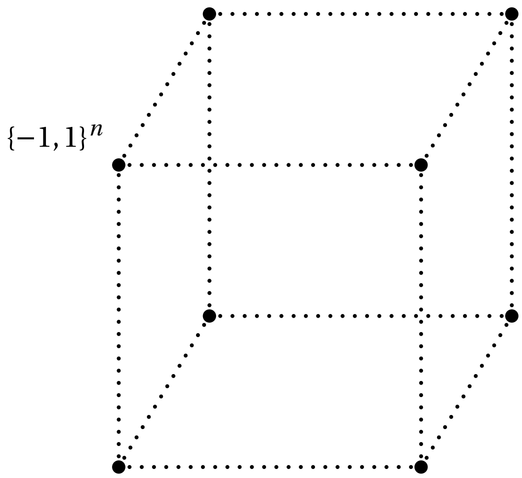
Analysis of Boolean Functions

Math 581A — Fall 2025

Thomas Rothvoss



UNIVERSITY *of*
WASHINGTON

Last changes: March 29, 2026

Contents

Chapter 1

Introduction to boolean functions

This course deals with the analysis of functions of the form $f : \{\pm 1\}^n \to \mathbb{R}$. The main tool will be *Fourier analysis* and we will see a rich set of applications to *theoretical computer science* and *combinatorics*. The main source for these notes is the terrific textbook by Ryan O'Donnell [O'D21] which is available for free on Arxiv[1]. The book was first published in 2014 and we add some more recent results that appeared later. Another excellent source are the lecture notes of Hamed Hatami[2].

Inspiration for the selection of additional material comes from the course *Analysis of Boolean Functions* by given by Avishay Tal in Spring 2023 at UC Berkeley[3] as well as the Spring 2021 course *Topics in Combinatorics: Analysis of Boolean Functions* given by Dor Minzer at MIT[4]. Moreover, we rely on the survey by Arturs Backurs[5].

1.1 The basics

As mentioned earlier the goal is to study functions of the form $f : \{\pm 1\}^n \to \mathbb{R}$.

---

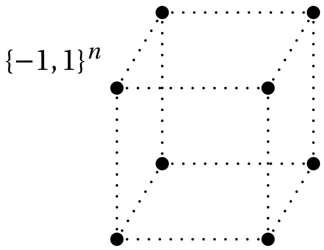
[1] See `https://arxiv.org/abs/2105.10386`

[2] See `https://cs.mcgill.ca/~hatami/Boolean_function_analysis.pdf`.

[3] See `https://www.avishaytal.org/cs294-analysis-of-boolean-functions`

[4] See `https://ocw.mit.edu/courses/18-218-topics-in-combinatorics-analysis-of-boolean-functions-s`

[5] See `https://www.scottaaronson.com/showcase2/report/arturs-backurs.pdf`



For two such functions $f, g : \{\pm 1\}^n \to \mathbb{R}$ we define an inner product

$$\langle f, g\rangle_E := \mathop{\mathbb{E}}_{x\sim\{\pm1\}^n}[f(x)\cdot g(x)] = \frac{1}{2^n}\sum_{x\in\{\pm1\}^n} f(x)\cdot g(x)$$

that is sometimes called the *expectation inner product* and we use the unusual notation $\langle\cdot,\cdot\rangle_E$ to remind ourselfs of the factor $\frac{1}{2^n}$ that is not present in the standard inner product. Here we write $x \sim \{-1,1\}^n$ to indicate that x is a vector that is drawn uniformly at random from $\{-1,1\}^n$. For a set $S \subseteq [n]$, consider the special function

$$\chi_S : \{\pm1\}^n \to \{\pm1\} \quad \text{with} \quad \chi_S(x) := \prod_{i\in S} x_i \ \ \forall x \in \{-1,1\}^n$$

The function χ_S is also called the *character function*. We denote $S\Delta T := (S\setminus T)\cup(T\setminus S)$ as the *symmetric difference* of sets $S, T \subseteq [n]$. We show a convinient fact for these special character functions:

Lemma 1.1. *For $S, T \subseteq [n]$ one has*

$$\langle\chi_S,\chi_T\rangle_E = \begin{cases} 1 & \text{if } S = T \\ 0 & \text{otherwise.}\end{cases}$$

Proof. We write

$$\langle\chi_S,\chi_T\rangle_E = \mathop{\mathbb{E}}_{x\sim\{\pm1\}^n}[\chi_S(x)\cdot\chi_T(x)] = \mathop{\mathbb{E}}_{x\sim\{\pm1\}^n}[\chi_{S\Delta T}(x)] = \prod_{i\in S\Delta T}\underbrace{\mathop{\mathbb{E}}_{x_i\sim\{\pm1\}}[x_i]}_{=0} = \begin{cases} 0 & \text{if } |S\Delta T| > 0 \\ 1 & \text{if } |S\Delta T| = 0\end{cases}$$

Here we use that $\chi_S(x)\cdot\chi_T(x) = \chi_{S\Delta T}(x)$. We also use that for independent random variables X and Y one has $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. □

We note that the set

$$V_n := \left\{f \mid f : \{\pm1\}^n \to \mathbb{R}\right\}$$

is a *vector space* of dimension 2^n and Lemma 1.1 says that the family of 2^n many functions $\{\chi_S\}_{S\subseteq[n]}$ is pairwise orthogonal and even orthonormal. Hence $\{\chi_S\}_{S\subseteq[n]}$ must be an *orthonormal basis* for that vector space. It then makes sense to consider the *coordinates* that an element $f : \{\pm1\}^n \to \mathbb{R}$ has with respect to that basis:

Definition 1.2. For $f : \{\pm 1\}^n \to \mathbb{R}$ and $S \subseteq [n]$ we denote the S*-th Fourier coefficient* as

$$\hat{f}(S) := \langle f, \chi_S \rangle_E = \mathop{\mathbb{E}}_{x \sim \{\pm 1\}^n} [f(x) \cdot \chi_S(x)].$$

By orthonormality we know the following:

Theorem 1.3 (Fourier Expansion Theorem)**.** *For every function* $f : \{\pm 1\}^n \to \mathbb{R}$ *there is a unique linear combination in terms of the character functions which is* $f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(x)$ *for* $x \in \{\pm 1\}^n$*.*

We make the following definition.

Definition 1.4. For $f : \{-1, 1\}^n \to \mathbb{R}$ we define the *degree* as[6] $\deg(f) := \max\{|S| : \hat{f}(S) \neq 0\}$.

Theorem 1.3 represents f as a multivariate multi-linear polynomial and $\deg(f)$ denotes its total degree. The following can be obtained by applying Theorem 1.3 and using the orthonormality of the characters.

Theorem 1.5. *For any* $f, g : \{\pm 1\}^n \to \mathbb{R}$ *one has*

(i) Plancharel's Theorem: $\langle f, g \rangle_E = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \hat{g}(S)$

(ii) Parsival's identity: $\langle f, f \rangle_E = \sum_{S \subseteq [n]} \hat{f}(S)^2$

Proof. For (i) we use Theorem 1.3 and linearity of $\langle \cdot, \cdot \rangle_E$ to write

$$\langle f, g \rangle_E = \sum_{S \subseteq [n]} \sum_{T \subseteq [n]} \hat{f}(S) \hat{g}(T) \underbrace{\langle \chi_S, \chi_T \rangle_E}_{=1 \text{ if } S=T, =0 \text{ o.w.}} = \sum_{S \subseteq [n]} \hat{f}(S) \hat{g}(S)$$

Then (ii) is a special case of (i). □

One should think of Plancharel's Theorem as the basic fact that for two elements f and g in a vector space one can obtain their inner product by summing up the coordinate-wise products with respect to any orthonormal basis. That brings us to the question why actually we have picked $\{\chi_S\}_{S \subseteq [n]}$ as a basis and not any other basis such as the standard basis which in this case would be $e_y : \{-1, 1\}^n \to \{0, 1\}$ with

$$e_y(x) := \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

The answer is that the Fourier basis takes the geometry of the hypercube into account and many statements become easier when being considered in the Fourier basis.

---

[6] We can make the convention that the zero-everywhere function has degree -1.

1.2 Fourier weights

For a function $f:\{-1,1\}^n \to \{-1,1\}$ we know by Parsival's identity that $\sum_{S\subseteq[n]} \hat{f}(S)^2 = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)^2] = 1$. So it makes sense to think of the values $\hat{f}(S)^2$ as a probabilities:

Definition 1.6. For $f:\{-1,1\}^n \to \{-1,1\}$ we denote $\mathcal{S}_f$ as the distribution that returns a set $S \subseteq [n]$ with probability $\hat{f}(S)^2$. We call $\mathcal{S}_f$ the *spectral sample* for f.

Often it will be important whether most of the Fourier weight of a function f lies on large sets S or on small sets.

Definition 1.7. For $f:\{-1,1\}^n \to \mathbb{R}$ and $k \in \{0,\dots,n\}$ we define the *Fourier weight at level* k as

$$W^k[f] := \sum_{S\subseteq[n]:|S|=k} \hat{f}(S)^2$$

We also define $f^{=k}$ as the part of f coming from level k, i.e.

$$f^{=k}(x) := \sum_{|S|=k} \hat{f}(S)\chi_S(x)$$

1.3 Relationship of $\{-1,1\}^n$ to $\{0,1\}^n$

In many settings it would be more natural to study functions of the form $g:\{0,1\}^n \to \{0,1\}$, rather than $f:\{-1,1\}^n \to \{-1,1\}$, for example when we want to work with addition modulo 2 or subspaces in $\mathbb{F}_2^n$. But one can always map a vector $x \in \{0,1\}^n$ to the vector $((-1)^{x_1},\dots,(-1)^{x_n}) \in \{-1,1\}^n$ and then do the analysis in the $\{-1,1\}^n$ cube where the addition modulo 2 (denoted by $\oplus$) is replaced by the coordinate-wise multiplication $\odot$. Mathematically speaking, for each coordinate we have the two 2-element groups $(\{0,1\},\oplus)$ and $(\{-1,1\},\odot)$ and we map the neutral element of one to the neutral element of the other (and the non-neutral element to the non-neutral element).

We should remark that the book by O'Donnell [O'D21] rather freely switches back and forth between both cubes. Instead we will for the most of it stick with the $\{-1,1\}^n$-cube which possibly helps reduce confusion while it means we will work with somewhat less intuitive notions of convolution and $\{-1,1\}$-linearity.

Still, we want to give some background on functions on the $\{0,1\}^n$-cube. We will aim to use g as symbol for such functions while we reserve f for the $\{\pm1\}^n$-cube. We will leave the proofs for some of the stated claims as an exercise.

Any multilinear polynomial $F:\mathbb{R}^n \to \mathbb{R}$ is of the form

$$F(x) = \sum_{S\subseteq[n]} a_S \prod_{i\in S} x_i \quad \forall x \in \mathbb{R}^n \tag{1.1}$$

We know that there is a unique choice for the parameters a_S which are the Fourier coefficients $\hat{F}(S) = \mathbb{E}_{x\sim\{\pm 1\}^n}[F(x)\cdot\chi_S(x)]$. But that means we have expressed the numbers a_S in terms of the function values on $\{\pm 1\}^n$. So the first question should be: how can the numbers a_S be expressed in terms of the function values on $\{0,1\}^n$?

Proposition 1.8. *For a function $g : \{0,1\}^n \to \mathbb{R}$, there is a unique choice of values $\{a_S\}_{S\subseteq[n]} \subseteq \mathbb{R}$ so that*

$$g(y) = \sum_{S\subseteq[n]} a_S \prod_{i\in S} y_i \quad \forall y \in \{0,1\}^n$$

In fact, the values are

$$a_S = \sum_{T\subseteq S} (-1)^{|S\setminus T|} g(\mathbf{1}_T) \tag{1.2}$$

where $\mathbf{1}_T \in \{0,1\}^n$ is the characteristic vector of set T.

For a multilinear polynomial $F : \mathbb{R}^n \to \mathbb{R}$ as in (1.1) we know that the degree $\max\{|S| : a_S \neq 0\}$ coincides with the degree $\deg(F)$ as defined in Def 1.4. The degree is invariant under shifts:

Proposition 1.9. *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial. Then for $\alpha,\beta \in \mathbb{R}$ with $\beta \neq 0$ let $G : \mathbb{R}^n \to \mathbb{R}$ be defined by $G(x) := F(\alpha + \beta x)$. Then G is a multilinear polynomial with $\deg(F) = \deg(G)$. Moreover for all $S \subseteq [n]$ with $|S| = \deg(F)$ one has $\hat{G}(S) = \beta^{|S|}\hat{F}(S)$.*

In particular we can consider a function $f : \{-1,1\}^n \to \mathbb{R}$ and let $g : \{0,1\}^n \to \mathbb{R}$ be the function defined by $g(\frac{x+1}{2}) := f(x)$ for $x \in \{-1,1\}^n$. Then both functions will have the same degree.

1.4 Convolution

For two vectors $x, y \in \{\pm 1\}^n$ we write $x \odot y \in \{\pm 1\}^n$ as the vector with entries $(x \odot y)_i := x_i \cdot y_i$. As explained above, the $\odot$-operation is the analogue to addition in $\mathbb{F}_2$.

Definition 1.10. For functions $f, g : \{-1,1\}^n \to \mathbb{R}$ we define their *convolution* as the function $f * g : \{-1,1\}^n \to \mathbb{R}$ defined by

$$(f * g)(x) := \mathop{\mathbb{E}}_{y\sim\{-1,1\}^n} \left[f(x\odot y)\cdot g(y)\right] \quad \forall x \in \{-1,1\}^n$$

We want to describe an important application of convolution.

Definition 1.11. A function $f : \{-1,1\}^n \to \mathbb{R}$ is called a *(probability) density function* if $f(x) \geq 0$ for all $x \in \{-1,1\}^n$ and $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)] = 1$.

Note that for a probability density function f, according to our definition one has $\sum_{x\in\{-1,1\}^n} f(x) = 2^n$ which might be somewhat unintuitive but this scaling will work well for us.

Proposition 1.12. *If $f, g : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$ are density functions, then also $f * g$ is a density function. Moreover if $x \sim f$ and $y \sim g$ independently then $(x \odot y) \sim f * g$.*

Proof. Clearly

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[(f*g)(x)] = \mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}\Big[\underbrace{\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x\odot y)]}_{=1}\cdot g(y)\Big] = \mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}[g(y)] = 1$$

and so $f * g$ is indeed a density function. For the moreover part, for any fixed $z \in \{-1,1\}^n$ we have

$$\Pr_{x\sim f, y\sim g}[x\odot y = z] = \sum_{x\in\{-1,1\}^n} \Pr_f[x]\Pr_g[z\odot x] = \frac{1}{2^n}\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x)g(z\odot x)] = \frac{1}{2^n}\cdot(f*g)(z)$$

as claimed. □

For the sake of completeness we want to mention that for all $f, g, h : \{-1,1\}^n \to \mathbb{R}$ one has commutativity in the form of $f * g = g * f$ and associativity, i.e. $f * (g * h) = (f * g) * h$. Finally we will prove the important fact that the Fourier coefficient of the convolution is simply the product of the two Fourier coefficients of the original functions.

Theorem 1.13. *For all $f, g : \{-1,1\}^n \to \mathbb{R}$ and $S \subseteq [n]$ one has $\widehat{(f*g)}(S) = \hat{f}(S)\cdot\hat{g}(S)$.*

Proof. We have

$$\begin{aligned}
\widehat{(f*g)}(S) &= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\big[(f*g)(x)\cdot\chi_S(x)\big] \\
&\overset{\text{Def} *}{=} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\Big[\mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}[f(y)\cdot g(x\odot y)]\cdot\chi_S(x)\Big] \\
&\overset{(*)}{=} \mathop{\mathbb{E}}_{y,z\sim\{-1,1\}^n}\big[f(y)\cdot g(z)\cdot\chi_S(y\odot z)\big] \\
&= \underbrace{\mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}\big[f(y)\cdot\chi_S(y)\big]}_{=\hat{f}(S)}\cdot\underbrace{\mathop{\mathbb{E}}_{z\sim\{-1,1\}^n}\big[g(z)\cdot\chi_S(z)\big]}_{=\hat{g}(S)} = \hat{f}(S)\cdot\hat{g}(S)
\end{aligned}$$

In $(*)$ we make the substitution $z := x \odot y$ and we use that for fixed y, $x \odot y$ is uniform from $\{-1,1\}^n$. □

1.5 Restrictions

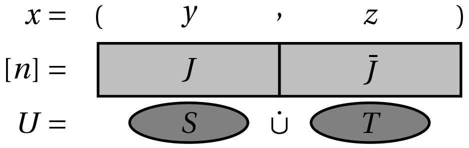
For a set $J \subseteq [n]$ of coordinates we will denote the complement as $\bar{J} := [n] \setminus J$.

Definition 1.14. For a function $f : \{-1,1\}^n \to \mathbb{R}$, an index set $J \subseteq [n]$ and $z \in \{-1,1\}^{\bar{J}}$, we define the *restriction of f to J using z* as the function $f_{J|z} : \{-1,1\}^J \to \mathbb{R}$ with $f_{J|z}(y) := f(y,z)$.

Intuitively speaking $f_{J|z}$ is the restriction of f to a subcube. It will be useful to determine the Fourier coefficients for the function $f_{J|z} : \{-1,1\}^J \to \mathbb{R}$ in terms of the original Fourier coefficients.

Proposition 1.15. *Let $f : \{-1,1\}^n \to \mathbb{R}$, $J \subseteq [n]$ and $z \in \{-1,1\}^{\bar{J}}$. Then for any $S \subseteq J$ one has $(\widehat{f_{J|z}})(S) = \sum_{T \subseteq \bar{J}} \hat{f}(S \cup T) \cdot \chi_T(z)$.*

Proof. For each $U \subseteq [n]$ there is a unique decomposition as $U = S \dot{\cup} T$ with $S \subseteq J$ and $T \subseteq \bar{J}$. Morever we can decompose $x \in \{-1,1\}^n$ as $x = (y,z)$ with $y \in \{-1,1\}^J$ and $z \in \{-1,1\}^{\bar{J}}$ so that $\chi_U(x) = \chi_S(y) \cdot \chi_T(z)$.



This can be used to write

$$f_{J|z}(y) = f(x) = \sum_{U \subseteq [n]} \hat{f}(U) \cdot \chi_U(x) = \sum_{S \subseteq J} \Big(\sum_{T \subseteq \bar{J}} \hat{f}(S \cup T) \cdot \chi_T(z) \Big) \cdot \chi_S(y)$$

Then by Theorem 1.3, the term $\sum_{T \subseteq \bar{J}} \hat{f}(S \cup T) \cdot \chi_T(z)$ has to be the Fourier coefficient $(\widehat{f_{J|z}})(S)$. □

We could also ask how the Fourier coefficients $(\widehat{f_{J|z}})(S)$ change as we vary z.

Proposition 1.16. *Let $f : \{-1,1\}^n \to \mathbb{R}$ and let $S \subseteq J \subseteq [n]$. Define*

$$F : \{-1,1\}^{\bar{J}} \to \mathbb{R} \quad \textit{with} \quad F(z) := (\widehat{f_{J|z}})(S)$$

Then the following holds

(a) One has $F(z) = \sum_{T \subseteq \bar{J}} \hat{f}(S \cup T) \cdot \chi_T(z)$.

(b) For all $T \subseteq \bar{J}$ one has $\hat{F}(T) = \hat{f}(S \cup T)$.

(c) One has $\mathbb{E}_{z \sim \{-1,1\}^{\bar{J}}}[F(z)] = \hat{f}(S)$.

(d) *One has* $\mathbb{E}_{z\sim\{-1,1\}^{\bar{J}}}[F(z)^2] = \sum_{T\subseteq\bar{J}} \hat{f}(S\cup T)^2$.

Proof. From Prop 1.15 we know that indeed

$$F(z) = \widehat{(f_{J|z})}(S) = \sum_{T\subseteq\bar{J}} \hat{f}(S\cup T)\cdot\chi_T(z)$$

which gives (a). Then again by Theorem 1.3, the Fourier coefficient $\hat{F}(T)$ has to be $\sum_{T\subseteq\bar{J}}\hat{f}(S\cup T)$ which gives (b). For (c) we use that $\mathbb{E}_{z\sim\{-1,1\}^{\bar{J}}}[F(z)] = \hat{F}(\emptyset) = \hat{f}(S)$ using (b). For (d) we use Parsival's Inequality (Theorem 1.5) to get $\mathbb{E}_{z\sim\{-1,1\}^{\bar{J}}}[F(z)^2] = \sum_{T\subseteq\bar{J}}\hat{F}(T)^2 = \sum_{T\subseteq\bar{J}}\hat{f}(S\cup T)^2$ making use of (b). ☐

1.6 Norms for functions on the hypercube

Occasionally it is useful to use ℓ_p-norms for boolean functions. Traditionally one would treat a function $f:\{-1,1\}^n\to\mathbb{R}$ simply as an 2^n-dimensional vector and define $\|f\|_p := (\sum_{x\in\{-1,1\}^n}|f(x)|^p)^{1/p}$ for $1\le p<\infty$ and $\|f\|_\infty := \max_{x\in\{-1,1\}^n}|f(x)|$. Standard comparison estimates give that for $1\le p\le q\le\infty$ one has $\|f\|_q \le \|f\|_p \le (2^n)^{1/p-1/q}\|f\|_q$. But since as inner product we use $\langle\cdot,\cdot\rangle_E$, it will make sense to define an ℓ_p-norm using the expectation as well:

Definition 1.17. For $f:\{-1,1\}^n\to\mathbb{R}$ and $1\le p<\infty$ we define

$$\|f\|_{E,p} := \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|f(x)|^p]^{1/p} = \frac{1}{(2^n)^{1/p}}\cdot\|f\|_p$$

and $\|f\|_{E,\infty} = \|f\|_\infty$.

Then we obtain the following comparison inequality:

Proposition 1.18. *For* $1\le p\le q<\infty$ *and* $f:\{-1,1\}^n\to\mathbb{R}$ *one has*

$$\left(\frac{1}{2^n}\right)^{1/p-1/q}\|f\|_{E,q} \le \|f\|_{E,p} \le \|f\|_{E,q}$$

Proof. We fix $1\le p\le q<\infty$. It will be convinient to prove the upper bound $\|f\|_{E,p}\le\|f\|_{E,q}$ and the lower bound in the sum form $\|f\|_q\le\|f\|_p$. For the upper bound we can see that

$$\|f\|_{E,p}^q = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[|f(x)|^p\right]^{q/p} \overset{\text{Jensen}}{\le} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[|f(x)|^q\right] = \|f\|_{E,q}^q$$

where we use Jensen's inequality (Theorem 1.45) together with the fact that the map $z\mapsto z^{q/p}$ is convex as $\frac{q}{p}\ge 1$.

Next we prove the lower bound. We can scale both sides of the inequality and just prove that $\|f\|_p = 1 \Rightarrow \|f\|_q \leq 1$. From $\|f\|_p = 1$ we know that $\|f\|_\infty \leq 1$. Then

$$\|f\|_q^q = \sum_{x\in\{-1,1\}^n} |f(x)|^q \leq \sum_{x\in\{-1,1\}^n} |f(x)|^p = 1$$

because for $0 \leq z \leq 1$ one has $z^q \leq z^p$. $\square$

Similarly we can consider ℓ_p-norms of the Fourier coefficients:

Definition 1.19. Let $1 \leq p < \infty$. The *Fourier p-norm* (or *spectral p-norm*) of $f : \{-1,1\}^n \to \mathbb{R}$ is

$$\hat{\|} f \hat{\|}_p := \Big(\sum_{S\subseteq[n]} |\hat{f}(S)|^p \Big)^{1/p}$$

Moreover $\hat{\|} f \hat{\|}_\infty := \max_{S\subseteq[n]} |\hat{f}(S)|$.

For example by Parsivals Theorem we know that $\|f\|_{E,2} = \hat{\|} f \hat{\|}_2$.

1.7 Noise stability

A recurrent theme in analysis of boolean function is to analyze how functions change under perturbations.

