.1), the component that was negative becomes positive and its magnitude increases by $2C/\sqrt{N}$.

Now, in the algorithm of subsection 3.3, in the loop of step (ii), first the amplitude in the selected state, S_ν, is inverted. Then, the inversion about average operation is carried out, giving an increase in the amplitude of S_ν by $2C/\sqrt{N}$ in each iteration. Therefore, as long as the magnitude of the amplitude in S_ν is less than $1/\sqrt{2}$, i.e. $\sqrt{1-C^2} \leq \frac{1}{\sqrt{2}}$, $C \geq \frac{1}{\sqrt{2}}$ and the increase in its magnitude is greater than $1/\sqrt{2N}$. Thus, there exists an $M \leq \sqrt{N}$ such that in M repetitions of the loop in step (ii), the magnitude of the amplitude in S_ν will exceed $1/\sqrt{2}$. Thus, measuring now the state of the system we get S_ν with a probability $\geq 1/2$.

3.5 Implementation

As mentioned in subsection 3.1 quantum mechanical operations that can be carried out in a controlled way are unitary operations that act on a small number of bits in each step, like for example the Walsh-Hadamard transformation with matrix say W and the phase rotation with matrix say R. We show that the diffusion transform $D = -I + 2P$, where $P_{ij} = 1/N$, $\forall i, j$, can be implemented as a product of three such unitary transformations, namely, $D = WRW$, where

$$R_{ij} = 0 \text{ if } i \neq j, \quad R_{ii} = 1 \text{ if } i = 0, \quad R_{ii} = -1 \text{ if } i \neq 0,$$

and $W_{ij} = 2^{-n/2}(-1)^{\overline{i}\cdot\overline{j}}$ as discussed in subsection 3.1.

Writing $R = R_1 + R_2$, where $R_1 = -I$ and $R_{2,00} = 2$, $R_{2,ij} = 0$ if $(i,j) \neq (0,0)$, we show that $D = WR_1W + WR_2W$. But since $MM = I$, where M is the matrix defined in subsection 3.1, we have $WW = I$ and hence $D_1 := WR_1W = -I$. Next, we evaluate $D_2 := WR_2W$ by standard matrix multiplication and get

$$D_{2,ad} = \sum_{b,c} W_{ab}R_{2,bc}W_{c,d} = 2W_{a0}W_{0d} = \frac{2}{2^n}(-1)^{\overline{a}\overline{0}+\overline{0}\overline{d}} = \frac{2}{N}.$$

Therefore, we get $WR_1W + WR_2W = -I + 2P = D$.

Thus, the only operations required for this quantum search algorithm are the Walsh-Hadamard transform and the conditional phase shift operation, and this makes the algorithm rather simple compared to many other known algorithms.

The author wishes to acknowledge Peter Shor, Ethan Bernstein, Gilles Brassard, Norm Margolus, and John Preskill for helpful comments.

Bibliography

[1] Lov K. Grover, *Quantum Mechanics Helps in Searching for a Needle in a Haystack* Physical Review Letters, Volume 79, Number 2 (1997).

Valeria Fragkiadaki, TAMU
email: **valeria96@tamu.edu**

Chapter 4

On the distribution of the Fourier spectrum of Boolean functions

after J. Bourgain [1]
A summary written by Christina Giannitsi

Abstract. We present a summary of Bourgain's result on the tail distribution of the Fourier spectrum of Boolean functions f defined on $\{-1,1\}^N$. Specifically, for a fixed positive integer k, and as long as f is not determined by a bounded number of variables, we have that

$$\sum_{|S|>k} |\hat{f}(S)|^2 \gtrsim k^{-\frac{1}{2}-\varepsilon}.$$

At the end we discuss the sharpness of the result by examining the majority function.

4.1 Introduction

We are studying the Fourier transform of Boolean functions, which are functions which, in one dimension, assume values from a two element set $\{-1,1\}$, and are generalized in higher dimensions as $f:\{-1,1\}^N \to \{0,1\}$. N is often called the arity of the function, [2].

The study of Boolean functions was popularized by their contributions to areas like complexity theory and computer science, where major results were discovered by studying their Fourier transforms. Bourgain references the work of Friedgut [3], who discovered a sharp bound for indicators of monotone subsets of $\{-1,1\}^N$, as well as the works of Kahn, Kalai, and Linial who studied how variables can influence these functions. Again, here, the analysis of the Fourier transform is crucial to obtaining the desired results.

Heuristically, the idea is that the higher the complexity of the property that f defines, the more spread out the support of the Fourier transform $\operatorname{supp}\hat{f}$ has to be. An application of this idea is the topic of Bourgain's paper [1], where we particularly see that the tail distribution of the Fourier transform of a function f which is not essentially determined by a few variables is bounded below, as described in Theorem 4.1.

Johan Håstad, and his work with Boolean functions in [4] and [5], was the one to initially raise the question about the tail distributions, however his original estimate was of the order C^{-k}.

Throughout this summary we shall use $\mathbb{1}_A$ to denote the usual indicator function of a subset A of the real numbers $\mathbb{R}$ or the integers $\mathbb{Z}$. We also use $[1,N] = \{1,2,\dots,N\}$ for the interval of integers. Moreover, any logarithms that appear are base 2. Finally, $\hat{f}$ to denote the Fourier transform of a Boolean function f. Specifically, for a real function $f:\{-1,1\}^N \to \{-1,1\}$ let

$$f = \sum \hat{f}(S) w_S$$

be its Fourier expansions, where

$$w_S(x_1,x_2,\dots,x_N) = (-1)^{\sum_{i\in S} x_i}$$

4.2 The main result

Bourgain's main result is Proposition 1 of [1], which is presented below.

Theorem 4.1. *Let $f = \mathbb{1}_A$ be the indicator of a set $A \subseteq \{-1,1\}^N$ and fix $\varepsilon > 0$. Let k be a positive integer and γ a fixed constant. Assume that*

$$\sum_{|\hat{f}(S)|<\gamma 4^{-k^2}} |\hat{f}(S)|^2 > \gamma^2 \tag{4.1}$$

Then

$$\sum_{|S|>k} |\hat{f}(S)|^2 \gtrsim k^{-\frac{1}{2}-\varepsilon} \tag{4.2}$$

The implied constant C_ε in (4.2) depends on ε but is independent of k.

We shall devote the rest of the section to a brief sketch of the proof.

The first step is to define a subinterval I_0 of [1,N] that contains integers for which the quantity $\sum_{|S|>k} |\hat{f}(S)|^2$ is "large" as long as S contains that integer. We can then bound the size of I_0 and use that bound to show that, on I_0, and provided that (4.1) holds,

$$\sum_{|S|>k,\ S\subset I_0} |\hat{f}(S)|^2 < \gamma^2/100.$$

We then focus on its complement $I_0' = [1,N] \setminus I_0$, and use the aforementioned estimate to show that when restricted on I_0',

$$\sum_{|S|>k,\ S\cap I_0'\neq\emptyset} |\hat{f}(S)|^2 > \gamma^2/2.$$

The next step is to consider dyadic sums defined as

$$\rho_t := \sum_{2^t\le|S\cap I_0'|<2^{t+1}} |\hat{f}(S)|^2,$$

and to show that for an arbitrary $0 \le t_0 \le \log k$ and $1 < p < 2$ there holds

$$\sum_{|S|>k} |\hat{f}(S)|^2 \gtrsim \min\left\{\rho_{t_0}^{2/p},\ (p-1)^{\frac{p}{2-p}}\left(\frac{2^{t_0}\rho_{t_0}}{\sum_{t\le\log k} 2^t\rho^t}\right)^{\frac{p}{2-p}}\rho_{t_0}\right\}. \tag{4.3}$$

Proving (4.3) is the biggest part of the proof, and involves carefully decomposing $x = (x_1, x_2)$ so that $x_1 \in \{-1,1\}^{I_1}$, for a subset I_1 of I_0' that satisfies certain growth criteria, as well as bounds for the expectations of the sizes of various intersection with S. Now, considering f as a function of two variables, we study its Fourier transforms with respect to each of them and attempt to obtain bounds for them. Working on the Fourier side, it is possible to ultimately obtain (4.3).

The last step is to consider two cases, one for $\sum_{t\le\log k} 2^t\rho_t < \sqrt{k}$ and one for $\sum_{t\le\log k} 2^t\rho_t \ge \sqrt{k}$, and show that in either case, the right hand-side of (4.3) satisfies the desired lower bound of (4.2), which completes the proof.

4.3 The majority function as an example of sharpness

Bourgain presents the following Corollary as an immediate consequence of Theorem 4.1:

Corollary 4.2. *Let $f = \mathbb{1}_A$ be the indicator of a set $A \subseteq \{-1,1\}^N$ that satisfies*

$$|A|\,(1-|A|) > \frac{1}{10}$$

Let K be a positive integer and assume that

$$\max_{|S|\leq k} |\hat{f}(S)| < 4^{-k^2-1}.$$

Then

$$\sum_{|S|>k} |\hat{f}(S)|^2 \gtrsim k^{-\frac{1}{2}-\varepsilon} \tag{4.4}$$

We now discuss how Bourgain established that the lower bound as presented in (4.4) is a sharp one. Indeed, consider the $\{-1,1\}$-valued majority function on $\{-1,1\}^N$, defined as

$$f(\varepsilon) := \text{sign}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_N)$$

It is shown in [7] that the majority function satisfies

$$|\hat{f}(S)|^2 \sim \binom{N}{n}^{-1} n^{-3/2}, \quad \text{for } |S| = n > 0,$$

and therefore one can easily see that

$$\sum_{|S|=k} |\hat{f}(S)|^2 \sim k^{-3/2},$$
$$\sum_{|S|>k} |\hat{f}(S)|^2 \sim k^{-1/2}.$$

Bibliography

[1] Bourgain, J. *On the distribution of the Fourier spectrum of Boolean functions.* Isr. J. Math. 131, 269–276 (2002). https://doi.org/10.1007/BF02785861

[2] Comtet, L. *Advanced Combinatorics: The Art of Finite and Infinite Expansions.* Dordrecht : Springer Netherlands : Imprint: Springer 1974

[3] Friedgut, E. *Sharp threshold of graph properties, and the k-sat problem.* Journal of the American Mathematical Society 12 (1999), 1017-1054

[4] Håstad, J. *Some optimal inapproximability results.* Proceedings of the 29th Annual ACM Symposium on Theory of Computing, pages 1–10, 1997.

[5] Håstad, J. *A slight sharpening of LMN.* J. Comput. Syst. Sci., 63(3):498–508, 2001.

[6] Kahn, J., Kalai, G. and Linial, N. *The influence of variables on Boolean functions* Proc. 29 th IEEE FOCS 58-80, IEEE, New York, 1988.

[7] Karpovsky, M. G. *Finite OrthogonM Series in the Design of Digital Devices.* Wiley, New York, 1976.

Christina Giannitsi, Georgia Tech
email: **cgiannitsi3@gatech.edu**

Chapter 5

Noise Stability of Weighted Majority

after Y. Peres [Y04]
A summary written by Felipe Gonçalves

Abstract. It is shown in [Y04] that in a threshold activation control system with a linear weighted majority function, boolean noise with probability ε produces a difference in decision with probability $O(\sqrt{\varepsilon})$.

5.1 Main Results

Noise sensibility of boolean functions has attracted a lot of attention in the last decades. In [Y04], Y. Peres studies noise sensitivity of the majority function

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} w_i x_i - t\right)$$

defined for $x \in \{-1,1\}^N$, where $w_i \in \mathbb{R}$ are given weights, $t \in \mathbb{R}$ is some given threshold and $\operatorname{sgn}(y) = y/|y|$ for $y \neq 0$ and normalized so that $\operatorname{sgn}(0) = 0$. The main result of the paper [Y04] is the following.

Theorem 5.1. *Let $X = (X_1, ..., X_N)$ be uniformly distributed in $\{-1,1\}^N$. Let $\sigma = (\sigma_1, ..., \sigma_N)$ be i.i.d. real random variables, all independent of X, and such that $\varepsilon = \mathbb{P}[\sigma_1 = -1] = 1 - \mathbb{P}[\sigma_1 = 1]$ for $0 < \varepsilon \leq 1/2$. Then*

$$\mathbb{P}[f(X) \neq f(\sigma X)] \leq 1.92\sqrt{\varepsilon}, \tag{5.1}$$

where $\sigma X = (\sigma_1 X_1, ..., \sigma_N X_N)$. In fact we have the stronger estimate

$$\mathbb{P}[f(X) \neq f(\sigma X)] \leq \frac{2}{m}\mathbb{E}[|B_m - m/2|] + (1 - (1-\varepsilon)^N)\binom{N}{N/2}2^{-N} \tag{5.2}$$

where $m = \lfloor \varepsilon^{-1} \rfloor$ and $B_m \equiv \operatorname{Bin}(m, 1/2)$.

