The Mysteries of Low-Degree Boolean Functions

Nathan Keller
Bar Ilan University

Based on joint work with Ohad Klein

UCLA-Caltech Analysis Seminar – April 2021

Contents

- Discrete Fourier analysis
  - Motivation, basic ideas and theorems
- General low degree functions
  - First degree functions: Proof of Tomaszewski's conjecture
  - Open problems and d-degree functions
- Bounded low degree functions
  - The Aaronson-Ambainis conjecture
- [Time pemitting]: (Almost) low degree Boolean functions
  - The Fourier entropy/influence conjecture

Discrete Fourier Analysis I

- Origin: Initiated by Kahn, Kalai, and Linial (1988).

- Basic object of study: We study properties of functions on the discrete cube: $f:\Omega \to R$, where $\Omega = \{-1, 1\}^n$, using analytic tools.

- Basic observation: Each such function has a unique expansion of the form $f = \sum_{S \subset [n]} \hat{f}(S) x_S$, where $x_S = \prod_{i \in S} x_i$. The $\hat{f}(S)$ are called *Fourier(-Walsh) coefficients* and the level of $\hat{f}(S)$ is defined as $|S|$.

Discrete Fourier Analysis II

- Meta question: What can we say about a function on Ω, given some information on its Fourier expansion?

- Applications: Social choice, machine learning, metric embedding, percolation, extremal combinatorics, hardness of approximation, phase transitions, and many more…

Discrete Fourier Analysis III

- **Basic tool:** Noise and hypercontractivity
  - The noise operator transforms f into $T_\rho f$, defined as $T_\rho f(x) = E[f\left(N_\rho x\right)]$, where $N_\rho(x)$ is obtained by randomizing each x_i with prob. $1-\rho$.
  - Theorem [B70]: The noise operator is hypercontractive: $\left\|T_\rho f\right\|_2 \leq \|f\|_{1+\rho^2}$
- **Relation to Fourier levels:** For any f,

$$T_\rho\left(\sum \hat{f}(S)x_S\right) = \sum \rho^{|S|}\hat{f}(S)x_S$$

- and thus, noise suppresses the high level coefficients.

Influences

- Definition: For $f:\Omega \to \{-1,1\}$, the *influence* of the i'th coordinate on f is $I_i[f] = \Pr[f(x) \neq f(x \cdot e_i)]$.
- The *total influence* of f is $I[f] = \sum_{i\in[n]} I_i[f]$.

- Natural interpretations:
  - $I_i[f]$ is the influence of a voter in an election.
  - $I[f]$ is the edge boundary size of the set $\{x: f(x)=1\}$ in the discrete cube (viewed as a graph).
  - $I[f]$ is the derivative of the function $p \mapsto \mu_p(\{x: f(x)=1\})$, where μ_p is the p-biased measure on the discrete cube.

- Relation to Fourier levels: $I[f] = \sum_S |S| \hat{f}(S)^2$

Friedgut's Junta Theorem

- Definition: A j-*junta* is a function that depends on at most j coordinates.

- Theorem [F98]: If $f:\Omega \to \{-1,1\}$ and $I[f] \le k$, then f can be ϵ-approximated by an $\exp(ck/\epsilon)$-junta.
  - Meaning: Functions with a low total influence essentially depend on a few coordinates.
  - Tight for the `address' function.

- Relation to Fourier levels [B99]: If most of the Fourier weight of a Boolean f is on low levels, then f is approximately a junta.

Low Degree Functions

- Definition: A function f is of *degree* d if all its Fourier coefficients are at level $\leq d$.
  - Alternatively, f is a multilinear polynomial of degree d.
  - For $d = 1$, $f = \sum a_i x_i$ can be viewed as a weighted sum of *Rademacher random variables*.

- Meta questions: Assume f is low-degree.
  - What does this tell us on the structure of f?
  - What if, in addition, f is bounded?
  - What if, in addition, f is Boolean – i.e., assumes only two values?