Definition 1.20. For $-1 \leq \rho \leq 1$ and $x \in \{-1,1\}^n$ we write $y \sim N_\rho(x)$ if $y \in \{-1,1\}^n$ is a random vector so that independently for each coordinate $i \in [n]$,

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1}{2}+\frac{\rho}{2} \\ -x_i & \text{with probability } \frac{1}{2}-\frac{\rho}{2} \end{cases}$$

It is useful to note that for $0 \leq \rho \leq 1$ we could have equivalently defined

$$y_i = \begin{cases} x_i & \text{with probability } \rho \\ \text{uniform from } \{-1,1\} & \text{with probability } 1-\rho \end{cases}$$

For all $-1 \leq \rho \leq 1$ one has

$$\mathop{\mathbb{E}}_{\substack{x\sim\{-1,1\}^n \\ y\sim N_\rho(x)}} [x_i \cdot y_i] = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n} \Big[x_i \cdot \underbrace{\Big(\Big(\frac{1}{2}+\frac{\rho}{2}\Big)\cdot x_i + \Big(\frac{1}{2}-\frac{\rho}{2}\Big)\cdot(-x_i)\Big)}_{=\rho x_i} \Big] = \rho$$

In other words, the *correlation* between x_i and y_i is exactly ρ. We also call (x,y) with $x \sim \{-1,1\}^n$ and $y \sim N_\rho(x)$ a *ρ-correlated pair.* One can think of y as a perturbation of the vector x.

Definition 1.21. For $f : \{-1,1\}^n \to \mathbb{R}$ and $-1 \leq \rho \leq 1$ we define the *noise stability*

$$\mathrm{Stab}_\rho[f] := \mathop{\mathbb{E}}_{\substack{x\sim\{-1,1\}^n\\ y\sim N_\rho(x)}}[f(x)\cdot f(y)]$$

In other words, the noise stability tells how much the function value at x correlates with the function value at a perturbation y. If $f : \{-1,1\}^n \to \{-1,1\}$ is boolean, then it is useful to note that

$$\mathrm{Stab}_\rho[f] = 2 \Pr_{\substack{x\sim\{-1,1\}^n,\\ y\sim N_\rho(x)}}[f(x) = f(y)] - 1$$

and $-1 \leq \mathrm{Stab}_\rho[f] \leq 1$.

For example for the character functions we have

$$\mathrm{Stab}_\rho[\chi_S] = \mathop{\mathbb{E}}_{\substack{x\sim\{-1,1\}^n,\\ y\sim N_\rho(x)}}\Big[\prod_{i\in S} x_i y_i\Big] = \prod_{i\in S} \underbrace{\mathop{\mathbb{E}}_{\substack{x\sim\{-1,1\}^n,\\ y\sim N_\rho(x)}}[x_i y_i]}_{=\rho} = \rho^{|S|}$$

That means for $0 \leq \rho \leq 1$, the smaller $|S|$ is the higher the stability of χ_S.

Definition 1.22. For $-1 \leq \rho \leq 1$ we define $T_\rho : V_n \to V_n$ as the linear operator that maps a function $f : \{-1,1\}^n \to \mathbb{R}$ to $T_\rho f : \{-1,1\}^n \to \mathbb{R}$ with

$$T_\rho f(x) = \mathop{\mathbb{E}}_{y\sim N_\rho(x)}[f(y)]$$

Intuitively, $T_\rho f$ is perturbed version of f. As usually we describe the Fourier expansion of $T_\rho f$:

Proposition 1.23. *For $f : \{-1,1\}^n \to \mathbb{R}$ and $-1 \leq \rho \leq 1$ one has*

$$(T_\rho f)(x) = \sum_{S\subseteq[n]} \rho^{|S|} \hat{f}(S)\cdot\chi_S(x) = \sum_{k=0}^{n} \rho^k f^{=k}(x)$$

Proof. Since T_ρ is a linear operator, it suffices to verify the claim for χ_S with $S \subseteq [n]$. Indeed for any $x \in \{-1,1\}^n$ one has

$$(T_\rho \chi_S)(x) = \prod_{i\in S} \underbrace{\mathop{\mathbb{E}}_{y\sim N_\rho(x)}[y_i]}_{=\rho x_i} = \prod_{i\in S}(\rho x_i) = \rho^{|S|}\cdot\chi_S(x)$$

□

In other words, the operator T_ρ "dampens" the Fourier coefficients and the effect is stronger, the larger $|S|$ is. We can use the operator to express the stability of a function in terms of its Fourier coefficients.

Proposition 1.24. *For any* $f : \{-1,1\}^n \to \mathbb{R}$ *and* $-1 \le \rho \le 1$ *one has*

$$Stab_\rho[f] = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S)^2$$

Moreover, for $f : \{-1,1\}^n \to \{-1,1\}$ *one also has* $Stab_\rho[f] = \mathbb{E}_{S \sim \mathcal{S}_f}\left[\rho^{|S|}\right]$.

Proof. Using the T_ρ operator we can write

$$\begin{aligned}\mathrm{Stab}_\rho[f] &\overset{\text{Def stability}}{=} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[f(x)\cdot \mathop{\mathbb{E}}_{y\sim N_\rho(x)}[f(y)]\right] \\ &\overset{\text{Def } T_\rho}{=} \langle f, T_\rho f\rangle_E \\ &\overset{\text{Plancharel}}{=} \sum_{S\subseteq[n]} \hat{f}(S)\cdot\widehat{(T_\rho f)}(S) \overset{\text{Prop 1.23}}{=} \sum_{S\subseteq[n]} \hat{f}(S)\cdot\hat{f}(S)\cdot\rho^{|S|}\end{aligned}$$

Then the "moreover" part is clear. □

From this claim we can draw the conclusion that the stability of a function f is high if much of its Fourier weight lies on the lower levels. Also we can see that for any $f : \{-1,1\}^n \to \mathbb{R}$ and $0 \le \rho \le 1$ one has $\mathrm{Stab}_\rho[f] \ge 0$, which is not obvious from the definition itself.

Noise sensitivity. We also introduce somewhat opposite quantity to stability:

Definition 1.25. Let $f : \{-1,1\}^n \to \{-1,1\}$ and $0 \le \delta \le 1$. Draw $x \sim \{-1,1\}^n$ and obtain y by flipping each bit independently with probability δ. Then the *noise sensitivity* of f is defined as

$$NS_\delta[f] := \Pr[f(x) \ne f(y)]$$

One can see that the distribution (x, y) that is produced in the definition corresponds to a ρ-correlated pair if $\delta = \frac{1}{2} - \frac{\rho}{2} \Leftrightarrow \rho = 1-2\delta$. Moreover if for a boolean function one has $\mathrm{Stab}_\rho[f] \approx 1$ then this corresponds to $NS_\delta[f] \approx 0$. The exact dependence is as follows:

Lemma 1.26. *For* $f : \{-1,1\}^n \to \{-1,1\}$ *and* $0 \le \delta \le 1$ *one has*

$$NS_\delta[f] = \frac{1}{2} - \frac{1}{2} Stab_{1-2\delta}[f]$$

1.8 Derivatives and Influences

We want to introduce the notion of a derivative for a function $f : \{-1,1\}^n \to \mathbb{R}$ in the coordinate directions. For a vector $x \in \{-1,1\}^n$ and $b \in \{-1,1\}$ we define

$$x^{(i \mapsto b)} := (x_1, \dots, x_{i-1}, b, x_{i+1}, \dots, x_n)$$

as the vector x where the ith bit is set to b (no matter what it was before). We also define

$$x^{\oplus i} := (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n)$$

as the vector x where the ith bit is flipped.

Definition 1.27. For $i \in \{1, \dots, n\}$, we define $D_i : V_n \to V_n$ as the operator that maps a function $f : \{\pm 1\}^n \to \mathbb{R}$ to the function $D_i f : \{\pm 1\}^n \to \mathbb{R}$ with

$$(D_i f)(x) := \frac{1}{2} \cdot \left(f(x^{i \mapsto 1}) - f(x^{i \mapsto -1}) \right)$$

Intuitively this gives the change of f at x in coordinate direction i. As always it will be usful to know the Fourier expansion of $D_i f$ in terms of the Fourier coefficients of the original function f. Note that by construction, $(D_i f)(x)$ does not depend on x_i and hence we already know that $\widehat{(D_i f)}(S) = 0$ whenever $i \in S$.

Proposition 1.28. *For any $f : \{-1,1\}^n \to \mathbb{R}$ and coordinate $i \in [n]$ one has*

$$(D_i f)(x) = \sum_{\{i\} \subseteq S \subseteq [n]} \hat{f}(S) \cdot \chi_{S \setminus \{i\}}(x) \quad \forall x \in \{-1,1\}^n$$

Hence for $S \subseteq [n]$ one has

$$\widehat{(D_i f)}(S) = \begin{cases} 0 & \text{if } i \in S \\ \hat{f}(S \cup \{i\}) & \text{if } i \notin S \end{cases}$$

Proof. One can check that for any $S \subseteq [n]$ one has

$$(D_i \chi_S)(x) = \begin{cases} \chi_{S \setminus \{i\}}(x) & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases}$$

Then by linearity

$$(D_i f)(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot (D_i \chi_S)(x) = \sum_{S \subseteq [n] : i \in S} \hat{f}(S) \cdot \chi_{S \setminus \{i\}}(x)$$

□

Summing up the squared change gives another useful quantity called influence.

Definition 1.29. For a function $f : \{-1,1\}^n \to \mathbb{R}$ and a coordinate $i \in [n]$ we define

$$\mathrm{Inf}_i[f] := \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[(D_i f)(x)^2\right] = \|D_i f\|_{E,2}^2$$

as the *influence of coordinate* i.

Often we are interested in boolean functions with values in $\{-1,1\}$ in which case the derivative and influence notions simplify:

Lemma 1.30. *For a function $f : \{-1,1\}^n \to \{-1,1\}$ one has*

$$(D_i f)(x)^2 = \begin{cases} 1 & \textit{if } f(x^{i\to 1}) \neq f(x^{i\to -1}) \\ 0 & \textit{otherwise.} \end{cases}$$

Moreover

$$\mathit{Inf}_i[f] = \Pr_{x\sim\{-1,1\}^n}\left[f(x) \neq f(x^{\oplus i})\right]$$

In other words, for a function $f : \{-1,1\}^n \to \{-1,1\}$, the influence $\mathrm{Inf}_i[f] \in [0,1]$ gives the fraction of edges of the hypercube with direction e_i where both endpoints have different values.

Definition 1.31. For $f : \{-1,1\}^n \to \mathbb{R}$ we define the *total influence* as

$$I[f] := \sum_{i=1}^n \mathrm{Inf}_i[f].$$

Note that for a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ one has $0 \leq I[f] \leq n$.

Theorem 1.32. *Let $f : \{-1,1\}^n \to \mathbb{R}$. Then*

(i) For any $i \in [n]$ one has $\mathit{Inf}_i[f] = \sum_{S\subseteq[n]: i\in S} \hat{f}(S)^2$.

(ii) One has $I[f] = \sum_{S\subseteq[n]} |S| \cdot \hat{f}(S)^2$.

Proof. For (i) we apply Prop 1.28 to get

$$\mathrm{Inf}_i[f] = \hat{\|} D_i f \hat{\|}_2^2 \overset{\text{Prop } 1.28}{=} \sum_{S\subseteq[n]: i\in S} \hat{f}(S)^2$$

For (ii) we sum over all coordinates to get

$$I[f] = \sum_{i=1}^n \mathrm{Inf}_i[f] \overset{(i)}{=} \sum_{i=1}^n \sum_{S\subseteq[n]: i\in S} \hat{f}(S)^2 = \sum_{S\subseteq[n]} |S| \cdot \hat{f}(S)^2$$

as the double sum counts every set S exactly $|S|$ times. □

For later use we record the fact that low degree bounded functions have a small total influence:

Lemma 1.33. *For any* $f:\{-1,1\}^n \to [-1,1]$ *one has* $I[f] \le \deg(f)$.

Proof. Abbreviate $d := \deg(f)$. We verify that

$$I[f] \overset{\text{Thm 1.32}}{=} \sum_{S\subseteq[n]} |S|\cdot \hat{f}(S)^2 \le d \sum_{S\subseteq[n]} \hat{f}(S)^2 \overset{\text{Parsival}}{=} d \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n} \underbrace{[f(x)^2]}_{\le 1} \le d$$

□

ρ-stable influence. We introduce a concept that connects noise stability from the previous section with influence:

Definition 1.34. Let $f:\{-1,1\}^n \to \mathbb{R}$, $0 \le \rho \le 1$ and $i \in [n]$. Then the *ρ-stable influence* of i on f is

$$\mathrm{Inf}_i^{(\rho)}[f] := \mathrm{Stab}_\rho[D_i f]$$

Moreover, $I^{(\rho)}[f] := \sum_{i=1}^n \mathrm{Inf}_i^{(\rho)}[f]$ is the *ρ-stable total influence* of f.

These quantities might be less intuitive, but we can also obtain their Fourier representation:

Lemma 1.35. *For any* $f:\{-1,1\}^n \to \mathbb{R}$ *and* $0 \le \rho \le 1$ *the following holds:*

(i) One has $\mathit{Inf}_i^{(\rho)}[f] = \sum_{S\subseteq[n]:i\in S} \rho^{|S|-1}\hat{f}(S)^2$.

(ii) One has $I^{(\rho)}[f] = \sum_{S\subseteq[n]} |S|\rho^{|S|-1}\hat{f}(S)^2$.

(iii) One has $I^{(\rho)}[f] = \frac{d}{d\rho}\mathit{Stab}_\rho[f]$.

Proof. For (i), we use that

$$\mathrm{Inf}_i^{(\rho)}[f] \overset{\text{Def}}{=} \mathrm{Stab}_\rho[D_i f] \overset{\text{Prop 1.24}}{=} \sum_{S\subseteq[n]} \rho^{|S|}\cdot \widehat{(D_i f)}(S)^2 \overset{\text{Prop 1.28}}{=} \sum_{S\subseteq[n]:i\notin S} \rho^{|S|}\cdot \hat{f}(S\cup\{i\})^2$$

We leave (ii) and (iii) as an exercise. □

Degree-d influences. We make another definition:

Definition 1.36. For $f:\{-1,1\}^n \to \mathbb{R}$ and $d \in \mathbb{Z}_{\ge 0}$ we define the *degree-d influences* as

$$\mathrm{Inf}_i^{\le d}[f] := \sum_{S\subseteq[n]:i\in S \text{ and } |S|\le d} \hat{f}(S)^2$$

One can think of the degree-d influences as an alternative to the ρ-stable influences. In the former we cut off any coefficients larger than d, in the latter we just discount those at an exponential rate. We record a lemma for later use:

Lemma 1.37. *For $f : \{\pm 1\}^n \to [-1,1]$, $d \in \mathbb{Z}_{\geq 0}$ and $\varepsilon > 0$ let $I := \{i \in [n] \mid \mathrm{Inf}_i^{\leq d}[f] \geq \varepsilon\}$ be the influential coordinates. Then $|I| \leq \frac{d}{\varepsilon}$.*

Proof. We have

$$\varepsilon |I| \leq \sum_{i=1}^{n} \mathrm{Inf}_i^{\leq d}[f] = \sum_{i=1}^{n} \sum_{S \subseteq [n] : i \in S \text{ and } |S| \leq d} \hat{f}(S)^2 = \underbrace{\sum_{|S| \leq d} \hat{f}(S)^2}_{\leq 1} \cdot \underbrace{|S|}_{\leq d} \leq d$$

which can be rearranged to $|I| \leq \frac{d}{\varepsilon}$. □

1.9 Variance of functions

Occasionally the following notion will be useful:

Definition 1.38. For a function $f : \{-1,1\}^n \to \mathbb{R}$ we abbreviate $\mathrm{Var}[f]$ as the *variance* of the random variable $f(x)$ where $x \sim \{-1,1\}^n$. In other words

$$\mathrm{Var}[f] := \mathop{\mathbb{E}}_{x \sim \{-1,1\}^n} \left[f(x)^2 \right] - \mathop{\mathbb{E}}_{x \sim \{-1,1\}^n} [f(x)]^2$$

Lemma 1.39. *For any $f : \{-1,1\}^n \to \mathbb{R}$ one has $\mathrm{Var}[f] = \sum_{\emptyset \subset S \subseteq [n]} \hat{f}(S)^2$.*

Proof. Follows from Parsival's identity and the fact that $\mathbb{E}_{x \sim \{-1,1\}^n}[f(x)] = \hat{f}(\emptyset)$. □

1.10 Useful inequalities

We collect a few standard inequalities that will turn out to be useful during this course.

1.10.1 Probabilistic inequalities

Lemma 1.40 (Reverse Markov Inequality)**.** *Let $0 \leq X \leq M$ be a random variable. Then $\Pr[X \geq t] \geq \frac{\mathbb{E}[X] - t}{M}$.*

Proof. We have

$$\mathbb{E}[X] \leq t \cdot \underbrace{\Pr[X < t]}_{\leq 1} + M \cdot \Pr[X \geq t] \leq t + M \cdot \Pr[X \geq t]$$

Rearranging gives the claim. □

Lemma 1.41 (Paley-Zygmund)**.** *Let X be a real-valued random variable with $X \geq 0$ and $0 < \mathbb{E}[X^2] < \infty$. Then for any $0 \leq t \leq 1$,*

$$\Pr[X > t\mathbb{E}[X]] \geq (1-t)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$$

Proof. We bound

$$\mathbb{E}[X] = \underbrace{\mathbb{E}[X \cdot \mathbf{1}_{X \leq t\mathbb{E}[X]}]}_{\leq t\mathbb{E}[X]} + \mathbb{E}[X \cdot \mathbf{1}_{X > t\mathbb{E}[X]}] \overset{\text{Cauchy-Schwarz}}{\leq} t\mathbb{E}[X] + \sqrt{\mathbb{E}[X^2] \cdot \Pr[X > t\mathbb{E}[X]]}$$

Rearranging gives the claim. □

Theorem 1.42 (Hoeffding Inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables so that $a_i \leq X_i \leq b_i$ for all $i \in [n]$ and let $X = \frac{1}{n}(X_1 + \ldots + X_n)$. Then for any $\varepsilon > 0$,*

$$\Pr[|X - \mathbb{E}[X]| \geq \varepsilon] \leq 2 \exp\Big(-\frac{2n^2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big)$$

1.10.2 Combinatorial inequalities

Theorem 1.43 (Generalized Binomial Theorem)**.** *For any $x, r \in \mathbb{R}$ with $|x| < 1$ one has*

$$(1+x)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^k$$

where

$$\binom{r}{k} := \frac{r \cdot (r-1) \cdot \ldots \cdot (r-k+1)}{k!}$$

1.10.3 Analytic inequalities

Next, we derive a few analytic inequalities.

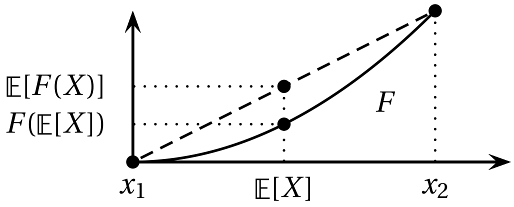
Lemma 1.44 (Cauchy-Schwarz)**.** *For any real-valued random variables X, Y one has*

$$\mathbb{E}[|X \cdot Y|] \leq \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}$$

We will actually prove a more general statement later.

Theorem 1.45 (Jensen Inequality for Convex Functions). *Let* $X : \Omega \to \mathbb{R}$ *be a random variable and* $F : \mathbb{R} \to \mathbb{R}$ *be a convex function. Then* $F(\mathbb{E}[X]) \leq \mathbb{E}[F(X)]$.

The inequality follows immediately from the definition of convexity.



Example of convex function F and
distribution X over only two values x_1, x_2

If the function F is rather concave then convex, then the inequality holds with reversed relation:

Theorem 1.46 (Jensen Inequality for Concave Functions). *Let* $X : \Omega \to \mathbb{R}$ *be a random variable and* $F : \mathbb{R} \to \mathbb{R}$ *be a concave function. Then* $F(\mathbb{E}[X]) \geq \mathbb{E}[F(X)]$.

The next inequality is due to Young:

Theorem 1.47 (Young's Inequality). *For* $x, y \geq 0$ *and* $0 \leq \lambda \leq 1$ *one has*

$$x \cdot y \leq (1-\lambda) \cdot x^{1/(1-\lambda)} + \lambda \cdot y^{1/\lambda}$$

Proof. Simply note that

$$\begin{aligned} \ln\Big((1-\lambda)x^{1/(1-\lambda)} + \lambda y^{1/\lambda}\Big) &\overset{\text{Jensen+concavity of ln}}{\geq} (1-\lambda)\ln(x^{1/(1-\lambda)}) + \lambda \ln(y^{1/\lambda}) \\ &= \ln(x) + \ln(y) = \ln(x \cdot y) \end{aligned}$$

□

The inequality of Hölder is basically a generalization of Cauchy-Schwarz to general $\|\cdot\|_p$-norms:

Theorem 1.48 (Hölder's Inequality I). *Let* $X, Y : \Omega \to \mathbb{R}_{\geq 0}$ *be jointly distributed non-negative random variables. Then for all* $0 \leq \lambda \leq 1$ *one has* $\mathbb{E}[X^{1-\lambda}Y^{\lambda}] \leq \mathbb{E}[X]^{1-\lambda}\mathbb{E}[Y]^{\lambda}$.

Proof. Scaling X by $s > 0$ scales both sides of the inequality by the same factor of $s^{1-\lambda}$. Hence we may assume w.l.o.g. that $\mathbb{E}[X] = 1$; similarly assume $\mathbb{E}[Y] = 1$. Then applying Young's Inequality gives

$$\begin{aligned} \mathbb{E}[X^{1-\lambda}Y^{\lambda}] &\overset{\text{Young}}{\leq} \mathbb{E}\left[(1-\lambda)\cdot(X^{1-\lambda})^{\frac{1}{1-\lambda}} + \lambda\cdot(Y^{\lambda})^{\frac{1}{\lambda}}\right] = \mathbb{E}[(1-\lambda)X + \lambda Y] \\ &= (1-\lambda)\,\mathbb{E}[X] + \lambda\,\mathbb{E}[Y] = 1 = \mathbb{E}[X]^{1-\lambda}\cdot\mathbb{E}[Y]^{\lambda} \end{aligned}$$

□

An equivalent statement is as follows:

Theorem 1.49 (Hölder's Inequality II)**.** *Let $X, Y : \Omega \to \mathbb{R}$ be jointly distributed random variables. Let $p, q \geq 1$ be a pair with $\frac{1}{p} + \frac{1}{q} = 1$. Then $\mathbb{E}[|X\cdot Y|] \leq \mathbb{E}[|X|^p]^{1/p}\cdot \mathbb{E}[|Y|^q]^{1/q}$.*

Theorem 1.50 (Littlewood's Inequality / L_p-interpolation inequality)**.** *Let X be a random variable. If $p, q, r \geq 1$ and $0 < \theta < 1$ are values so that $\frac{1}{p} = \frac{\theta}{q} + \frac{1-\theta}{r}$, then*

$$\mathbb{E}[|X|^p]^{1/p} \leq \mathbb{E}[|X|^q]^{\theta/q}\cdot\mathbb{E}[|X|^r]^{(1-\theta)/r}$$

Proof. We can write

$$\begin{aligned} \mathbb{E}[|X|^p]^{1/p} &= \mathbb{E}\left[\left(|X|^q\right)^{p\theta/q}\left(|X|^r\right)^{p(1-\theta)/r}\right]^{1/p} \\ &\overset{\text{Hölder I}}{\leq} \left(\mathbb{E}\left[|X|^q\right]^{p\theta/q}\mathbb{E}\left[|X|^r\right]^{p(1-\theta)/r}\right)^{1/p} \\ &= \mathbb{E}\left[|X|^q\right]^{\theta/q}\cdot\mathbb{E}\left[|X|^r\right]^{(1-\theta)/r} \end{aligned}$$

Here we crucially use that θ is chosen so that $\frac{p\theta}{q} + \frac{p(1-\theta)}{r} = 1$ in order to apply Hölder's Inequality I (Theorem 1.49). □

We note that necessarily p needs to lie between q and r so that $\theta \in (0,1)$. We would like to remark that Littlewood's Inequality can be restated as follows:

Theorem 1.51 (Littlewood's Inequality II)**.** *Let X be a random variable. Then the function $\phi : (0,1] \to \mathbb{R}_{\geq 0}$ with $\phi(t) := \mathbb{E}[|X|^{1/t}]^t$ is log convex.*

Proof. Let $t = \theta a + (1-\theta)b$ with $a < t < b$. Then $p := \frac{1}{t} \geq 1$, $q := \frac{1}{a} \geq 1$, $r := \frac{1}{b} \geq 1$. Then $\frac{1}{p} = \frac{\theta}{q} + \frac{1-\theta}{r}$. Hence

$$\phi(t) \leq \phi(a)^{\theta}\cdot\phi(b)^{(1-\theta)} \quad\Longleftrightarrow\quad \mathbb{E}[|X|^p]^{1/p} \leq \mathbb{E}[|X|^q]^{\theta/q} + \mathbb{E}[|X|^r]^{(1-\theta)/r}$$

which is exactly the statement in Theorem 1.50. □

This can be conviniently rewritten for the norm of functions:

Theorem 1.52 (Littlewood's Inequality III)**.** *Let* $f : \{\pm 1\}^n \to \mathbb{R}$*. If* $p, q, r \geq 1$ *and* $0 < \theta < 1$ *are values so that* $\frac{1}{p} = \frac{\theta}{q} + \frac{1-\theta}{r}$*, then*

$$\|f\|_{E,p} \leq \|f\|_{E,q}^{\theta} \cdot \|f\|_{E,r}^{1-\theta}$$

Chapter 2

Linearity testing

Recall that a function $F : \mathbb{F}_2^n \to \mathbb{F}_2$ is *linear* if $F(x \oplus y) = F(x) \oplus F(y)$ for all $x, y \in \mathbb{F}_2^n$. The following topic is typically phrased using the cube $\mathbb{F}_2^n$ but as we explained above, we prefer not to have to switch back and forth and will do the exposition and proof fully with the cube $\{-1,1\}^n$.

