2024–2025 学年春季学期　　**调 和 分 析**　　主讲：吴耀琨老师

Testing of Hypercube Functions: Booleanity and Sparsity

江语　　522070910030

Abstract

This reading report explores the problem of property testing for functions defined on the hypercube, focusing on spectral sparsity and Booleanity.

In the first part, we present the task of Testing Booleanity: determining whether a real-valued function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ maps to $\{-1, 1\}$, given its sparse discrete Fourier expansion. We show that functions with sparse Fourier support must either be Boolean or far from Boolean. Its proof uses the Discrete Uncertainty Principle, and an augmented version of this result relies on Hirschman's entropic version of Heisenberg's Uncertainty Principle.

In the second part, we take a closer look at the sparsity of Boolean functions. Our proof relies on random coset hashing and an analysis of a new truncation method for sparse Boolean functions.

Keywords: discrete Fourier analysis, property testing, Boolean functions

1 Introduction

One of the central goals of property testing is to understand the minimal conditions under which a property is testable. Analyzing the Fourier spectrum to characterize the testability of function properties on the hypercube is a natural approach. For example, in our class we studied linearity testing for Boolean functions, which leads to the Blum–Luby–Rubinfeld linearity testing algorithm. In this reading report, we study property testing problems for functions defined on the hypercube, focusing on properties such as sparsity and Booleanity.

In the first part, we aim to investigate whether a function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ is a Boolean function under the assumption of a sparse Fourier expansion. The core of our proof is: f is Boolean if and only if the convolution of $\hat{f}$ with itself equals the delta function, as we show below in Proposition 2.1. Equipped with this simple characterization of Boolean functions, we apply the uncertainty principle to estimate the size of the support of $f^2 - 1$ [GT12].

In the second part, we study the inverse problem of testing Booleanity: given a Boolean function, we test its sparsity. The high-level idea is to hash Fourier coefficients into random cosets to test whether the function is low-dimensional [Fel+06]. However, this is not sufficient to conclude that sparsity holds. A common way to obtain an s-sparse approximation is to keep the largest s Fourier coefficients and then take the sign, but taking the sign may destroy the desired sparsity. Instead, we leverage a more structural analysis of sparse Boolean functions and use a different truncation scheme to overcome this issue [Gop+11].

2 Preliminary

2.1 The Fourier transform on hypercube

Every function defined on hypercube $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ has a unique representation as

$$f(x) = \sum_{\alpha \in \mathbb{Z}_2^n} \hat{f}(\alpha)\chi_\alpha(x), \quad \text{where } \chi_\alpha(x) \stackrel{\text{def}}{=} (-1)^{\langle \alpha, x \rangle} = (-1)^{\sum_{i=1}^n \alpha_i x_i}. \tag{1}$$

The functions $\chi_\alpha(\cdot)$ are character functions. The Fourier coefficients are given by

$$\hat{f}(\alpha) = \mathbb{E}_{y \in \mathbb{Z}_2^n}[f(y)\chi_\alpha(y)]. \tag{2}$$

We write $\mathrm{Spec}(f)$ for the Fourier spectrum of f, i.e. the set $\mathrm{supp}(f) = \{\alpha \in \mathbb{Z}_2^n : \hat{f}(\alpha) \neq 0\}$. The sparsity of f is $\mathrm{sp}(f) = |\mathrm{Spec}(f)|$.

We define $\boldsymbol{\delta} : \mathbb{Z}_2^n \mapsto \mathbb{R}$ by

$$\boldsymbol{\delta}(x) = \begin{cases} 1 & \text{when } x = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Let $\mathbf{1}(x) : \mathbb{Z}_2^n \mapsto \mathbb{R}$ denote the constant function such that $\mathbf{1}(x) = 1$ for all $x \in \mathbb{Z}_2^n$, then it is easy to verify that

$$\hat{\mathbf{1}} = \boldsymbol{\delta}. \tag{3}$$

Given functions $f, g : \mathbb{Z}_2^n \mapsto \mathbb{R}$, their convolution $f * g$ is also a function from $\mathbb{Z}_2^n$ to $\mathbb{R}$, which is defined by

$$[f * g](x) = \sum_{y \in \mathbb{Z}_2^n} f(y) g(x + y).$$

We denote $f^{(2)} = f * f$, and more generally $f^{(k)}$ is the convolution of f with itself $k - 1$ times. $f^{(0)}$ is taken to equal $\boldsymbol{\delta}$, since $f * \boldsymbol{\delta} = f$. The convolution theorem for $\mathbb{Z}_2^n$ states that, up to multiplying by a constant, the Fourier transform of the pointwise product of two functions is equal to the convolution of their Fourier transforms, and that likewise the Fourier transform of a convolution is the product of the Fourier transforms:

$$\widehat{f \cdot g} = \hat{f} * \hat{g}, \quad \text{and} \quad \widehat{f * g} = 2^n \hat{f} \cdot \hat{g}. \tag{4}$$

We say function f is Boolean if its image is contained in $\{-1, 1\}$. The following result can be directly derived from Eqs. (3) and (4) [GT12, Proposition 3.1].