Contents

- Discrete Fourier analysis
  - Motivation, basic ideas and theorems
- General low degree functions
  - First degree functions: Proof of Tomaszewski's conjecture
  - Open problems and d-degree functions
- Bounded low degree functions
  - The Aaronson-Ambainis conjecture
- (Almost) low degree Boolean functions
  - The Fourier entropy/influence conjecture

First Degree Functions I

- Definition: A *Rademacher random variable* assumes each of the values ± 1 with probability $\frac{1}{2}$.
- A Rademacher sum is $X = \sum a_i x_i$, where $\{x_i\}$ are independent Rademacher r.v.'s. Usually, $\sum a_i^2 = 1$.
  - A first degree function is essentially a Rademacher sum.
- Meta question: How do Rademacher sums look like?
- Meta answer: In many aspects, like a Gaussian
  - Motivating example: $\sum \frac{1}{\sqrt{n}} x_i \to N(0,1)$

First Degree Functions II

- Small coefficients - Berry-Esseen theorem:
- Let $X = \sum a_i x_i$ be as above, and $F = CDF(X)$. Then:

  $$\forall x: \left| F(x) - \Phi(x) \right| \leq 0.56 \sum a_i^3$$
- where $\Phi(x)$ is the CDF of a $N(0, 1)$ random variable.
  - Consequently, if $\forall i: \left| a_i \right| \leq m$, then for any interval I,

    $$|\Pr[X \in I] - \Pr[N(0,1) \in I]| \leq 1.12m$$

- Tail for general coefficients [BD15]: Let X be as above.

  $$\forall t: \Pr[X > t] \leq 3.17 \cdot \Pr[N(0,1) > t]$$
  - Tight, for $X = (x_1 + x_2)/\sqrt{2}$ and $t = \sqrt{2}$.

- Question: What happens “near the middle”?

How to Get a Free Lunch? I

- Excerpt from "Probabilistic Methods in Combinatorics" course, Hebrew University, 2005:

3. (*) Show that there is a positive constant c such that the following holds. For any n reals $a_1, \ldots, a_n$ satisfying $\sum a_i^2 = 1$, if $(\epsilon_1 \ldots, \epsilon_n)$ is chosen uniformly at random from $\{-1, 1\}^n$ then

$$Pr\left(\left|\sum \epsilon_i a_i\right| \leq 1\right) \geq c.$$

(**) If you can prove the above for $c = 1/2$ your grade in this course will be 100. (And I will buy you lunch.)

- In other words: Let X be a Rademacher sum. Can we prove that with prob. $\geq 1/2$, it lies within a single standard deviation of its mean?

Basic Observations

- Chebyshev's inequality: $\Pr\left[\left|X - E[X]\right| \geq \lambda\sigma\right] \leq 1/\lambda^2$
  - Yields nothing for $\lambda = 1$!
- Simple argument for a weaker bound:
  - Arrange the a_i's in decreasing order and let k be minimal s.t. $|\sum_{i\leq k} a_i x_i| \geq 1/2$.
  - With probability ½, the sign of $\sum_{i>k} a_i x_i$ is opposite from that of $\sum_{i\leq k} a_i x_i$.
  - Hence, by Chebyshev's inequality,

$$\Pr\left[\left|X\right| \leq 1\right] \geq \frac{1}{2} - \Pr\left[\sum_{i>k} a_i x_i \geq 1.5\right] \geq \frac{5}{18} > 0.27$$

  - Small further improvements possible, but ½ is far…

Tomaszewski's Conjecture I

- Origin of the problem: Denote $c = \Pr\left[\,|X| \leq 1\right]$. The claim $c \geq 1/2$ is a well-known conjecture, raised as a question by Tomaszewski (1986) and conjectured by Holzman and Kleitman (1992).