We say that a function $f : \{-1,1\}^n \to \{-1,1\}$ is $\{-1,1\}$-*linear* if

$$f(x \odot y) = f(x) \cdot f(y) \quad \forall x, y \in \{-1,1\}^n$$

We have already seen that for any $S \subseteq [n]$ one has $\chi_S(x \odot y) = \chi_S(x) \cdot \chi_S(y)$ for all $x, y \in \{\pm 1\}^n$, meaning that the character functions are $\{-1,1\}$-linear. In fact, one can show that the character functions are the only $\{-1,1\}$-linear functions, which we leave as an exercise.

In 1990, Blum, Luby and Rubinfeld [BLR90] studied approximately linear functions. In particular they considered the following test which can be done using only query access to 3 random points.

BLR LINEARITY TEST

Input: Query access to a function $f : \{-1,1\}^n \to \{-1,1\}$

(1) Draw $x, y \sim \{-1,1\}^n$ independently at random.

(2) Accept if $f(x \odot y) = f(x) \cdot f(y)$.

Suppose this test passes with 99%, then would this imply some structure on f? Still, f might not be an actual $\{-1,1\}$-linear function but maybe it is close to one. In fact, it will be within 1% of a character function. Here, for functions $f, g : \{-1,1\}^n \to \{-1,1\}$ we define the *distance* between them

$$\mathrm{dist}(f, g) := \Pr_{x \sim \{-1,1\}^n}[f(x) = g(x)]$$

Note that always $0 \leq \mathrm{dist}(f, g) \leq 1$.

Theorem 2.1. *Suppose a function Let $f : \{\pm 1\}^n \to \{\pm 1\}$ be a function that passes the BLR Linearity Test with probability at least $\frac{1}{2} + \varepsilon$ for $0 \le \varepsilon \le \frac{1}{2}$. Then there is a set $S \subseteq [n]$ so that $\hat{f}(S) \ge 2\varepsilon$ as well as dist$(f, \chi_S) \le \frac{1}{2} - \varepsilon$.*

Proof. We write

$$\begin{aligned}
2\varepsilon &= \left(\frac{1}{2} + \varepsilon\right) - \left(\frac{1}{2} - \varepsilon\right) \\
&\overset{\text{BLR test}}{\le} \mathop{\mathbb{E}}_{x,y \sim \{\pm 1\}^n} [f(x \odot y) \cdot f(x) \cdot f(y)] \\
&\overset{\text{Thm 1.3}}{=} \mathop{\mathbb{E}}_{x,y \sim \{\pm 1\}^n} \left[\Big(\sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x \odot y) \Big) \Big(\sum_{T \subseteq [n]} \hat{f}(T) \chi_T(x) \Big) \Big(\sum_{R \subseteq [n]} \hat{f}(R) \chi_R(y) \Big) \right] \\
&\overset{\chi_S(x \odot y) = \chi_S(x)\chi_S(y)}{=} \sum_{S,T,R \subseteq [n]} \hat{f}(S) \hat{f}(R) \hat{f}(T) \mathop{\mathbb{E}}_{x,y \sim \{\pm 1\}^n} [\chi_S(x) \cdot \chi_S(y) \cdot \chi_T(x) \cdot \chi_R(y)] \\
&\overset{\text{indep.}}{=} \sum_{S,T,R \subseteq [n]} \hat{f}(S) \hat{f}(T) \hat{f}(R) \underbrace{\mathop{\mathbb{E}}_{x \sim \{\pm 1\}^n} [\chi_S(x) \chi_T(x)]}_{=1 \text{ if } S=T, \, =0 \text{ o.w.}} \underbrace{\mathop{\mathbb{E}}_{y \sim \{\pm 1\}^n} [\chi_S(y) \chi_R(y)]}_{=1 \text{ if } S=R, \, =0 \text{ o.w.}} \\
&= \sum_{S \subseteq [n]} \hat{f}(S)^3 \\
&\le \max_{S \subseteq [n]} \{\hat{f}(S)\} \cdot \underbrace{\sum_{S \subseteq [n]} \hat{f}(S)^2}_{= \langle f, f \rangle_E = 1} \\
&\le \max_{S \subseteq [n]} \{\hat{f}(S)\}
\end{aligned}$$

Now fix the set S maximizing $\hat{f}(S)$. Then

$$2\varepsilon \le \hat{f}(S) = \langle f, \chi_S \rangle_E = 1 - 2\text{dist}(f, \chi_S)$$

which can be rearranged to dist$(f, \chi_S) \le \frac{1}{2} - \varepsilon$. ☐

The original result is due to [BLR90] while we have presented a later proof due to Bellare.

Chapter 3

The Goldreich Levin algorithm

In this chapter we discuss boolean functions from the view point of *learning theory.* In general terms, there is a function $f : \{-1,1\}^n \to \{-1,1\}$ and we want to learn certain facts about it. For example we could be interested in learning the large Fourier coefficients of f. The only access that we have to the function is via function values $f(x)$ at a few points x. In principle there are three ways how this access could be given:

- *Adaptive queries:* We compute a sequence of points $x_1,\dots,x_N \in \{-1,1\}^n$ and an oracle gives us the values $f(x_1),\dots,f(x_N)$. Here x_i can depend on the previous queried values $f(x_1),\dots,f(x_{i-1})$.

- *Non-adaptive queries:* We compute a sequence of points $x_1,\dots,x_N \in \{-1,1\}^n$ and an oracle gives us the values $f(x_1),\dots,f(x_N)$. Here x_i may not depend on $f(x_1),\dots,f(x_{i-1})$.

- *Random examples:* An oracle draws $x_1,\dots,x_N \sim \{-1,1\}^n$ uniformly and independently. Then we are provided with the pairs $(x_1,f(x_1),\dots,(x_N,f(x_N))$.

We note that adaptive queries are more powerful than non-adaptive queries which in turn are more powerful than random examples. When designing a learning algorithm, there are 3 often conflicting goals that one might want to achieve:

- Use an oracle that is as weak as possible.

- Minimize the number of queries.

- Minimize the running time.

3.1 Estimating Fourier coefficients

To warm up, we discuss how to estimate a single Fourier coefficient using random examples.

Lemma 3.1. *Given* $S \subseteq [n]$ *and* $\delta, \varepsilon > 0$ *and access to* $N := \Theta(\frac{1}{\varepsilon^2} \cdot \log \frac{1}{\delta})$ *many random examples from* $f : \{-1,1\}^n \to \{-1,1\}$ *one can compute a value* $\alpha \in \mathbb{R}$ *so that*

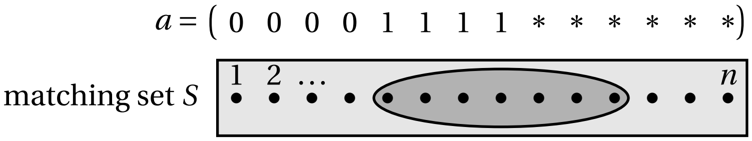
$$\Pr[|\hat{f}(S) - \alpha| \leq \varepsilon] \geq 1 - \delta$$

Proof. Note that $\hat{f}(S) = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x) \cdot \chi_S(x)]$. So if we draw $x_1, \ldots, x_N \sim \{-1,1\}^n$ and set $X_i := f(x_i) \cdot \chi_S(x_i)$ then $-1 \leq X_i \leq 1$ and $\mathbb{E}[X_i] = \hat{f}(S)$. Our estimate is $\alpha := \frac{1}{N} \sum_{i=1}^N X_i$ and by Hoeffding's Inequality (Theorem 1.42), $N = \Theta(\frac{1}{\varepsilon^2} \log(\frac{1}{\delta}))$ samples suffice. $\square$

Next, we will extend this argument and show that also "groups" of Fourier coefficients can be estimated. For a vector $a \in \{0,1,*\}^n$ we define

$$W_a(f) := \sum_{\substack{S \subseteq [n]: \\ a_i = 0 \Rightarrow i \notin S, \\ a_i = 1 \Rightarrow i \in S}} \hat{f}(S)^2$$

Here one can think of a as a pattern and $W_a(f)$ is the sum of the squared Fourier weight over all sets S that match the pattern.



That means if a contains k many placeholders $*$, then $W_a(f)$ is the sum over 2^k many squared Fourier coefficients of f. In particular, for all $S \subseteq [n]$ one has $W_{\mathbf{1}_S}(f) = \hat{f}(S)^2$ and assuming $f : \{-1,1\}^n \to \{-1,1\}$ one has $W_{(*,\ldots,*)}(f) = \sum_{S \subseteq [n]} \hat{f}(S)^2 = 1$. Of course one could estimate $W_a(f)$ via Lemma 3.1 by estimating each Fourier coefficient separately. But then the running and number of samples would be polynomial in 2^k. It turns out that this can be done much more elegantly:

Lemma 3.2. *Given* $a \in \{0,1,*\}^n$ *and* $\varepsilon, \delta > 0$, *with* $N := \Theta(\frac{1}{\varepsilon^2} \log \frac{1}{\delta})$ *many non-adaptive queries to* $f : \{-1,1\}^n \to \{-1,1\}$, *one can produce a value* $\beta \in \mathbb{R}$ *so that*

$$\Pr[|W_a(f) - \beta| \leq \varepsilon] \geq 1 - \delta$$

The running time is polynomial in n, $\frac{1}{\varepsilon}$ *and* $\log(\frac{1}{\delta})$.

Proof. Let $[n] = I_0 \dot\cup I_1 \dot\cup I_*$ be the partition of the coordinates according to where a has 0's, 1's and $*$'s. Let $F : \{-1,1\}^{I_*} \to \mathbb{R}$ be the function which for $z \in \{-1,1\}^{I_*}$ is defined as

$$F(z) := (\widehat{f_{I_0 \cup I_1 | z}})(I_1) = \mathop{\mathbb{E}}_{y \sim \{-1,1\}^{I_0 \cup I_1}} \left[f(y,z) \cdot \chi_{I_1}(y) \right] \qquad (*)$$

Here we use the notation of restrictions from Section 1.5. Then we can express

$$\begin{aligned} W_a(f) &\overset{\text{Def}}{=} \sum_{T \subseteq I_*} \hat{f}(I_1 \cup T)^2 \\ &\overset{\text{Parsival}}{=} \mathop{\mathbb{E}}_{z \sim \{-1,1\}^{I_*}} [F(z)^2] \\ &\overset{(*)}{=} \mathop{\mathbb{E}}_{z \sim \{-1,1\}^{I_*}} \Big[\mathop{\mathbb{E}}_{y \sim \{-1,1\}^{I_0 \cup I_1}} \left[f(y,z) \cdot \chi_{I_1}(y) \right]^2 \Big] \\ &= \mathop{\mathbb{E}}_{z \sim \{-1,1\}^{I_*}} \Big[\mathop{\mathbb{E}}_{y,y' \sim \{-1,1\}^{I_0 \cup I_1}} \left[f(y,z) \cdot \chi_{I_1}(y) \cdot f(y',z) \cdot \chi_{I_1}(y') \right] \Big] \end{aligned}$$

Here we use that for any random variable $Y \sim \mathcal{D}$ coming from some distribution $\mathcal{D}$ one has $\mathbb{E}[Y]^2 = \mathbb{E}[Y]\mathbb{E}[Y'] = \mathbb{E}[Y \cdot Y']$ where $Y, Y' \sim \mathcal{D}$ are independent copies of that same distribution $\mathcal{D}$. Similar to Lemma 3.1 we can draw independent samples (z_i, y_i, y_i') for $i = 1, \dots, N$ and set β as the unweighted average of the random variables $X_i := f(y_i, z_i)\chi_{I_1}(y_i) \cdot f(y_i', z_i)\chi_{I_1}(y_i')$. Again $-1 \leq X_i \leq 1$ and $\mathbb{E}[X_i] = W_a(f)$ and so $N = O(\frac{1}{\varepsilon^2}\log(\frac{1}{\delta}))$ samples suffice. □

We would like to point out a subtle difference in the argument to Lemma 3.1. We do not simply query f at $2N$ uniform random points because we need the product $f(y_i, z_i) \cdot f(y_i', z_i)$. That means we need correlated random samples where pairwise the coordinates of I_* are the same!

3.2 The Goldreich-Levin algorithm

Now we come to the main part of this chapter. Intuitively, the Goldreich-Levin algorithm uses a bounded number of queries to compute the set $\mathcal{F} = \{S \subseteq [n] \mid |\hat{f}(S)| \geq \varepsilon\}$ of large Fourier coefficients. Now, there is the slight technicality that if $|\hat{f}(S)| \approx \varepsilon$ then using an estimate as in Lemma 3.1 we could not be certain whether the S-th Fourier coefficient was slightly above or below the threshold of ε. So the precise statement that we prove is as follows:

Theorem 3.3. *Let $\varepsilon > 0$. With $N := \Theta(\frac{n}{\varepsilon^2}\log(\frac{n}{\varepsilon}))$ adaptive queries to a function $f : \{-1,1\}^n \to \{-1,1\}$, one can with high probability compute a family $\mathcal{F} \subseteq 2^{[n]}$ of size $|\mathcal{F}| \leq \frac{4}{\varepsilon^2}$ so that* *(i)* $|\hat{f}(S)| \geq \varepsilon \Rightarrow S \in \mathcal{F}$

(ii) $S \in \mathcal{F} \Rightarrow |\hat{f}(S)| \geq \frac{\varepsilon}{2}$.

The running time is polynomial in n and $\frac{1}{\varepsilon}$.

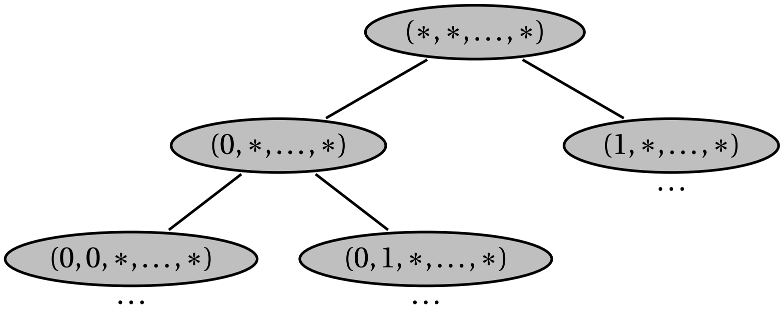
Proof. We first state the algorithm formally:

```
GOLDREICH-LEVIN ALGORITHM
Input: Adaptive query access to a function $f : \{-1,1\}^n \to \{-1,1\}$ and $\varepsilon > 0$
  (1) Initialize $\mathcal{F} := \{(*,\dots,*)\}$
  (2) WHILE $\mathcal{F}$ contains a vector containing a $*$ DO
      (3) Select and remove a vector $a$ from $\mathcal{F}$ that contains a $*$; let $i$ be the index with $a_i = *$
      (4) Create two vectors $a^{(0)} := (a_1,\dots,a_{i-1},0,a_{i+1},\dots,a_n)$ and $a^{(1)} := (a_1,\dots,a_{i-1},1,a_{i+1},\dots,a_n)$
      (5) For $\ell \in \{0,1\}$, compute an estimate $\beta_\ell \in \mathbb{R}$ with $|W_{a^{(\ell)}}(f) - \beta_\ell| \leq \frac{\varepsilon^2}{4}$.
      (6) FOR $\ell \in \{0,1\}$ IF $\beta_\ell \leq \frac{\varepsilon^2}{2}$ then discard $a^{(\ell)}$, otherwise add it to $\mathcal{F}$.
  (7) Return $\mathcal{F}$
```

Note that the algorithm initializes $\mathcal{F}$ as a single "bucket" corresponding to the pattern $(*,\dots,*)$ containing all Fourier coefficients $S \subseteq [n]$. In each iteration we remove a bucket containing more than one set S and split it into two buckets containing half the sets. Then we measure its squared Fourier weight; if the measured value is below $\frac{\varepsilon^2}{2}$ then we discard it, otherwise we keep it. At the end all remaining buckets will contain a single set S. If we run the test in Lemma 3.2 with accuracy $\varepsilon' := \frac{\varepsilon^2}{4}$ (and high enough confidence $1-\delta$) then we know that no set $S \subseteq [n]$ with $|\hat{f}(S)| \geq \varepsilon$ will ever be discarded and all sets that we have left at the end have $\hat{f}(S)^2 \geq \frac{\varepsilon^2}{2} - \frac{\varepsilon^2}{4} \Rightarrow |\hat{f}(S)| \geq \frac{\varepsilon}{2}$. The final set $\mathcal{F}$ returned in (7) only contains singletons satisfying $1 \geq \hat{f}(S)^2 \geq \frac{\varepsilon^2}{4}$ implying that $|\mathcal{F}| \leq \frac{4}{\varepsilon^2}$. One can arrange the set of all considered vectors a as a binary (but not necessarily balanced) tree which has $O(\frac{1}{\varepsilon^2})$ leaves and $O(\frac{n}{\varepsilon^2})$ nodes total[1]. Hence a confidence of $\delta := \Theta(\frac{\varepsilon^2}{n})$ suffices.

---

[1] One way to see this is by arguing that all interior nodes have a weight of at least $\frac{\varepsilon^2}{4}$ and the total weight of nodes in the tree can be at most the depth which is n.



□

3.3 Application to List Decoding of the Hadamard Code

The *Walsh Hadamard code* is an error correcting code that maps $S \subseteq [n]$ to the code words $\mathtt{WH}(S) := (\chi_S(x))_{x \in \{-1,1\}^n}$. Note that this is an extremely inefficient code as it encodes the n-bits represented by $S \subseteq [n]$ with the 2^n-bits needed to encode the boolean function χ_S. But the code has man useful properties. Recall that in the notation from Chapter 2 we write $\mathrm{dist}(f,g)$ as the fraction in which boolean functions $f, g : \{-1,1\}^n \to \{-1,1\}$ differ. First, note that distinct code words differ in exactly half the bits:

Lemma 3.4. *For all distinct $S, T \subseteq [n]$ one has dist$(\chi_S, \chi_T) = \frac{1}{2}$.*

Proof. As used earlier $\mathrm{dist}(f,g) = \frac{1}{2} - \frac{1}{2}\langle \chi_S, \chi_T \rangle_E = \frac{1}{2}$. □

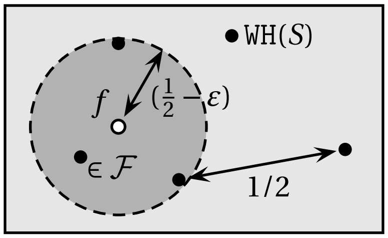
Then certainly, if we have a function $f : \{-1,1\}^n \to \{-1,1\}$ with $\min_{S \subseteq [n]}\{\mathrm{dist}(f, \chi_S)\} < \frac{1}{4}$, then by the triangle inequality there has to be a *unique* set $S^* \subseteq [n]$ with $\mathrm{dist}(f, \chi_{S^*}) < \frac{1}{4}$. But one can easily pick two distinct set S_1, S_2 and construct f as the “mid point” between χ_{S_1}, χ_{S_2} so that $\mathrm{dist}(f, \chi_{S_1}) = \frac{1}{4} = \mathrm{dist}(f, \chi_{S_2})$. In other words, the unique decoding property is lost once we reach a radius of $\frac{1}{4}$. But between distance $\frac{1}{4}$ and $\frac{1}{2} - \varepsilon$, the Walsh Hadamard code is still *list decodable*. In particular for any given f and $\varepsilon > 0$ there is a bounded number of sets S with $\mathrm{dist}(f, \chi_S) \leq \frac{1}{2} - \varepsilon$ and one can even compute these efficiently.

Theorem 3.5 (List decoding of Walsh Hadamard)**.** *Given $\varepsilon > 0$ and query access to a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ one can compute a list $\mathcal{F} \subseteq 2^{[n]}$ so that*

(i) For all $S \in \mathcal{F}$ one has dist$(f, \chi_S) \leq \frac{1}{2} - \frac{\varepsilon}{2}$.

(ii) If dist$(f, \chi_S) \leq \frac{1}{2} - \varepsilon$, then $S \in \mathcal{F}$.

The list can be computed from poly$(n, \frac{1}{\varepsilon})$ many queries to f and $|\mathcal{F}| \leq O(\frac{1}{\varepsilon^2})$.



Proof. We use the Goldreich Levin algorithm (Theorem 3.3 plus removing those sets S with negative $\hat{f}(S)$) to compute a set $\mathcal{F}$ with $|\mathcal{F}| \leq O(\frac{1}{\varepsilon^2})$ so that $\hat{f}(S) \geq 2\varepsilon \Rightarrow S \in \mathcal{F}$ and $S \in \mathcal{F} \Rightarrow \hat{f}(S) \geq \varepsilon$. We consider the cases:

(i) For $S \in \mathcal{F}$ we have $\mathrm{dist}(f, \chi_S) = \frac{1}{2} - \frac{1}{2}\langle f, \chi_S \rangle_E \leq \frac{1}{2} - \frac{\varepsilon}{2}$.

(ii) If $\frac{1}{2} - \frac{1}{2}\langle f, \chi_S \rangle_E = \mathrm{dist}(f, \chi_S) \leq \frac{1}{2} - \varepsilon$ then $\hat{f}(S) \geq 2\varepsilon$ and so by construction $S \in \mathcal{F}$.

□

Chapter 4

Hardness of Approximation I (via PCP Theorem + Parallel Repetition)

In this chapter we discuss a very important application of boolean functions to derive hardness of approximation results. We will make a small detour and explain some background on the PCP Theorem and the Parallel Repetition Theorem first (even if it does not contain any boolean functions). After that we prove that for any $\varepsilon > 0$, there is no $(\frac{1}{2} + \varepsilon)$-approximation algorithm for maximizing the number of satisfied linear equations in $\mathbb{F}_2$ with 3 variables per equation. For this chapter, we follow the notes of Minzer [Min22]. Much of the covered material can also be found in Chapter 7 of O'Donnell's book [O'D21].