Proposition 2.1. *$f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ is Boolean iff $\hat{f} * \hat{f} = \boldsymbol{\delta}$.*

2.2 The discrete uncertainty principle

The discrete uncertainty principle for $\mathbb{Z}_2^n$ is a straightforward consequence of harmonic analysis on finite Abelian groups.

Theorem 2.2. *for any non-zero function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ (i.e., $\|f\| > 0$) it holds that*

$$|\operatorname{supp}(f)| \cdot |\operatorname{supp}(\hat{f})| \geq 2^n. \tag{5}$$

A stronger uncertainty principle can be derived from the information entropy inequality in [DCT02, Theorem 23]. We define the entropy of a function f by

$$H[f] = -\sum_{x \in \mathbb{Z}_2^n} f(x)^2 \log f(x)^2,$$

where logarithms are base two and $0 \log 0 = 0$.

Theorem 2.3. *for any non-zero function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ (i.e., $\|f\| > 0$) it holds that*

$$H\left[\frac{f}{\|f\|}\right] + H\left[\frac{\hat{f}}{\|\hat{f}\|}\right] \geq n, \tag{6}$$

where the L_2-norm is defined as

$$\|f\| = \sqrt{\sum_{x \in \mathbb{Z}_2^n} f(x)^2}.$$

A well-known fact that follows from the Kullback-Leibler divergence is $H[f/\|f\|] \leq \log |\operatorname{supp}(f/\|f\|)|$. Therefore, the following result is directly derived from this and Theorem 2.3.

Theorem 2.4. *for any non-zero function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ (i.e., $\|f\| > 0$) it holds that*

$$|\operatorname{supp}(f)| \cdot 2^{H\left[\hat{f}/\|\hat{f}\|\right]} \geq 2^n. \tag{7}$$

We note that for the proof of Theorem 3.2, we use Theorem 2.2, whereas for the more general case of Theorem 3.4, we must use Theorem 2.4.

2.3 Projection and hashing of the Fourier Spectrum

Definition 2.5. Given a subspace $H \leq \mathbb{Z}_2^n$ and a coset $r+H$, define the projection operator P_{r+H} on function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ as:

$$\widehat{P_{r+H}f}(\alpha) = \begin{cases} \hat{f}(\alpha) & \text{if } \alpha \in r+H \\ 0 & \text{otherwise} \end{cases}.$$

We have $P_{r+H}f = \frac{1}{2^n} \cdot A_{r+H} * f$, where $A_{r+H} = \sum_{\alpha \in r+H} \chi_\alpha = \chi_r \cdot \sum_{h \in H} \chi_h$.

With a simple fact $\sum_{h \in H} \chi_h = |H| \cdot \mathbf{1}_{H^\perp}$, we have:

$$P_{r+H}f(x) = \mathbb{E}_{y \in H^\perp}\left[\chi_r(y) f(x+y)\right]. \tag{8}$$

By applying Chernoff bound, we can empirically estimate the expectation in Eq. (8) [Gop+11, Proposition 1].

Proposition 2.6. *For any $x \in \mathbb{Z}_2^n$, the value $P_{r+H}f(x)$ can be estimated to within $\pm\tau$ with confidence $1-\delta$ using $O(\log(1/\delta)/\tau^2)$ queries to f.*

Denote the Fourier weight of f on a set $\mathcal{A}$ by $\mathrm{wt}(\mathcal{A}) = \sum_{\alpha \in \mathcal{A}} \hat{f}(\alpha)^2$. Using Parseval's identity, we have

$$\mathrm{wt}(r+H) = \sum_{\alpha \in r+H} \hat{f}(\alpha)^2 = \mathbb{E}_{w \in \mathbb{Z}_2^n}\left[P_{r+H}f(w)^2\right] = \mathbb{E}_{w \in \mathbb{Z}_2^n, y_1, y_2 \in H^\perp}\left[\chi_r(y_1)\chi_r(y_2) f(w+y_1) f(w+y_2)\right].$$

Letting $x = w + y_1$ and $z = y_1 + y_2$, we have [Gop+11, Proposition 2]:

$$\mathrm{wt}(r+H) = \mathbb{E}_{x \in \mathbb{Z}_2^n, z \in H^\perp}\left[\chi_r(z) f(x) f(x+z)\right].$$

Proposition 2.7. *The value $\mathrm{wt}(r+H)$ can be estimated to within $\pm\tau$ with confidence $1-\delta$ using $O(\log(1/\delta)/\tau^2)$ queries to f.*

We introduce pairwise independent hashing of the Fourier coefficients to generate a random coset structure.

Definition 2.8. For $t \in \mathbb{N}$, we define a random t-dimensional coset structure $(\mathcal{H}, \mathcal{C})$ as follows: We choose vectors $\beta_1, \ldots, \beta_t \in \mathbb{Z}_2^n$ independently and uniformly and set $\mathcal{H} = \mathrm{span}\{\beta_1, \ldots, \beta_t\}^\perp$. For each $b \in \mathbb{Z}_2^t$ we define the "bucket"

$$C(b) = \{\alpha \in \mathbb{Z}_2^n : \langle \alpha, \beta_i \rangle = b_i, \forall i\}.$$

We take $\mathcal{C}$ to be the multiset of $C(b)$'s, which has cardinality 2^t.