- Previous results:
  - Holzman and Kleitman (1992): $c \geq 0.375$
  - Boppana and Holzman (2017): $c \geq 0.406$
  - Boppana, Hendriks, and van Zuijlen (2020): $c \geq 0.428$
  - Dvorak, van Hintum, and Tiba (2020): $c \geq 0.46$
  - Various results for specific types of Rademacher sums

Tomaszewski's Conjecture II

- Why ½? A few examples (always assume the a_i's are positive and in descending order):

- $a_1 + \min \left| \sum_{i>1} a_i x_i \right| > 1$ (e.g., $(x_1 + x_2)/\sqrt{2}$)
  - In this case, $\Pr\left[|X| \le 1 \right] = \Pr\left[|X| < 1 \right] = 1/2$.

- $X = (x_1 + x_2 + x_3 + x_4)/2$
  - In this case, $\Pr\left[|X| \le 1 \right] = 7/8$ and $\Pr\left[|X| < 1 \right] = 3/8$.

- $X = (x_1 + x_2 + \ldots + x_9)/3$
  - In this case, $\Pr\left[|X| \le 1 \right] \approx 0.82$ and $\Pr\left[|X| < 1 \right] \approx 0.493$.

How to Get a Free Lunch II

Ingredients

- A simple lemma allowing removing variables
  - At some price, of course
- Segment comparison theorem for Rademacher sums
- A semi-inductive argument
- Enhanced Berry-Esseen bound for Rademacher sums
- A generalized Chebyshev inequality

Basic Lemma – Eliminate Variables

- Lemma: Let $X = \sum a_i x_i$ be a Rademacher sum. Denote

$\sigma = \sqrt{1 - \sum_{i \le m} a_i^2}$, and $X' = \sum_{i=m+1}^{n} \frac{a_i}{\sigma} x_i$.

- Tomaszewski's conjecture is equivalent to:

$$\sum_j \Pr[X' > T_j] \le 2^{m-2}$$

- where $\{T_j\}$ ranges over $1 \pm a_1 \pm \ldots \pm a_m$.

- Special case: Substituting $m = 1$, one sees that Tomaszewski's conjecture is equivalent to

$$\Pr\left[|X| < t\right] \ge \Pr[|X| > 1/t]$$, for all $t > 0$.

Segment Comparison for Rademacher Sums

- Motivation: We want to compare probabilities of the form $\Pr[X \in I]$, $\Pr[X \in J]$, for intervals I, J.

- Theorem: Let X be a Rademacher sum, and write $M = \max a_i$. For any $A, B, C, D \in R$ with
  - $|A| \leq \min(B, C)$, $2M \leq C - A$, and
  - $D - C + \min(2M, D - B) \leq B - A$,
- we have $\Pr\left[X \in [C, D)\right] \leq \Pr[X \in [A, B)]$.

- Proof: By a direct bijection, or via local tail inequalities for Rademacher sums by Devroye and Lugosi (2008).

A Semi-Inductive Argument

- Lemma: Tomaszewski's statement for $X = \sum_{i \le n} a_i x_i$ with $a_1 + a_2 \ge 1$ follows from the same statement for $X' = \sum_{j \le m} b_j x_j$ for certain $m < n$ and b_j's.
  - However, $b_1 + b_2 \ge 1$ is not guaranteed.

- Thus, (only) if we prove the conjecture in the case $a_1 + a_2 < 1$ directly, we can complete proof by induction.

- Proof-of-Lemma: Elimination of variables and a "stopping time" argument

Enhanced Berry-Esseen for Rademacher Sums

- Reminder: Berry-Esseen allows deducing that if $\forall i: |a_i| \leq m$, then for any interval I, $|\Pr[X \in I] - \Pr[N(0,1) \in I]| \leq 1.12m$
  - As $\Pr\left[N(0,1) \in [-1,1]\right] \approx 0.68$, this implies Tomaszewski's statement in the case $\max a_i \leq 0.16$.
- Can we push this bound further?

- Result: Several improved Berry-Esseen type bounds for Rademacher sums, which allow deducing Tomaszewski's statement in the range $\max a_i \leq 0.31$.
  - Almost best possible, in view of $X = (x_1 + \ldots + x_9)/3$.
  - Proof uses a strategy of Prawitz (1972).