4.1 Probabilistically checkable proofs

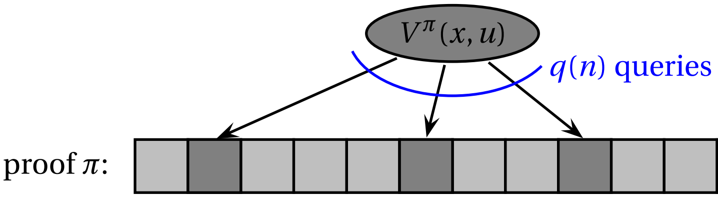
Consider a language $L \subseteq \{0,1\}^*$. We are given an input $x \in \{0,1\}^*$ and our goal is to decide whether $x \in L$ or not. Suppose there is an all powerful *prover* that wants to convince us of the former. The prover presents a proof string π. But we cannot read the whole proof, we can merely read a few randomly chosen entries of the proof but it has to suffice to convince us that x should be accepted when $x \in L$ while we should likely reject if $x \notin L$. This is called a **p**robabilistically **c**heckable **p**roof.

Definition 4.1. Let Σ be a finite set, $1 \geq c > s \geq 0$. A $\mathbf{PCP}_{\Sigma}^{[c,s]}(r(n), q(n))$*-verifier* is a deterministic polynomial time Turing machine $V^{\pi}(x, u)$ that receives an input $x \in \{0,1\}^*$ and uniform random bits $u \sim \{0,1\}^{r(|x|)}$ and can make non-adaptive queries to $q(|x|)$ positions of a proof string $\pi \in \Sigma^*$. More precisely, the Turing machine can write indices $i_1, \ldots, i_{q(n)}$ $(n := |x|)$ on a special tape and then receive the symbols $\pi_{i_1},\dots,\pi_{i_{q(n)}}$ — but it can make such a query only once, in particular the queries are *non-adaptive*.

We say that such a verifier V^π *decides* a language $L \subseteq \{0,1\}^*$ if

$$\begin{aligned} x \in L \quad &\Rightarrow \quad \exists \pi : \Pr_{u \sim \{0,1\}^{r(n)}}[V^\pi(x,u) \text{ accepts}] \geq c \\ x \notin L \quad &\Rightarrow \quad \forall \pi : \Pr_{u \sim \{0,1\}^{r(n)}}[V^\pi(x,u) \text{ accepts}] \leq s \end{aligned}$$

We also denote $\mathbf{PCP}_\Sigma^{s,c}(r(n),q(n))$ as the set of languages that can be decided by a $\mathbf{PCP}_\Sigma^{s,c}(r(n),q(n))$-verifier.



Here c is the *completeness parameter* ans s is the *soundness parameter*. One of the deepest results in all of theoretical computer science is the following:

Theorem 4.2 (PCP Theorem — Arora, Feige, Goldwasser, Lund, Lovász, Motwani, Safra, Sudan, Szegedy 1992[1])**.** *There are constants $\varepsilon > 0$ and $|\Sigma|$ so that one has* $\mathbf{PCP}_\Sigma^{[1,1-\varepsilon]}(O(\log n), O(1)) = \mathbf{NP}$.

The reader should appreciate at this point that it is mindblowing how just checking a constant number of symbols could suffice for **NP**-hard problems. This has dramatic consequences for the approximability of **NP**-hard problems as we will discuss here. Proving Theorem 4.2 is far beyond the scope of this lecture. For an excellent exposition of the original algebraic proof of the PCP Theorem we recommend the notes of Minzer [Min22]. A more recent proof using a *gap-amplification argument* was found by Dinur [Din07]. The latter proof can also be found in Chapter 22 of the textbook of Arora and Barak [AB09]. The weaker statement of $\mathbf{NP} \subseteq \mathbf{PCP}_\Sigma^{[1,1-\varepsilon]}(\text{poly}(n), O(1))$ has a much simpler proof that be found for example in Chapter 11 of [AB09]. In fact, the proof is heavily based on the linearity test that we have seen in Chapter 2.

First, it would be a simple observation that one can encode each symbol Σ by bits and hence enforce that $\Sigma = \{0,1\}$. But we will take a different route instead that is more useful for hardness results.

---

[1] Really this is a combination of several works and we cite the set of authors that received the 2001 Gödel prize.

4.1.1 Constraint Satisfaction Problems

The following problem provides a useful alternative view of the functionality of a PCP verifier.

Definition 4.3. The input to the *constraint satisfaction problem* $\textsc{Csp}_{\Sigma,q}$ consists of a q-uniform[2] hypergraph $\mathcal{H} = ([n], \mathcal{E})$ and functions $\Phi_e : \Sigma^e \to \{0,1\}$ that depend only on the values assigned to elements in e. An *assignment* $x \in \Sigma^n$ satisfies constraint e if $\Phi_e(x) = 1$ (where we really mean $\Phi_e((x_i)_{i \in e}) = 1$). The goal is to find an assignment $x \in \Sigma^n$ that maximizes the number of satisfied constraints with $\Phi_e(x) = 1$. We write $\mathsf{val}(\mathcal{H}) \in [0,1]$ as the optimum fraction of satisfiable constraints. We write $\textsc{Csp}_{\Sigma,q}^{[c,s]}$ as the corresppoding gap version of the problem were one needs to distinguish whether $\mathsf{val}(\mathcal{H}) \geq c$ or $\mathsf{val}(\mathcal{H}) \leq s$.

Example 4.4. Several well-studied problems appearing in combinatorial optimization are in fact a CSP. For example for *MaxCut* one is given an undirected graph $G = ([n], E)$ and the goal is to find a set $S \subseteq [n]$ maximizing the number $|\delta(S)|$ of edges crossing the cut. One can model this as a CSP by using the graph as constraint graph (i.e. $q := 2$), using alphabet $\Sigma = \{0,1\}$ and using constraints

$$\Phi_e(x) = \begin{cases} 1 & \text{if } x_u \neq x_v \\ 0 & \text{otherwise.} \end{cases} \quad \forall e = \{u,v\} \in E$$

We denote the *Karp reduction* between two languages by $\leq_p$. More formally, for two languages $A, B \subseteq \{0,1\}^*$ we write $A \leq_p B$ if there is a polynomial time computable function $f : \{0,1\}^* \to \{0,1\}^*$ so that $x \in A \Leftrightarrow f(x) \in B$.

Proposition 4.5. *Let $L \in \mathbf{NP}$ and fix Σ and $1 \geq c > s \geq 0$. Then the following is equivalent:*

- *(A)* $L \in \mathbf{PCP}_{\Sigma}^{c,s}(O(\log n), q)$
- *(B) One has* $L \leq_p \textsc{Csp}_{\Sigma,q}^{[c,s]}$

Proof. $(A) \Rightarrow (B)$. Consider a $\mathbf{PCP}_{\Sigma}^{[c,s]}(r,q)$-verifier $V^{\pi}(x,u)$ with a proof of length $|\pi| = n$. For each choice $u \in \{0,1\}^r$ of random bits the verifier reads q entries of the proof; we denote those entries by $e_u \in \binom{[n]}{q}$. Let $\Phi_{e_u} : \Sigma^{e_u} \to \{0,1\}$ be the function with $\Phi_{e_u}(\pi) = 1$ if and only if the verifier accepts π in case the random bits are u. Then we obtain a $\textsc{Csp}_{\Sigma,q}$ instance $\mathcal{H}$ whose value $\mathsf{val}(\mathcal{H})$ is exactly the maximum probability that the verifier accepts any proof. Note that if $r(n) \leq O(\log n)$, then the instance $\mathcal{H}$ has polynomial size.

---

[2] That means all edges $e \in \mathcal{E}$ have size $|e| = q$.

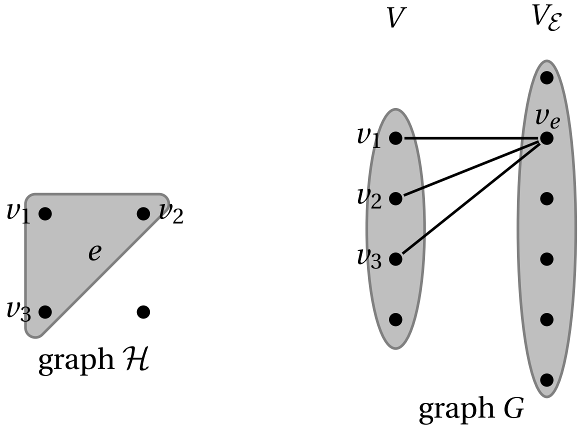
$(B) \Rightarrow (A)$. This reduction also works in reverse: suppose we have a polynomial time reduction from a language $L \in \mathbf{NP}$ to the gap version $\textsc{Csp}^{[c,s]}_{\Sigma,q}$. Then if $\mathcal{H}$ is the produced hypergraph with n vertices, then use $\pi \in \Sigma^n$ as the proof string. We define a verifier that picks a uniform random edge $e \sim \mathcal{E}$ and accepts if and only if $\Phi_e(\pi) = 1$. Then this is a $\mathbf{PCP}^{[c,s]}_{\Sigma}(O(\log|\mathcal{E}|), q)$ verifier. $\square$

4.1.2 Reducing to 2 queries

Next, we prove that in the PCP Theorem *two* queries suffice, i.e. $\mathbf{PCP}^{[1,1-\varepsilon']}_{\Sigma'}(O(\log n), 2) = \mathbf{NP}$ (while $\mathbf{PCP}^{[c,s]}_{\Sigma'}(O(\log n), 1) = \mathbf{BPP}$ for any $1 \geq c > s \geq 0$, meaning that single query does not help).

Proposition 4.6 (2-query PCP Theorem)**.** *There are constant $\varepsilon' > 0$ and $|\Sigma'|$ so that $\textsc{Csp}^{[1,1-\varepsilon']}_{\Sigma',2}$ is $\mathbf{NP}$-hard.*

Proof. Consider a $\textsc{Csp}_{\Sigma,q}$ instance $\mathcal{H} = (V, \mathcal{E})$. We define a bipartite graph $G = (V \dot\cup V_{\mathcal{E}}, F)$ with the original vertices V and a vertex v_e for every original hyperedge $e \in \mathcal{E}$, i.e. $V_{\mathcal{E}} = \{v_e : e \in \mathcal{E}\}$. We insert an edge (v, v_e) whenever $v \in e$. We use symbols Σ for V and symbols Σ^q for vertices in $V_{\mathcal{E}}$ corresponding to assignments to all the q many original nodes. An assignment $x : (V \cup U) \to (\Sigma \cup \Sigma^q)$ satisfies an edge (v, v_e) if (i) $x(v_e)$ satisfies e and (ii) the value of $x(v)$ is consistent with the entry in $x(v_e)$.

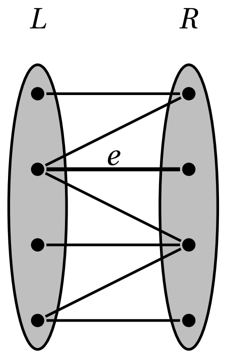


The following is simple and we skip the argument:
Claim I. $\mathsf{val}(\mathcal{H}) = 1 \Rightarrow \mathsf{val}(G) = 1$.
The next claim is the more interesting part and we leave it as homework:
Claim II. $\mathsf{val}(\mathcal{H}) \leq 1 - \varepsilon \Rightarrow \mathsf{val}(G) \leq 1 - \frac{\varepsilon}{q}$. $\square$

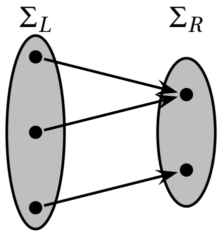
4.1.3 Label Cover

Instead of working with $\text{CSP}_{\Sigma',2}$ it is common to work with an different problem that is equivalent in terms of hardness.

Definition 4.7. A *label-cover* instance Ψ consists of a bipartite graph $G = (L \dot\cup R, E)$, an alphabet $\Sigma = \Sigma_L \dot\cup \Sigma_R$ and maps $\Phi_e : \Sigma_L \to \Sigma_R$ for all edges $e \in E$. The goal is to find an assignment $A : V \to \Sigma$ with $A(u) \in \Sigma_L$ for $u \in L$ and $A(v) \in \Sigma_R$ for $v \in R$ that maximizes the number of satisfied constraints. Here a constraint $e = (u, v) \in E$ is satisfied if $\Phi_e(A(u)) = A(v)$.



graph G



map Φ_e

We denote $\mathsf{val}(\Psi) \in [0, 1]$ as the *value* of the instance, which is the maximum fraction of satisfiable constraints. We would like to point out that the constraints Φ_e with $e = (u, v)$ are of a particular form in the sense that for any assignment for u there is *exactly one* assignment for v that makes the constraint Φ_e true. This is also called a *projection constraint*.

We write $\text{LABELCOVER}_k^{[1,1-\varepsilon]}$ as the gap version of the problem where we have to distinguish the cases $\mathsf{val}(\Psi) = 1$ from $\mathsf{val}(\Psi) \leq 1 - \varepsilon$ where Ψ has alphabet size k.

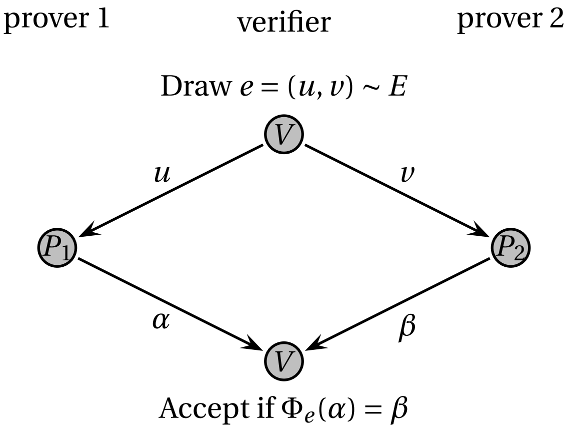
Theorem 4.8. *There are constants $\varepsilon > 0$ and $k \in \mathbb{N}$ so that $\text{LABELCOVER}_k^{[1,1-\varepsilon]}$ is* **NP***-hard.*

Proof. This follows from the **NP**-hardness of $\text{CSP}_{\Sigma',2}$ with the following additional observation concerning the proof of Prop 4.6: In the constructed $\text{CSP}_{\Sigma',2}$-instance $G = (V \cup V_{\mathcal{E}}, F)$ any assignment for v_e that satisfies the constraint for e, allows exactly one assignment to v that makes $(v, v_e) \in F$ true. So indeed this is a *projection* constraint (where $L := V_{\mathcal{E}}$ and $R := V$). $\square$

Naturally, for deriving hardness results it would be much more useful to have **NP**-hardness for $\text{LABELCOVER}_k^{[1,\varepsilon]}$ for *every* constant $\varepsilon > 0$ rather than hardness for $\text{LABELCOVER}_k^{[1,1-\varepsilon]}$ for *some* tiny ε.

4.2 The 2-prover 1-round game

Let $\Psi = (G, \Sigma = \Sigma_L \cup \Sigma_R, \{\Phi_e\}_{e \in E})$ be a label-cover instance and consider the following game: we have a verifier V that only has randomized polynomial time computation available and two all-powerful provers P_1 and P_2. The verifier draws a random edge $e = (u, v) \in E$, sends u to P_1 and v to P_2. Both provers need to output assignments $a \in \Sigma_L$ and $b \in \Sigma_R$ respectively without communicating with each other and their assignments should satisfy the chosen constraint of $\Phi_e(a) = b$. The goal of the provers is to maximize the probability to satisfy the chosen constraint.

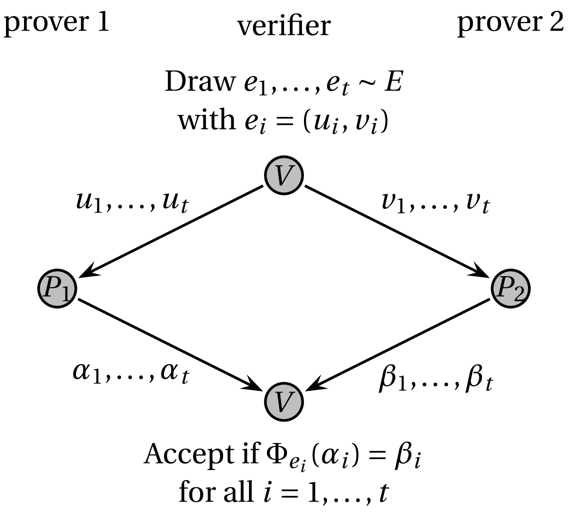


projection game corresponding to Ψ

It is an exercise to argue that this game is equivalent to label cover. This is the reason the game is also called a *projection game*.

Lemma 4.9. *The value of the game equals the value of the label-cover instance,* $\mathsf{val}(\Psi)$.

So by some abuse of notation we interpret $\mathsf{val}(\Psi)$ not just as the value of the label-cover instance but also the value of the equivalent 2-prover 1-round game. Now we want to generalize the game. Imagine the verifier wanted to increase its chances and sample independently edges $e_1, \ldots, e_t \sim E$ where $e_i = (u_i, v_i)$, sends $u_1, \ldots, u_t$ to prover 1 and $v_1, \ldots, v_t$ to prover 2. Then prover 1 sends assignments $\alpha_1, \ldots, \alpha_t \in \Sigma_L$ and prover 2 sends assignments $\beta_1, \ldots, \beta_t \in \Sigma_R$ to the prover. Then the verifier accepts if $\Phi_{e_i}(\alpha_i) = \beta_i$ for all $i = 1, \ldots, t$. We call this the *t-fold game* and denote it by $\Psi^{\otimes t}$.



projection game $\Psi^{\otimes t}$

Again one can show that this game corresponds to a label cover instance of size n^t. Clearly, $\mathsf{val}(\Psi^{\otimes t}) \geq \mathsf{val}(\Psi)^t$ for all t. One might suspect that in fact $\mathsf{val}(\Psi^{\otimes t}) = \mathsf{val}(\Psi)^t$, but that turns out to be false. However a weaker version holds.

Theorem 4.10 (Parallel Repetition Theorem — Raz [Raz95]). *For any $\delta > 0$ and $|\Sigma|$ there is a constant*[3] *$C > 0$ so that: For any label cover instance Ψ with $\mathsf{val}(\Psi) \leq 1 - \delta$ and any $t \in \mathbb{N}$ one has $\mathsf{val}(\Psi^{\otimes t}) \leq \exp\big(-C \cdot t\big)$.*

This is a fundamental result originally due to Raz [Raz95] and while there are several proofs known, they all are beyond the scope of this class. We recommend the simplifications due to Holenstein [Hol07] and Rao [Rao08] as well as the work of Moshkovitz [Mos14] which first modifies the game so that the modified games indeed has $\mathsf{val}(\Psi^{\otimes t}) \approx \mathsf{val}(\Psi)^t$. By going back to the Label Cover problem we can now derive the statement which will be the starting point for our hardness reductions.

Theorem 4.11 (Strong Label Cover Hardness)**.** *For each $\varepsilon > 0$ there is a $k \in \mathbb{N}$ so that* $\textsc{LabelCover}_k^{[1,\varepsilon]}$ *is* **NP***-hard.*

4.3 The 3Lin problem

The target problem for our hardness proof will be the following:

---

[3] One can choose $C := \Theta\left(\frac{\delta^3}{\log|\Sigma|}\right)$

Definition 4.12. For a $3\textsc{Lin}_2$ instance $\mathcal{I}$ we are given m equations of the form

$$x_{e_{i,1}} \oplus x_{e_{i,2}} \oplus x_{e_{i,3}} = c_i$$

where $e_{i,1}, e_{i,2}, e_{i,3} \in \{1, \ldots, n\}$ are distinct indices, $c_i \in \mathbb{F}_2$ and $\oplus$ is the addition modulo 2. The goal is to find an assignment $x \in \mathbb{F}_2^n$ that maximizes the fraction of satisfied equations. We denote the optimum value by $\mathsf{val}(\mathcal{I}) \in [0,1]$.

If $\mathsf{val}(\mathcal{I}) = 1$, then one can use *Gaussian elimination* to find a satisfying assignment x. We also note that for any instance, a random assignment will satisfy half the equations. In the remainder of this chapter, we will prove that remarkably this is already the best possible approximation algorithm:

Theorem 4.13. *For any constant $\varepsilon > 0$ the following holds: Given a $3\textsc{Lin}_2$ instance $\mathcal{I}$ it is* **NP***-hard to distinguish whether $\mathsf{val}(\mathcal{I}) \geq 1 - \varepsilon$ or $\mathsf{val}(\mathcal{I}) \leq \frac{1}{2} + \varepsilon$.*

Similar to earlier chapters it will be notationally more convinient to work with the $\{-1,1\}^n$ cube rather than $\{0,1\}^n$.

Definition 4.14. For a $3\textsc{Lin}_{\{-1,1\}}$ instance $\mathcal{I}$ we are given m weighted equations of the form

$$x_{e_{i,1}} \cdot x_{e_{i,2}} \cdot x_{e_{i,3}} = c_i$$

where $e_{i,1}, e_{i,2}, e_{i,3} \in \{1, \ldots, n\}$ are distinct indices, $c_i \in \{-1,1\}$ and $w_i \geq 0$ is the weight. The goal is to find an assignment $x \in \{-1,1\}^n$ that maximizes the cumulated weight of satisfied equations. Again denote the optimum value by $\mathsf{val}(\mathcal{I})$.

In this formulation we admit weights, but one could "simulate" weights by replacing each equation i with $\lfloor N w_i \rfloor$ many unweighted copies where N is big enough.

4.4 The Noisy Linearity Test

First we want to build up on the linearity test from Chapter 2.

Definition 4.15. A function $f : \{\pm 1\}^n \to \{\pm 1\}$ is called a *dictatorship function* if there is an index $i \in [n]$ so that $f(x) = x_i = \chi_{\{i\}}(x)$ for all $x \in \{-1,1\}^n$.

Note that there are only n dictator functions in dimension n. There is also a coding-theoretic interpretation:

Definition 4.16. The *long code* in dimension n is the set $\mathsf{LC} := \{(x_i)_{x \in \{-1,1\}^n} \mid i \in [n]\}$.

In other words, the long code contains the function tables of all the dictator ship functions. In particular the long code is a subset of the Hadamard code (see Section 3.3). The long code is called *long* code because — well — it is long. It uses 2^n bits to encode merely n code words (which could be encoded using only $\log_2 n$ bits). But it has so much redundancy that it is quite useful. The idea is that given a LABELCOVER instance and an assignment $A : L \to \Sigma_L$ for the left hand side nodes, we will encode the symbol $A(u)$ for $u \in L$ using the function table $(\chi_{A(u)}(x))_{x \in \{-1,1\}^{\Sigma_L}}$; analogously for the right hand side nodes R. Before we come to the actual reduction, we need to learn how to make use out of those dictatorship functions.

We recall that in Chapter 2 we have proven that any function $f : \{\pm 1\}^n \to \{\pm 1\}$ that passes the linearity test

$$f(x \odot y) = f(x) \cdot f(y) \quad \text{for } x, y \sim \{\pm 1\}^n$$

with probability at least $\frac{1}{2} + \delta$, must have a coefficient $S \subseteq [n]$ with $\hat{f}(S) \geq 2\delta$. But every function χ_S with $S \subseteq [n]$ passes this test with probability 1. Now we would like to modify this linearity test so that it still accepts dictatorship functions but is likely rejects functions χ_S with large $|S|$. It turns out that dictatorship functions are less sensitive to *noise* than functions χ_S with large $|S|$. This is the crucial property that we will use.

Definition 4.17. For $0 \leq \varepsilon \leq 1$, we define the *$\varepsilon$-biased distribution* $\mathcal{D}_\varepsilon([n])$ as the distribution over $\{-1,1\}^n$ with independent coordinates so that

$$\Pr_{x \sim \mathcal{D}_\varepsilon([n])}[x_i = 1] = 1 - \varepsilon \quad \text{and} \quad \Pr_{x \sim \mathcal{D}_\varepsilon([n])}[x_i = -1] = \varepsilon$$

for all $i \in [n]$.

If clear from context, then we drop the set $[n]$. Note that $\mathbb{E}_{x \sim \mathcal{D}_\varepsilon}[x_i] = 1 - 2\varepsilon$.

NOISY LINEARITY TEST

Input: Access to a function $f : \{\pm 1\}^n \to \{\pm 1\}$.

(1) Pick independent random $x, y \sim \{-1,1\}^n$ and $a \sim \mathcal{D}_\varepsilon([n])$
(2) Accept if $f(a \odot x \odot y) = f(x) \cdot f(y)$

We will now analyze the Noisy Linearity test; the arguments will extend the proof of Theorem 2.1.