Remark 5.2. *This theorem is an improvement of a result of Benjamini, G. Kalai and O. Schramm [BKS01], where they show that*

$$\mathbb{P}[f(X) \neq f(\sigma X)] \leq C\varepsilon^{1/4}$$

and asked if the exponent $1/4$ can be improved. Indeed, the exponent $1/2$ in Theorem 5.1 is optimal since for the classical majority function ($w_i = 1$ for all i) we have [G66]

$$\lim_{N\to\infty} \mathbb{P}[f(X) \neq f(\sigma X)] = \pi^{-1}\arccos(1-2\varepsilon) = \frac{2\sqrt{\varepsilon}}{\pi} + O(\varepsilon^{3/2}).$$

Remark 5.3. *Indeed inequality* (5.2) *implies* (5.1) *by the following argument. We can assume* $\varepsilon < 1/\sqrt{2}$. *Since* $\binom{N}{N/2}2^{-N} \le \sqrt{2/(\pi N)}$, $\lfloor \varepsilon^{-1} \rfloor \ge 4/(5\varepsilon)$ *and* $\mathbb{E}[|B_m - m/2|] \le \sqrt{\mathrm{Var}[B_m]} = \sqrt{m/4}$ *we obtain*

$$\begin{aligned}\mathbb{P}[f(X) \neq f(\sigma X)] &\le m^{-1/2} + (1-(1-\varepsilon)^N)\sqrt{2/(\pi N)} \\ &\le \sqrt{\varepsilon}\left(\sqrt{5/4} + \sqrt{2/\pi}\right) \\ &< 1.92\sqrt{\varepsilon},\end{aligned}$$

where above we used that $1-(1-\varepsilon)^N \le \min(1, \varepsilon N) \le \sqrt{N\varepsilon}$.

Remark 5.4. *The Central Limit Theorem implies that* $\frac{1}{\sqrt{m}}\mathbb{E}[B_m - m/2] \to 1/\sqrt{2\pi}$ *as* $\varepsilon \to 0$ *and so*

$$\limsup_{N\to\infty} \sup_{t, w_1, \ldots, w_N} \mathbb{P}[f(X) \neq f(\sigma X)] \le (\sqrt{2/\pi} + o_\varepsilon(1))\sqrt{\varepsilon}.$$

Hence we cannot replace 1.92 *by anything smaller than* $\sqrt{2/\pi} = 0.797....$

5.2 Proof of the Main Result

First note that we can assume that $t = 0$ and $w_i > 0$. The main idea is to write $\mathbb{P}[f(X) \neq f(\sigma X)]$ as an expectation. Let $m = \lfloor \varepsilon^{-1} \rfloor$, consider a partition of $[N]$ into $m+1$ sets A_j for $j = 0, ..., m$ and define the sums

$$f_{A_j}(x) = \sum_{i \in A_j} w_i x_i,$$

with the convention that $f_\emptyset = 0$. We then select the sets A_j randomly by setting $A_j = \{i \in [N] : \tau_i = j\}$, where $\tau_1, ..., \tau_N$ are i.i.d. distributions such that $\mathbb{P}[\tau_i = j] = \varepsilon$ and $\mathbb{P}[\tau_i = 0] = 1 - \varepsilon m$. Y. Peres then shows the key identity

$$\mathbb{P}[f(X) \neq f(\sigma X)] = \frac{2}{m}\mathbb{E}\left[\sum_{1 \le j \le m \,:\, f_{A_j}(X) \neq 0} \left(\frac{1}{2} - \mathbf{1}_{\{\operatorname{sgn} f(X) = -\operatorname{sgn} f_{A_j}(X)\}}\right)\right].$$

Let Q be the quantity above inside the expectation. If $f(X) = 0$ then

$$Q = \frac{1}{2}\#\{1 \le j \le m : f_{A_j}(X) \neq 0\} \le \frac{1}{2}\#\{1 \le j \le m : A_j \neq \emptyset\}$$

and if $\operatorname{sgn} f(X) = \pm 1$ then

$$\begin{aligned}Q &= -\frac{1}{2}\#\{1 \le j \le m : f_{A_j}(X) = \mp 1\} + \frac{1}{2}\#\{1 \le j \le m : f_{A_j}(X) = \pm 1\} \\ &\le \left|\#\{1 \le j \le m : f_{A_j}(X) = 1\} - \frac{1}{2}\#\{1 \le j \le m : f_{A_j}(X) \neq 0\}\right|.\end{aligned}$$

Observing that $\#\{1 \le j \le m : f_{A_j}(X) = 1\} \equiv \mathrm{Bin}(\ell, \frac{1}{2})$ with $\ell = \#\{1 \le j \le m : f_{A_j}(X) \neq 0\}$ we conclude

$$\begin{aligned}\mathbb{P}[f(X) \neq f(\sigma X)] &\le \frac{2}{m}\mathbb{E}[|\mathrm{Bin}(\ell, \tfrac{1}{2}) - \ell/2|] + \mathbb{P}[A_1 \neq \emptyset]\mathbb{P}[f(X) \neq 0] \\ &\le \frac{2}{m}\mathbb{E}[|\mathrm{Bin}(m, \tfrac{1}{2}) - m/2|] + (1-(1-\varepsilon)^N)\mathbb{P}[f(X) \neq 0],\end{aligned}$$

where above we use that $\mathbb{E}[|\mathrm{Bin}(\ell, \frac{1}{2}) - \ell/2|]$ increases with ℓ. Since $w_i > 0$ for all i, the collection of subsets $D \subset [N]$ such that $f(\mathbf{1}_D - \mathbf{1}_{D^c}) = 0$ is an anti-chain (no set is a proper subset of another) and Sperner's Theorem guarantees that this collection has at most $\binom{N}{N/2}$ elements. Finally, noting that $f(X) = 0$ iff $f(\mathbf{1}_D - \mathbf{1}_{D^c}) = 0$ for $D = \{i : X_i = 1\}$ we conclude $\mathbb{P}[f(X) \neq 0] \le 2^{-N}\binom{N}{N/2}$. This finishes the proof.

Bibliography

[BKS01] I. Benjamini, G. Kalai and O. Schramm (2001), Noise sensitivity of Boolean functions and applications to percolation. Inst. Hautes Etudes Sci. Publ. Math., 90, 5-43.

[G66] G. Guilbaud (1966), Theories of the general interest, and the logical problem of aggregation. In Readings in Mathematical Social Science, edited by P. F. Lazarsfeld and N. W. Henry, MIT Press, 262–307.

[Y04] Peres Y. (2021), Noise Stability of Weighted Majority. In: Vares M.E., Fernández R., Fontes L.R., Newman C.M. (eds) In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius. Progress in Probability 77.

Felipe Gonçalves, University of Bonn
email: **goncalve@math.uni-bonn.de**

Chapter 6

Quantum Lower Bounds by Polynomials

after R. Beals, H. Buhrman, R. Cleave, M. Mosca and R. De Wolf. [1]
A summary written by Dylan Langharst

Abstract. We analyze a black-box model and determine the number of input variables a quantum algorithm in said model requires to compute Boolean functions on $\{0,1\}^N$. We show the exponential speed increase for partial functions from certain algorithms cannot be obtained for any total function. In the exact, zero-error and bounded-error settings, asymptotic estimates for T are given. These results are a quantum extension of the polynomial method.

6.1 Introduction and Definitions

A boolean variable is a variable that takes on values $0, 1$; an N-tuple of Boolean variables shall be denoted $X = (x_0, x_1, \ldots, x_{N-1})$. A *black-box model* is a form of computation where, given the input i, the black-box outputs the bit x_i. Accessing the bits x_i only through a black-box is called a query. A function $f : \{0,1\}^N \to \{0,1\}$ is a Boolean function, and is called a property of X. The goal of this paper is to compute such properties using as few queries as possible. Quantum mechanics allows a drastic increase in the efficiency of algorithms design to accomplish this task. For example, the computation

$$\mathrm{OR}_N(X) = x_0 \vee x_1, \vee \cdots \vee x_{N-1}$$

determines if any of bits x_i of X contain a 1, and classically (i.e. deterministic-ally or probabilistic-ally) has a computation time of $\Theta(\sqrt{N})$. However, by using the concept of superposition, Grover [2] was able to construct a quantum algorithm using only $O(\sqrt{N})$ queries; different i can be in super-positions, and so a query can access different input bits x_i, each with some probability amplitude, simultaneously.

A *promise* is a model with some constraint. For example: consider the black-box model with $N = n2^n$, then, query X n times. This creates a function, $\tilde{X} : \{0,1\}^n \to \{0,1\}^n$. Suppose we have the constraint, or promise, that there exists an $s \in \{0,1\}^n$ such that $\tilde{X}(i) = \tilde{X}(j)$ if, and only if, $i = j + s \bmod 2$ component wise. *Simon's problem* is, given this scenario, one must compute if s is the n-tuple of 0's. The quantum algorithm for such a task requires $O(n)$ applications of $\tilde{X}$, while classically, $\Omega(\sqrt{2^n})$ queries are required. The fact that there is a promise means that Simon's problem is *partial*, as the associated $f : \{0,1\}^N \to \{0,1\}$ is not defined on all $X \in \{0,1\}^N$, but only on those that satisfy the promise.

The goal of this paper is to establish upper and lower bounds for the black-box complexity of several functions and classes of functions in the quantum computing setting. In particular, it will be shown that the exponential speed-up, discussed in the example of the Simon problem, cannot be obtained for a quantum algorithm for an arbitrary total function; like in the $\mathrm{OR}_N(X)$ example, only a polynomial speed-up is possible in general. The main step is the translation of quantum algorithms that make T queries into multi-linear polynomials of degree at most $2T$ over N variables; this is a quantum extension

of the polynomial method. Three different settings for computing f on $\{0,1\}^N$ in the black-box model will be discussed:

The *exact setting* where an algorithm must return $f(X)$ with certainty for every X; the *zero-error setting* where, for every X, the result "inconclusive" can have probability at most $1/2$; when a result is returned, it must be exact; and the *two-sided bounded-error setting*, or Monte Carlo algorithm, where, for every X, an algorithm must return the correct answer with probability $> 2/3$.

Throughout, X will be an N-tuple, and with N an arbitrary positive integer unless specified. The *Hamming weight* of X is the number of 1's of X, denoted $|X|$. We say f is symmetric if $f(X)$ depends only on $|X|$. We will be interested in symmetric functions, non-symmetric functions and the functions AND, OR, PARITY, and MAJORITY. These functions are defined as follows: $\mathrm{OR}_N(X) = 1$ iff $|X| > 0$, $\mathrm{AND}_N(X) = 1$ iff $|X| = N$, $\mathrm{PARITY}_N(X) = 1$ iff $|X| = 1 \bmod 2$, and $\mathrm{MAJ}_N(X) = 1$ iff $|X| > N/2$.

A multilinear N-variate polynomial $p : \mathbb{R}^N \to \mathbb{R}$ represents a function f if $p(X) = f(X)$ for all $X \in \{0,1\}^N$. If such a p exists then it is unique and has degree $\leq N$; the degree is $\deg(f)$. If $|p(X) - f(X)| \leq \frac{1}{3}$ for all $X \in \{0,1\}^N$, then we say p approximates f, and $\widetilde{\deg}(f)$ is the degree of a minimum-degree polynomial p that approximates f. If S_N is the symmetry group of $\{0,1,\ldots,N-1\}$ and π is any permutation, then $\pi(X) = \{x_{\pi(0)},\ldots,x_{\pi(N-1)}\}$ and the symmetrization of a polynomial p is given by

$$p^{sym}(X) = \frac{\sum_{\pi \in S_N} p(\pi(X))}{N!} \tag{6.1}$$

6.2 Quantum Networks

Throughout we will assume f is a Boolean function on N-tuples X, and a black-box on i returns the bit x_i of X. A classical algorithm that computes f using black-box queries is a decision tree. The cost is the number of queries made on the worst-case input X. A *quantum network* with T queries is a string of unitary operations that changes the state of a quantum bit, or qubit, in the form

$$U_0, O_1, U_1, O_2, \ldots U_{T-1}, O_T, U_T$$

where U_i are arbitrary unitary transformations and O_j are unitary transformations corresponding to queries on X. If there are m qubits, and each qubit has base states $|0\rangle$ and $|1\rangle$, then there are 2^m basis states for each basis state of computation, denoted $|0\rangle, |1\rangle, \ldots, |2^m - 1\rangle$. If $K = \{0,1,\ldots,2^m-1\}$, then a superposition state ϕ is given by $\phi\rangle = \sum_{k\in K} \alpha_k |k\rangle$, $\alpha_k \in \mathbb{C}$ and $\sum_{k\in K} |\alpha_k|^2 = 1$; the probability of measuring $|k\rangle$ is $|\alpha_k|^2$. The initial state will always be taken to be $|0\rangle$. Unitary operations act in the following way: let $\bigoplus =$ addition mod 2 = exclusive-or. Then, if i is $\lceil \log N \rceil$ bits, b is one bit and z is $m - \lceil \log N \rceil - 1$ bits, O_j sends $|i,b,z\rangle$ to $O_j|i,b,z\rangle = |i, b \bigoplus x_i, z\rangle$. All O_j are equal.