Generalized Chebyshev Inequality

- Proposition: Let X be a symmetric r.v. with $Var[X] = 1$, and let $c_0, \ldots, c_n, d_0, \ldots, d_n$ be such that

$$0 = c_0 \leq c_1 \leq \ldots \leq c_n = 1 = d_0 \leq d_1 \leq \ldots \leq d_m$$

- Then

$$\sum_i (1 - c_i^2)\Pr[X \in [c_i, c_{i+1})] \geq \sum_j \left(d_j^2 - d_{j-1}^2\right)\Pr\left[X \geq d_j\right]$$

- Special case: For X as above,

$$\Pr\left[X \in [0,1)\right] \geq \Pr\left[X \geq \sqrt{2}\right] + \Pr\left[X \geq \sqrt{3}\right] + \Pr\left[X \geq \sqrt{4}\right] + \ldots$$

- compared to

$$\Pr\left[X \in [0,1)\right] \geq (t^2 - 1)\Pr[X \geq t]$$

- of original Chebyshev's inequality.

How to Get a Free Lunch III

Recipe [KK20]

- Mix all above ingredients:
  - Cover $\max a_i \leq 0.31$ with enhanced Berry-Esseen,
  - Cover $a_1 + a_2 \geq 1$ with the semi-inductive argument,
  - Divide cases in the middle to sub-cases and cover them with combinations of variable removal, segment comparison, and generalized Chebyshev

- Blend for a year or so…

and the free lunch is yours!

Contents

- Discrete Fourier analysis
  - Motivation, basic ideas and theorems
- General low degree functions
  - First degree functions: Proof of Tomaszewski's conjecture
  - Open problems and d-degree functions
- Bounded low degree functions
  - The Aaronson-Ambainis conjecture
- (Almost) low degree Boolean functions
  - The Fourier entropy/influence conjecture

An Anti-Concentration Lower Bound

- Motivation: Above results assert that $\Pr[|X| \leq 1]$ cannot be too small. What about $\Pr[|X| \geq 1]$?

- Conjecture (Hitczenko and Kwapien, 1994): For any Rademacher sum X, we have $\Pr\left[|X| \geq 1\right] \geq 7/32$.
  - Tight, for $X = (x_1 + x_2 + \ldots + x_6)/\sqrt{6}$
  - Best previously known bound (Oleszkiewicz, 1996): $\Pr\left[|X| > 1\right] \geq 1/10$

- New result (Dvorak and Klein, 2021)
  - $\Pr\left[|X| > 1\right] \geq 1/8$ (tight, for $X = (x_1 + \ldots + x_4)/2$)
  - $\Pr\left[|X| \geq 1\right] \geq 6/32$

The General Problem

- Definition: Let $M(t) = \sup \Pr[X > t]$ be the supremum on the tail probabilities of Rademacher sums.
  - Example: Tomaszewski asserts $M(1) = 1/4$, and Hitczenko-Kwapien assert $M(-1) = 57/64$.
  - Various previous results on $M(t)$ can be improved by our methods.
- Goal: Understand how $M(t)$ looks.
- Natural conjecture (Edelman, 1991): For any t, the supremum $M(t)$ is attained for some Binomial

$$X = (x_1 + x_2 + \ldots + x_n)/\sqrt{n}$$

  - Complies with all previous results and conjectures
  - Unfortunately, false! (Zhubr, 2012; Pinelis, 2015)

More Open Questions

- Robust version: What can be said on Rademacher sums whose tail probability is close to the extremum?
  - Our methods give a strong robust version of Tomaszewski

- Signed sums of vectors: Let $X' = \sum v_i x_i$, where $v_i \in R^d$ and $\sum ||v_i||_2^2 = 1$. What can be said on $\Pr[||X'||_2 \leq t]$?
  - For $t = 1$, a lower bound of $e^{-2}/4$ can be derived from a result of Ivanisvili and Tkocz (2019) on comparison of norms of low-degree functions on Ω.
  - Improves over the bound 0.03 proved by Veraar (2008)

d-Degree Functions

- Definition: A d-degree Rademacher chaos is $X = \sum_{|S|=d} a_S x_S$, where $\{x_i\}$ are independent Rademacher r.v.'s, and $x_S = \prod_{i \in S} x_i$. Usually, $\sum a_i^2 = 1$.
  - A homogeneous d-degree function on Ω is essentially a d-degree Rademacher chaos.