Theorem 4.18 (Noisy Linearity Test)**.** *Let $f : \{\pm 1\}^n \to \{\pm 1\}$ and $0 < \varepsilon \leq \frac{1}{2}$.*

(A) If f is a dictatorship function, then it passes the Noisy Linearity test with probability $1 - \varepsilon$.

(B) If f passes the Noisy Linearity test with probability $\frac{1}{2}+\delta$, then there is an $S \subseteq [n]$ so that $(1-2\varepsilon)^{|S|}\hat{f}(S) \geq 2\delta$ (in particular $|S| \leq \frac{\ln(1/\delta)}{2\varepsilon}$ and $\hat{f}(S) \geq 2\delta$).

Proof. For (A). If $f(x) = x_i$, then for $x, y \sim \{-1,1\}^n$ and $a \sim \mathcal{D}_\varepsilon$ one has

$$\Pr[f(a \odot x \odot y) = f(x)f(y)] = \Pr[a_i x_i y_i = x_i y_i] = \Pr[a_i = 1] = 1-\varepsilon$$

For (B). Now assume that f is an arbitrary function that passes the test with probability $\frac{1}{2}+\delta$. We note that

$$\mathop{\mathbb{E}}_{\substack{x,y\sim\{\pm1\}^n\\ a\sim\mathcal{D}_\varepsilon}}[f(a\odot x\odot y)f(x)f(y)] = 2\underbrace{\Pr[f(a\odot x\odot y)f(x)f(y)=1]}_{\geq 1/2+\delta}-1 \geq 2\delta$$

Writing out the Fourier expansion (as we did in the proof of Theorem 2.1) we obtain

$$\begin{aligned}
2\delta &\leq \mathop{\mathbb{E}}_{x,y\sim\{\pm1\}^n, a\sim\mathcal{D}_\varepsilon}[f(a\odot x\odot y)f(x)f(y)] \\
&\overset{\text{Thm 1.3}}{=} \mathop{\mathbb{E}}_{x,y\sim\{\pm1\}^n, a\sim\mathcal{D}_\varepsilon}\Big[\Big(\sum_{S\subseteq[n]}\hat{f}(S)\chi_S(a\odot x\odot y)\Big)\Big(\sum_{T\subseteq[n]}\hat{f}(T)\chi_T(x)\Big)\Big(\sum_{R\subseteq[n]}\hat{f}(R)\chi_R(y)\Big)\Big] \\
&\overset{\text{indep.}}{=} \sum_{S,T,R\subseteq[n]}\hat{f}(S)\hat{f}(T)\hat{f}(R)\mathop{\mathbb{E}}_{a\sim\mathcal{D}_\varepsilon}[\chi_S(a)]\underbrace{\mathop{\mathbb{E}}_{x\sim\{\pm1\}^n}\big[\chi_S(x)\chi_T(x)\big]}_{=1 \text{ if } S=T,\, =0 \text{ o.w.}}\underbrace{\mathop{\mathbb{E}}_{y\sim\{\pm1\}^n}\big[\chi_S(y)\chi_R(y)\big]}_{=1 \text{ if } S=R,\, =0 \text{ o.w.}} \\
&= \sum_{S\subseteq[n]}\hat{f}(S)^3\underbrace{\prod_{i\in S}\underbrace{\mathbb{E}[a_i]}_{=1-2\varepsilon}}_{=(1-2\varepsilon)^{|S|}} \\
&\leq \max_{S\subseteq[n]}\Big\{(1-2\varepsilon)^{|S|}\hat{f}(S)\Big\}\cdot\underbrace{\sum_{S\subseteq[n]}\hat{f}(S)^2}_{=1}
\end{aligned}$$

Here we have used independence of x, y, a and the fact that $\chi_S(x \odot y) = \chi_S(x)\cdot\chi_S(y)$. ☐

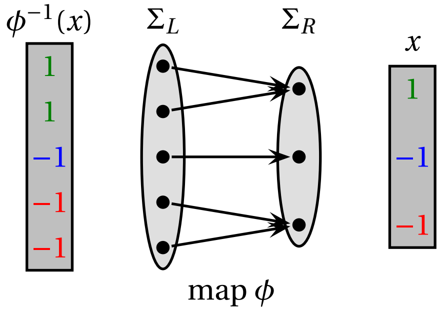
4.5 A combined Noisy Linearity + constraint test

The next step is to develop a variant of the noisy linearity test that can incorporate one label cover constraint $\phi : \Sigma_L \to \Sigma_R$. Recall that a pair $(a,b) \in \Sigma_L \times \Sigma_R$ satisfies the constraint ϕ if $\phi(a) = b$. We encode $a \in \Sigma_L$ with a dictatorship function $f := \chi_{\{a\}} : \{-1,1\}^{\Sigma_L} \to \{\pm1\}$ and $b \in \Sigma_R$ is encoded by the dictatorship function $g := \chi_{\{b\}} : \{\pm1\}^{\Sigma_R} \to \{\pm1\}$. In the instructive special case where $\Sigma_L = \Sigma_R = \Sigma$ and ϕ

is the identity, the right test would be to draw $x, z \sim \{-1,1\}^{\Sigma}$ and $w \sim \mathcal{D}_{\varepsilon}(\Sigma)$ and check whether

$$f(w \odot x \odot z) = f(z) \odot g(x)$$

The case where ϕ is an arbitrary function needs some modification. For $x \in \{-1,1\}^{\Sigma_R}$ we define $\phi^{-1}(x) \in \{-1,1\}^{\Sigma_L}$ as the vector with $\phi^{-1}(x)_i := x_{\phi(i)}$ for $i \in \Sigma_L$. We also call ϕ^{-1} the *pull-back function* of ϕ.



Consider the following test:

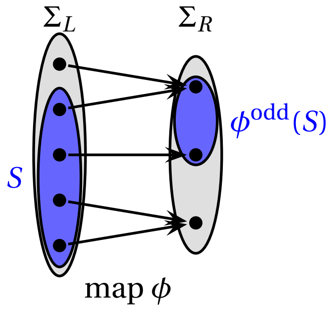
NOISY LINEARITY + CONSTRAINT TEST

Input: Constraint $\phi : \Sigma_L \to \Sigma_R$. Access to functions $f : \{\pm 1\}^{\Sigma_L} \to \{\pm 1\}$ and $g : \{-1,1\}^{\Sigma_R} \to \{\pm 1\}$.

(1) Sample $x \sim \{-1,1\}^{\Sigma_R}$ and set $y := \phi^{-1}(x)$
(2) Sample $z \sim \{-1,1\}^{\Sigma_L}$ and $w \sim \mathcal{D}_{\varepsilon}(\Sigma_L)$
(3) Accept if $f(w \odot y \odot z) \cdot f(z) = g(x)$

For a set $S \subseteq \Sigma_L$ we define

$$\phi^{\text{odd}}(S) := \left\{ b \in \Sigma_R \mid \text{there is an odd number of } a \in S \text{ with } \phi(a) = b \right\}$$



Actually the only propery that is going to be relevant from this definition is that $\phi^{\text{odd}}(S) \subseteq \phi(S)$. We will now analyze the Noisy Linearity + Constraint Test.

Theorem 4.19. *The Noisy Linearity + Constraint Test satisfies the following:*

(A) *If* $f = \chi_{\{a\}}$ and $g = \chi_{\{b\}}$ *with* $\phi(a) = b$, *then the test accepts with probability* $1-\varepsilon$.

(B) *If the test accepts with probability at least* $\frac{1}{2} + \delta$, *then*

$$\sum_{S\subseteq\Sigma_L : |S| \le \frac{\ln(1/\delta)}{\varepsilon}} \hat{f}(S)^2 \cdot \hat{g}(\phi^{\text{odd}}(S))^2 \ge \delta^2$$

Proof. For (A). In this case, the equation

$$f(w\odot y\odot z)\cdot f(z) = w_a \underbrace{y_a}_{=x_b} \underbrace{z_a^2}_{=1} = w_a x_b \overset{!}{=} x_b = g(x)$$

is true iff $w_a = 1$, which happens with probability $1-\varepsilon$.

For (B). As before in the proof of Theorem 4.18 we expand the bias of $f(w\odot y\odot z)f(z)g(x)$ into the Fourier basis and simplify the terms:

$$\begin{aligned}
2\delta \quad &\le \quad \mathop{\mathbb{E}}_{x,y,z,w}[f(w\odot y\odot z)f(z)g(x)] \\
&= \quad \mathop{\mathbb{E}}_{x,y,z,w}\Big[\sum_{S\subseteq\Sigma_L}\hat{f}(S)\chi_S(w\odot y\odot z)\sum_{T\subseteq\Sigma_L}\hat{f}(T)\chi_T(z)\sum_{R\subseteq\Sigma_R}\hat{g}(R)\chi_R(x)\Big] \\
&= \quad \sum_{S,T,R}\hat{f}(S)\hat{f}(T)\hat{g}(R)\mathop{\mathbb{E}}_{x\sim\{-1,1\}^{\Sigma_R}}[\chi_R(x)\chi_S(y)]\underbrace{\mathop{\mathbb{E}}_{z\sim\{-1,1\}^{\Sigma_L}}[\chi_S(z)\chi_T(z)]}_{=1\text{ if }S=T,\ =0\text{ o.w.}}\underbrace{\mathop{\mathbb{E}}_{w\sim\mathcal{D}_\varepsilon(\Sigma_L)}[\chi_S(w)]}_{=(1-2\varepsilon)^{|S|}} \\
&\overset{(*)}{=} \quad \sum_{S\subseteq\Sigma_L, R\subseteq\Sigma_R}(1-2\varepsilon)^{|S|}\hat{f}(S)^2\hat{g}(R)\cdot\underbrace{\mathop{\mathbb{E}}_{x\sim\{-1,1\}^{\Sigma_R}}[\chi_R(x)\chi_{\phi^{\text{odd}}(S)}(x)]}_{=1\text{ if }R=\phi^{\text{odd}}(S),0\text{ o.w.}} \\
&= \quad \sum_{S\subseteq\Sigma_L}(1-2\varepsilon)^{|S|}\hat{f}(S)^2\hat{g}(\phi^{\text{odd}}(S)) \\
&\overset{\text{Cauchy-S.}}{\le} \quad \Big(\underbrace{\sum_{S\subseteq\Sigma_L}\hat{f}(S)^2}_{=1}\cdot\sum_{S\subseteq\Sigma_L}(1-2\varepsilon)^{2|S|}\hat{f}(S)^2\hat{g}(\phi^{\text{odd}}(S))^2\Big)^{1/2} \\
&\le \quad \Big(\underbrace{\sum_{\substack{S\subseteq\Sigma_L:\\ |S|>\frac{\ln(1/\delta)}{\varepsilon}}}\hat{f}(S)^2}_{\le 1}\underbrace{(1-2\varepsilon)^{2|S|}}_{\le\delta^4\text{ by }(**)}\underbrace{\hat{g}(\phi^{\text{odd}}(S))^2}_{\le 1}+\sum_{\substack{S\subseteq\Sigma_L:\\ |S|\le\frac{\ln(1/\delta)}{\varepsilon}}}\underbrace{(1-2\varepsilon)^{2|S|}}_{\le 1}\hat{f}(S)^2\hat{g}(\phi^{\text{odd}}(S))^2\Big)^{1/2} \\
&\le \quad \Big(\delta^4+\sum_{S\subseteq\Sigma_L:|S|\le\frac{\ln(1/\delta)}{\varepsilon}}\hat{f}(S)^2\hat{g}(\phi^{\text{odd}}(S))^2\Big)^{1/2}
\end{aligned}$$

Then rearranging gives $\sum_{S\subseteq\Sigma_L:|S|\leq\frac{\ln(1/\delta)}{\varepsilon}} \hat{f}(S)^2\cdot\hat{g}(\phi^{\text{odd}}(S))^2 \geq (2\delta)^2-\delta^4\geq\delta^2$. Here we use in $(*)$ that

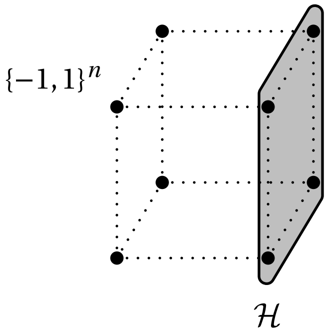
$$\chi_S(y)=\chi_S(\phi^{-1}(x))=\prod_{i\in S}x_{\phi(i)}=\chi_{\phi^{\text{odd}}(S)}(x)$$

as pairs of distinct indices $i_1,i_2\in S$ with $\phi(i_1)=\phi(i_2)$ have $x_{\phi(i_1)}\cdot x_{\phi(i_2)}=1$. This is the reason why we have the term $\phi^{\text{odd}}(S)$ appearing in the statement in the first place. Finally note that in $(**)$ we use that for $|S|>\frac{\ln(1/\delta)}{\varepsilon}$ one has $(1-2\varepsilon)^{2|S|}\leq\exp(-4\varepsilon|S|)\leq\exp(-4\ln(\frac{1}{\delta}))=\delta^4$. $\square$

In particular, Theorem 4.19 shows that if f and g pass the test with probability $\frac{1}{2}+\delta$, then there is a significant Fourier coefficient $\hat{f}(S)^2\cdot\hat{g}(\phi^{\text{odd}}(S))^2$ for small S. Intuitively this should be helpful to extract a good labelling from S. But there is one obstacle for this in order to be useful. We need to make sure that the large Fourier coefficient does not come from the set $S=\emptyset$. For that the following definition will be useful:

Definition 4.20. A function $f:\{-1,1\}^n\to\{\pm1\}$ is *odd* if $f(-x)=-f(x)$ for all $x\in\{-1,1\}^n$.

In particular an odd function has the Fourier coefficient $\hat{f}(\emptyset)=\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)]=0$. Now, let us go back to the test where we have functions $f:\{-1,1\}^{\Sigma_L}\to\{-1,1\}$. Let $\mathcal{H}\subseteq\{-1,1\}^{\Sigma_L}$ be any subset of the hypercube so that each pair $\{x,-x\}$ of antipodal points has $|\mathcal{H}\cap\{x,-x\}|=1$. A canonical choice would be $\mathcal{H}:=\{x\in\{-1,1\}^n\mid x_1=1\}$.



We can demand that the test only has a table for the partial function $f:\mathcal{H}\to\{-1,1\}$ and whenever the test adresses an entry $f(x)$ with $x\in\{-1,1\}^n\setminus\mathcal{H}$ then we define that entry as $f(x):=-f(-x)$. This was we can enforce that the function f is odd. Now we have all ingredients for a reduction.

4.6 Hardness for 3LIN

Now we will reduce LABELCOVER to $3\textsc{Lin}_{\{-1,1\}}$. The crucial ingredient to that reduction is the fact that the equations $f(w \odot y \odot z) \cdot f(z) = g(x)$ from the Noisy Linearity + Constraint Test are in fact $3\textsc{Lin}_{\{-1,1\}}$-equations. Now we come to the actual reduction:

Proposition 4.21. *For any* $0 < \varepsilon \leq 1$ *there is a* $\gamma := \gamma(\varepsilon) > 0$ *so that the following holds. Given a label cover instance* $\Psi = (G, \Sigma = \Sigma_L \dot\cup \Sigma_R, (\Phi_e)_{e \in E})$ *one can construct a* $3\textsc{Lin}_{\{-1,1\}}$ *instance* $\mathcal{I}$ *of size polynomial in* $|V|$ *and* $2^{|\Sigma|}$ *so that:*

- *Completeness:* $\mathsf{val}(\Psi) = 1 \Rightarrow \mathsf{val}(\mathcal{I}) \geq 1 - \varepsilon$.
- *Soundness:* $\mathsf{val}(\Psi) \leq \gamma \Rightarrow \mathsf{val}(\mathcal{I}) \leq \frac{1}{2} + \varepsilon$.

Proof. We create a $3\textsc{Lin}_{\{-1,1\}}$ instance $\mathcal{I}$ that contains a variable $f_u(z) \in \{-1,1\}$ for all $u \in L$ and $z \in \{-1,1\}^{\Sigma_L}$. Moreover we have a variable $g_v(x)$ for all $v \in R$ and $x \in \{-1,1\}^{\Sigma_R}$. For each edge $e = (u,v) \in E$ in the label cover instance, each $x \in \{-1,1\}^{\Sigma_R}$, $z \in \{-1,1\}^{\Sigma_L}$ and $a \in \{-1,1\}^{\Sigma_L}$ we insert the equation

$$f_u(a \odot y \odot z) \cdot f_u(z) \cdot g_v(x) = 1$$

where $y := \phi_e^{-1}(x)$. The weight of that equation is $\frac{1}{|E|}$ times the probability/density of the tuple (x,z,a), which is $2^{-|\Sigma_R|} \cdot 2^{-|\Sigma_L|} \cdot (1-\varepsilon)^{\#i:a_i=1} \cdot \varepsilon^{\#i:a_i=-1}$. Note that the sum of all the weights is exactly 1. As explained above, we enforce that the functions f_u and g_v are odd (which really means we only have half the variables that we listed).

Claim I. *One has* $\mathsf{val}(\Psi) = 1 \Rightarrow \mathsf{val}(\mathcal{I}) \geq 1 - \varepsilon$.

Proof of Claim I. Let $A : V \to \Sigma$ be a satisfying assignment for Ψ. Then we set the variables for $u \in L$ and $v \in L$ to the corresponding dictatorship functions $f_u := \chi_{\{A(u)\}}$ and $g_v := \chi_{\{A(v)\}}$. As proven in Theorem 4.19.(A), for each single constraint $e = (u,v)$, the weight of the associated $3\textsc{Lin}_{\{-1,1\}}$-equations that are satisfied is at least $\frac{1-\varepsilon}{|E|}$. Here we also use that dictatorship functions are odd. □

Now we can prove soundness:

Claim II. *For any* $\delta, \varepsilon > 0$, *there is a* $\gamma := \gamma(\delta, \varepsilon) > 0$ *so that* $\mathsf{val}(\mathcal{I}) \geq \frac{1}{2} + \delta \Rightarrow \mathsf{val}(\Psi) \geq \gamma$.

Proof of Claim II. We fix the functions f_u and g_v that satisfy a $\frac{1}{2} + \delta$ fraction of equations. For an edge $e = (u,v) \in E$ we abbreviate

$$\delta_e := \mathop{\mathbb{E}}_{x,z,a,y:=\phi_e^{-1}(x)} \left[f_u(a \odot y \odot z) \cdot f_u(z) \cdot g_v(x) \right]$$

Equivalently, the fraction of equations in $\mathcal{I}$ that arise from e and are satisfied is $\frac{1}{2} + \frac{\delta_e}{2}$. One should think of δ_e as the *advantage* that the functions f_u and g_v

provide over a random assignment (which would satisfy half of the equations). Note that $\mathsf{val}(\mathcal{I}) = \frac{1}{2} + \frac{1}{2}\mathbb{E}_{e\sim E}[\delta_e] \geq \frac{1}{2} + \delta$ and so $\mathbb{E}_{e\sim E}[\delta_e] \geq 2\delta$. We call an edge e *good* if $\delta_e \geq \delta$ and denote those good edges by $E_{\text{good}} := \{e \in E \mid e \text{ is good}\}$. By the Reverse Markov inequality (Lemma 1.40) we know that $|E_{\text{good}}| \geq \delta|E|$. So we have a constant fraction of edges where f_u, g_v provide a constant advantage.

Next, we construct an assignment $A : V \to \Sigma$ that satisfies a constant fraction of good edges. For each vertex $u \in L$ we consider the function $f : \{\pm 1\}^{\Sigma_L} \to \{\pm 1\}$ that is supposed to encode the label for u. Recall that $\sum_{S\subseteq\Sigma_L} \hat{f}_u(S)^2 = 1$. We draw a set $S_u \subseteq \Sigma_L$ at random with probability $\hat{f}_u(S)^2$. Then we draw $A(u) \sim S_u$ uniformly at random. Similarly we assign labels to vertices on the right: for $v \in R$ we draw $S_v \subseteq \Sigma_R$ with probability $\hat{g}_v(S)^2$ and then sample $A(v) \sim S_v$. It remains to prove that this is a decent assignment:

Subclaim II.A. *For each $e \in E_{good}$ one has $\Pr_A[A$ satisfies $e] \geq \frac{\delta^2}{\ln(1/\delta)^2}\varepsilon^2$.*

Proof of Subclaim II.A. Let $e = (u, v)$. First let us condition that we choose a set S_u and $S_v := \phi_e^{\text{odd}}(S_u)$. If these events have happened with positive probability, then $\hat{f}_u(S_u)^2 > 0$ and $\hat{g}_v(\phi_e^{\text{odd}}(S_v))^2 > 0$. Since by construction f_u and g_v are odd, we know that $S_u \neq \emptyset$ and $\phi_e^{\text{odd}}(S_v) \neq \emptyset$. Any $b \in \phi_e^{\text{odd}}(S_v)$ has at least one $a \in S_v$ so that $\phi_e(a) = b$. Hence the probability to satisfy the edge e is

$$\Pr_A\left[A \text{ satisfies } e \mid S_u \text{ and } S_v := \phi_e^{\text{odd}}(S_u)\right] \geq \frac{1}{|S_u|\cdot|\phi_e^{\text{odd}}(S_u)|} \overset{|\phi_e^{\text{odd}}(S_u)|\leq|S_u|}{\geq} \frac{1}{|S_u|^2}$$

Now, let us uncondition. Then only summing over the small sets S_u guaranteed in Theorem 4.19.(B) we get a lower bound of

$$\Pr_A[A \text{ satisfies } e] \geq \underbrace{\sum_{S\subseteq\Sigma_L:|S|\leq\frac{\ln(1/\delta)}{\varepsilon}} \hat{f}_u(S)^2\hat{g}_v(\phi_e^{\text{odd}}(S))^2}_{\geq\delta^2 \text{ by Thm 4.19.(B)}} \cdot \underbrace{\frac{1}{|S|^2}}_{\geq\frac{\varepsilon^2}{\ln(1/\delta)^2}} \geq \frac{\varepsilon^2\delta^2}{\ln(1/\delta)^2}$$

That finishes Subclaim II.A. Since at least a δ-fraction of edges is good, we have that $\mathsf{val}(\Psi) \geq \frac{\delta^3}{\ln(1/\delta)^2}\varepsilon^2$ and Claim II follows. □

For the conclusion we can set for example $\varepsilon := \delta$. □

We can conclude that gap version $3\textsc{Lin}^{[1-\varepsilon,\frac{1}{2}+\varepsilon]}_{\{-1,1\}}$ is **NP**-hard:

Theorem 4.22. *For any constant $\varepsilon > 0$ the following holds: Given a $3\textsc{Lin}_{\{-1,1\}}$ instance $\mathcal{I}$ it is* **NP***-hard to distinguish whether* $\mathsf{val}(\mathcal{I}) \geq 1 - \varepsilon$ *or* $\mathsf{val}(\mathcal{I}) \leq \frac{1}{2} + \varepsilon$.

Proof. Follows from combining the hardness of $\textsc{LabelCover}_k^{[1,\gamma]}$ for any $\gamma > 0$ (with $k := k(\gamma)$ large enough) from Theorem 4.11 with the reduction in Prop 4.21. □

Arguably a better known problem is MAX-3-SAT where we are given a SAT formula $\mathcal{C}$ with 3 literals per clause[4] and the goal is to find an assignment to the variables that satisfies as many clauses as possible. Using similar techniques as for 3LIN one can prove the following:

Theorem 4.23 (Håstad [H01]). *For any* $\varepsilon > 0$*, given a* 3SAT *instance* $\mathcal{C}$*, it is* **NP**-*hard to distinguish between* $\text{val}(\mathcal{C}) = 1$ *and* $\text{val}(\mathcal{C}) \leq \frac{7}{8} + \varepsilon$.

Here $\text{val}(\mathcal{C})$ denote the maximum fraction of satisfiable clauses. Recall that for any instance one already has $\text{val}(\mathcal{C}) \geq \frac{7}{8}$ via a random assignment.

---

[4] For example $\mathcal{C}$ could be $(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee x_4) \wedge (x_2 \vee \neg x_3 \vee \neg x_4)$.