Remark 2.9. Given such a random coset structure, if β_i's are linearly independent, then the buckets $C(b)$ are precisely the cosets in $\mathbb{Z}_2^n/\mathcal{H}$, and the coset-projection function $P_{C(b)}$ is defined as Definition 2.5. In the (usually unlikely) case that the β_i's are linearly dependent, then some of the $C(b)$ will be cosets in $\mathbb{Z}_2^n/\mathcal{H}$ and the others will be empty. For the empty bucket $C(b)$, we define $P_{C(b)}f$ to be 0. ♢

Next, we derive several properties of this random hashing structure [Gop+11, Proposition 3].

Proposition 2.10. *Let $(\mathcal{H}, \mathcal{C})$ be a random t-dimensional coset structure. Define random variable $I_{\alpha \to b}$ to be the indicator for the event that $\alpha \in C(b)$.*

(a) For each $\alpha \in \mathbb{Z}_2^n \backslash \{0\}$ and each $b \in \mathbb{Z}_2^t$, we have $\mathbb{P}_{\beta_1, \ldots, \beta_t}[\alpha \in C(b)] = 2^{-t}$.

(b) Let $\alpha, \alpha' \in \mathbb{Z}_2^n$ be distinct. Then $\mathbb{P}_{\beta_1, \ldots, \beta_t}[\alpha, \alpha'$ belong to the same bucket$] = 2^{-t}$.

(c) Fix any set $S \subseteq \mathbb{Z}_2^n$ with $|S| \leq s+1$. If $t \geq 2\log s + \log(1/\delta)$ then except with probability at most δ, all vectors in S fall into different buckets.

(d) For each b, the collection of random variables $(I_{\alpha \to b})_{\alpha \in \mathbb{Z}_2^n}$ is pairwise independent.

Proof. Part a is because for any $\alpha \neq 0$, each $\langle \alpha, \beta_i \rangle$ is an independent uniformly random bit. Part b is because each $\langle \alpha - \alpha', \beta_i \rangle$ is an independent uniformly random bit, so that the probability that $\langle \alpha, \beta_i \rangle = \langle \alpha', \beta_i \rangle$ for all i is 2^{-t}. Part c comes from Part b by taking a bound that there are at most $\binom{s+1}{2} \leq s^2$ distinct pairs in S. For part d, assume first that $\alpha \neq \alpha'$ are both nonzero. Then from the fact that α and α' are linearly independent, then $\mathbb{P}_{\beta_1,\ldots,\beta_t}[\alpha, \alpha' \in C(b)] = 2^{-2t}$ as required. On the other hand, if one of $\alpha \neq \alpha'$ is zero, then $\mathbb{P}_{\beta_1,\ldots,\beta_t}[\alpha, \alpha' \in C(b)] = \mathbb{P}_{\beta_1,\ldots,\beta_t}[\alpha \in C(b)]\mathbb{P}_{\beta_1,\ldots,\beta_t}[\alpha' \in C(b)]$, which can be verified by checking the two cases $b = 0$ and $b \neq 0$. □

There is still something to be amended in Proposition 2.10(a) that $\alpha = 0$ is always hashed to $C(0)$. We can easily handle this problem by renaming buckets with a random permutation.

Definition 2.11. In a random permuted t-dimensional coset structure, we additionally choose a random $z \in \mathbb{Z}_2^n$ and rename $C(b)$ by $C(b + z)$.

Proposition 2.12. *For a random permuted t-dimensional coset structure, Proposition 2.10 still holds, with Proposition 2.10(a) even holding for $\alpha = 0$.*

2.4 Analysis of s-sparse functions

In this part, we present theorems about close-to-sparse Boolean functions, which are crucial to s-sparsity testing.

Definition 2.13. Let $B_f = \{\alpha_1, \ldots, \alpha_s\}$ denote the s-largest Fourier coefficients of f, and let $S_f = \overline{B_f}$ be its complement. We say that f is μ-close to s-sparse if

$$\sum_{\alpha \in S_f} \hat{f}(\alpha)^2 \leq \mu^2.$$

Definition 2.14. We say a rational number is k-granular if it is of the form $(integer)/2^k$. We say a function f is k-granular if $\hat{f}(\alpha)$ is k-granular for every α. We say a number v is μ-close to k-granular if $|v - j/2^k| \leq \mu$ for some integer j.

The following theorem is key to establishing the completeness of the s-sparsity test; it states that for any function that is close to being sparse, all large Fourier coefficients are close to being granular [Gop+11, Theorem 1].

Theorem 2.15. *If Boolean function f is μ-close to s-sparse, then each $\hat{f}(\alpha)$ for $\alpha \in B_f$ is $\frac{\mu}{\sqrt{s}}$-close to $\lceil \log s \rceil$-granular.*

Proof. Let $t = \lceil \log s \rceil + 1$ and construct a random permuted t-dimensional coset structure $(\mathcal{H}, \mathcal{C})$. We have

$$P_{C(b)} f(x) = \sum_{\alpha \in C(b)} \hat{f}(\alpha) \chi_\alpha(x).$$

Select any $\alpha_i \in B_f$. We will show that there exists a choice of $\beta_1, \ldots, \beta_t$ and b, such that the following two events happen together with non-zero probability:

1. α_i is the unique coefficient in $B_f \cap C(b)$,
2. $\mathrm{wt}(S_f \cap C(b)) \leq \mu^2/s$.