The right-most qubit of the final state of a network is the output bit. If this output equals $f(X)$ with certainty for every X, then the network computes f exactly. If the output equals $f(X)$ with probability at least $2/3$, then the bounded error probability is said to be at most $1/3$. For the zero-error setting, the two rightmost qubits are observed. If the first qubit is 0, then the network outputs "inconclusive". Otherwise, the second qubit should contain $f(X)$ with certainty. The minimum number of queries required by a quantum network to compute f will be denoted $Q_E(f)$, $Q_0(f)$ and $Q_2(f)$ for exact, zero-error and bounded-error settings respectively.

6.3 General Lower Bounds on the Number of Queries

6.3.1 Peremptory Lemmas

Lemma 6.1. *Let $\mathcal{N}$ be a quantum network that makes T queries to a black-box X. Then, there exists complex-valued N-variate multilinear polynomials $p_0,\ldots,p_{2^m-1}$, each of degree at most T, such that the final state of the network is the superposition state*

$$\sum_{k\in K} p_k(X)|k\rangle$$

for any black-box X.

The main argument is that, if the amplitude of $|i,0,z\rangle$ is α and the amplitude of $|i,1,z,\rangle$ is β before a query, then after a query the amplitudes are $(1-x_1)\alpha + x_i\beta$ and $x_1\alpha + (1-x_i)\beta$ respectively (which are polynomials of degree 1); then, continue counting after each query. Furthermore, by splitting each polynomial p_k into real and imaginary parts, one obtains the following lemma:

Lemma 6.2. *Let $\mathcal{N}$ be a quantum network that makes T queries to a black-box X, and B be a set of basis states. Then, there exists a real-valued multilinear polynomial $P(X)$ of degree at most $2T$ which equals the probability that observing the final state of the network with black-box X yields a state from B.*

6.3.2 The exact and zero-error settings

Theorem 6.3. *If f is a Boolean function, then $Q_E(f) \geq \deg(f)/2$*

Proof. Suppose a quantum network computes f using exactly $T = Q_E(f)$ queries. Then, its acceptance polynomial has degree $deg(f)$. But from Lemma 6.2, this is bounded above by $2T$, and the conclusion follows. □

In [5], Nisan and Szegedy showed that, if f is a Boolean function that depends on N variables, then $deg(f) \geq \log N - O(\log\log N)$. Combining this and Theorem 6.3, we obtain the following.

Corollary 6.4. *If f depends on N variables, then $Q_E(f) \geq \log N/2 - O(\log\log N)$.*

Suppose $p : \mathbb{R}^n \to \mathbb{R}$ is a multilinear polynomial. Then, it was shown in [3] that there exists a polynomial $q : \mathbb{R} \to \mathbb{R}$ of degree at most $deg(p)$ such that $p^{sym}(X) = q(|X|)$ for all $X \in \{0,1\}^N$. Letting $T = Q_0(f)$ and using this fact and Lemma 6.2 on the set of basis states that have 11 as the rightmost bits, one obtains the following.

Theorem 6.5. *If f is non-constant and symmetric, then $Q_0(f) \geq (N+1)/4$.*

We conclude this section by remarking that the above yields like OR_N, AND_N, etc. require at least $(N+1)/4$ queries to be computed exactly or with zero-error on a quantum network. Thus, since N queries always suffice (even classically) one has, for all non-constant symmetric f that $Q_E(f), Q_O(f) \in \Theta(N)$.

6.3.3 Lower Bounds for Bounded-Error Quantum Computation

From the definition of bounded error, one immediately obtains the following.

Theorem 6.6. *If f is a Boolean function, then $Q_2(f) \geq \tilde{deg}(f)/2$.*

In the case of symmetric f, we can do better. Let f be symmetric, and denote $f_k = f(X)$ for $|X| = k$. Define

$$\Gamma(f) = \min\{|2k - N + 1| : f_k \neq f_{k+1} \text{ and } 0 \leq k \leq N-1\}. \tag{6.2}$$

Then, in [6], Paturi showed that, if f is a non-constant symmetric Boolean function on $\{0,1\}^N$, then $\tilde{deg}(f) \in \Theta(\sqrt{N(N-\Gamma(f))})$. Using this, one can show the following:

Theorem 6.7. *If f is non-constant and symmetric, then $Q_2(f) \in \Theta(\sqrt{N(N-\Gamma(f))})$.*

6.3.4 Lower Bounds from Block Sensitivity

In Section 6.3, we saw that the minimum number of queries for a quantum network in various settings is bounded below by degrees of polynomials. Instead of polynomials, one can introduce another method for bounding these quantities. Let $f : \{0,1\}^N \to \{0,1\}$ be a Boolean function, $X \in \{0,1\}^N$ and $B \subset \{0,\ldots,N-1\}$ a set of indices. Let X^B denote the string obtained from X by flipping the variables in B. We say f is sensitivity to B is $f(X) \neq f(X^B)$. The *block sensitivity* $b_{S_x}(f)$ of f on X is the maximum number t for which there exist t disjoint sets of indices $B_1,\ldots,B_t$ such that f is sensitive to each B_i on X. The block sensitivity of f is

$$b_s(f) = \max_{X\in\{0,1\}^N} b_{s_X}(f).$$

Through a series of counting arguments, one can show the following.

Theorem 6.8. *If f is a Boolean function, then*

$$Q_E(f) \geq \sqrt{\frac{b_s(f)}{8}} \quad \textit{and} \quad Q_2(f) \geq \sqrt{\frac{b_s(f)}{16}}.$$

6.4 Polynomial Relation for Classical and Quantum Complexity and Specific Functions

Let $D(f)$ be the decision tree complexity $D(f)$ of f, that is the cost of the best decision tree that (classically) computes f. Similarly, let $R(f)$ be the worst-case number of queries for randomized algorithms that computes (classically) $f(X)$ with error probability $\leq 1/3$ for all X. We will state, without proof, various relations between $D(f), R(f), Q_E(f), Q_0(f)$, and $Q_2(f)$.

Theorem 6.9. *Let $f : \{0,1\}^N \to \{0,1\}$ be a Boolean function, $X \in \{0,1\}^N$. Then, the following hold (some of which are already known):*

1. $D(f) \in O(R(f)^3)$, *shown by Nisan in [4]*
2. $D(f) \leq b_s(f)^3$
3. $b_s(f) \leq 16Q_2(f)^2$ *(Theorem 6.8)*
4. $D(f) \leq 4096Q_2(f)^6$
5. *If f is monotone, then* $D(f) \leq 256Q_2(f)^4$
6. $D(f) \leq 32Q_E(f)^4$
7. $\tilde{deg}(f) \leq D(f) \leq 216\tilde{deg}(f)^6$
8. $Q_2(f) \leq Q_0(f) \leq Q_E(f) \leq D(f) \leq N$
9. $Q_2(f) \leq R(f) \leq D(f) \leq N$

We remark that Item Four in Theorem 6.9 implies that if a quantum algorithm computes f with bounded-error probability using T queries, then the corresponding classical algorithm needs at most $O(T^6)$ queries. Item Five states, if f is monotonically increasing (decreasing), that is changing any input bit from 0 to 1 causes an increase (decrease), then one only needs $O(T^4)$ queries. Furthermore, if f is symmetric, then Theorem 6.7 yields $Q_2(f) \in \Omega(\sqrt{N})$ and so the classical algorithm only needs $O(T^2)$ queries. We conclude by stating the following calculations for specific functions.

	Exact ($Q_E(f)$)	Zero-error ($Q_0(f)$)	Bounded-error ($Q_2(f)$)
$\text{OR}_N, \text{AND}_N$	N	N	$\Theta(\sqrt{N})$
PARITY_N	$\lceil N/2 \rceil$	$\lceil N/2 \rceil$	$\lceil N/2 \rceil$
MAJ_N	$\Theta(\sqrt{N})$	$\Theta(\sqrt{N})$	$\Theta(\sqrt{N})$

Bibliography

[1] Beals, R., Buhrman, H., Cleave, R., Mosca, M., and De Wolf, R., *Quantum Lower Bounds by Polynomials*, ACM. 48 (4), pp. 778-797 (2001),

[2] Grover, L.K., *A fast quantum mechanical algorithm for database search*, In *Proceedings of 28th Annual ACM Symposium on Theory of Computing* (Philadelphia, PA, May 22-24). ACM, New York, pp. 212-219 (1996),

[3] Minsky, M., and Papert, S., *Perceptrons*. MIT Press, Cambridge, Mass., Second, expanded edition 1988.

[4] Nisan, N., *CREW PRAMs and decision trees*, SIAM J. Comput. 20,6, 999-1007 (1991),

[5] Nisan, N., and Szegedy, M., *On the degree of Boolean functions as real polynomials*, Computat. Complex. 4,4 301-313 (1994),

[6] Paturi, R., *On the degree of polynomials that approximate symmetric Boolean functions (preliminary version).* In *Proceedings of 24th Annual ACM Symposium on Theory of Computing* (Victoria, B.C., Canada, May 4-6). ACM, New York, pp. 468-474 (1992),

Dylan Langharst, KSU
email: **dlanghar@kent.edu**

Chapter 7

Complexity measures and decision tree complexity: a survey

after H. Buhrman and R. de Wolf [1]
A summary written by Haojian Li

Abstract. We introduce various complexity measures of Boolean functions and compare their relations. We survey how they give bounds to the complexity measures of deterministic, randomized, and quantum decision trees of Boolean functions.

7.1 Complexity Measures of Boolean Functions

A Boolean function is a function $f : \{0,1\}^n \to \{0,1\}$. We call $x = x_1 \dots x_n \in \{0,1\}^n$ a Boolean input. Let x_i and $|x|$ denote the i-th bit and the Hamming weight of x (the number of 1's), respectively. For $S \subset \{1,\dots,n\}$, we use the notation x^S to mean x with the i-th input flipped for $i \in S$ and abbreviate $x^{\{i\}}$ to x^i. The certificate for an input $x \in \{0,1\}^n$ to a Boolean function is an index set $S \subset \{1,\dots,n\}$ such that $x_i = y_i$ for all $i \in S$ implies that $f(x) = f(y)$. Certificate complexity $C(f)$ of f is defined as $C(f) = \max_x C_x(f)$, where $C_x(f)$ is the size of the smallest certificate S for x. Certificate complexity captures how many input bits that one must query to ascertain the output of the function. Sensitivity $s(f)$ of f is the maximum of the numbers of i-th bits such that $f(x^i) \neq f(x)$ for any Boolean input $x \in \{0,1\}^n$. Sensitivity describes how unstable the output of the function is to perturbations (changes) to the bits in the input. Block sensitivity $bs(f)$ of f is defined as $bs(f) = \max_x bs_x(f)$, where $bs_x(f)$ is the maximum number b of disjoint sets $B_1,\dots,B_b \subset \{1,\dots,n\}$ for which $f(x) \neq f(x^{B_j})$. A simple relationship between certificate complexity and sensitivity is that

$$s(f) \leq bs(f) \leq C(f). \tag{7.1}$$

It was also proved in [2] that

$$C(f) \leq s(f)\, bs(f). \tag{7.2}$$

A long-standing conjecture is whether block sensitivity can bounded by a polynomial in sensitivity.