Chapter 5

Hypercontractivity

Recall that for any $f : \{-1,1\}^n \to \mathbb{R}$ and $1 \le p \le q < \infty$ we have $\|f\|_{E,p} \le \|f\|_{E,q}$ by a Jensen inequality argument (see Prop 1.18). The goal of this chapter will be how to bound $\|f\|_{E,q}$ in terms of $\|f\|_{E,p}$.

5.1 Bonami's Lemma

First we prove that for any function f the ratio $\frac{\|f\|_{E,q}}{\|f\|_{E,p}}$ (the "Jensen gap") can be bounded dependent on the degree of f. Recall that for any random variable X, by Jensen inequality one has $\mathbb{E}[X^4] \ge \mathbb{E}[X^2]^2$. On the other hand, for well concentrated random variables one would expect that the gap between both quantities is not large.

Definition 5.1. We say that a random variable X is B*-reasonable* if $\mathbb{E}[X^4] \le B \cdot \mathbb{E}[X^2]^2$.

It is not hard to verify that the random variables $x \sim \{-1,1\}$, $g \sim N(0,1)$ and $u \sim [-1,1]$ are B-reasonable for some constant B. Reasonable random variables satisfy some (weak) concentration:

Lemma 5.2. *If X is B-reasonable, then for all $t > 0$, $\Pr[|X| > t\,\mathbb{E}[X^2]^{1/2}] < \frac{B}{t^4}$.*

Proof. Using monotonicity of $z \to z^4$ and Markov's Inequality we get

$$\Pr\left[|X| > t\,\mathbb{E}[X^2]^{1/2}\right] = \Pr\left[X^4 > t^4\,\mathbb{E}[X^2]^2\right] < \frac{\mathbb{E}[X^4]}{t^4\,\mathbb{E}[X^2]^2} \le \frac{B}{t^4}.$$

□

It is a well known fact that concentration of a random variable also implies some form of *anti-concentration*:

Proposition 5.3. *Let X be B-reasonable. Then for any $0 \le t \le 1$ one has $\Pr[|X| \ge t\mathbb{E}[X^2]^{1/2}] \ge \frac{(1-t^2)^2}{B}$.*

Proof. Using the Paley-Zygmund inequality (Lemma 1.41) we obtain

$$\Pr\left[|X| \ge t\,\mathbb{E}[X^2]^{1/2}\right] = \Pr\left[X^2 \ge t^2\,\mathbb{E}[X^2]\right] \overset{\text{Paley-Zygmund}}{\ge} (1-t^2)^2\frac{\mathbb{E}[X^2]^2}{\mathbb{E}[X^4]} \overset{B\text{-reasonable}}{\ge} \frac{(1-t^2)^2}{B}$$

□

Next, we prove an important result telling us that low degree boolean functions correspond to reasonable random variables.

Theorem 5.4 (Bonami Lemma)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function with $\deg(f) \le k$. Then*

(i) The random variable $f(x)$ (where $x \sim \{-1,1\}^n$) is 9^k-reasonable.

(ii) One has $\|f\|_{E,4} \le \sqrt{3}^k \|f\|_{E,2}$.

Proof. We quickly show that (ii) follows from (i) as

$$\|f\|_{E,4} = \mathbb{E}[f(x)^4]^{1/4} \overset{(i)}{\le} \left(9^k \cdot \mathbb{E}[f(x)^2]^2\right)^{1/4} = \sqrt{3}^k \cdot \|f\|_{E,2}$$

Now we prove (i) by induction over n. For $n = 0$, the random variable $f(x)$ is constant and the claim is true. Now assume $n \ge 1$. We write $x = (\bar{x}, x_n)$ with $\bar{x} = (x_1, \dots, x_{n-1})$ and pull out the variable x_n to obtain

$$f(x) = x_n g(\bar{x}) + h(\bar{x})$$

where g and h depend on at most $n-1$ variables with $\deg(g) \le k-1$ and $\deg(h) \le k$. The goal is to prove that $\mathbb{E}[f(x)^4] \le 9^k \cdot \mathbb{E}[f(x)^2]^2$ where $x \sim \{-1,1\}^n$. First we can rewrite the right hand side as

$$\begin{aligned}
\mathbb{E}\left[f(x)^2\right]^2 &= \mathbb{E}\left[\left(x_n g(\bar{x}) + h(\bar{x})\right)^2\right]^2 \\
&= \Big(\underbrace{\mathbb{E}[x_n^2]}_{=1}\mathbb{E}[g(\bar{x})^2] + 2\underbrace{\mathbb{E}[x_n]}_{=0}\mathbb{E}[g(\bar{x})h(\bar{x})] + \mathbb{E}[h(\bar{x})^2]\Big)^2 \\
&= \Big(\mathbb{E}[g(\bar{x})^2] + \mathbb{E}[h(\bar{x})^2]\Big)^2 =: (*)
\end{aligned}$$

using that x_n and $\bar{x}$ are independent. Now we do the main argument and bound the left hand side as

$$\begin{aligned}
\mathbb{E}\left[f(x)^4\right] &= \mathbb{E}\left[\left(x_n g(\bar{x})+h(\bar{x})\right)^4\right] \\
&\overset{\substack{\text{indep.+binom}\\\text{formula}}}{=} \underbrace{\mathbb{E}[x_n^4]}_{=1}\mathbb{E}\left[g(\bar{x})^4\right]+4\underbrace{\mathbb{E}[x_n^3]}_{=0}\mathbb{E}\left[g(\bar{x})^3h(\bar{x})\right]+6\underbrace{\mathbb{E}[x_n^2]}_{=1}\mathbb{E}\left[g(\bar{x})^2h(\bar{x})^2\right] \\
&\qquad +4\underbrace{\mathbb{E}[x_n]}_{=0}\mathbb{E}\left[g(\bar{x})h(\bar{x})^3\right]+\mathbb{E}\left[h(\bar{x})^4\right] \\
&= \mathbb{E}\left[g(\bar{x})^4\right]+6\mathbb{E}\left[g(\bar{x})^2h(\bar{x})^2\right]+\mathbb{E}\left[h(\bar{x})^4\right] \\
&\overset{\text{Cauchy-Schwarz}}{\leq} \mathbb{E}\left[g(\bar{x})^4\right]+6\sqrt{\mathbb{E}\left[g(\bar{x})^4\right]\mathbb{E}\left[h(\bar{x})^4\right]}+\mathbb{E}\left[h(\bar{x})^4\right] \\
&\overset{\text{induction}}{\leq} 9^{k-1}\mathbb{E}\left[g(\bar{x})^2\right]^2+6\sqrt{9^{k-1}\mathbb{E}\left[g(\bar{x})^2\right]^2\cdot 9^k\mathbb{E}\left[h(\bar{x})^2\right]^2}+9^k\mathbb{E}\left[\bar{h}(x)^2\right]^2 \\
&\overset{(**)}{\leq} 9^k\cdot\left(\mathbb{E}\left[g(\bar{x})^2\right]^2+2\mathbb{E}[g(\bar{x})^2]\mathbb{E}[h(\bar{x})^2]+\mathbb{E}[h(\bar{x})^4]\right) \\
&\overset{\text{bin.formula}}{=} 9^k\cdot\underbrace{\left(\mathbb{E}[g(\bar{x})^2]+\mathbb{E}[h(\bar{x})^2]\right)^2}_{=(*)}=9^k\cdot\mathbb{E}\left[f(x)^2\right]^2
\end{aligned}$$

In $(**)$ we use that $6\sqrt{9^{k-1}9^k}=6\cdot\frac{9^k}{3}=2\cdot 9^k$.

□

Remark 1. Let B_k be the minimum value so that every degree-k function f is B_k-reasonable. We have proven in Theorem 5.4 that $B_k \leq 9^k$. For a simple exponential lower bound, let $1 \leq k \leq n$ and consider the function $f:\{-1,1\}^n \to \{0,1\}$ with

$$f(x) := \begin{cases} 1 & \text{if } x_1 = \ldots = x_k = 1 \\ 0 & \text{otherwise} \end{cases}$$

for $x \in \{-1,1\}^n$. Then $\deg(f) = k$ as one can verify. One can also check that $\mathbb{E}[f(x)^2] = 2^{-k} = \mathbb{E}[f(x)^4]$ which implies that $B_k \geq 2^k$ is needed. A more complex example shows that $B_k \geq \Omega(\frac{9^k}{k^2})$, see Exercise 9.3 in [O'D21].

5.2 The FKN Theorem

Next, we see an application of Bonami's Lemma to derive the FKN Theorem which says that any boolean function $f:\{-1,1\}^n \to \{-1,1\}$ with most weight on level-1 must be close to $\pm\chi_{\{i\}}$ for some coordinate i. Recall that for two functions $f,g:\{-1,1\}^n \to \{-1,1\}$, their *distance* is denoted by $\text{dist}(f,g) := \Pr_{x\sim\{-1,1\}^n}[f(x) \neq g(x)] \in [0,1]$ and the *level-1 weight* of f is $W^1[f] = \sum_{i=1}^n \hat{f}(\{i\})^2$.

Theorem 5.5 (Friedgut-Kalai-Naor (FKN) Theorem)**.** *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be a function with* $W^1[f] = 1-\delta$ *for some* $0 \leq \delta \leq 1$*. Then there is an index* $i \in [n]$ *and a sign* $\sigma \in \{-1,1\}$ *so that* $dist(f, \sigma\chi_{\{i\}}) \leq O(\delta)$*.*

Proof. We may assume that $\delta \leq \frac{1}{C}$ for some large universal constant $C > 0$; otherwise the claim would be trivially true. Let $g(x) := f^{=1}(x) = \sum_{i=1}^n \hat{f}(\{i\})\chi_{\{i\}}(x)$ be the linear part of f. It will be useful to study the function g^2 which is a quadratic function of the form

$$g(x)^2 = \Big(\sum_{i=1}^n \hat{f}(\{i\})\chi_{\{i\}}(x)\Big)^2 = \sum_{i=1}^n \hat{f}(\{i\})^2 + \sum_{i=1}^n \sum_{j \neq j} \hat{f}(\{i\})\hat{f}(\{j\})\chi_{\{i,j\}}(x)$$

Note that by assumption, we have $\mathbb{E}_{x\sim\{-1,1\}^n}[g(x)^2] = W^1[f] = 1-\delta$. It will be crucial to prove that the variance of g^2 is small[1].

Claim I. *One has Var*$[g^2] \leq O(\delta)$*.*

Proof of Claim I. Since $\deg(g^2) \leq 2$ we know by the Bonami Lemma (Theorem 5.4) that g^2 is 9^2-reasonable. As $\mathbb{E}_{x\sim\{-1,1\}^n}[g(x)^2] = 1-\delta$ we can use Prop 5.3 to obtain that $\Pr\big[|g(x)^2 - (1-\delta)| \geq \frac{1}{2}\sqrt{\mathrm{Var}[g^2]}\big] \geq \Omega(1)$. For the sake of contradiction, let us assume that $\frac{1}{2}\sqrt{\mathrm{Var}[g^2]} \geq \delta + C\sqrt{\delta}$ (since otherwise $\mathrm{Var}[g^2] \leq O(\delta)$ and we are done). Then $\Pr_{x\sim\{-1,1\}^n}[|g(x)^2 - 1| \geq C\sqrt{\delta}] \geq \Omega(1)$. Since $|z^2-1| \leq 4\big||z|-1\big|$ for $-2 \leq z \leq 2$, this implies that $\Pr_{x\sim\{-1,1\}^n}[4||g(x)|-1| \geq C\sqrt{\delta}] \geq \Omega(1)$. Then

$$\begin{aligned}
\delta &= \sum_{S\subseteq[n]} \widehat{(g-f)}(S)^2 \\
&= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\big[\big(g(x)-f(x)\big)^2\big] \\
&\geq \Omega(1)\cdot \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\Big[\underbrace{\big(g(x)-f(x)\big)^2}_{\geq (C\sqrt{\delta}/4)^2} \mid \big||g(x)|-1\big| \geq \frac{C}{4}\sqrt{\delta}\Big] \geq \Omega(C^2\delta)
\end{aligned}$$

Choosing C large enough results in a contradiction. ☐

---

[1] Which shouldn't be surprising as g as close to f and $f^2 = 1$ is constant.

Now inspecting the Fourier representation of the variance of g^2 we see that

$$\begin{aligned}\Omega(\delta) \geq \mathrm{Var}[g^2] &\overset{\text{Lem } 1.39}{=} \sum_{|S|=2} \widehat{g^2}(S)^2 \qquad (*) \\ &= \sum_{i=1}^{n} \sum_{j\neq i} \hat{f}(\{i\})^2 \hat{f}(\{j\})^2 \\ &= \underbrace{\Big(\sum_{i=1}^{n} \hat{f}(\{i\})^2\Big)}_{=1-\delta}^2 - \sum_{i=1}^{n} \hat{f}(\{i\})^4 \\ &\geq \underbrace{(1-\delta)^2}_{\geq 1-2\delta} - \max\big\{\hat{f}(\{i\})^2 : i \in [n]\big\} \cdot \underbrace{\sum_{i=1}^{n} \hat{f}(\{i\})^2}_{\leq 1} \\ &\geq 1 - 2\delta - \hat{f}(\{i^*\})^2\end{aligned}$$

where i^* is the index attaining the maximum. Then rearranging $(*)$ gives $\hat{f}(\{i^*\})^2 \geq 1-\Theta(\delta)$. Let $\sigma \in \{-1,1\}$ be the sign with $\sigma \hat{f}(\{i^*\}) \geq 1-\Theta(\delta)$. Recalling the relation between distance and inner product from Chapter 2 we then conclude that

$$\mathrm{dist}(f, \sigma\chi_{\{i^*\}}) = \frac{1}{2} \cdot \big(1 - \underbrace{\langle f, \sigma\chi_{\{i^*\}}\rangle_E}_{\geq 1-\Theta(\delta)}\big) \leq O(\delta)$$

□

The FKN Theorem also has a probabilistic variant that does not rely on boolean functions. Recall that a random variable X is *symmetric* if X has the same distribution as $-X$.

Theorem 5.6 ([JOW15]). *Let $X := X_1 + \ldots + X_n$ where $X_1, \ldots, X_n \in \mathbb{R}$ are independent random variables with finite variance. Then there is an index $k \in [n]$ so that*

$$\mathit{Var}[X - X_k] \leq 6 \cdot \mathit{Var}[|X|]$$

5.3 The KKL Theorem

In this section we discuss an important application of hypercontractivity to analyze the *influence* of boolean functions that we introduced in Section 1.8. Recall that for a function $f : \{-1,1\}^n \to \{-1,1\}$, the influence of the ith coordinate is the probability that flipping the ith bit changes the value, i.e.

$$\mathrm{Inf}_i[f] = \Pr_{x\sim\{-1,1\}^n}\big[f(x) \neq f(x^{\oplus i})\big] \overset{\text{Thm } 1.32.(i)}{=} \sum_{S\subseteq[n]: i\in S} \hat{f}(S)^2$$

A function would have $\mathrm{Inf}_i[f] = 0$ for all i if f is constant, so let us focus on functions that are balanced (i.e. $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)] = 0$) or almost balanced (i.e. $\mathrm{Var}[f] = \Theta(1)$). Clearly $0 \leq \mathrm{Inf}_i[f] \leq 1$, but how small can the influence of coordinates actually be? We can estimate that the sum of the influences of a balanced function (i.e. $\hat{f}(\emptyset) = 0$) is

$$I[f] = \sum_{i=1}^{n} \mathrm{Inf}_i[f] \overset{\text{Thm 1.32.}(ii)}{=} \underbrace{\sum_{\emptyset \subset S \subseteq [n]} \hat{f}(S)^2}_{=1-\hat{f}(\emptyset)^2} \cdot \underbrace{|S|}_{\geq 1} \geq 1 - \underbrace{\hat{f}(\emptyset)^2}_{=0} = 1$$

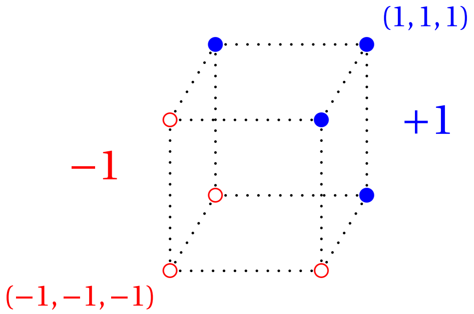
and so there has to be some coordinate i with $\mathrm{Inf}_i[f] \geq \frac{1}{n}$. Next, we discuss two non-trivial constructions and analyze their influence.

5.3.1 The Majority Function

Consider an odd n and consider the *majority function* $f : \{-1,1\}^n \to \{-1,1\}$ with

$$f(x) := \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i > 0 \\ -1 & \text{if } \sum_{i=1}^n x_i < 0 \end{cases}$$

The function is symmetric, hence the influence of all coordinates must be the same and it suffices to determine $\mathrm{Inf}_1[f]$. Let us draw $x_2, \ldots, x_n \sim \{-1,1\}$ at random. Then the outcome of f depends on the first coordinate if and only if $\sum_{i=2}^n x_i = 0$. It is a well known fact in probability that $\Pr_{x_2,\ldots,x_n\sim\{-1,1\}}[\sum_{i=2}^n x_i = 0] = \Theta(\frac{1}{\sqrt{n}})$ and so $\mathrm{Inf}_i[f] = \Theta(\frac{1}{\sqrt{n}})$ for all n.



majority function for $n = 3$

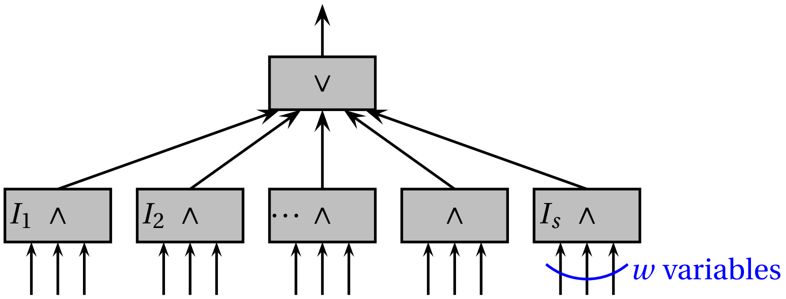
5.3.2 The Tribes Function

Next, we discuss a more complex function. We fix integers $s, w \in \mathbb{N}$ and set $n := s \cdot w$. We partition the coordinates as $[n] = I_1 \dot\cup \ldots \dot\cup I_s$ with $|I_j| = w$ for $j = 1, \ldots, s$.

Then we define the function $\text{Tribes}_{w,s} : \{-1,1\}^n \to \{-1,1\}$ by

$$\text{Tribes}_{w,s}(x) := \begin{cases} -1 & \text{if } \exists j \in [s] : x_{I_j} = (-1,\dots,-1) \\ 1 & \text{otherwise.} \end{cases}$$

One can imagine that one has s many tribes of size w each and $\text{Tribes}_{w,s}(x)$ is a voting function that rejects if at least one tribe unanimously rejects. If one sets $-1 \equiv$ `TRUE` and $1 \equiv$ `FALSE` then $\text{Tribes}_{w,s}$ corresponds to a DNF of s clauses containing w many variables each:



We observe that

$$\Pr_{x\sim\{-1,1\}^n}[\text{Tribes}_{w,s}(x) = 1] = \prod_{j=1}^{s} \Pr[x_{I_j} \neq (-1,\dots,-1)] = (1-2^{-w})^s$$

We are interested in the parameter regime where this function is approximately balanced and setting $s := 2^w$ gives[2] $(1-2^{-w})^s \approx \frac{1}{e}$. Now we can prove that using this choice of parameters, every variable has very low influence.

Lemma 5.7. *For $w \in \mathbb{N}$, set $s := 2^w$ and $n := sw$. Then $Var[\text{Tribes}_{w,s}] = \Theta(1)$ and $Inf_i[\text{Tribes}_{w,s}] = \Theta(\frac{\ln(n)}{n})$ for all $i = 1,\dots,n$.*

Proof. First, from $n = sw = w2^w$ we can get that $2^w = \Theta(\frac{n}{\log(n)})$. By symmetry all coordinates have the same influence, so consider coordinate 1 and assume $1 \in I_1$. If we draw $x_2,\dots,x_n \sim \{-1,1\}$, then $f(x)$ depends on x_1 if and only if both of the following is satisfied:

(A) One has $x_i = -1$ for all $i \in I_1 \setminus \{1\}$.

(B) One has $x_{I_j} \neq (-1,\dots,-1)$ for all $j \in \{2,\dots,s\}$

---

[2] One could also choose s more careful to get a probability very close to 1/2 but this choice will suffice for us.

The probability of this happening is then

$$\mathrm{Inf}_1[\mathrm{Tribes}_{w,s}] = \underbrace{2^{-(w-1)}}_{=\Pr[(A)]} \cdot \underbrace{(1-2^{-w})^{s-1}}_{=\Pr[(B)]=\Theta(1)} = \Theta(2^{-w}) = \Theta\left(\frac{\log(n)}{n}\right)$$

$\square$

This construction gives a function whose maximum influence is within a $\Theta(\log(n))$ factor from the trivial lower bound of $\frac{1}{n}$. In the remainder of this section, we close the gap.

5.3.3 Proof of the KKL Theorem

In this section, we will prove the Kahn-Kalai-Linial (KKL) Theorem which in particular says that any balanced function $f : \{-1,1\}^n \to \{-1,1\}$ must have a coordinate i with $\mathrm{Inf}_i[f] \geq \Omega(\frac{\log n}{n})$, matching the influence of the tribes function. For this part, we will follow the exposition by Minzer [Min21].

First, consider a "partial boolean" function $f : \{-1,1\}^n \to \{-1,0,1\}$ and let $\alpha := \Pr_{x\sim\{-1,1\}^n}[|f(x)| = 1]$. Note that the total Fourier weight of such a function is simply $\sum_{S\subseteq[n]} \hat{f}(S)^2 = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)^2] = \alpha$. Surprisingly, if α is small, then only a small fraction of the Fourier weight can be on low levels.

Lemma 5.8. *Let $f : \{-1,1\}^n \to \{-1,0,1\}$ be a function with $\alpha := \Pr_{x\sim\{-1,1\}^n}[|f(x)| = 1]$. Then for any $d \in \mathbb{N}$,*

$$\sum_{|S|\leq d} \hat{f}(S)^2 \leq \sqrt{3}^d \cdot \alpha^{5/4}$$

Proof. Since $|f(x)| \in \{0,1\}$ we have the convinient fact that for any $p \geq 1$ one has $\|f\|_{E,p} = \mathbb{E}_{x\sim\{-1,1\}^n}[|f(x)|^p]^{1/p} = \alpha^{1/p}$. Now consider the function $f^{\leq d} : \{-1,1\}^n \to \mathbb{R}$ with $f^{\leq d}(x) := \sum_{|S|\leq d} \hat{f}(S)\chi_S(x)$ which is the low-degree part of f. Then using Hölder's Inequality (Theorem 1.49) and the Bonami Lemma (Theorem 5.4) we can bound

$$\begin{aligned}
\|f^{\leq d}\|_{E,2}^2 &= \langle f^{\leq d}, f^{\leq d}\rangle_E \\
&\overset{(*)}{=} \langle f^{\leq d}, f\rangle_E \\
&\overset{\text{Hölder}}{\leq} \|f^{\leq d}\|_{E,4} \cdot \|f\|_{E,4/3} \\
&\overset{\text{Bonami}}{\leq} \sqrt{3}^d \underbrace{\|f^{\leq d}\|_{E,2}}_{\leq \|f\|_{E,2}} \cdot \|f\|_{E,4/3} \\
&= \sqrt{3}^d \underbrace{\|f\|_{E,2}}_{=\alpha^{1/2}} \underbrace{\|f\|_{E,4/3}}_{=\alpha^{3/4}} = \sqrt{3}^d \alpha^{5/4}
\end{aligned}$$

In $(*)$ we use that we could write $f = f^{\leq d} + f^{>d}$ with the high degree part $f^{>d}$ and $\langle f^{\leq d}, f^{>d}\rangle_E = 0$ by orthogonality of the character functions. □

Now we prove the following statement which basically is a restatement of the KKL Theorem.