It is easy to say that $\mathbb{P}_{\beta_1,\ldots,\beta_t,b}[\alpha_i \in C(b)] = 2^{-t}$. Let us condition on this event. Because of pairwise independence, for any $j \neq i$, $\mathbb{P}_{\beta_1,\ldots,\beta_t,b}[\alpha_j \in C(b) | \alpha_i \in C(b)] = 2^{-t} \leq \frac{1}{2s}$. Thus,

$$\mathbb{E}_{\beta_1,\ldots,\beta_t,b}[\#\{j \neq i : \alpha_j \in B \cap C(b)\} | \alpha_i \in C(b)] = \frac{s-1}{2^k} < \frac{1}{2}.$$

Hence, by Markov's inequality:

$$\mathbb{P}_{\beta_1,\ldots,\beta_t,b}[\exists j \neq i \text{ such that } \alpha_j \in B \cap C(b) | \alpha_i \in C(b)] < \frac{1}{2}. \tag{9}$$

Now consider the second event. We have

$$\mathbb{E}_{\beta_1,\ldots,\beta_t,b}\left[\sum_{\beta\in S_f\cap C(b)}\hat{f}(\beta)^2|\alpha_i\in C(b)\right]=\sum_{\beta\in S_f}\mathbb{P}[\beta\in C(b)|\alpha_i\in C(b)]\hat{f}(\beta)^2\le 2^{-t}\mu^2\le\frac{\mu^2}{2s}.$$

Hence, by Markov's inequality,

$$\mathbb{P}_{\beta_1,\ldots,\beta_t,b}\left[\sum_{\beta\in S_f\cap C(b)}\hat{f}(\beta)^2\ge\frac{\mu^2}{s}|\alpha_i\in C(b)\right]\le\frac{1}{2}. \tag{10}$$

Thus, by applying the bound to Eq. (9) and **??**, we have both the desired events happening with non-zero probability over the choice of $\beta_1,\ldots,\beta_t$ and b. Fixing this choice, we have

$$P_{C(b)}f(x)=\hat{f}(\alpha_i)\chi_{\alpha_i}(x)+\sum_{\beta\in S_f\cap C(b)}\hat{f}(\beta)\chi_\beta(x),\quad\text{where}\sum_{\beta\in S\cap C(b)}\hat{f}(\beta)^2\le\frac{\mu^2}{s}.$$

But by Eq. (8), we also have $P_{C(b)}f(x)=\mathbb{E}_{y\in\mathcal{H}^\perp}[\chi_b(y)f(x+y)]$. Thus, the value of the function $P_{C(b)}f(x)$ is the average of a Boolean function over 2^t points, hence it is $(t-1)$-granular.

Now we consider the function

$$g(x)=\sum_{\beta\in S_f\cap C(b)}\hat{f}(\beta)\chi_\beta(x).$$

We have know that $\mathbb{E}_{x\in\mathbb{Z}_2^n}\left[g(x)^2\right]\le\frac{\mu^2}{s}$, then there exist $x_0\in\mathbb{Z}_2^n$ such that $g(x_0)^2\le\frac{\mu^2}{s}$, hence $g(x_0)\le\frac{\mu}{\sqrt{s}}$. Fixing this x_0, we have $P_{C(b)}(x_0)=\hat{f}(\alpha_i)\chi_{\alpha_i}(x_0)+g(x_0)$, hence

$$|\hat{f}(\alpha_i)|=|P_{C(b)}f(x_0)-g(x_0)|.$$

Since $P_{C(b)}f(x_0)$ is $(t-1)$-granular and $|g(x_0)|\le\frac{\mu}{\sqrt{s}}$, we have $\hat{f}(\alpha_i)$ is $\frac{\mu}{\sqrt{s}}$-close to $\lceil\log s\rceil$-granular. □

Thus, if a Boolean function $f:\mathbb{Z}_2^n\mapsto\{-1,1\}$ has its Fourier spectrum concentrated on s coefficients, then it is close to an s-sparse, $\lceil\log s\rceil$-granular real-valued function. The next theorem shows that this real-valued function must be Boolean [Gop+11, Theorem 6].