Conjecture 7.1 (Sensitivity conjecture)**.** *Does there exist a universal constant $k > 0$ such that for all Boolean functions f,*

$$bs(f) = \mathcal{O}(s(f)^k).$$

The monomial X_S of the index set $S \subset \{1,\dots,n\}$ is defined as the product of variables $X_S = \prod_{i \in S} x_i$. If a function $p : \mathbb{R}^n \to \mathbb{C}$ can be written $p(x) = \sum_S c_S X_S$ for $c_S \in \mathbb{C}$, then we call p a multilinear polynomial with degree $deg(p) = \max\{|S| \; | c_S \neq 0\}$. A polynomial $p : \mathbb{R}^n \to \mathbb{R}$ represents f if $p(x) = f(x)$ on all Boolean inputs x. Every Boolean function can be represented by a unique multilinear polynomial

$p : \mathbb{R}^n \to \mathbb{R}$, and we define the degree $deg(f)$ as the degree $deg(p)$ of the multilinear polynomial p that represents f. It was proved in [3] that

$$deg(f) \geq \log n - \mathcal{O}(\log \log n) \tag{7.3}$$

if f depends on all n variables. The approximate degree $\widetilde{deg}(f)$ of f is defined as the minimum degree of any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ such that $|p(x) - f(x)| \leq \frac{1}{3}$ for any Boolean input $x \in \{0,1\}^n$. Ambainis ([4]) proved that almost all functions f have high approximate degree

$$\widetilde{deg}(f) \geq n/2 - \mathcal{O}(\sqrt{n} \log n). \tag{7.4}$$

In 1994, Nisan and Szegedy ([3]) pointed out that

$$bs(f) \leq 2\, deg(f)^2 \tag{7.5}$$

and

$$bs(f) \leq 6\, \widetilde{deg}(f)^2 \tag{7.6}$$

for any Boolean function f. They also put forth the conjecture that sensitivity is bounded below by a polynomial of any other complexity measure, which has been resolved recently in [5].

Theorem 7.2. *For any Boolean function f, $s(f) \geq \sqrt{deg(f)}$.*

7.2 Decision Trees

A deterministic decision tree for a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is a rooted ordered binary tree, where each internal node is assigned with an input bit x_i and each leaf is assigned with either 0 or 1. We proceed the computation by querying the input bit assigned to the root, which lead to the left (right) sub-tree if the returned value is 0 (1, respectively). We repeat the procedure recursively till we reach the leaf. A decision tree is said to compute f if the outputs of the tree coincide with the outputs of f for all Boolean inputs. Decision tree complexity $D(f)$ of f is the minimal depth of trees that compute f. We can add randomness to the decision tree by including a coin flip node with bias $p \in (0,1)$. We reach the left (right) sub-tree if the outcome of the coin flip is head (tail). Such a tree is said to compute f with bounded-error if the outcome of the tree equals $f(x)$ with probability at least $2/3$ for all Boolean inputs x. The corresponding tree complexity $R_2(f)$ is the minimal depth of trees that compute f with bounded-error. The quantum decision tree is usually referred to as *quantum query algorithm* or *quantum black-box algorithm* in the literature, where we work with qubits instead of classical binary bits. A T-quantum decision tree is defined by a initial state $|0\rangle$ and a series unitary transformations $U_0, O, U_1, \ldots, 0, U_T$, where O is the query unitary transformation. Here U_i are independent of the choice of the Boolean input. The output of the quantum decision tree only depend on querying the input T times via O. The quantum decision tree is said to compute the function f exactly if the output of the quantum algorithm coincide with $f(x)$ for any Boolean input x. It is said to compute the function f with bounded-error if the output of the quantum algorithm equals $f(x)$ with probability at least $2/3$ for any Boolean input x. Let $Q_E(f)$ and $Q_2(f)$ be the minimal number of queries if a quantum decision tree that compute f exactly and with bounded-error, respectively. Every T-query deterministic decision tree can be simulated by a T-query quantum decision tree without error, and every T-query randomized decision tree can be simulated by a T-query quantum decision tree with bounded-error. Thus we have

$$Q_2(f) \leq R_2(f) \leq D(f) \leq n \tag{7.7}$$

and

$$Q_2(f) \leq Q_E(f) \leq D(f) \leq n \tag{7.8}$$

for any Boolean function f.

7.3 Applications to Decision Tree Complexity

A natural question is how the complexity measures $C(f)$, $s(f)$, $bs(f)$, $deg(f)$, and $\widetilde{deg}(f)$ of Boolean functions are related to the decision trees complexity $D(f)$, $R_2(f)$, $Q_E(f)$, and $Q_2(f)$. It has been proved that those complexity measures are all polynomially related. We first summarize relationship between the deterministic decision tree complexity $D(f)$ and the complexity measures of Boolean functions f. For any Boolean function f, we have

1. $s(f) \leq bs(f) \leq D(f)$ and $deg(f) \leq D(f)$;
2. $D(f) \leq s(f)bs(f)^2 \leq bs(f)^3$ ([2]);
3. $D(f) = \mathcal{O}(\widetilde{deg}(f)^6)$ ([3]).

It still remains unknown whether we can bound $D(f)$ by block sensitivity quadratically. Let $f : \{0,1\}^{k^2} \to \{0,1\}$ be the AND of k ORs of k variables each, then $D(f) = bs(f)^2 = n$. Thus the optimal scenario is $D(f) = bs(f)^2$ for any Boolean function f.

Conjecture 7.3. *Does $D(f) = \mathcal{O}(bs(f)^2)$ for every Boolean function f?*

Degree of any Boolean function f is bounded above by $R_2(f)$ cubically: ([2])

$$D(f) \leq 27R_2(f)^3. \tag{7.9}$$

This gap is not optimal, and biggest gap between $D(f)$ and $R_2(f)$ still remains a conjecture. $R_2(f)$ is bounded below by approximate degree and block sensitivity ([2]):

$$\widetilde{deg}(f) \leq R_2(f) \tag{7.10}$$

and

$$bs(f) \leq 3\, R_2(f). \tag{7.11}$$

$Q_E(f)$ and $Q_2(f)$ are the quantum analogue of $D(f)$ and $R_2(f)$, respectively, and the biggest gap between $Q_E(f)$ and $Q_2(f)$ remains unknown as well.

Conjecture 7.4. *What are the biggest gaps between the classical $D(f)$, $R_2(f)$ and their quantum analogue $Q_E(f)$, $Q_2(f)$?*

Following the spirit of the proof of $deg(f) \leq D(f)$, we can show that

$$deg(f) \leq 2Q_E(f) \tag{7.12}$$

and

$$\widetilde{deg}(f) \leq 2Q_2(f) \tag{7.13}$$

for any Boolean function. Another conjecture is whether $Q_E(f) = \mathcal{O}(deg(f))$ and $Q_2(f) = \mathcal{O}(\widetilde{deg}(f))$ for any Boolean function f.

Bibliography

[1] Buhrman, H. and de Wolf, R., *Complexity measures and decision tree complexity: a survey.* Theoretical Computer Science, 288(1), 21-43 (2002).

[2] Nisan, N., *CREW PRAMs and decision trees.* SIAM Journal on Computing, 20(6), 999-1007 (1991).

[3] Nisan, N. and Szegedy, M., *On the degree of Boolean functions as real polynomials.* Computational complexity, 4(4), 301-313 (1994).

[4] Ambainis, A., *A note on quantum black-box complexity of almost all Boolean functions.* Information Processing Letters, 71(1), 5-7(1999).

[5] Huang, H., *Induced subgraphs of hypercubes and a proof of the sensitivity conjecture.* Annals of Mathematics, 190(3), 949-955 (2019).

Haojian Li, Baylor University
email: **lihaojianmath@gmail.com**

Chapter 8

Vector-valued Talagrand influence inequalities

after D. Cordero-Erausquin and A. Eskenazis [2]
A summary written by Sang Woo Ryoo

Abstract. Talagrand's influence inequality is an enhancement of the discrete Poincaré inequality for real-valued functions on the discrete hypercube. We state and prove Talagrand-type inequalities for functions on the discrete hypercube taking values in Banach spaces of Rademacher or martingale type 2. The proof builds upon the work of Ivanisvili, van Handel, and Volberg (2020), who proved the discrete Poincaré inequality for functions taking values in Banach spaces of Rademacher type 2, and uses Bonami's hypercontractive inequality and a vector-valued Littlewood-Paley-Stein due to Xu (2020).

8.1 Introduction

Let $C_n = \{-1,1\}^n$ be the discrete hypercube, and let σ_n be the uniform probability measure on C_n. If $(E, \|\cdot\|_E)$ is a Banach space and $p \geq 1$, then we denote the $L_p(\sigma_n; E)$ norm of a function $f : C_n \to E$ by

$$\|f\|_{L_p(\sigma_n;E)} = \left(\int_{C_n} \|f(\varepsilon)\|_E^p d\sigma_n(\varepsilon)\right)^{1/p}.$$

We define the i-th partial discrete derivative of f by

$$\partial_i f(\varepsilon) = \frac{f(\varepsilon) - f(\varepsilon_1, \cdots, -\varepsilon_i, \cdots, \varepsilon_n)}{2}.$$

When $E = \mathbb{C}$, the discrete Poincaré inequality tells us that for $f : C_n \to \mathbb{C}$,

$$\|f - \mathbb{E}_{\sigma_n} f\|_{L_2(\sigma_n;\mathbb{C})}^2 \leq \sum_{i=1}^n \|\partial_i f\|_{L_2(\sigma_n;\mathbb{C})}^2. \tag{8.1}$$

Talagrand's influence inequality [5] provides an asymptotic improvement over the discrete Poincaré inequality: there exists $C > 0$ such that for all $f : C_n \to \mathbb{C}$,

$$\|f - \mathbb{E}_{\sigma_n;\mathbb{C}} f\|_{L_2(\sigma_n;\mathbb{C})}^2 \leq C \sum_{i=1}^n \frac{\|\partial_i f\|_{L_2(\sigma_n;\mathbb{C})}^2}{1 + \log(\|\partial_i f\|_{L_2(\sigma_n;\mathbb{C})} / \|\partial_i f\|_{L_1(\sigma_n;\mathbb{C})})}. \tag{8.2}$$

One may inquire whether analogous phenomena happen for general Banach spaces E. At the very least, we should require (8.1) to be true (up to constant factors) for linear functions $f(\varepsilon) = \sum_{i=1}^n \varepsilon_i x_i$, $x_i \in E$: there should exist $T > 0$ such that

$$\int_{C_n} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_E^2 d\sigma_n(\varepsilon) \leq T^2 \sum_{i=1}^n \|x_i\|_E^2, \quad \forall n \in \mathbb{N},\ x_1, \cdots, x_n \in E. \tag{8.3}$$

We say that E has *Rademacher type* 2 *with constant* T if (8.3) holds. The recent breakthrough of Ivanisvili, van Handel, and Volberg [3] asserts that then the discrete Poincaré inequality holds:

Theorem 8.1 ([3])**.** *There is a universal constant $C > 0$ such that the following is true. Let $(E, \|\cdot\|_E)$ be a Banach space having Rademacher type 2 with constant T. Then for any $n \in \mathbb{N}$, $f : C_n \to E$,*

$$\|f - \mathbb{E}_{\sigma_n} f\|^2_{L_2(\sigma_n;E)} \le CT^2 \sum_{i=1}^n \|\partial_i f\|^2_{L_2(\sigma_n;E)}.$$

Cordero-Erausquin and Eskenazis [2] enhance the approach of [3] to prove a near-optimal analogue of Talagrand's influence inequality.