Proposition 5.9. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a function with $I[f] \leq K \cdot \mathrm{Var}[f]$ for $K \geq 1$. Then there is an index $i \in [n]$ with $\mathrm{Inf}_i[f] \geq e^{-\Theta(K)}$.*

Proof. Just for the sake of simpler notation, we prove this claim for *balanced* functions, i.e. $\mathbb{E}_{x\sim\{-1,1\}}[f(x)] = 0$ and so $\mathrm{Var}[f] = 1$ — the mechanics of the general proof would be the same. Then the assumption says that $I[f] \leq K$ and we need to find a coordinate i with $\mathrm{Inf}_i[f] \geq e^{-\Theta(K)}$. Note that if $K \geq \Omega(\log n)$ then this statement is dominated by using the bound of $\mathbb{E}_{i\sim[n]}[\mathrm{Inf}_i[f]] = \frac{I[f]}{n}$. So one should think of K as a small quantity between $\Theta(1)$ and $\Theta(\log n)$.

We prove the claim by contradiction and assume that for all $i \in [n]$ one has $\mathrm{Inf}_i[f] \leq \alpha := e^{-CK}$ where we choose $C > 0$ large enough. For each coordinate $i \in [n]$ we abbreviate the derivative by $F_i(x) := D_i f(x) = \frac{1}{2}\cdot\big(f(x^{i\mapsto 1}) - f(x^{i\mapsto -1})\big)$. Since $f(x) \in \{-1,1\}$ we have $F_i(x) \in \{-1,0,1\}$. Then $\Pr_{x\sim\{-1,1\}^n}[|F_i(x)| = 1] = \mathrm{Inf}_i[f]$ and so Lemma 5.8 we can upper bound the low-degree Fourier weight involving coordinate i by

$$\sum_{|S|\leq d+1, i\in S} \hat{f}(S)^2 \overset{\text{Prop } 1.28}{=} \sum_{|S|\leq d} \hat{F}_i(S)^2 \overset{\text{Lem } 5.8}{\leq} \sqrt{3}^d \mathrm{Inf}_i[f]^{5/4} \tag{5.1}$$

Summing over all coordinates we can upper bound the low-degree Fourier weight by

$$\sum_{|S|\leq d+1} \hat{f}(S)^2 \leq \sum_{i=1}^n \sum_{|S|\leq d+1: i\in S} \hat{f}(S)^2 \overset{(5.1)}{\leq} \sqrt{3}^d \sum_{i=1}^n \mathrm{Inf}_i[f]^{5/4} \leq \sqrt{3}^d \alpha^{1/4} I[f] \tag{5.2}$$

On the other hand, the high degree Fourier weight can also be bounded by

$$\sum_{|S|>d+1} \hat{f}(S)^2 \leq \sum_{|S|>d+1} \underbrace{\frac{|S|}{d+1}}_{\geq 1} \hat{f}(S)^2 \overset{\text{Thm } 1.32}{\leq} \frac{I[f]}{d+1} \tag{5.3}$$

Combining both gives

$$\begin{aligned}
1 &= \sum_{S\subseteq[n]} \hat{f}(S)^2 \\
&\overset{(5.2)+(5.3)}{\leq} \left(\sqrt{3}^d \alpha^{1/4} + \frac{1}{d+1}\right) \cdot \underbrace{I[f]}_{\leq K} \\
&\leq \left(\underbrace{\sqrt{3}^{2K} \cdot e^{-CK/4}}_{\leq e^{-K} \text{ for } C\geq 12} + \frac{1}{2K+1}\right) \cdot K \\
&\overset{C:=12}{\leq} \left(e^{-K} + \frac{1}{2K+1}\right) \cdot K < 1
\end{aligned}$$

where we make the choice of $d := 2K$ and $C := 12$. This is a contradiction which then proves the claim. □

Finally we prove the main result of this section:

Theorem 5.10 (Kahn-Kalai-Linial (KKL) Theorem)**.** *For any function* $f : \{-1,1\}^n \to \{-1,1\}$ *there is a coordinate* $i \in [n]$ *with*

$$\mathit{Inf}_i[f] \geq \Omega\left(\frac{\log(n)}{n} \cdot \mathit{Var}[f]\right)$$

Proof. Again, let $\mathrm{Var}[f] = 1$ for simplicity. If $I[f] \geq c \log n$ for some constant $c > 0$, then $\mathbb{E}_{i\sim[n]}[\mathrm{Inf}_i[f]] = \frac{I[f]}{n} \geq \frac{c\log(n)}{n}$ and we are done. On the other hand, if $I[f] \leq c\log n$, then by Prop 5.9 there is a coordinate i with $\mathrm{Inf}_i[i] \geq e^{-\Theta(c\log(n))} \geq n^{-0.1} \geq \frac{n}{\log(n)}$ for c small enough. □

5.4 Introduction to hypercontractivity

We again abbreviate $V_n := \{f \mid f : \{-1,1\}^n \to \mathbb{R}\}$ as the vector space of all functions on the n-dimensional hypercube. We make a few definitions:

Definition 5.11. For a (linear) operator $M : V_n \to V_n$ and $p, q \in [1,\infty)$, we define the *p-to-q operator norm* as

$$\|M\|_{p\to q} := \sup_{f\in V_n} \frac{\|Mf\|_{E,q}}{\|f\|_{E,p}}$$

We call M a *contraction from* $\|\cdot\|_{E,p}$ *to* $\|\cdot\|_{E,q}$ if $\|M\|_{p\to q} \leq 1$, i.e. if

$$\|Mf\|_{E,q} \leq \|f\|_{E,p} \quad \forall f \in V_n$$

If $1 \leq p < q < \infty$ and $\|M\|_{p\to q} \leq 1$ then M is called *hypercontractive*.

Recall that by Jensen's inequality, for $1 \leq p < q < \infty$ one has $\|f\|_{E,p} \leq \|f\|_{E,q}$ but in general this inequality is strict. So in order for an operator M to be hypercontractive it must shrink the length of f enough so that the length decreases even if measured in the stricter norm $\|\cdot\|_{E,q}$ that punishes peaks more than $\|\cdot\|_{E,p}$ does.

The only operator that we will be considering for this purpose will be the noise operator T_ρ from Section 1.7. Recall that for $-1 \leq \rho \leq 1$ and $x \in \{-1,1\}^n$ we write $y \sim N_\rho(x)$ as the distribution over $y \in \{-1,1\}^n$ with

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1}{2} + \frac{\rho}{2} \\ -x_i & \text{with probability } \frac{1}{2} - \frac{\rho}{2} \end{cases}$$

independently for all coordinates $i \in [n]$. Morever we define $T_\rho : V_n \to V_n$ as the linear operator that maps a function $f \in V_n$ to $T_\rho f \in V_n$ with

$$T_\rho f(x) = \mathop{\mathbb{E}}_{y \sim N_\rho(x)}[f(y)]$$

Recall that $(T_\rho f)(x) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \cdot \chi_S(x)$, so the operator indeed shrinks all Fourier coefficients — but it does not do that at the same rate and it is not obvious what the effect should be on various $\|\cdot\|_{E,p}$-norms. To warm up, we give a hypercontractivity result that can be proven very similar to Bonami's Lemma (Theorem 5.4). In fact, if f had all Fourier weight on the same level k, then $T_\rho f = \rho^k f$ and by Bonami's Lemma (Theorem 5.4.(ii), f is 9^k-reasonable so that $\|T_\rho f\|_{E,4} = \rho^k \|f\|_{E,4} \leq \rho^k \sqrt{3}^k \|f\|_{E,2}$ implying that $\rho = \frac{1}{\sqrt{3}}$ suffices as noise factor.

Theorem 5.12 ((2,4)-Hypercontractivity Theorem)**.** *For any $f : \{-1,1\}^n \to \mathbb{R}$ one has*

$$\|T_{1/\sqrt{3}} f\|_{E,4} \leq \|f\|_{E,2}$$

Proof. We abbreviate $\rho := \frac{1}{\sqrt{3}}$ from now on. We will prove by induction over n that

$$\mathop{\mathbb{E}}_{x \sim \{-1,1\}^n}\left[T_\rho f(x)^4\right] \leq \mathop{\mathbb{E}}_{x \sim \{-1,1\}^n}\left[f(x)^2\right]^2 \tag{5.4}$$

The claim is true with equality for $n = 0$ when the function f is constant, so suppose $n \geq 1$. We write $x = (\bar{x}, x_n)$ with $\bar{x} \in \{-1,1\}^{n-1}$ and $x_n \in \{-1,1\}$. Pulling out the variable x_n from all terms of f gives $f(x) = x_n g(\bar{x}) + h(\bar{x})$ for two functions $g, h : \{-1,1\}^{n-1} \to \mathbb{R}$.

Then

$$(T_\rho f)(x) = \mathop{\mathbb{E}}_{\bar{y} \sim N_\rho(\bar{x})}\Big[\underbrace{\mathop{\mathbb{E}}_{y_n \sim N_\rho(x_n)}[y_n]}_{=\rho x_n} g(\bar{y}) + h(\bar{y})\Big] = \rho x_n \cdot T_\rho g(\bar{x}) + T_\rho h(\bar{x})$$

Now let $x \sim \{-1,1\}^n$ be uniform at random. We first verify that the right hand side of (5.4) is

$$\begin{aligned}
\mathbb{E}\left[f(x)^2\right]^2 &= \mathbb{E}\left[\left(x_n g(\bar{x}) + h(\bar{x})\right)^2\right]^2 \qquad (*) \\
&= \Big(\underbrace{\mathbb{E}[x_n^2]}_{=1}\mathbb{E}\left[g(\bar{x})^2\right] + 2\underbrace{\mathbb{E}[x_n]}_{=0}\mathbb{E}\left[g(\bar{x})h(\bar{x})\right] + \mathbb{E}\left[h(\bar{x})^2\right]\Big)^2 \\
&= \Big(\mathbb{E}\left[g(\bar{x})^2\right] + \mathbb{E}\left[h(\bar{x})^2\right]\Big)^2
\end{aligned}$$

On the other hand, the left hand side of (5.4) is

$$\begin{aligned}
\mathbb{E}\left[\left(T_\rho f(x)\right)^4\right] &= \mathbb{E}\Big[\left(x_n \rho T_\rho g(\bar{x}) + T_\rho h(\bar{x})\right)^4\Big] \\
&\overset{(**)}{=} \underbrace{\rho^4}_{\leq 1}\underbrace{\mathbb{E}[x_n^4]}_{=1}\mathbb{E}\left[\left(T_\rho g(\bar{x})\right)^4\right] + \underbrace{6\rho^2}_{=2}\underbrace{\mathbb{E}[x_n^2]}_{=1}\mathbb{E}\left[\left(T_\rho g(\bar{x})\right)^2\left(T_\rho h(\bar{x})\right)^2\right] + \mathbb{E}\left[\left(T_\rho h(\bar{x})\right)^4\right] \\
&\overset{\text{Cauchy-S.}}{\leq} \mathbb{E}\left[\left(T_\rho g(\bar{x})\right)^4\right] + 2\sqrt{\mathbb{E}\left[\left(T_\rho g(\bar{x})\right)^4\right]\mathbb{E}\left[\left(T_\rho h(\bar{x})\right)^4\right]} + \mathbb{E}\left[\left(T_\rho h(\bar{x})\right)^4\right] \\
&\overset{\text{induction}}{\leq} \mathbb{E}\left[g(\bar{x})^2\right]^2 + 2\mathbb{E}\left[g(\bar{x})^2\right]\mathbb{E}\left[h(\bar{x})^2\right] + \mathbb{E}\left[h(\bar{x})^2\right]^2 \\
&= \left(\mathbb{E}[g(\bar{x})^2] + \mathbb{E}[h(\bar{x})^2]\right)^2 \overset{(*)}{=} \mathbb{E}\left[f(x)^2\right]^2
\end{aligned}$$

Here in $(**)$ we drop the odd terms as $\mathbb{E}[x_n] = 0 = \mathbb{E}[x_n^3]$. ☐

5.5 The General Hypercontractivity Theorem

In this section, we will prove a hypercontractivity theorem for general parameters p and q.

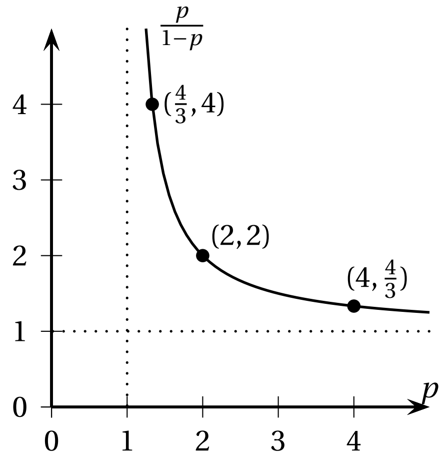
5.5.1 Functional analysis

Our goal will be to prove hypercontractivity for parameters $1 \leq p < q \leq 2$ and then transfer the result to the parameter ranges $p < 2 < q$ and $2 \leq p < q$. For that transfer we need to review a few facts from functional analysis. For convinience we restate Hölder's inequality (see Theorem 1.49) specialized for functions on the hypercube:

Theorem 5.13 (Hölder's Inequality for functions on $\{-1,1\}^n$)**.** *Let $p, p' \geq 1$ be a pair with $\frac{1}{p} + \frac{1}{p'} = 1$. Then for any $f, g \in V_n$ one has*

$$|\langle f, g\rangle_E| \leq \|f\|_{E,p} \cdot \|g\|_{E,p'}$$

The numbers (p, p') with $\frac{1}{p} + \frac{1}{p'} = 1$ are also called *conjugate (Hölder) indices*. Note that $p' = \frac{p}{p-1}$ is the conjugate index to p.



For example $(2,2)$ are conjugate pairs and $(1,\infty)$ are. We also require the following fact:

Lemma 5.14. *Let $p, p' \geq 1$ so that $\frac{1}{p} + \frac{1}{p'} = 1$. Then $\|\cdot\|_{E,p}$ is the* dual norm *to $\|\cdot\|_{E,p'}$, i.e. for all $f \in V_n$,*[3]

$$\|f\|_{E,p} = \sup_{g \in V_n : \|g\|_{E,p'} = 1} \langle g, f \rangle_E$$

We can rephrase Lemma 5.14 as follows: fix any conjugate pair (p, p') and any $f \in V_n$. If we let $g \in V_n$ with $\|g\|_{E,p'} = 1$ denote the function attaining the maximum in Lemma 5.14, then

$$\langle f, g \rangle_E = \|f\|_{E,p} \cdot \|g\|_{E,p'}$$

In other words, each $f \in V_n$ has a *dual element* $g \in V_n$ that satisfies Hölder's Inequality with equality.

Proposition 5.15. *Let $1 \leq p \leq q < \infty$ and let $p', q' > 1$ be their conjugate Hölder indices, i.e. $\frac{1}{p} + \frac{1}{p'} = 1$ and $\frac{1}{q} + \frac{1}{q'} = 1$. Then for any fixed $0 \leq \rho \leq 1$ and $C > 0$ the following is equivalent:*

(A) One has $\|T_\rho f\|_{E,q} \leq C\|f\|_{E,p}$ for all $f \in V_n$.

(B) One has $\|T_\rho f\|_{E,p'} \leq C\|f\|_{E,q'}$ for all $f \in V_n$.

---

[3] By compactness of V_n, the supremum is always attained. However it seems more common in the literature to use sup instead of max in this context.

Proof. By symmetry it suffices to prove that $(A) \Rightarrow (B)$. First we observe that the linear operator T_ρ is *self-adjoint*, i.e. for any functions $f, g \in V_n$ one has

$$\langle T_\rho f, g\rangle_E = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S)\hat{g}(S) = \langle f, T_\rho g\rangle_E$$

using for both equations Plancharel's Theorem (Theorem 1.5) and Prop 1.23. Then

$$\begin{aligned} \|T_\rho f\|_{E,p'} &\overset{\text{Lem 5.14}}{=} \sup_{\|g\|_{E,p}=1} \langle g, T_\rho f\rangle_E \\ &\overset{T_\rho \text{ self-adj.}}{=} \sup_{\|g\|_{E,p}=1} \langle T_\rho g, f\rangle_E \overset{\text{Thm 5.13}}{\leq} \sup_{\|g\|_{E,p}=1} \underbrace{\|T_\rho g\|_q}_{\leq C \text{ by } (A)} \|f\|_{q'} \leq C\|f\|_{q'} \end{aligned}$$

□

5.5.2 Hypercontractivity for $n = 1$

Next, we will show prove hypercontractivity for 1-dimensional random variables. While this sounds modestly exciting, this is where much of the work needs to be done. We make the following crucial definition:

Definition 5.16. Let $1 \leq p \leq q \leq \infty$ and $0 \leq \rho < 1$. Let X be a real-valued random variable with $\mathbb{E}[|X|^q] < \infty$. Then X is called (p, q, ρ)*-hypercontractive* if

$$\mathbb{E}[|a + \rho b X|^q]^{1/q} \leq \mathbb{E}[|a + bX|^p]^{1/p} \quad \forall a, b \in \mathbb{R}$$

Hypercontractive random variables satisfy a range of nice properties (we leave the proof as homework).

Proposition 5.17 (Properties of (p, q, ρ)-hypercontractivity)**.** *Let X and Y be independent random variables that are (p, q, ρ)-hypercontractive.*

(i) One has $\mathbb{E}[X] = 0$.

(ii) For any constant $c \in \mathbb{R}$, cX is (p, q, ρ)-hypercontractive.

(iii) X is (p, q, ρ')-hypercontractive for all $0 \leq \rho' \leq \rho$.

(iv) The sum $X + Y$ is (p, q, ρ)-hypercontractive.

Lemma 5.18 (Two-Point Inequality)**.** *Let $1 \leq p < q \leq \infty$ and let $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$. Then*

(i) *The uniform random bit* $X \sim \{-1,1\}$ *is* (p,q,ρ)*-hypercontractive.*

(ii) *For* $f \in V_1$ *one has* $\|T_\rho f\|_{E,q} \le \|f\|_{E,p}$*.*

Proof. First we argue that for any given triple (p,q,ρ), (i) and (ii) are equivalent. In fact, any function $f \in V_1$ is of the form $f(X) = \hat{f}(\emptyset) + X \cdot \hat{f}(\{1\})$ while $T_\rho f(X) = \hat{f}(\emptyset) + \rho X \cdot \hat{f}(\{1\})$. Then $\|T_\rho f\|_{E,q} \le \|f\|_{E,p}$ is equivalent to

$$\mathop{\mathbb{E}}_{X\sim\{-1,1\}}[|\hat{f}(\emptyset) + \rho X \cdot \hat{f}(\{1\})|^q]^{1/q} \le \mathop{\mathbb{E}}_{X\sim\{-1,1\}}[|\hat{f}(\emptyset) + X \cdot \hat{f}(\{1\})|^p]^{1/p}$$

which indeed is the statement of (i) and obviously the reduction works the other way around.

Now fix a triple (p,q,ρ) where by Prop 5.17.(iii) we may assume that $\rho = \sqrt{\frac{p-1}{q-1}}$. We consider three regimes of parameters where we will prove either (i) or (ii) depending which view is more convinient.

- *Case* $1 \le p < q \le 2$. First we make the observation that it suffices to prove the inequality $\|T_\rho f\|_{E,q} \le \|f\|_{E,p}$ for non-negative functions f since replacing f by the function $F(x) := |f(x)|$ would leave the right hand side invariant while it can only increase the left hand side. Now we switch to the view of (i). By scaling the pair (a,b) from Def 5.16 it suffices to prove that for any $\varepsilon \in \mathbb{R}$ one has

  $$\mathop{\mathbb{E}}_{X\sim\{-1,1\}}[|1+\rho\varepsilon X|^q]^{1/q} \le \mathop{\mathbb{E}}_{X\sim\{-1,1\}}[|1+\varepsilon X|^p]^{1/p} \quad (*)$$

  By the non-negativity assumption we may assume $|\varepsilon| < 1$[4]. Then we continue

  $$\begin{aligned}(*) \quad &\Leftrightarrow \quad \left(\frac{1}{2}(1+\rho\varepsilon)^q + \frac{1}{2}(1-\rho\varepsilon)\right)^{p/q} \le \frac{1}{2}(1+\varepsilon)^p + \frac{1}{2}(1-\varepsilon)^p\\ &\overset{(**)}{\Leftrightarrow} \quad \left(1 + \sum_{k=1}^{\infty}\binom{q}{2k}\rho^{2k}\varepsilon^{2k}\right)^{p/q} \le 1 + \sum_{k=1}^{\infty}\binom{p}{2k}\varepsilon^{2k}\\ &\overset{(1+t)^\theta \le 1+\theta t \forall t\ge 0, 0\le\theta\le 1}{\Leftarrow} \quad 1 + \sum_{k=1}^{\infty}\frac{p}{q}\binom{q}{2k}\rho^{2k}\varepsilon^{2k} \le 1 + \sum_{k=1}^{k}\binom{p}{2k}\varepsilon^{2k}\end{aligned}$$

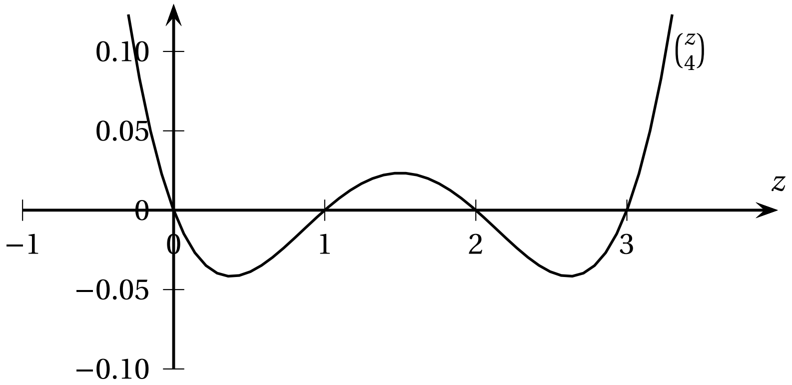
  In $(**)$ we apply the Generalized Binomial Theorem (Theorem 1.43) on both sides separately and use that the odd terms cancel while the even terms are identical. We note that

  $$\binom{p}{2k} = \frac{p(p-1)(p-2)(p-3)\cdot(p-(2k-1))}{(2k)!} \ge 0$$

---

[4] We skip the case $|\varepsilon| = 1$ which follows by continuity.

because each consecutive pair of factors is non-negative, i.e. $p(p-1) \geq 0$, $(p-2)(p-3) \geq 0$. This crucially uses that $1 \leq p \leq 2$. By the same argument $\binom{q}{2k} \geq 0$.



Finally by an elementary but tedious calculation one can do a term-wise comparison (we refer to [O'D21], page 287 for details).

Claim. *For* $1 \leq p < q \leq 2$, $k \in \mathbb{N}$ *and* $\rho = \sqrt{\frac{p-1}{q-1}}$ *one has* $\frac{p}{q}\binom{q}{2k}\rho^{2k} \leq \binom{p}{2k}$.

Proof of Claim. We write

$$\begin{aligned}
& \frac{p}{q}\binom{q}{2k}\rho^{2k} \leq \binom{p}{2k} \\
\Leftrightarrow \quad & \frac{p}{q}\left(\frac{p-1}{q-1}\right)^k \frac{1}{(2k)!}\prod_{i=0}^{2k-1}(q-i) \leq \frac{1}{(2k)!}\prod_{i=0}^{2k-1}(p-i) \\
\Leftrightarrow \quad & \prod_{i=2}^{2k-1}\frac{|q-i|}{\sqrt{q-1}} \leq \prod_{i=2}^{2k-1}\frac{|p-i|}{\sqrt{p-1}} \\
\Leftarrow \quad & \frac{i-q}{\sqrt{q-1}} \leq \frac{i-p}{\sqrt{p-1}} \qquad \forall i \geq 2
\end{aligned}$$

The last line is true because for all $i \geq 2$, the function $g(z) := \frac{i-z}{\sqrt{z-1}}$ is positive and decreasing for $z \geq 1$ as $g'(z) = \frac{z-i}{2(z-1)^{3/2}} - \frac{1}{\sqrt{z-1}} = -\frac{z+i-2}{2(z-1)^{3/2}} \leq 0$.