Theorem 2.16. *Let $f:\mathbb{Z}_2^n\mapsto\{-1,1\}$ be μ-close to s-sparse, where $\mu\le\frac{1}{20s^2}$. Then there is an s-sparse Boolean function $F:\mathbb{Z}_2^n\mapsto\{-1,1\}$ within Hamming distance $\frac{\mu^2}{2}$.*

Proof. By Theorem 2.15 and let $k=\lceil\log s\rceil$, each $\hat{f}(\alpha_i)$ for $\alpha_i\in B_f=\{\alpha_1,\ldots,\alpha_i\}$ is $\frac{\mu}{\sqrt{s}}$-close to k-granular. So we write $\hat{f}(\alpha_i)$ as

$$\hat{f}(\alpha_i)=\hat{F}(\alpha_i)+\hat{G}(\alpha_i),$$

where $\hat{F}(\alpha_i)$ is k-granular and $|\hat{G}(\alpha_i)|\le\frac{\mu}{\sqrt{s}}$. For $\beta\in S_f$, we set $\hat{F}(\beta)=0$ and $\hat{G}(\beta)=\hat{f}(\beta)$. Thus we have $f(x)=F(x)+G(x)$, and F is s-sparse and k-granular. While we have

$$\mathbb{E}\left[G(x)^2\right]=\sum_{\alpha\in\mathbb{Z}_2^n}\hat{G}(\alpha)^2\le s\cdot\frac{\mu^2}{s}+\mu^2=2\mu^2.$$

We will show that F is Boolean and our claim follows. In this case, G's value must be in $\{-2,0,2\}$, so that $G(x)^2=4$ whenever $f(x)\ne F(x)$. So that

$$\mathbb{P}_x\left[f(x)\ne F(x)\right]=\mathbb{P}_x\left[G(x)^2=4\right]=\frac{1}{4}\mathbb{E}\left[G(x)^2\right]\le\frac{\mu^2}{2}.$$

We rewrite $f^2=\mathbf{1}$ as:

$$\mathbf{1}=f^2=F2+2FG+G^2=F^2+\underbrace{G(2f-G)}_{H}.$$

There must be $\widehat{F^2}(0) + \hat{H}(0) = 1$ and $\widehat{F^2}(\alpha) + \hat{H}(\alpha) = 0$ for all $\alpha \neq 0$. As we know that F is k-granular, so F^2 is $2k$-granular. Hence $|\widehat{F^2}(\alpha)|$ is either an integer, or at least $2^{-2k} \geq \frac{1}{4s^2}$-far from being an integer. If $|\hat{H}(\alpha)| < \frac{1}{4s^2}$ holds, then we must have $\widehat{F^2}(0) = 1$ and $\widehat{F^2}(\alpha) = 0$ for all $\alpha \neq 0$. That is $F^2 = 1$ and so F is Boolean.

Applying Cauchy-Schwartz and Parseval, we have

$$\begin{aligned}|\hat{H}(\alpha)| &= \left|\sum_\beta \hat{G}(\beta)\widehat{2f - G}(\alpha+\beta)\right| \leq \sqrt{\sum_\beta \hat{G}(\beta)^2}\sqrt{\sum_\beta \widehat{2f-G}(\alpha+\beta)^2} \\ &= \frac{1}{2^n}\|G\|\|2f - G\| \leq \frac{1}{2^n}\|G\|(2\|f\| + \|G\|) \leq 2\sqrt{2}\mu + 2\mu^2 < 5\mu \leq \frac{1}{4s^2}.\end{aligned}$$

□

3 Testing Booleanity

3.1 Main results

We say function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$ is k-sparse if $|\operatorname{supp}(\hat{f})| \leq k$. The following result tells us that a function with a sparse Fourier expansion is either Boolean or far from Boolean [GT12, Theorem 1.1].

Theorem 3.1. *Every k-sparse function f is either Boolean or satisfies*

$$\mathbb{P}_x[f(x) \notin \{-1, 1\}] \geq \frac{2}{(k+2)^2},$$

where $\mathbb{P}_x[\cdot]$ denotes the uniform distribution over the domain of f.

We'll prove a more general result [GT12, Theorem 1.2]:

Theorem 3.2. *Let $D \subset \mathbb{R}$ be a set with d elements. Then for any k-sparse function f, either $\mathbb{P}_x[f(x) \in D] = 1$, or satisfies*

$$\mathbb{P}_x[f(x) \notin D] \geq \frac{d!}{(k+d)^d},$$

In particular, for $D = \{-1, 1\}$ or any other set of size two, this theorem reduces to Theorem 3.1. An immediate consequence of Theorem 3.1 is that there exists a non-adaptive algorithm with $O(k^2 \log(1/\epsilon))$ queries for ϵ-testing whether f is Boolean. Furthermore, any algorithm (adaptive or non-adaptive, even with two-sided error allowed) must make at least $\Omega(k)$ queries [GT12, Theorem 1.4].

Theorem 3.3. *Let A be a randomized algorithm that, given k and oracle access to a k-sparse function f,*

- *returns* true *with probability at least $2/3$ if f is Boolean, and*
- *returns* false *with probability at least $2/3$ if f is not Boolean.*

Then A has query complexity $\Omega(k)$.

A natural extension is to consider the Fourier transform $\hat{f}$ of the function not constrained to have a support set of size k, but instead to require that its entropy is $\log k$. Unfortunately, these results do not generalize to this relaxation. However, another augmentation is to consider the entropy of $\hat{f} * \hat{f}$, which in fact is at most $2 \log k$, as this is a natural consequence of Proposition 3.5. We say f is ϵ-close to Boolean if

$$\sqrt{\frac{1}{2^n}\sum_{x \in \mathbb{Z}_2^n}(f(x)^2 - 1)^2} \leq \epsilon.$$

This is simply the L_2 distance of f^2 from the constant function $\mathbf{1}$. In the following theorem, we test for ϵ-closeness to Booleanity [GT12, Theorem 1.5].