Theorem 8.2 ([2], Theorem 1)**.** *Let $(E, \|\cdot\|_E)$ be a Banach space with Rademacher type 2. Then, there exists $C(E) \in (0,\infty)$ such that for every $\epsilon \in (0,1)$, $n \in \mathbb{N}$, and $f : C_n \to E$,*

$$\|f - \mathbb{E}_{\sigma_n} f\|^2_{L_2(\sigma_n,E)} \le \frac{C(E)}{\epsilon} \sum_{i=1}^n \frac{\|\partial_i f\|^2_{L_2(\sigma_n;E)}}{1 + \log^{1-\epsilon}\left(\|\partial_i f\|_{L_2(\sigma_n;E)}/\|\partial_i f\|_{L_1(\sigma_n;E)}\right)}. \tag{8.4}$$

The choice of $\epsilon = 1/\sigma(f)$, where $\sigma(f) := \max_i \log\log\left(e + \|\partial_i f\|_{L_2(\sigma_n;E)}/\|\partial_i f\|_{L_1(\sigma_n;E)}\right)$, gives

$$\|f - \mathbb{E}_{\sigma_n} f\|^2_{L_2(\sigma_n,E)} \le C(E)\sigma(f) \sum_{i=1}^n \frac{\|\partial_i f\|^2_{L_2(\sigma_n;E)}}{1 + \log\left(\|\partial_i f\|_{L_2(\sigma_n;E)}/\|\partial_i f\|_{L_1(\sigma_n;E)}\right)}.$$

It is unknown whether the proper Talagrand influence inequality (8.2) holds for Banach spaces with Rademacher type 2. It does hold, however, under the stronger assumption that E has *martingale type 2*, i.e., there exists $M > 0$ such that for every $n \in \mathbb{N}$, probability space $(\Omega, \mathcal{F}, \mu)$, and filtration $\{\mathcal{F}_i\}_{i=0}^n$ of σ-algebras, every E-valued martingale $\{\mathcal{M}_i : \Omega \to E\}_{i=0}^n$ adapted to $\{\mathcal{F}_i\}_{i=0}^n$ satisfies

$$\|\mathcal{M}_n - \mathcal{M}_0\|^2_{L_2(\mu;E)} \le M^2 \sum_{i=1}^n \|\mathcal{M}_i - \mathcal{M}_{i-1}\|^2_{L_2(\mu;E)}.$$

Theorem 8.3 ([2], Theorem 2)**.** *Let E be a Banach space with martingale type 2. Then, there exists $C(E) \in (0,\infty)$ such that for every $n \in \mathbb{N}$ and $f : C_n \to E$,*

$$\|f - \mathbb{E}_{\sigma_n} f\|^2_{L_2(\sigma_n,E)} \le C(E) \sum_{i=1}^n \frac{\|\partial_i f\|^2_{L_2(\sigma_n;E)}}{1 + \log\left(\|\partial_i f\|_{L_2(\sigma_n;E)}/\|\partial_i f\|_{L_1(\sigma_n;E)}\right)}. \tag{8.5}$$

We will prove Theorem 8.2 in section 8.2 and Theorem 8.3 in section 8.3.

8.2 Proof of Theorem 8.2

We will first sketch the proof of Theorem 8.1 given by [3], and then describe the modifications made by [2] which lead to Theorem 8.2.

We consider the heat flow on C_n relative to the Laplacian $\Delta = -\sum_{i=1}^n \partial_i^2 = \sum_{i=1}^n \partial_i$. The heat kernel at time t is given by the random vector

$$\xi(t) = (\xi_1(t), \cdots, \xi_n(t)) \in C_n, \quad \mathbb{P}\{\xi_i(t) = \pm 1\} = \frac{1 \pm e^{-t}}{2}$$

whose coordinates are independent, so that the time-t evolute of $f : C_n \to E$ is $P_t f(\varepsilon) = \mathbb{E}_{\xi(t)} f(\varepsilon\xi)$. We also denote the centered normalization $\delta(t) = (\delta_1(t), \cdots, \delta_n(t))$ of $\xi(t)$:

$$\delta_i(t) = \frac{\xi_i(t) - \mathbb{E}\xi_i(t)}{\sqrt{\mathrm{Var}\xi_i(t)}} = \frac{\xi_i(t) - e^{-t}}{\sqrt{1 - e^{-2t}}}, \quad i = 1, \cdots, n.$$

The key idea of [2] is as follows. First, we have the identity

$$\frac{\partial}{\partial t} P_t f(\varepsilon) = -\frac{1}{\sqrt{e^{2t} - 1}} \mathbb{E}_{\xi(t)} \left[\sum_{i=1}^n \delta_i(t) \partial_i f(\varepsilon\xi(t))\right]$$

by a straightforward computation, and so by convexity

$$\left\|\frac{\partial}{\partial t}P_t f\right\|_{L_2(\sigma_n;E)} = \frac{1}{\sqrt{e^{2t}-1}}\left(\mathbb{E}_{\varepsilon,\xi(t)}\left\|\sum_{i=1}^n \delta_i(t)\partial_i f(\varepsilon)\right\|_E^2\right)^{1/2},$$

where we used that $(\varepsilon,\xi(t)) \stackrel{d}{=} (\varepsilon,\varepsilon\xi(t))$. Due to a result by Ledoux and Talagrand [4, Proposition 9.11], since δ_i are centered and normalized, we may use the type condition on the expectation with a slightly worse constant:

$$\mathbb{E}_{\varepsilon,\xi(t)}\left\|\sum_{i=1}^n \delta_i(t)\partial_i f(\varepsilon)\right\|_E^2 \le (2T)^2\sum_{i=1}^n \|\partial_i f\|^2_{L_2(\sigma_n;E)},$$

and so

$$\left\|\frac{\partial}{\partial t}P_t f\right\|_{L_2(\sigma_n;E)} \le \frac{2T}{\sqrt{e^{2t}-1}}\left(\sum_{i=1}^n \|\partial_i f\|^2_{L_2(\sigma_n;E)}\right)^{1/2}. \tag{8.6}$$

Therefore

$$\begin{aligned}\|f-\mathbb{E}_{\sigma_n}f\|_{L_2(\sigma_n;E)} &\le \int_0^\infty \left\|\frac{\partial}{\partial t}P_t f\right\|_{L_2(\sigma_n;E)} dt\\ &\le 2T\left(\sum_{i=1}^n \|\partial_i f\|^2_{L_2(\sigma_n;E)}\right)^{1/2}\int_0^\infty \frac{dt}{\sqrt{e^{2t}-1}}\\ &= CT\left(\sum_{i=1}^n \|\partial_i f\|^2_{L_2(\sigma_n;E)}\right)^{1/2}.\end{aligned}$$

The idea of [2] is to replace f by $P_t f$ in (8.6):

$$\|\Delta P_{2t} f\|_{L_2(\sigma_n;E)} \le \frac{2T}{\sqrt{e^{2t}-1}}\left(\sum_{i=1}^n \|P_t\partial_i f\|^2_{L_2(\sigma_n;E)}\right)^{1/2}$$

(we replace $\frac{\partial}{\partial_t}$ by Δ to avoid confusion, and we used the fact that P_t and ∂_t commute), and then apply Bonami's hypercontractive inequality [1]

$$\|P_t g\|_{L_2(\sigma_n;E)} \le \|g\|_{L_{1+e^{-2t}}(\sigma_n;E)}, \quad \forall g: C_n \to E,$$

to obtain

$$\|\Delta P_{2t} f\|_{L_2(\sigma_n;E)} \le \frac{2T}{\sqrt{e^{2t}-1}}\left(\sum_{i=1}^n \|\partial_i f\|^2_{L_{1+e^{-2t}}(\sigma_n;E)}\right)^{1/2}. \tag{8.7}$$

Then

$$\begin{aligned}&\|f-\mathbb{E}_{\sigma_n}f\|_{L_2(\sigma_n;E)}\\ &\le 2\int_0^\infty \|\Delta P_{2t}f\|_{L_2(\sigma_n;E)}\,dt\\ &\le 4T\int_0^\infty \left(\sum_{i=1}^n \|\partial_i f\|^2_{L_{1+e^{-2t}}(\sigma_n;E)}\right)^{1/2}\frac{dt}{\sqrt{e^{2t}-1}}\\ &\le 4T\left(\int_0^\infty \sum_{i=1}^n \|\partial_i f\|^2_{L_{1+e^{-2t}}(\sigma_n;E)}\frac{dt}{(e^{2t}-1)^\epsilon}\right)^{1/2}\left(\int_0^\infty \frac{dt}{(e^{2t}-1)^{1-\epsilon}}\right)^{1/2}\\ &\lesssim \frac{T}{\sqrt{\epsilon}}\left(\sum_{i=1}^n \int_0^\infty \|\partial_i f\|^2_{L_{1+e^{-2t}}(\sigma_n;E)}t^{-\epsilon}e^{-\epsilon t}dt\right)^{1/2} \qquad (\because e^{2t}-1 \ge te^t).\end{aligned}$$

One can show by calculus that for any $g : C_n \to E$,

$$\int_0^\infty \|g\|^2_{L_{1+e^{-2t}}(\sigma_n;E)} t^{-\epsilon} e^{-\epsilon t} dt \lesssim \frac{\|g\|^2_{L_2(\sigma_n;E)}}{1 + \log^{1-\epsilon}(\|g\|_{L_2(\sigma_n;E)}/\|g\|_{L_1(\sigma_n;E)})}.$$

Thus (8.4) follows.

8.3 Proof of Theorem 8.3

The starting point of the proof is the following vector-valued Littlewood-Paley-Stein inequality due to Xu [6]:

Theorem 8.4 ([6], Theorem 2)**.** *Let $(E, \|\cdot\|_E)$ be a Banach space with martingale type 2. Then there exists $C(E) > 0$ such that for a symmetric diffusion semigroup $\{T_t\}_{t\geq 0}$ on a probability space (Ω, μ), every function $f : \Omega \to E$ satisfies*

$$\|f - \mathbb{E}_\mu f\|^2_{L_2(\mu;E)} \leq C(E)^2 \int_0^\infty \|t\partial_t T_t f\|^2_{L_2(\mu;E)} \frac{dt}{t}.$$

We now proceed with (8.7):

$$\begin{aligned}\|f - \mathbb{E}_{\sigma_n} f\|^2_{L_2(\sigma_n;E)} &\leq 4C(E)^2 \int_0^\infty \|t\Delta P_{2t} f\|^2_{L_2(\sigma_n;E)} \frac{dt}{t} \\ &\overset{(8.7),\ 2t\leq e^t - e^{-t}}{\leq} 8C(E)^2 T^2 \int_0^\infty e^{-t} \sum_{i=1}^n \|\partial_i f\|^2_{L_{1+e^{-2t}}(\sigma_n;E)} dt.\end{aligned}$$

One can show by calculus that for any $g : C_n \to E$,

$$\int_0^\infty \|g\|^2_{L_{1+e^{-2t}}(\sigma_n;E)} e^{-t} dt \lesssim \frac{\|g\|^2_{L_2(\sigma_n;E)}}{1 + \log(\|g\|_{L_2(\sigma_n;E)}/\|g\|_{L_1(\sigma_n;E)})}.$$

Thus (8.5) follows.

Bibliography

[1] Bonami, A., *Étude des coefficients de Fourier des fonctions de $L^p(G)$*. Ann. Inst. Fourier (Grenoble), 20(fasc.2):335-402 (1971)

[2] Cordero-Erausquin, D. and Eskenazis, A., *Talagrand's influence inequality revisited.* To appear in Analysis & PDE.

[3] Ivanisvili, P., van Handel, R., and Volberg, A., *Rademacher type and Enflo type coincide.* Ann. of Math. (2), 192(2):665-678 (2020)

[4] Ledoux, M. and Talagrand, M., *Probability in Banach spaces: Isoperimetry and processes*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin (1991)

[5] Talagrand, M., *On Russo's approximate zero-one law.* Ann. Probab., 22(3):1576-1587 (1994)

[6] Xu, Q., *Vector-valued Littlewood-Paley-Stein theory for semigroups II.* Int. Math. Res. Not., 2020(21):7769–7791 (2020)

Sang Woo Ryoo, Princeton University
email: **sryoo@princeton.edu**

Chapter 9

On Russo's Approximate Zero One Law

after M. Talagrand [1]
A summary written by Yonathan Stone

Abstract. We outline a paper by Michel Talagrand in which he proves the existence of a 'threshold' effect for the measures of sufficiently nice subsets of the discrete cube as the mass of the cube becomes more concentrated towards a single vertex. Some of the more informative proofs are explained in detail while others are more tersely summarized.

Given $p \in [0,1]$, consider the product measure μ_p on the discrete cube $\{0,1\}^n$ in which 0 is given weight $1-p$ and 1 is given weight p, i.e., considering $x = (x_1, ..., x_n) \in \{0,1\}^n$, and writing $|x| = \sum x_i$, we have that $\mu_p(\{x\}) = (1-p)^{n-|x|}p^{|x|}$. In his paper "On Russo's Approximate Zero-One Law," Michel Talagrand investigates a so-called *"threshhold effect"* in this measure, namely that for specific types of subsets A of the discrete cube, the measure $\mu_p(A)$ increases from near 0 to near 1 as p varies within a very small neighborhood of $[0,1]$. In all cases, we assume A to be a *monotone* subset of the discrete cube, that is that for any point $x \in A$, any other point $y \in \{0,1\}^n$ whose coordinates pointwise dominate those of x must also be in A. On a purely intuitive level, this threshold effect has been demonstrated to exist for subsets that are essentially determined by very few coordinates[1]. The author expands on a result by Russo in which the threshold effect exists as soon as A depends little on any given coordinate, although he adapts Russo's definition as follows. Given $x = (x_1, ..., x_n) \in \{0,1\}^n$, let $U_i(x) = (x_1, ..., x_{i-1}, 1-x_i, x_{i+1}, ..., x_n)$ and set $A_i = \{x \in \{0,1\}^n; x \in A, U_i(x) \notin A\}$. Since by definition a monotone subset A of the discrete cube must contain $U_i(x)$ if $x_i = 0$ and $x \in A$, the set A_i gives us some idea of which points in A are in A without their existence being required by the presence of the point "directly underneath x" in the i^{th} coordinate direction. This encodes the idea of "points in A that don't depend on the i^{th} coordinate". This brings us to the primary result Talagrand presents in the paper

Theorem. *There exists a universal constant K, such that, for any p and any monotone subset A of $\{0,1\}^n$, we have*

$$\mu_p(A)(1-\mu_p(A)) \leq K(1-p)\log\frac{2}{p(1-p)}\sum_{i\leq n}\frac{\mu_p(A_i)}{\log[1/((1-p)\mu_p(A_i))]} \tag{9.1}$$

which gives rise to the following corollaries:

Corollary. *Let $\varepsilon = \sup_i \mu_p(A_i)$. Then*

$$\frac{d\mu_p(A)}{dp} \frac{\log(1/\varepsilon)}{Kp(1-p)\log[2/(p(1-p))]}\mu_p(A)(1-\mu_p(A)).$$

[1]It is useful to think about this in terms of the rule regarding monotone subsets and then finding the minimal number of points needed to generate the entire subset using this rule.