- Meta question: Let $M_d(t) = \sup \Pr[X > t]$ be the supremum on the tail probabilities of d-degree Rademacher chaoses. What can be said on $M_d(t)$?
  - Example: Ben Tal, Nemirovsky and Roos (2001) conjectured that $M_2(0) \geq 1/4$. The best known result is $M_2(0) \geq 0.03$, due to Veraar (2008).

Local Tail Inequalities for d-Degree Functions

- Back to $d = 1$: The key to our results on Rademacher sums was a local tail inequality:
- Theorem (Devroye and Lugosi, 2008): Let $X = \sum a_i x_i$ be a Rademacher sum. If $\Pr[X > t] = \epsilon$, then

  $\Pr\left[X > t + \delta\right] \leq \epsilon/2$
- for some $\delta \leq c/\sqrt{\log(1/\epsilon)}$, c a universal constant.

- Conjecture: Let $X = \sum_{|S|=d} a_S x_S$ be a d-degree Rademacher chaos. If $t \geq 0$ and $\Pr\left[X > t^d\right] = \epsilon$, then

  $\Pr\left[X > (t + \delta)^d\right] \leq \epsilon/2$
- for some $\delta \leq c/\sqrt{\log(1/\epsilon)}$, $c = c(d)$.

Target Application I

- Definition: A linear threshold function (LTF) is $1\{\sum a_i x_i > t\}$. A d-degree polynomial threshold function (PTF) is $1\{\sum_{|S|\le d} a_S x_S\}$.

- Observation: If $f = 1\{\sum a_i x_i > t\}$ is an LTF, then the i'th coordinate is influential on f if and only if
$\sum_{j\neq i} a_j x_j \in (t - a_i, t + a_i]$

- Application of our method [KK19]: For f as above, with $E[f] = \epsilon$,
$\max_i I_i(f) = \Theta(\epsilon \cdot \min(1, a_1\sqrt{\log(\frac{1}{\epsilon})})$.
  - Proves (in a strong form) a conjecture of [MORS10]
  - Applications to learning, noise sensitivity, correlation,…

Target Application II

- Our hope: Use conjectured local tail inequality for d -degree Rademacher chaos, to study influences of d -degree PTFs.
  - Potential application to testing and learning of PTFs.

- Conjecture (Gotsman and Linial, 1994): Let f be a d -degree PTF. Then $I(f) = O(d\sqrt{n})$.

- Potential complication: No simple relation between influences and the probability of a Rademacher chaos to lie in a segment.

Comparison of moments

- Question: Let $X = \sum_{|S|=d} a_S x_S$ be a d-degree Rademacher chaos. What is the smallest $C(d)$ s.t.
$||f||_2 \le C(d)||f||_1$?

- Known results:
  - Khinchine-Kahane asserts $C(1) = \sqrt{2}$
  - Hypercontractive inequality implies $C(d) \le e^d$
  - Ivanisvili-Tkocz (2019): $C(d) \le e^{d/2}$

- Conjecture: $C(d) \le 2^{d/2}$
  - Tight
  - Will improve the aforementioned bound $e^{-2}/4$ to 1/16.