Theorem 3.4. *Let* $H\left[\frac{\hat{f}*\hat{f}}{\|\hat{f}*\hat{f}\|}\right] \le 2\log k$*, and let* $\|f\|^2 = 2^n$*. Then* f *is either* ϵ*-close to Boolean, or satisfies*

$$\mathbb{P}_x[f(x) \notin \{-1,1\}] = \Omega\left(\frac{1}{k^{2(\epsilon^2+1)/\epsilon^2}}\right).$$

3.2 Testing Booleanity with oracle access

We begin by proving the following fact, which relates the support of functions f and g with the support of their convolution [GT12, Proposition 3.4].

Proposition 3.5. *Let* $f, g : \mathbb{Z}_2^n \mapsto \mathbb{R}$*. Then*

$$\operatorname{supp}(f*g) \subseteq \operatorname{supp}(f) \oplus \operatorname{supp}(g).$$

Here, $\operatorname{supp}(f) \oplus \operatorname{supp}(g)$ *is the Minkowski sum, which is the set of elements of* $\mathbb{Z}_2^n$ *that can be written as the sum of an element in* $\operatorname{supp}(f)$ *and an element in* $\operatorname{supp}(g)$*.*

Proof. Let $x \in \operatorname{supp}(f*g)$. Then, by the definition of convolution, there exist y and z such that $f(y) \neq 0$ and $g(z) \neq 0$ and $x = y + z$. Hence $x \in \operatorname{supp}(f) \oplus \operatorname{supp}(g)$. □

Proof of Theorem 3.2. Let $D = \{y_1, \dots, y_d\}$. Let

$$g = \prod_{i=1}^{d}(f - y_i),$$

so that $g(x) = 0$ means $f(x) \in D$. We now compute the Fourier transform $\hat{g}$.

$$\hat{g} = (\hat{f} - y_1\boldsymbol{\delta}) * \cdots * (\hat{f} - y_d\boldsymbol{\delta}) = \hat{f}^{(d)} + a_{d-1}\hat{f}^{(d-1)} + \cdots + a_1\hat{f} + a_0\boldsymbol{\delta},$$

for some coefficients $(a_{d-1}, \dots, a_0)$. Therefore,

$$\operatorname{supp}(\hat{g}) \subseteq \bigcup_{i=1}^{d} \operatorname{supp}(\hat{f}^{(i)}) \cup \{0\}.$$

Let $A = \operatorname{supp}(\hat{f}) \cup \{0\}$. Then by Proposition 3.5 $\operatorname{supp}(\hat{f}^{(i)}) \subset iA = A \oplus \cdots \oplus A$, where the sum is taken $i-1$ times. Since $0 \in A$, then for all $i \le d$ we have $iA \subset dA$. Hence

$$\operatorname{supp}(\hat{g}) \subseteq dA.$$

Therefore, $\operatorname{supp}(\hat{g})$ is contained in the set of elements that can be written as the sum of at most d elements of A. The size of this set is bounded by the number of ways to choose d elements of A with replacement, disregarding order. Since $|A| \le |\operatorname{supp}(\hat{f})| + 1 = k + 1$, we have

$$|\operatorname{supp}(\hat{g})| \le \binom{|A| - 1 + d}{d} \le \frac{(k+d)^d}{d!}.$$

Now, if $\mathbb{P}_x[f(x) \in D] = 1$, then clearly $\|g\| = 0$. Otherwise, $\|g\| > 0$. We apply Theorem 2.2, which implies that

$$|\operatorname{supp}(g)| \ge \frac{2^n d!}{(k+d)^d}.$$

Since the support of g is precisely the set $\{x \in \mathbb{Z}_2^n : f(x) \notin D\}$. Then it follows that

$$\mathbb{P}_x[f(x) \notin D] \ge \frac{d!}{(k+d)^d}.$$

□

Assuming oracle access to f, the algorithm samples f at random $\frac{1}{2}(k+2)^2\ln(1/\epsilon)$ times and will find an x such that $f(x) \notin \{-1,1\}$ with probability at least $1-\epsilon$, unless f is Boolean. We also show an $\Omega(k)$ lower bound using an indistinguishable case.

Proof of Theorem 3.3. Let A be an algorithm that is given oracle access to a function $f : \mathbb{Z}_2^n \mapsto \mathbb{R}$, together with the guarantee that $|\operatorname{supp}(\hat{f})| \leq k$. Denote by B_k the set of Boolean functions that depend only on the first $\log k$ coordinates, and denote by C_k the set of functions that likewise depend only on the first $\log k$ coordinates, return values in $\{-1,1\}$ for some $k-1$ of the k possible values of the first $\log k$ coordinates, but return 2 otherwise. Note that functions in both B_k and C_k have sparsity at most k.