Corollary. *Let* $\varepsilon' = \sup_{0\le p\le 1} \sup_i \mu_p(A_i)$. *Then, for* $p_1 < p_2$, *we have*

$$\mu_{p_1}(A)(1-\mu_{p_2}(A)) \le (\varepsilon')^{(p_2-p_1)/K'},$$

where K' *is universal.*

Corollary. *We have*

$$\sup_{i\le n} \mu_p(A_i) \ge \frac{1}{K'(1-p)} U \log\frac{1}{U},$$

where K' is universal and where $U = \mu_p(A)(1-\mu_p(A))/(n\log(2/p(1-p)))$

In essence all these corrolaries illustrate that the presence of a threshold effect (encoded as the restrictions on the quantity $\mu_p(A)(1-\mu_p(A))$ is more quantifiable the less A depends on any given coordinate (the dependence on the i-th coordinate is itself encoded by the magnitudes $\mu_p(A_i)$. In the case where $p = \frac{1}{2} = \mu_p(A)$, one can prove Corollary 3 using harmonic analysis. Talagrand adapts these ideas in order to prove an analogous result in this more general setting where the techniques of harmonic analysis are unavailable. Moreover, since Theorem 1 doesn't concern itself with the specifics of the set A too much, it can be derived from the following more general result concerning functions on $\{0,1\}^n$: Given $f : \{0,1\}^n \to \mathbb{R}$, set $\Delta_i f(x) = (1-p)(f(x) - f(U_i(x)))$ if $x_i = 1$, and $\Delta_i f(x) = p(f(x) - f(U_i(x)))$ if $x_i = 0$. Then

Theorem. *For some numerical constant* K *and each function* $f : \{0,1\}^n \to \mathbb{R}$ *such that* $\int f d\mu_p = 0$, *we have*

$$||f||_2^2 \le K \log\frac{2}{p(1-p)} \sum_i \frac{||\Delta_i f||_2^2}{\log(e||\Delta_i f||_2/||\Delta_i f||_1)}. \tag{9.2}$$

Here $||f||_q$ denotes the $L_q(\mu_p)$ norm. Theorem 1 is an immediate consequence of Theorem 2 as soon as one observes that for $f = 1_A - \mu_p(A)$, we have $||f||_2^2 = \mu_p(A)(1-\mu_p(A))$ and $||\Delta_i f||_q^q = p^{-1}\mu_p(A_i)(p(1-p)^q + (1-p)p^q)$. In addition Talagrand proves another estimate that improves upon the result in Theorem 2, although Theorem 2 is still included for its ease of understanding and sufficiency in deducing Theorem 1. For this, let $\varphi(x) = \frac{x^2}{\log(e+x)}$ for $x \ge 1$. For a function f we will consider the following Orlicz norm:

$$||f||_\varphi = \inf\left\{c > 0; \int \varphi\left(\frac{f}{c}\right) \le 1\right\}$$

which is used in the following result:

Theorem. *There is a universal constant* K *such that for each* $f : \{0,1\}^n \to \mathbb{R}$ *with* $\int f d\mu_p = 0$, *we have*

$$||f||_2^2 \le K \log\frac{2}{p(1-p)} \sum_i ||\Delta_i f||_\varphi^2.$$

The introduction concludes with a proof of the following claim:

Claim. *For each p, the estimate in Theorem 1 is sharp.*

Proof. Case 1: $p < \frac{1}{2}$. Let $k \ge 1$ and assume that $r = p^{-k}$ is an integer. For $n = kr$ consider points in $\{0,1\}^n$ as r k-tuples of coordinates. Let A denote the set of points in $\{0,1\}^n$ such that at least one k-tuple of coordinates consists of 1's only. We can compute that $\mu_p(A) = (1-p^k)^r$, which we note approximates e^{-1} closely for sufficiently large r. This tells us that the left hand side of (9.1) is of constant order. Furthermore, we have that for each i, $\mu_p(A_i) = p^k(1-p^k)^{r-1}$, which by the same logic as above approximates $\frac{p^k}{e}$, which means that $n\mu_p(A_i)$ is of order k. Furthermore, given that $\log(1/(1-p)\mu_p(A_i)) \simeq k\log(1/p)$, which gives us that the RHS of (9.1) is also of order 1.

□

As a tool in some of proofs, Talagrand introduces the following collection of functions on $L^2(\mu_p)$. Given a subset $S \subseteq \{1, ..., n\}$, write

$$r_S(x) = \prod_{i \in S} r_i(x)$$

where

$$r_i(x) = \begin{cases} \sqrt{\frac{1-p}{p}} & \text{if } x_i = 1 \\ -\sqrt{\frac{p}{1-p}} & \end{cases}.$$

Noting that $r_\emptyset \equiv 1$, we have that $\{r_S\}_{S \subseteq \{1,...,n\}}$ forms an orthogonal basis for $L^2(\mu_p)$. Given $g = \sum a_S r_S$ such that $a_\emptyset = \int g d\mu_p = 0$ define

$$M(g)^2 = \sum_S \frac{a_S^2}{|S|}.$$

The quantity $M(g)$ is important to the results presented in the paper (specifically Theorem 2) insofar that for f on $\{0,1\}^n$ with $\int f d\mu_p = 0$, we can write $f = \sum_S b_S r_S$, $b_\emptyset = 0$. We note that Δ_i has been defined in such a way that $\Delta_i(r_S) = 0$ if $i \notin S$ and $\Delta_i(r_S) = r_S$ if $i \in S$. Thus, a series of computations allows us to deduce that

$$||f||_2^2 = \sum_S b_S^2 = \sum_i M(\Delta_i f)^2 \tag{9.3}$$

Talagrand then proceeds to prove the following important property for the basis $\{r_S\}_{S \subseteq \{1,...,n\}}$.

Lemma. *For $q \geq 2$ and set $\theta = \frac{1}{\sqrt{p(1-p)}}$. Then for any k and numbers $\{a_S\}_{|S|=k}$, we have*

$$\left|\left| \sum_{|S|=k} a_S r_S \right|\right|_q \leq (q-1)^{k/2} \theta^k \left(\sum_{|S|=k} a_S^2 \right)^{\frac{1}{2}} \tag{9.4}$$

Proof. Step 1: Consider the space $\{-1, 1\}$ equipped with the uniform measure λ, and for $S \subseteq \{1, ..., n\}$, set

$$w_S(\varepsilon) = \prod_{i \in S} \varepsilon_i.$$

We have that the w_S form an orthonormal basis for $L^2(\lambda)$, and moreover, we have by results from Fourier analysis that the operator

$$T_\delta : \sum b_S w_S \to \sum b_S \delta^{|S|} w_S \tag{9.5}$$

is of norm 1 from $L^2(\lambda)$ to $L^q(\lambda)$ for $\delta = \frac{1}{\sqrt{q-1}}$.
Step 2: Equip the space $H = \{0,1\}^n \times \{0,1\}^n \times \{-1,1\}^n$ with the measure $\nu = \mu_p \otimes \mu_p \otimes \lambda$ and consider the function

$$h_S(x, y, \varepsilon) = g_S(x, y) w_S(\varepsilon) \prod_{i \in S} (r_i(x) - r_i(y)) \varepsilon_i.$$

By some simple computations, it follows that

$$\left|\left| \sum a_S g_S \right|\right|_{L^q(\mu_p \otimes \mu_p)} = \left|\left| \sum a_S h_S \right|\right|_{L^q(\nu)} \tag{9.6}$$

Furthermore, applying the results from Step 1 to $b_S = a_S g_S(x, y) = a_S \prod_{i \in S} (r_i(x) - r_i(y))$, noting that $|r_i(x) - r_i(y)| \leq \theta$, and some further computations yield that

$$\left|\left| \sum_{|S|=k} a_S h_S \right|\right|_{L^q(\nu)} \leq \theta^k (q-1)^{k/2} \left(\sum_{|S|=k} a_S^2 \right)^{1/2}. \tag{9.7}$$

Step 3: The result follows by combining the estimates obtained in Part 2 as well as the observation that

$$\left|\left|\sum a_S r_S\right|\right|_{L^q(\mu_p)} \leq \left|\left|\sum a_S g_S\right|\right|_{L^q(\mu_p \otimes \mu_p)}.$$

□

Using duality via Hölder's inequality, one can obtain the following result from the previous Lemma:

Proposition. *Let $g : \{0,1\}^n \to \mathbb{R}$ and set $a_S = \int r_S g d\mu$. Then*

$$\sum_{|S|=k} a_S^2 \leq (q-1)^k \theta^{2k} ||g||_{q'}^2,$$

where q' is the conjugate exponent of q.

Equation (3) combined with the next statement are sufficient to prove Theorem 2 (and by extension Theorem 1):

Proposition. *For some universal constant K, if $\int g d\mu_p = 0$, we have*

$$M(g)^2 \leq K \log \frac{2}{p(1-p)} \frac{||g||_2^2}{\log(||g||_2/(e||g||_1))} \tag{9.8}$$

Proof. The proof of this proposition involves considering the result from Proposition 1 for the case $q = 3, q' = \frac{3}{2}$. We also observe that for the sequence $x_k = \frac{(2\theta^2)^k}{k}$ and for any integer m, $\sum\limits_{k \leq m} x_k \leq 2x_m$ by previous observations. Thus, combining the results from our application of Prop 1 and this observation, we obtain the following:

$$M(g)^2 \leq \sum_{k \leq m} x_k ||g||_{3/2}^2 + \sum_{|S|>m} \frac{a_S^2}{|S|} \leq \frac{1}{m+1} \left(4m x_m + ||g||_2^2\right)$$

One can then cleverly choose m as the largest integer such that $(2\theta^2)^m ||g||_{3/2}^2 \leq ||g||_2^2$.[2] This results in the observation that $(2\theta^2)^{m+1} ||g||_{3/2}^2 \geq ||g||_2^2$, i.e.

$$m + 1 \geq \frac{2 \log(||g||_2 / ||g||_{3/2})}{\log 2\theta^2}$$

Using both this and our initial constraint on m, we can plug this into what we have so far for $M(g)^2$ to get

$$M(g)^2 \leq \frac{K \log 2\theta^2}{\log(e||g||_2/||g||_{3/2})} ||g||_2^2$$

We finally obtain the desired result by noting that

$$\frac{||g||_2}{||g||_1} \leq \left(\frac{||g||_2}{||g||_{3/2}}\right)^3$$

which is in itself a simple consequence of the Cauchy-Schwarz Inequality. □

The remainder of the section is dedicated to the proof of Theorem 3, which itself involves exploiting a few key properties of the Orlicz norm $||\cdot||_\varphi$. These are the following:

Lemma. *For a function f:*

$$||f||_\varphi^2 \leq \frac{K||f||_2^2}{\log(e||f||_2/||f||_1)}.$$

[2]Note that since the total measure of $\{0,1\}^n$ for μ_p is 1, the L^p norm of a function is monotone increasing in p. This combined with the observation that for $m = 0$, $(2\theta^2)^m = 1$ and that the $(2\theta^2)^m$ are increasing in m, we know we can find such an integer m

This allows us to improve upon Proposition 2 as follows:

Proposition. *For a universal constant K, we have*

$$M(g)^2 \leq K||g||_\varphi^2 \log \frac{2}{p(1-p)}. \tag{9.9}$$

The proof of this proposition involves manipulating the following family of seminorms. Given $h = \sum h_S r_S$, with $h_\emptyset = 0$, define:

$$M_l(h)^2 = \sum_{2^l \leq k < 2^{l+1},\, |S|=k} \frac{h_S^2}{|S|}.$$

The remaining proof proceeds similarly to using Proposition 1 to prove Proposition 2. Theorem 3 immediately follows via an application of Propsosition 3 to Equation (3).
The remainder of the paper concerns deriving the corollaries of Theorem 1.