Contents

- Discrete Fourier analysis
  - Motivation, basic ideas and theorems
- General low degree functions
  - First degree functions: Proof of Tomaszewski's conjecture
  - Open problems and d-degree functions
- Bounded low degree functions
  - The Aaronson-Ambainis conjecture
- (Almost) low degree Boolean functions
  - The Fourier entropy/influence conjecture

Bounded Low Degree Functions I

- Question: Does any d-degree function essentially depend on $O_d(1)$ coordinates?
- Answer: Of course, no! Example: $f = \frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}}$.
- Question: What if, in addition, the function is bounded?
- Proposition: Any d-degree function on Ω can be ϵ-approximated by a junta on $2^{O(d)}/\epsilon^2$ coordinates.
  - Tightness example: The address function

$$f\left(x_0, \ldots, x_{d-1}, y_0, \ldots, y_{2^d-1}\right) = y_{Bin(x)}$$

Bounded Low Degree Functions II

- Question: In the address function, only d coordinates have non-negligible influence. Moreover, it can be computed by a decision tree of depth $d+1$. Does the same hold for any bounded low-degree function?

- Conjecture (Aaronson and Ambainis, 2008): Let f be a d-degree bounded function. Then:
  - There exists i, such that $I_i(f) \geq poly(Var[f]/d)$.
  - f can be ϵ-approximated by a decision tree of depth at most $poly(d/Var[f])$.

- Previous results: Conj. holds for Boolean functions. For bounded functions, best known bd. $\exp(-d/Var(f))$.

Potential Application to Quantum Computing

- Consequence: If correct, conjecture would imply:

- Conjecture (Folklore, 1999): Let Q be a quantum algorithm that makes T queries to a Boolean input. There exists a deterministic classical algorithm that makes $poly(T, 1/\epsilon, 1/\delta)$ queries and approximates Q's acceptance probability to within an additive error ϵ on a $1-\delta$ fraction of inputs.
  - Meaning: Any quantum algorithm can be simulated on most inputs by a classical algorithm which is only polynomially slower, in terms of query complexity.

Contents

- Discrete Fourier analysis
  - Motivation, basic ideas and theorems
- General low degree functions
  - First degree functions: Proof of Tomaszewski's conjecture
  - Open problems and d-degree functions
- Bounded low degree functions
  - The Aaronson-Ambainis conjecture
- (Almost) low degree Boolean functions
  - The Fourier entropy/influence conjecture

Almost Low-Degree Boolean Functions

- Theorem (Gotsman-Linial, 1994): Any d-degree Boolean (i.e., two-valued) function on Ω depends on at most $d2^{d-1}$ coordinates.
  - [CHS20] The exact bound is $\Theta(2^d)$.
  - [KS03] Same (up to ϵ-approximation) holds for almost d -degree Boolean functions.

- Consequence: The Fourier weight of any such function is concentrated on $2^{O(2^d)}$ coefficients.

- Question: For the address function, all weight is concentrated on $O(2^d)$ coefficients. Maybe the same holds for any almost d-degree function?

Fourier Entropy/Influence Conjecture I

- Definition: The Fourier entropy of a function f on Ω is

$$E(f) = -\sum \hat{f}(S)^2 \log \hat{f}(S)^2.$$

- Conjecture (Friedgut and Kalai, 1996): For any f,

$$E(f) \le cI(f)$$

  - Meaning: The Fourier weight is essentially concentrated on $2^{cI(f)} \le 2^{O(\deg f)}$ coefficients.

- Remarks:
  - Conjecture fails for bounded functions. Example:

    $$f(x) = \min\left(\left| (x_1 + \ldots + x_n)/\sqrt{n} \right|, 1 \right) \cdot sign(x_1 + \ldots + x_n)$$

  - Conjecture has far-reaching consequences in learning.

Fourier Entropy/Influence Conjecture II

- Conjecture: For any f, $E(f) \leq cI(f)$
  - Fourier concentrated on $2^{cI(f)} \leq 2^{O(\deg f)}$ coefficients

- A few selected results:
  - [Easy] $E(f) = O(d)$ for any d-degree function.
  - [BK00] Same holds for almost d-degree functions.
  - [KMS12] If conj. true, it can be generalized to a biased measure on Ω; result tight for graph properties.

- Recent breakthrough [KKLMS20]: Fourier weight is concentrated on $2^{cI(f)\log I(f)}$ coefficients!
  - Are we close to a solution?

Thanks for listening!