We prove the lower bound on the query complexity by showing that any randomized algorithm that makes at most $o(k)$ queries would not be able to distinguish between these two distributions with non-negligible probability. Observe that an arbitrary query to f in either distribution would output a non-Boolean value with probability at most $1/k$, independently of previous queries with different values of the first $\log k$ coordinates. Therefore, any algorithm that makes $o(k)$ queries would find an input such that $f(x) = 2$ with probability $o(1)$, and thus would be unable to distinguish between B_k and C_k with noticeable probability. □

3.3 Proof of Theorem 3.4

We begin by proving a proposition related to information entropy [GT12, Proposition 3.5].

Proposition 3.6. *Let X be a discrete random variable, and let x_0 be a value for which $\mathbb{P}[X = x_0] > 0$. Then*

$$H(X|X \neq x_0) \leq \frac{H(X)}{\mathbb{P}[X \neq x_0]}.$$

Proof. Let A be the indicator of the event $X = x_0$. Then

$$\begin{aligned} H(X) &\geq H(X|A) \\ &= \mathbb{P}[X = x_0]H(X|X = x_0) + \mathbb{P}[X \neq x_0]H(X|X \neq x_0) \\ &= \mathbb{P}[X \neq x_0]H(X|X \neq x_0). \end{aligned}$$

□

Proof of Theorem 3.4. Assume that f is ϵ-far from being Boolean. Using Parseval's equation, we have

$$\|\hat{f}^{(2)}\|^2 = \frac{1}{2^n}\|f^2\|^2 = \frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n} f(x)^4 = \frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n}(f(x)^2-1)^2 + 1 \geq 1+\epsilon^2. \tag{11}$$

Let X be a $\mathbb{Z}_2^n$-valued random variable, such that $\mathbb{P}[X = x] = \hat{f}^{(2)}(x)^2/\|\hat{f}^{(2)}\|^2$. Since $\|f\|^2 = 2^n$, then $\hat{f}^{(2)}(0) = 1$, and

$$\mathbb{P}[X = 0] = \frac{\hat{f}^{(2)}(0)^2}{\|\hat{f}^{(2)}\|^2} \leq \frac{1}{\epsilon^2+1}.$$

Let $g = f^2 - 1$. Then $\hat{g} = \hat{f}^{(2)} - \boldsymbol{\delta}$, and $\hat{g}(0) = \hat{f}^{(2)}(0) - 1 = 0$. We have $\mathbb{P}[X = x|X \neq 0] = \hat{g}(x)^2/\|\hat{g}\|^2$. Apply Proposition 3.6 and Eq. (11), we obtain

$$H\left[\frac{\hat{g}}{\|\hat{g}\|}\right] \leq H\left[\frac{\hat{f}^{(2)}}{\|\hat{f}^{(2)}\|}\right]\cdot\frac{1+\epsilon^2}{\epsilon^2} \leq 2\cdot\frac{1+\epsilon^2}{\epsilon^2}\cdot\log k.$$

By Theorem 2.4, it follows that $|\operatorname{supp}(f^2-1)|\cdot k^{2(\epsilon^2+1)/\epsilon^2} \geq 2^n$, so that

$$\mathbb{P}_x[f(x) \notin \{-1,1\}] = \Omega\left(\frac{1}{k^{2(\epsilon^2+1)/\epsilon^2}}\right).$$

□

4 Testing sparsity

4.1 Main results

We present an algorithm for testing whether a Boolean function $f : \mathbb{Z}_2^n \mapsto \{-1, 1\}$ is s-sparse [Gop+11].

Algorithm 1 Testing s-sparsity

Inputs: f, s, ϵ.

1. Let $\mu = \min\left(\sqrt{2\epsilon}, \frac{1}{20s^2}\right)$, $t = \lceil 2\log s + \log 100 \rceil$, $\tau = \frac{\mu^2}{100 \cdot 2^t}$.
2. Choose a random permuted t-dimensional coset structure $(\mathcal{H}, \mathcal{C})$.
3. For each bucket $C \in \mathcal{C}$, estimate $\mathrm{wt}(C) = \sum_{\alpha \in C} \hat{f}(\alpha)^2$ to accuracy $\pm\tau$ with confidence $1 - \frac{1}{100 \cdot 2^t}$, using Proposition 2.7.
4. Let $\mathcal{L}$ be the set of buckets where the estimation is at least 2τ. If $|\mathcal{L}| \geq s + 1$, reject.

We briefly explain the algorithm. Step 2 pairwisely independently hashes the Fourier spectrum of f into $\Theta(s^2)$ buckets. If f is s-sparse, then at most s buckets will have non-zero weight and pass the test. On the other hand, if f passes this test with high probability, we show that almost all of the Fourier spectrum is concentrated on at most s coefficients (one from each bucket). Theorem 2.16 tells us that f is close to a sparse Boolean function. Our algorithm satisfies [Gop+11, Theorem 8]:

Theorem 4.1. *Algorithm 1 takes function $f : \mathbb{Z}_2^n \mapsto \{-1, 1\}$, s and ϵ as inputs, satisfies the following conditions while making $O\left(\frac{s^6 \log s}{\epsilon^2} + s^{14} \log s\right)$ nonadaptive queries.*

(a) If f is s-sparse, it passes with probability at least $2/3$;

(b) If f is ϵ-far from any s-sparse Boolean function, it passes with probability at most $1/3$.