Proof of Corollary 1. The only thing required beyond Theorem 1 is what is commonly referred to as "Russo's formula", that is:

$$\frac{d\mu_p(A)}{dp} = \frac{1}{p}\sum_{i \leq n} \mu_p(A_i),$$

from which the remaining derivations are straightforward. □

The computations to derive Corollary 2 follows from an application of Corollary 1 to the expression

$$\frac{d}{dp}\left(g(\mu_p(A))\right)$$

where $g(x) = \log(x/(1-x))$.
Finally, Corollary 3 is a consequence of Theorem 1 and the observation that $x\log(1/x)$ is increasing for $x < 1$ as well as the fact that for $x \leq \frac{1}{2}$

$$\frac{y}{\log(1/y)} \geq x \implies y \geq \frac{x}{K}\log(1/x).$$

Bibliography

[1] Talagrand, M., *On Russo's Approximate Zero-One Law.* he Annals of Probability, Ann. Probab. 22(3), 1576-1587, (July, 1994).

Yonathan Stone, UCI
email: **ystone@uci.edu**

Chapter 10

On the Fourier tails of bounded functions over the discrete cube

after I. Dinur, E. Friedgut, G. Kindler and R. O'Donnell [1]
A summary written by Alberto Takase

Abstract. This is a terse summary. The main theorem is stated along with related theorems. The proof of the main theorem is outlined.

10.1 Main Theorem and Related Theorems

Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function. Define $\widehat{f} : \mathscr{P}(\{1,\dots,n\}) \to \mathbb{R}$ by

$$f = \sum_S \widehat{f}(S)\chi_S,$$

where $\chi_S : \{-1,1\}^n \to \{-1,1\} : x \mapsto \prod_{i\in S} x_i$. Here $\widehat{f}$ is called the Fourier transform of f. Also here χ_S is called the character of S. The following theorem is the main theorem, and the related theorems are listed afterwards.

Theorem 10.1 (2007 [1])**.** *Assume f has codomain $[-1,1]$ and*

$$\sum_{\#(S)>k} |\widehat{f}(S)|^2 \leq e^{-O(k^2 \log k)/r}$$

for some $k \in \{1,\dots,n\}$ and for some $r > 0$. Then there exists a function $g : \{-1,1\}^n \to \mathbb{R}$ such that g depends on at most $2^{O(k)}/r^2$ coordinates and

$$\sum_S |\widehat{f}(S) - \widehat{g}(S)|^2 \leq r.$$

Furthermore, this theorem is tight, except, possibly, for the $\log k$ in the exponent; see `Theorem 2` within [1].

Theorem 10.2 (2002 [2])**.** *Assume f has codomain $\{-1,1\}$ and*

$$\sum_{\#(S)>k} |\widehat{f}(S)|^2 > (k/r)^{-1/2-o(1)}$$

for some $k \in \{1,\dots,n\}$ and for some $r > 0$. Then there exists a function $g : \{-1,1\}^n \to \mathbb{R}$ such that g depends on at most $2^{O(k)}/r^{O(1)}$ coordinates and

$$\sum_S |\widehat{f}(S) - \widehat{g}(S)|^2 \leq r.$$

Theorem 10.3 (2002 [3])**.** *Assume f has codomain $\{-1,1\}$ and*

$$\sum_{\#(S)>1} |\widehat{f}(S)|^2 \leq r$$

for some $r > 0$. Then there exists a function $g : \{-1,1\}^n \to \mathbb{R}$ such that g depends on at most 1 coordinate and

$$\sum_S |\widehat{f}(S) - \widehat{g}(S)|^2 \leq O(r).$$

Theorem 10.4 (1998 [4]). *Assume f has codomain $\{-1,1\}$ and*

$$\sum_S |\widehat{f}(S)|^2 \#(S) \leq k$$

for some $k \in \{1,\ldots,n\}$. Then for each $r > 0$, there exists a function $g : \{-1,1\}^n \to \mathbb{R}$ such that g depends on at most $2^{O(k/r)}$ coordinates and

$$\sum_S |\widehat{f}(S) - \widehat{g}(S)|^2 \leq r.$$

10.2 Proof of Main Theorem

Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function. Define $\widehat{f} : \mathscr{P}(\{1,\ldots,n\}) \to \mathbb{R}$ by

$$f = \sum_S \widehat{f}(S)\chi_S,$$

where $\chi_S : \{-1,1\}^n \to \{-1,1\} : x \mapsto \prod_{i\in S} x_i$. Assume f has codomain $[-1,1]$ and

$$\sum_{\#(S)>k} |\widehat{f}(S)|^2 \leq e^{-O(k^2 \log k)/r}$$

for some $k \in \{1,\ldots,n\}$ and for some $r > 0$. Without loss of generality, $1 > r > 0$. Define $g : \{-1,1\}^n \to \mathbb{R}$ by

$$g = \sum_{\#(S)\leq k} \widehat{f}(S)\chi_S.$$

Lemma 10.5 (`Theorem 7` within [1]). *There exists $C > 0$ such that for each $T \subseteq \{1,\ldots,n\}$ with $\sum_{S\setminus T\neq\varnothing} |\widehat{g}(S)|^2 \geq r$ and for each $t \geq \sqrt{r}$, if*

$$\sum_{S\ni i} |\widehat{g}(S)|^2 \leq r^2 t^{-2} C^{-k}$$

for every $i \in \{1,\ldots,n\} \setminus T$, then

$$\mathbb{P}[|g| \geq t] \geq e^{-(Ct^2k^2 \log k)/r}.$$

Define $J = \{i : \sum_{S\ni i} |\widehat{g}(S)|^2 \geq r^2(4)^{-2}C^{-k}\}$. Define $h : \{-1,1\}^n \to \mathbb{R}$ by

$$h = \sum_{\#(S)\leq k, S\subseteq J} \widehat{f}(S)\chi_S.$$

Observe

$$\sum_{i\leq n, S\ni i} |\widehat{g}(S)|^2 = \sum_S |\widehat{g}(S)|^2 \#(S) = \sum_{\#(S)\leq k} |\widehat{f}(S)|^2 \#(S) \leq k.$$

Therefore $\#(J) \leq k/r^2(4)^{-2}C^{-k} \approx 2^{O(k)}/r^2$ and h depends on the coordinates of J. Observe

$$\sum_S |\widehat{f}(S) - \widehat{g}(S)|^2 = \sum_{\#(S)>k} |\widehat{f}(S)|^2 \leq e^{-O(k^2 \log k)/r} \leq e^{-O(1)/r} \leq r/2.$$

By Lemma 10.5,

$$\sum_S |\widehat{g}(S) - \widehat{h}(S)|^2 \leq r/2.$$

Indeed, suppose

$$\sum_S |\widehat{g}(S) - \widehat{h}(S)|^2 > r/2.$$

By Lemma 10.5 with parameters J and $r/2$ and 2,

$$\mathbb{P}[|g| \geq 2] \geq e^{-(8Ck^2 \log k)/r}$$

and

$$\sum_S |\widehat{f}(S) - \widehat{g}(S)|^2 = \mathbb{E}_x |f(x) - g(x)|^2 \geq 1 \cdot e^{-(8Ck^2 \log k)/r}$$

which is a contradiction when taking a large enough constant in the $O(\cdot)$.

Bibliography

[1] Dinur, I.; Friedgut, E.; Kindler, G.; O'Donnell, R., *On the fourier tails of bounded functions over the discrete cube*, Israel Journal of Mathematics **160** (2007), 389–412.

[2] Bourgain, J., *On the distribution of the Fourier spectrum of boolean functions*, Israel Journal of Mathematics **131** (2002), 269–276.

[3] Friedgut, E.; Kalai, G.; Naor, A., *Boolean functions whose fourier transform is concentrated on the first two levels and neutral social choice*, Advances in Applied Mathematics **29** (2002), 427–437.

[4] Friedgut, E., *Boolean functions with low average sensitivity depend on few coordinates*, Combinatorica **18** (1998), 27–36.

Alberto Takase, UCI
email: **atakase@uci.edu**

Chapter 11

On the Fourier spectrum of functions on Boolean cubes

after A. Defant, M. Mastyło, and A. Pérez [1]
A summary written by Haonan Zhang

Abstract. We discuss Bohnenblust–Hille type inequalities for n-dimensional Boolean cubes $\{\pm 1\}^n$ [1]. Similar to the result in [2] for n-dimensional torus $\mathbb{T}^n$, the Bohnenblust–Hille constant for Boolean cubes is also of subexponential growth. The main ideas and ingredients of the proof are presented.

11.1 Introduction

Let $\mathbb{T}^n$ be the n-dimensional torus. A classical inequality of Bohnenblust and Hille [3] says that for any $n \geq 1$ and any complex-valued degree-d polynomial

$$P(z) = \sum_{\alpha \in \mathbb{Z}_{\geq 0}^n : |\alpha| \leq d} a_\alpha z^\alpha, \quad z = (z_1, \ldots, z_n) \in \mathbb{T}^n,$$

there exists a constant $C(d) > 0$ depending only on d such that

$$\left(\sum_{\alpha \in \mathbb{Z}_{\geq 0}^n : |\alpha| \leq d} |a_\alpha|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}} \leq C(d) \|P\|_{\mathbb{T}^n}. \tag{11.1}$$

Here and in what follows, we use $\|f\|_K$ to denote the supremum norm of a scalar function on K. The best possible constant $C(d)$ in (11.1) is the *Bohnenblust–Hille constant*, and will be denoted by $\mathrm{BH}_{\mathbb{T}}^{\leq d}$. Similarly, we denote by $\mathrm{BH}_{\mathbb{T}}^{=d}$ the best constant such that (11.1) holds for all d-homogeneous polynomials. Clearly $\mathrm{BH}_{\mathbb{T}}^{=d} \leq \mathrm{BH}_{\mathbb{T}}^{\leq d}$. An easy trick shows that we actually have $\mathrm{BH}_{\mathbb{T}}^{=d} = \mathrm{BH}_{\mathbb{T}}^{\leq d}$. The upper bound of $\mathrm{BH}_{\mathbb{T}}^{\leq d}$ established in the original proof and its later improvements is essentially of order $\sqrt{d}^d$. Recent years have seen many improvements on the Bohnenblust–Hille constants $\mathrm{BH}_{\mathbb{T}}^{\leq d}$. Notably, Bayart, Pellegrino and Seaoane-Sepúlveda [2] proved that there exists a universal constant $C > 0$ such that

$$\mathrm{BH}_{\mathbb{T}}^{\leq d} \leq C^{\sqrt{d \log d}}. \tag{11.2}$$

This, in particular, implies the subexponential growth of $\mathrm{BH}_{\mathbb{T}}^{\leq d}$:

$$\limsup_{d \to \infty} \sqrt[d]{\mathrm{BH}_{\mathbb{T}}^{\leq d}} = 1. \tag{11.3}$$

In the following, we present a Boolean analog of this result.

The analysis of scalar-valued functions on the Boolean cube $\{\pm 1\}^n$ plays an important role in many areas such as theoretical computer sciences. Any function $f : \{\pm 1\}^n \to \mathbb{R}$ has a unique Fourier–Walsh expansion:

$$f(x) = \sum_{S \subset [n]} \widehat{f}(S)\chi_S(x), \quad x \in \{\pm 1\}^n,$$

where for each $S \subset [n] := \{1, \dots, n\}$, χ_S is defined as

$$\chi_S(x) = x^S := \prod_{k \in S} x_k, \quad x = (x_1, \dots, x_n) \in \{\pm 1\}^n.$$

In particular, when $S = \emptyset$, $\chi_\emptyset \equiv 1$. Endowing the uniform probability measure on $\{\pm 1\}^n$, we define the expectation of $f : \{\pm 1\}^n \to \mathbb{R}$ as

$$\mathbb{E}f := \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x).$$

We use $\|f\|_p := (\mathbb{E}|f|^p)^{1/p}, 1 \le p < \infty$ to denote the associated L_p-norms.

For each $S \subset [n]$, let $|S|$ be the cardinality of S. Then $f : \{\pm 1\}^n \to \mathbb{R}$ is of degree-d if $\widehat{f}(S) = 0$ for all $|S| > d$, and f is d-homogeneous if $\widehat{f}(S) = 0$ whenever $|S| \neq d$. We will need the following consequence of hypercontractivity [1]: for $f : \{\pm 1\}^n \to \mathbb{R}$ of degree-d we have

$$\|f\|_2 \le C(p, d)\|f\|_p, \quad 1 \le p \le 2, \tag{11.4}$$

where $C(p, d) = (p - 1)^{-d/2}$ if $1 < p \le 2$, and $C(1, d) = e^d$.

Now we can state the Bohnenblust–Hille type inequalities for the Boolean cube: for any $f : \{\pm 1\}^n \to \mathbb{R}$ of degree-d we have

$$\left(\sum_{S \in [n] : |S| \le d} |\widehat{f}(S)|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}} \le \mathrm{BH}^{\le d}_{\{\pm 1\}} \|f\|_{\{\pm\}^n}. \tag{11.5}$$

Here $\mathrm{BH}^{\le d}_{\{\pm 1\}}$ already denotes the best constant for Boolean cubes. One can understand $\mathrm{BH}^{=d}_{\{\pm 1\}}$ in an obvious way, i.e., the best constant such that (11.5) holds for all d-homogeneous functions. The trick for proving $\mathrm{BH}^{=d}_{\mathbb{T}} = \mathrm{BH}^{\le d}_{\mathbb{T}}$ does not work for Boolean cube anymore, so we only have $\mathrm{BH}^{=d}_{\{\pm 1\}} \le \mathrm{BH}^{\le d}_{\{\pm 1\}}$ in general. The main result of [1] is the following:

Theorem 11.1. *[1] There exists a universal constant $C > 0$ such that*

$$\mathrm{BH}^{\le d}_{\{\pm 1\}} \le C^{\sqrt{d \log d}}. \tag{11.6}$$

In particular, $\limsup_{d \to \infty} \sqrt[d]{\mathrm{BH}^{\le d}_{\{\pm 1\}}} = 1$.

This result is closely related to many other topics such as Sidon sets, Boolean radii and the Aaronson–Ambainis conjecture [3]. To compare, we also have Bohnenblust–Hille type inequalities for real polynomials on n-dimensional cubes $[-1, 1]^n$, with the best constants $\mathrm{BH}^{=d}_{[-1,1]}$ and $\mathrm{BH}^{\le d}_{[-1,1]}$ satisfying [1]

$$\limsup_{d \to \infty} \sqrt[d]{\mathrm{BH}^{=d}_{[-1,1]}} = 2, \text{ and } \limsup_{d \to \infty} \sqrt[d]{\mathrm{BH}^{\le d}_{[-1,1]}} = 1 + \sqrt{2}.$$

11.2 Proof of the d-homogeneous case

Let us sketch the proof of Theorem 11.1 in the d-homogeneous case so that one can easily grasp the point. For any $n \in \mathbb{N}$ and any finite set $A \subset \mathbb{N}$, define

$$\mathcal{I}(A, n) := \{\mathbf{i} : A \to [n]\}.$$

When $A = [d]$, we denote $\mathcal{I}(d, n) := \mathcal{I}([d], n)$ for simplicity. For two disjoint subsets A_1 and A_2 of $\mathbb{N}$ we may define the direct sum $\mathbf{i}_1 \oplus \mathbf{i}_2$ of $\mathbf{i}_1 \in \mathcal{I}(A_1, n)$ and $\mathbf{i}_2 \in \mathcal{I}(A_2, n)$ as an element of $\mathcal{I}(A_1 \cup A_2, n)$. In

particular, for any fixed d and any $S \subset [d]$, we use $\widehat{S} := [d] \setminus S$ to denote the complement of S in $[d]$. Any element $\mathbf{i}$ in $\mathcal{I}(d,n)$ can be uniquely decomposed into the direct sum of some $\mathbf{i}_1 \in \mathcal{I}(S,n)$ and some $\mathbf{i}_2 \in \mathcal{I}(\widehat{S},n)$.

The following inequality is crucial and will also be used in the proof of the degree-d case.

Proposition 11.2. *[2] Let $n \in \mathbb{N}$ and $1 \le k \le d$ be integers. Then for any scalar matrix $(a_{\mathbf{i}})_{\mathbf{i}\in\mathcal{I}(d,n)}$, we have*

$$\left(\sum_{\mathbf{i}\in\mathcal{I}(d,n)} |a_{\mathbf{i}}|^{\frac{2d}{d+1}}\right)^{\frac{d+1}{2d}} \le \left[\prod_{S\subset[d]:|S|=k} \left(\sum_{\mathbf{i}_1\in\mathcal{I}(S,n)} \left(\sum_{\mathbf{i}_2\in\mathcal{I}(\widehat{S},n)} |a_{\mathbf{i}_1\oplus\mathbf{i}_2}|^2\right)^{\frac{1}{2}\cdot\frac{2k}{k+1}}\right)^{\frac{k+1}{2k}}\right]^{\frac{1}{\binom{d}{k}}}. \tag{11.7}$$

The proof of (11.6) for $\mathrm{BH}^{=d}_{\{\pm1\}}$ is based on the following inductive inequality

$$\mathrm{BH}^{=d}_{\{\pm1\}} \le C(k,d)\mathrm{BH}^{=k}_{\{\pm1\}}, \tag{11.8}$$

for some constant $C(k,d) > 0$ and any $1 \le k \le d$. The desired bound (11.6) for $\mathrm{BH}^{=d}_{\{\pm1\}}$ will follow by applying (11.8) repeatedly to special k's. Now we use a simple example to illustrate the proof of (11.8). Let $(d,k) = (2,1)$. Then for any 2-homogeneous function $f(x) = \sum_{i<j} a_{ij}x_ix_j$ on $\{\pm1\}^n$ we need to show

$$\left(\sum_{i<j} |a_{ij}|^{4/3}\right)^{3/4} \le C\mathrm{BH}^{=1}_{\{\pm1\}}\|f\|_{\{\pm1\}^n}.$$

Here and in what follows, $C > 0$ is some constant that may differ from line to line. The proof consists of four steps.

Step 1: Put $a_{ii} := 0, i \in [n]$ and $a_{ji} := a_{ij}$ for $i < j$.

Step 2: Apply the inequality (11.7) to $(a_{ij})_{i,j\in[n]}$:

$$\left(\sum_{i<j} |a_{ij}|^{4/3}\right)^{3/4} \le C\sum_{i=1}^n \left(\sum_{j\ne i} |a_{ij}|^2\right)^{1/2}.$$

Step 3: The hypercontractivity result (11.4) (with $p = 1$) implies

$$\sum_{i=1}^n \left(\sum_{j\ne i} |a_{ij}|^2\right)^{1/2} = \sum_{i=1}^n \left\|\sum_{j\ne i} a_{ij}y_j\right\|_2 \le C\mathbb{E}_y \sum_{i=1}^n \left|\sum_{j\ne i} a_{ij}y_j\right|.$$

By definition of $\mathrm{BH}^{=1}_{\{\pm1\}}$, the last term is bounded from above by

$$C \sup_{y\in\{\pm1\}^n} \sum_{i=1}^n \left|\sum_{j\ne i} a_{ij}y_j\right| \le C\mathrm{BH}^{=1}_{\{\pm1\}} \sup_{x,y\in\{\pm1\}^n} \left|\sum_{i=1}^n \sum_{j\ne i} a_{ij}x_iy_j\right|.$$

Step 4: By a polarization result that we will discuss later, the right-hand side is bounded from above by $C\mathrm{BH}^{=1}_{\{\pm1\}}\|f\|_{\{\pm1\}^n}$. This finishes the proof.

Now let us recall the polarization result in the last step establishing

$$\sup_{x,y\in\{\pm1\}^n} \left|\sum_{i\ne j} a_{ij}x_iy_j\right| \le C\|f\|_{\{\pm1\}^n}. \tag{11.9}$$

Any degree-d function $f : \{\pm1\}^n \to \mathbb{R}$ is the restriction of a unique polynomial (the *tetrahedral*) $P = P_f : \mathbb{R}^n \to \mathbb{R}$ that is affine in each variable. Moreover, $\|f\|_{\{\pm1\}^n} = \|P_f\|_{[-1,1]^n}$. This polynomial P is associated to a unique d-affine symmetric form $L : (\mathbb{R}^n)^d \to \mathbb{R}$ such that $P(x) = L(x,\dots,x)$. When f is d-homogeneous, the form L is d-linear. For the above $f(x) = \sum_{i<j} a_{ij}x_ix_j$, $L(x,y) = \frac{1}{2}\sum_{i\ne j} a_{ij}x_iy_j$, which is (up to a scalar) what we need to estimate in (11.9). In general, one has to bound $L(x,\dots,x,y,\dots,y)$. For this, we need to use the following polarization result:

Proposition 11.3. *[1] Let $P : \mathbb{R}^n \to \mathbb{R}$ be a d-homogeneous polynomial. Let $L : (\mathbb{R}^n)^d \to \mathbb{R}$ be the associated unique d-linear symmetric form. Then for any $0 < k < d$, we have*

$$|L(\overbrace{x,\dots,x}^{k},\overbrace{y,\dots,y}^{d-k})| \le M_{k,d}\frac{d^d}{k^k(d-k)^{d-k}}\frac{k!(d-k)!}{d!}\|P\|_{[-1,1]^n}, \quad x,y\in\mathbb{R}^n.$$

Here $M_{k,d}$ is the Markov number that has an explicit form.

The proof relies on *Markov's inequality*, saying that for any real polynomial $p(t) := \sum_{k=0}^d a_k t^k$ of degree d, we have

$$|a_k| \le M_{k,d}\|p\|_{[-1,1]}, \quad 0\le k\le d.$$

The constant $M_{k,d}$ is optimal and can be captured by Chebyshev polynomial of degree d. Then Proposition 11.3 will follow by taking $p(t) = P(t\lambda x + (1-\lambda)y)$ for suitable $\lambda\in[0,1]$ and expanding p by *linearity* and symmetry of L.

11.3 Proof of the degree-d case

The proof of (11.6) is again based on an inductive inequality similar to (11.8). For this, we repeat the four-step argument. The first three steps can be easily adapted to the degree-d case. However, the proof of Proposition 11.3 used in the last step does not work for degree-d case, since it requires the linearity of L. Before stating the substitute of polarization result, we start with a variant of Markov's inequality:

Proposition 11.4. *For each $0\le k\le d$ set*

$$\psi_{d,k}(t) := \left(\frac{1+t}{2}\right)^k\left(\frac{1-t}{2}\right)^{d-k}, \quad t\in\mathbb{R}.$$

Then any degree-d polynomial $p:\mathbb{R}\to\mathbb{R}$ can be represented as

$$p(t) = \sum_{k=0}^d a_k\psi_{d,k}(t), \quad t\in\mathbb{R},$$

where each $a_k = a_k(p)\in\mathbb{R}$ satisfies

$$|a_k(p)| \le |a_k(T_d)|\|p\|_{[-1,1]}.$$

Here T_d denotes the Chebyshev polynomial of degree d.

Different from Proposition 11.3, here we replace the basis $\{t^k\}_k$ with $\{\psi_{k,d}\}_k$. Then one can prove the following polarization result by choosing $p(t) = P(\frac{1+t}{2}x + \frac{1-t}{2}y)$ and expanding p by *affinity* and symmetry of L.

Proposition 11.5. *Let $P:\mathbb{R}^n\to\mathbb{R}$ be a polynomial of degree d and $L = L_P$ the associated d-affine form. Then for any $0\le k\le d/2$, we have*

$$|L(\overbrace{x,\dots,x}^{k},\overbrace{y,\dots,y}^{d-k})| \le 2d^k\|P\|_{[-1,1]^n}, \quad x,y\in\mathbb{R}^n.$$

With this, one can complete the proof of Theorem 11.1 in the general case.

Bibliography

[1] Defant, A., Mastyło, M., Pérez, A. , *On the Fourier spectrum of functions on Boolean cubes.* Math. Ann., 374(1), 653-680 (2019).

[2] Bayart, F., Pellegrino, D., Seaoane-Sepúlveda, J.B., *The Bohr radius of the n-dimensional polydisk is equivalent to (log n)/n.* Adv. Math. 264, 726–746 (2014).

[3] Bohnenblust, H. F., Hille, E., *On the absolute convergence of Dirichlet series.* Ann. Math. 32(3), 600–622 (1931).

Haonan Zhang, IST Austria
email: **haonan.zhang@ist.ac.at**