The query complexity of Theorem 4.1 arises from Proposition 2.7, where the number of buckets is $2^t = O(s^2)$.

4.2 Proof of Theorem 4.1

Theorem 4.2. *If f is s-sparse, then it passes with probability at least 0.9.*

We begin with Theorem 4.1(a). This result is straightforward. Since there are 2^t buckets, and all of their estimates in Step 2 are τ-accurate, except with probability at most $1/100$. So if these estimates are accurate, the buckets in $\mathcal{L}$ are at most s in number. So f passes the test with probability at least 0.9.

We partition the Fourier coefficient of f into two sets: B_f os big coefficients and S_f of small coefficients.

$$B_f = \{\alpha : \hat{f}(\alpha) \geq 3\tau\}, \quad C_f = \{\alpha : \hat{f}(\alpha) < 3\tau\}.$$

We claim that if there are too many big coefficients, f will probably be rejected [Gop+11, Lemma 2].

Lemma 4.3. *If $|B_f| \geq s + 1$, then the test rejects with probability at least $3/4$.*

Proof. By Proposition 2.10(c), after Step 2, except with probability at most $1/100$, there are at least $s + 1$ buckets containing an element of set B_f. Assume we have accurate estimate, then $|\mathcal{L}|$ will be at least $s + 1$. Hence, the reject probability is at least $1 - 2/100$. □

Next, we show that if the weights on S_f are large enough, f will also be probably rejected [Gop+11, Lemma 3].

Lemma 4.4. *If $\mathrm{wt}(S_f) \geq \mu^2$, then the test rejects with probability at least $3/4$.*

Proof. For each bucket $C(b)$, we define a random variable $M_b = \mathrm{wt}(C(b) \cap S_f) = \sum_{\alpha \in S_f} \hat{f}(\alpha)^2 \cdot I_{\alpha \to b}$. We say a bucket $C(b)$ is good if $M_b \geq \frac{1}{2}\mathbb{E}[M_b]$, while we have

$$\mathbb{E}[M_b] = 2^{-t}\,\mathrm{wt}(S_f) \geq 100\tau > 0.$$

By pairwise independence, we have

$$\begin{aligned}\mathbf{Var}[M_b] &= \sum_{\alpha \in S_f} \mathbf{Var}\left[\hat{f}(\alpha)^2 \cdot I_{\alpha \to b}\right] \leq \sum_{\alpha \in S_f} \mathbb{E}\left[\left(\hat{f}(\alpha)^2 \cdot I_{\alpha \to b}\right)^2\right] \\ &\leq 3\tau \cdot \sum_{\alpha \in S_f} \mathbb{E}\left[\hat{f}(\alpha)^2 \cdot I_{\alpha \to b}\right] = 3\tau \mathbb{E}[M_b].\end{aligned}$$

By Chebyshev's inequality, we have

$$\mathbb{P}\left[M_b \leq \frac{1}{2}\mathbb{E}[M_b]\right] \leq \frac{3\tau}{(1/2)^2 \mathbb{E}[M_b]} \leq \frac{3}{25}.$$

Thus, the expected fraction of bad buckets is at most $3/25$. By Markov's inequality, there are at most $(3/5)2^t$ bad buckets except with probability at most $1/5$. However, if there are at least $(2/5)2^t \geq 40s^2 \geq s+1$ good buckets with $\mathrm{wt}(C(b) \cap S_f) \geq \frac{1}{2}\mathbb{E}[M_b] \geq 50\tau$, the test will reject. Thus, we reject except with probability at most $1/5 + 1/100 < 1/4$. □

Finally, we establish Theorem 4.1(b) [Gop+11, Lemma 4].

Theorem 4.5. *Suppose the test accepts f with probability exceeding $1/4$. Then f is ϵ-close to an s-sparse Boolean function.*

Proof. By Lemma 4.3, we have $|B_f| \leq s$, and by Lemma 4.4, we have $\mathrm{wt}(S_f) \leq \mu^2$. Thus f is μ-close to being s-sparse. We now apply Theorem 2.16 and $\mu \leq \frac{1}{20s^2}$ to conclude that f must be close in Hamming distance to an s-sparse Boolean function. □

References

[GT12] Tom Gur and Omer Tamuz. "Testing Booleanity and the Uncertainty Principle". In: *ArXiv* abs/1204.0944 (2012). URL: https://api.semanticscholar.org/CorpusID:14539523.

[Fel+06] Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. "New results for learning noisy parities and halfspaces". In: *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE. 2006, pp. 563–574.

[Gop+11] Parikshit Gopalan, Ryan O'Donnell, Rocco A Servedio, Amir Shpilka, and Karl Wimmer. "Testing Fourier dimensionality and sparsity". In: *SIAM Journal on Computing* 40.4 (2011), pp. 1075–1100.

[DCT02] Amir Dembo, Thomas M Cover, and Joy A Thomas. "Information theoretic inequalities". In: *IEEE Transactions on Information theory* 37.6 (2002), pp. 1501–1518.