- *Case* $2 \leq p < q$. Let p' and q' be the conjugate indices of p and q, i.e. $\frac{1}{p} + \frac{1}{p'} = 1$ and $\frac{1}{q} + \frac{1}{q'} = 1$. Note that $1 \leq q' < p' \leq 2$. Moreover one has $\frac{p-1}{q-1} = \frac{q'-1}{p'-1}$ which means the parameter ρ for the pairs (p,q) and (q',p') is the same. From the first case we know that $\|T_\rho f\|_{E,p'} \leq \|f\|_{E,q'}$ for all $f \in V_1$ which by Prop 5.15 implies that $\|T_\rho f\|_q \leq \|f\|_{E,p}$ for all $f \in V_1$.

- *Case* $p < 2 < q$. Set $\rho_1 := \sqrt{\frac{2-1}{q-1}}$ and $\rho_2 := \sqrt{\frac{p-1}{2-1}}$ and note that $\rho = \rho_1 \cdot \rho_2$. Then

  $$\|T_\rho f\|_{E,q} = \|T_{\rho_1} T_{\rho_2} f\|_{E,q} \overset{(2,q,\rho_1)\text{-hypercon.}}{\leq} \|T_{\rho_2} f\|_{E,2} \overset{(p,2,\rho_2)\text{-hypercon}}{\leq} \|f\|_{E,p}$$

  making use if the previous cases.

$\square$

5.5.3 Lifting to general dimension

Next, we want to prove hypercontractivity for functions in general dimension n. For that purpose it will be more useful to prove a more general result first which is more friendly towards a proof by induction.

Theorem 5.19 (Two-Function Hypercontractivity Theorem)**.** *Let* $p, q \geq 1$ *and* $0 \leq \rho \leq \sqrt{(p-1)(q-1)} \leq 1$. *Then for any* $f, g \in \{-1,1\}^n \to \mathbb{R}$ *one has*

$$\mathop{\mathbb{E}}_{\substack{x \sim \{-1,1\}^n, \\ y \sim N_\rho(x)}} [f(x) \cdot g(y)] = \langle f, T_\rho g \rangle_E \leq \|f\|_{E,p} \|g\|_{E,q}$$

Proof. We prove the claim by induction over n. First we begin with $n = 1$, which is actually the hard case, but fortunately we have done all the tedious work already in Lemma 5.18. Let $p' \geq 2$ be the conjugate index to p, i.e. $\frac{1}{p} + \frac{1}{p'} = 1$. Note that $p - 1 = \frac{1}{p'-1}$ and by Lemma 5.18, the triple (q, p', ρ) satisfies that $\|T_\rho h\|_{p'} \leq \|h\|_q$ for all $h \in V_1$. We use this to bound

$$\langle f, T_\rho g \rangle_E \overset{\text{Hölder}}{\leq} \|f\|_{E,p} \|T_\rho g\|_{E,p'} \overset{\text{Lemma 5.18}}{\leq} \|f\|_{E,p} \|g\|_{E,q}$$

which completes the induction base case.

Now consider $n \geq 2$. We write $x = (\bar{x}, x_n)$ and $y = (\bar{y}, y_n)$ where $x \sim \{-1,1\}^n$ and $y \sim N_\rho(x)$. Note that (x_n, y_n) is a ρ-correlated pair and $(\bar{x}, \bar{y})$ is also ρ-correlated. We denote $f_{x_n} : \{-1,1\}^{n-1} \to \mathbb{R}$ as the restriction with $f_{x_n}(\bar{x}) = f(x, x_n)$ where the last coordinate has been fixed to the value of x_n. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_{(x,y)} [f(x) \cdot g(y)] &= \mathop{\mathbb{E}}_{(x_n,y_n)} \left[\mathop{\mathbb{E}}_{(\bar{x},\bar{y})} [f_{x_n}(\bar{x}) \cdot g_{y_n}(\bar{y})] \right] \\
&\overset{\text{induction for dim } n-1}{\leq} \mathop{\mathbb{E}}_{(x_n,y_n)} \left[\|f_{x_n}\|_{E,p} \|g_{y_n}\|_{E,q} \right] \\
&\overset{\text{induction for dim 1}}{\leq} \mathop{\mathbb{E}}_{x_n} \left[\|f_{x_n}\|_{E,p}^p \right]^{1/p} \cdot \mathop{\mathbb{E}}_{x_n} \left[\|g_{x_n}\|_{E,q}^q \right]^{1/q} \\
&= \mathop{\mathbb{E}}_{x} \left[|f(x)|^p \right]^{1/p} \cdot \mathop{\mathbb{E}}_{x} \left[|g(x)|^q \right]^{1/q}
\end{aligned}$$

where we apply the inductive hypothesis twice, once for dimension $n-1$ and once for dimension 1. ☐

Finally we can derive the main result of this chapter:

Theorem 5.20 (Hypercontractivity Theorem)**.** *Let* $1 \le p \le q \le \infty$ *and* $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$. *Then for any* $f : \{-1,1\}^n \to \mathbb{R}$ *one has* $\|T_\rho f\|_{E,q} \le \|f\|_{E,p}$.

Proof. Let q' be the conjugate index to q, i.e. $\frac{1}{q} + \frac{1}{q'} = 1$. Again, $q' - 1 = \frac{1}{q-1}$ and so we have $0 \le \rho \le \sqrt{(p-1)(q'-1)}$ as required in Theorem 5.19. Let g with $\|g\|_{E,q'} = 1$ be the dual function to $T_\rho f$ (see Lemma 5.14). Then

$$\|T_\rho f\|_{E,q} = \langle T_\rho f, g\rangle_E \overset{\text{Thm 5.19}}{\le} \|f\|_{E,p} \underbrace{\|g\|_{E,q'}}_{=1} = \|f\|_{E,p}$$

☐

We record a fact that we will prove in the homework:

Corollary 5.21. For any $q \ge 1$ and $f : \{\pm 1\}^n \to \mathbb{R}$, the function $\rho \to \|T_\rho f\|_{E,q}$ is monotonically non-decreasing for $0 \le \rho \le 1$.

5.6 Small-set expansion of the hypercube

Next, we provide a simple but important application of hypercontractivity. First we small result that will be useful more than once, so we keep it general.

Lemma 5.22. *For any* $f : \{-1,1\}^n \to \{-1,0,1\}$ *and* $0 \le \rho \le 1$ *one has* $\langle f, T_\rho f\rangle_E \le \alpha^{2/(1+\rho)}$. *where* $\alpha := \Pr_{x\sim\{-1,1\}^n}[|f(x)| = 1]$.

Proof. For any $p \ge 1$, we have $\|f\|_{E,p} = \mathbb{E}_{x\sim\{-1,1\}^n}[|f(x)|^p]^{1/p} = \alpha^{1/p}$ as $|f(x)| \in \{0,1\}$. We want to apply Theorem 5.19 and we want to pick a parameter p so that[5] $\rho = \sqrt{(p-1)\cdot(p-1)}$ which can be rearranged to $p = 1+\rho$. Then

$$\langle f, T_\rho f\rangle_E \overset{\text{Theorem 5.19}}{\le} \|f\|_{1+\rho} \cdot \|f\|_{1+\rho} = \alpha^{2/(1+\rho)}.$$

☐

Then we can lower bound the probability to exit a set via perturbations:

---

[5] One can of course try general parameters p,q with $\rho = \sqrt{(p-1)(q-1)}$ and try to optimize. But it seems the choice of $p = q$ is already optimal.

Theorem 5.23. *Let $A \subseteq \{-1,1\}^n$ be any subset of the hypercube. Then for $0 \le \rho \le 1$ one has*

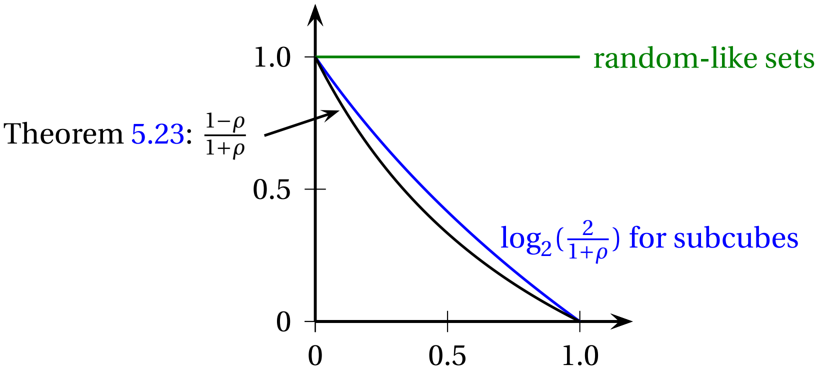
$$\Pr_{x\sim A, y\sim N_\rho(x)}[y \in A] \le \left(\frac{|A|}{2^n}\right)^{(1-\rho)/(1+\rho)}$$

Proof. Let $\alpha := \frac{|A|}{2^n}$ be the volume of the set A. The proof works by analyzing the characteristic function $\mathbf{1}_A : \{-1,1\}^n \to \{0,1\}$ of the set A. Then using conditional probability and Lemma 5.22 we obtain

$$\begin{aligned}\Pr_{x\sim A, y\sim N_\rho(x)}[y \in A] &= \frac{\Pr_{x\sim\{-1,1\}^n, y\sim N_\rho(x)}[\mathbf{1}_A(x)\cdot\mathbf{1}_A(y)]}{\Pr_{x\sim\{-1,1\}^n}[x \in A]} \\ &= \frac{\langle \mathbf{1}_A, T_\rho \mathbf{1}_A\rangle_E}{\alpha} \\ &\overset{\text{Lem } 5.22}{\le} \frac{\alpha^{2/(1+\rho)}}{\alpha} = \alpha^{(1-\rho)/(1+\rho)}\end{aligned}$$

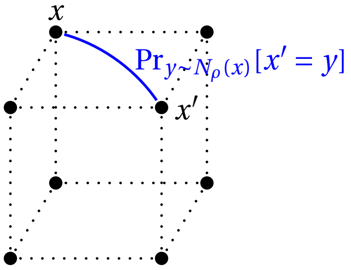
□

The bound proven in Theorem 5.23 appears to not be tight. If we choose $A \subseteq \{-1,1\}^n$ as a $(n-k)$-dimensional subcube with density $\alpha = 2^{-k}$ for $k \in \{0,\dots,n\}$ then $\Pr_{x\sim A, y\sim N_\rho(x)}[y \in A] = \alpha^{\log_2(\frac{2}{1+\rho})}$ for all $0 \le \rho \le 1$. I believe that this is the extremal case.



Now we discuss a more combinatorial application. For $0 \le \rho \le 1$, let the *ρ-noisy hypercube* be the complete weighted undirected graph $G_{n,\rho} := (\{-1,1\}^n, E, w)$ where for each $x, x' \in \{-1,1\}^n$, the edge has a weight of

$$w(x,x') := \Pr_{y\sim N_\rho(x)}[x' = y]$$



We note that the probability to move from x to x' is the same as to move from x' to x and hence we can indeed think of $G_{n,\rho}$ as undirected. For a set $A \subseteq \{\pm 1\}^n$ we define $\delta(A)$ as the edges with exactly one endpoint in A while $E(A)$ are the edges with both endpoints in A. For a subset $F \subseteq E$ of edges we use the shorthand notation $w(F) := \sum_{e\in F} w(e)$. We note that for each node $x \in \{\pm 1\}^n$ we have $w(\delta(x)) = 1$ (since probabilities add up to 1). In particular $w(\delta(A)) \leq |A|$. But for small sets, this bound is almost attained with equality — we also say that $G_{n,\rho}$ is a *small-set expander.*

Lemma 5.24. *Let $G_{n,\rho} = (\{\pm 1\}^n, E, w)$ be the ρ-noisy hypercube. Then for any $A \subseteq \{\pm 1\}^n$ one has*

$$\frac{w(\delta(A))}{|A|} \geq 1 - \left(\frac{|A|}{2^n}\right)^{(1-\rho)/(1+\rho)}$$

Proof. We abbreviate the density of A by $\alpha := \frac{|A|}{2^n}$. By Theorem 5.23 we have

$$\frac{w(E(A))}{|A|} = \Pr_{x\sim A, y\sim N_\rho(x)}[y \in A] \overset{\text{Thm } 5.23}{\leq} \alpha^{\frac{1-\rho}{1+\rho}}$$

Then

$$\frac{w(\delta(A))}{|A|} = 1 - \frac{w(E(A))}{|A|} \geq 1 - \alpha^{\frac{1-\rho}{1+\rho}}$$

□

5.7 Friedgut's Junta Theorem

In this section, we prove *Friedgut's Junta Theorem* which says that any boolean function f with very small total influence $I[f]$ is close to a *junta*, which is a function that only depends on a few coordinates. Before we come to the formal statement and its proof, we make a small detour.

In Section 1.8 we defined the *ρ-stable influence* of a function $f : \{-1, 1\}^n \to \mathbb{R}$ as

$$\mathrm{Inf}_i^{(\rho)}[f] := \mathrm{Stab}_\rho[D_i f] = \langle D_i f, T_\rho D_i f\rangle_E = \sum_{S\subseteq[n]: i\in S} \rho^{|S|-1} \hat{f}(S)^2 \tag{5.5}$$

Recall that for $\rho = 1$, this quantity is simply equal to $\mathrm{Inf}_i[f]$. We can prove that for any function f with $f(x) \in \{-1,1\}$, the ρ-stable influence is a lot smaller than the influence itself.

Proposition 5.25. *Let $f : \{-1,1\}^n \to \{-1,1\}$, $0 \le \rho \le 1$ and $i \in [n]$ one has*

$$Inf_i^{(\rho)}[f] \le Inf_i[f]^{2/(1+\rho)}$$

Proof. We fix the index i and abbreviate the derivative of f in coordinate direction i has $g(x) := D_i f(x) = \frac{1}{2} \cdot \big(f(x^{i\mapsto 1}) - f(x^{i\mapsto -1})\big)$. Since $f(x) \in \{-1,1\}$, we have $g(x) \in \{-1,0,1\}$. Note that $\mathrm{Inf}_i[f] = \Pr_{x\sim\{-1,1\}^n}[|g(x)| = 1]$. Hence we can apply Lemma 5.22 and get

$$\mathrm{Inf}_i^{(\rho)}[f] \overset{(5.5)}{=} \langle g, T_\rho g\rangle_E \overset{\text{Lem 5.22}}{\le} \Pr_{x\sim\{-1,1\}^n}[|g(x)| = 1]^{2/(1+\rho)} = \mathrm{Inf}_i[f]^{2/(1+\rho)}.$$

□

Recall that for two functions $f, g : \{-1,1\}^n \to \{-1,1\}$ we denote their distance as $\mathrm{dist}(f,g) := \Pr_{x\sim\{-1,1\}^n}[f(x) \ne g(x)]$. We will also need a simple rounding argument to make functions $\{-1,1\}$-valued. For $z \in \mathbb{R}$ we define the *sign function* as

$$\mathrm{sign}(z) := \begin{cases} +1 & \text{if } z \ge 0 \\ -1 & \text{if } z < 0 \end{cases}$$

Lemma 5.26. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and $g : \{-1,1\}^n \to \mathbb{R}$. Define $h : \{-1,1\}^n \to \{-1,1\}$ by $h(x) := sign(g(x))$. Then $dist(f,h) \le \|f - g\|_{E,2}^2$.*

Proof. We have

$$\mathrm{dist}(f,h) = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\mathbf{1}_{f(x)\ne\mathrm{sign}(g(x))}\right] \overset{(*)}{\le} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[|f(x) - g(x)|^2\right] = \|f - g\|_{E,2}^2$$

where we use in $(*)$ that $(f(x) \ne \mathrm{sign}(g(x))) \Rightarrow |f(x) - g(x)| \ge 1$. □

Now we make the formal definition that gives the junta theorem its name:

Definition 5.27. A function $f : \{-1,1\}^n \to \mathbb{R}$ is called a *k-junta* if there are coordinates $J \subseteq [n]$ with $|J| \le k$ so that for all $x \in \{-1,1\}^n$, the value $f(x)$ only depends on $(x_i)_{i\in J}$.

Now we can prove the main results of this section. Note that the statement is only non-trivial if $I[f] \le \varepsilon \log(n)$, so one should think of the total influence $I[f]$ as tiny here.

Theorem 5.28 (Friedgut's Junta Theorem)**.** *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and* $0 < \varepsilon \leq 1$. *Then there exists a* $e^{O(I[f]/\varepsilon)}$*-junta* $h : \{-1,1\}^n \to \{-1,1\}$ *with* $\text{dist}(f,h) \leq \varepsilon$.

Proof. For a parameter $\delta > 0$ that we decide later, we denote

$$J := \left\{ i \in [n] \mid \mathrm{Inf}_i[f] \geq \delta \right\}$$

as all the influential coordinates. We define

$$g(x) := \sum_{S \subseteq J} \hat{f}(S) \cdot \chi_S(x)$$

which by construction is a $|J|$-junta, though it is only a function of the form $g : \{-1,1\}^n \to \mathbb{R}$. But the rounded function $h : \{-1,1\}^n \to \{-1,1\}$ with $h(x) := \mathrm{sign}(g(x))$ is still a $|J|$-junta and by Lemma 5.26 we have $\mathrm{dist}(f,h) \leq \|f-g\|_{E,2}^2$. So it suffices to prove that for a suitable choice of parameters one has $\|f-g\|_{E,2}^2 \leq \varepsilon$.

First, set $d := \frac{2I[f]}{\varepsilon}$ and note that similar to the proof of Prop 5.9 one has

$$\sum_{|S|>d} \hat{f}(S)^2 \leq \frac{I[f]}{d} = \frac{\varepsilon}{2}, \tag{5.6}$$

implying that we will be able to ignore the higher order Fourier coefficients.

The main idea of the remaining proof is to analyze the ρ-stable influences for the coordinates outside of J. Here we can make a choice of say $\rho := \frac{1}{3}$. On one hand we can upper bound

$$\sum_{i \notin J} \mathrm{Inf}_i^{(1/3)}[f] \overset{\text{Prop } 5.25}{\leq} \sum_{i \notin J} \mathrm{Inf}_i[f]^{3/2} \leq \sqrt{\delta} \underbrace{\sum_{i \notin J} \mathrm{Inf}_i[f]}_{\leq I[f]} \leq \sqrt{\delta} \cdot I[f] \tag{5.7}$$

On the other hand, using the Fourier representation of the ρ-stable influence, the same quantity can be lower bounded as

$$\begin{aligned}
\sum_{i \notin J} \mathrm{Inf}_i^{(1/3)}[f] &\overset{(5.5)}{=} \sum_{i \notin J} \sum_{S \subseteq [n] : i \in S} (1/3)^{|S|-1} \hat{f}(S)^2 && (5.8)\\
&= \sum_{S \subseteq [n]} |S \cap \bar{J}| \cdot (1/3)^{|S|-1} \hat{f}(S)^2 \\
&\geq \sum_{|S| \leq d \text{ and } |S \cap \bar{J}| \geq 1} \underbrace{|S \cap \bar{J}|}_{\geq 1} \cdot \underbrace{(1/3)^{|S|-1}}_{\geq 3^{-d}} \hat{f}(S)^2 \\
&\geq 3^{-d} \sum_{|S| \leq d \text{ and } |S \cap \bar{J}| \geq 1} \hat{f}(S)^2
\end{aligned}$$

Now the distance between f and g is

$$\|f-g\|_{E,2}^2 \leq \underbrace{\sum_{|S|>d} \hat{f}(S)^2}_{\leq \varepsilon/2 \text{ by (5.6)}} + \underbrace{\sum_{|S|\leq d \text{ and } |S\cap \bar{J}|\geq 1} \hat{f}(S)^2}_{\leq 3^d\sqrt{\delta} I[f] \text{ by (5.7)+(5.8)}} \leq \frac{\varepsilon}{2} + 3^{2I[f]/\varepsilon}\sqrt{\delta} I[f] \overset{!}{\leq} \varepsilon$$

where the last inequality holds if we choose $\delta := e^{-\Theta(I[f]/\varepsilon)}$. Note that $I[f] \geq \sum_{i\in J} \text{Inf}_i[J] \geq \delta |J|$ and so $|J| \leq e^{O(I[f]/\varepsilon)}$ which completes the proof. □

In case that $f : \{-1,1\}^n \to \{-1,1\}$ has degree $d := \deg(f)$ we know that $I[f] \leq d$ and hence by Theorem 5.28 we know that there is a $e^{O(d/\varepsilon)}$-junta that approximates f up to an ε error. However, in the homework we have seen that a boolean degree-d function can only depend on at most $d \cdot 2^d$ many coordinates which supercedes this bound anyway.

5.8 A generalization of the Bonami Lemma

In the Bonami Lemma (Theorem 5.4.(ii)) we have seen that for any degree-k function one has $\|f\|_{E,4} \leq 3^{k/2} \cdot \|f\|_{E,2}$. It will be useful to have a generalization to parameters q other than $q = 4$, for example in order to prove stronger concentration bounds. Rather than proving these generalizations from scratch, we can derive them from hypercontractivity.

Theorem 5.29 (Generalized Bonami Lemma)**.** *For any function $f : \{-1,1\}^n \to \mathbb{R}$ of degree at most k and any $q \geq 2$ one has*

$$\|f\|_{E,q} \leq (q-1)^{k/2} \cdot \|f\|_{E,2}$$

Proof. The original definition of the noise operator T_ρ only makes sense if $-1 \leq \rho \leq 1$. But we could agree to define the operator instead by the identity $T_\rho f(x) = \sum_{S\subseteq[n]} \rho^{|S|} \hat{f}(S)\chi_S(x)$ which makes sense for all $\rho \in \mathbb{R}$. While many of the theorems we have proven for the noise operator only work for $-1 \leq \rho \leq 1$, other properties still hold. For example for all $\rho_1, \rho_2 \in \mathbb{R}$ one has that $T_{\rho_1} T_{\rho_2} f = T_{\rho_1 \cdot \rho_2} f$. Then

$$\begin{aligned} \|f\|_{E,q}^2 &= \|T_{1/\sqrt{q-1}}(T_{\sqrt{q-1}} f)\|_{E,q}^2 \\ &\overset{\text{hypercontr.}}{\leq} \|T_{\sqrt{q-1}} f\|_{E,2}^2 \\ &= \sum_{|S|\leq k} (q-1)^{|S|} \hat{f}(S)^2 \\ &\leq (q-1)^k \|f\|_{E,2}^2 \end{aligned}$$

Here we have used the General Hypercontractivity Theorem (Theorem 5.20) with parameters $(2,q)$, $q \geq 2$, which tells us that for any function g one has $\|T_{1/\sqrt{q-1}} g\|_{E,q} \leq \|g\|_2$. $\square$

We note that the inequality from Theorem 5.29 can be written out to

$$\mathbb{E}[|f(x)|^q] \leq (q-1)^{qk/2} \cdot \mathbb{E}[f(x)^2]^{q/2}$$

where $x \sim \{-1,1\}^n$.

We also prove an inequality for the regime $[1,2]$.

Theorem 5.30 (Generalized Bonami Lemma II)**.** *For any function* $f : \{-1,1\}^n \to \mathbb{R}$ *of degree at most* k *and any* $1 \leq p \leq 2$ *one has*

$$\|f\|_{E,2} \leq (e^{\frac{2}{p}-1})^k \cdot \|f\|_{E,p}$$

Proof. For the sake of simplicity we consider the case of $p = 1$, i.e. we prove that $\|f\|_{E,2} \leq e^k \cdot \|f\|_{E,1}$. We want to compare $\|f\|_{E,2}$ with $\|f\|_{E,1}$ and $\|f\|_{E,2+\varepsilon}$ where we determine $\varepsilon > 0$ later. For that purpose, let $\theta \in (0,1)$ be the unique value so that

$$\frac{1}{2} = \frac{\theta}{1} + \frac{1-\theta}{2+\varepsilon}$$

as required by Littlewood's Inequality (Theorem 1.52). One can easily check that $\theta = \frac{1}{2} \cdot \frac{\varepsilon}{1+\varepsilon}$. Then combining this with the Generalized Bonami Lemma that we just proved, we obtain

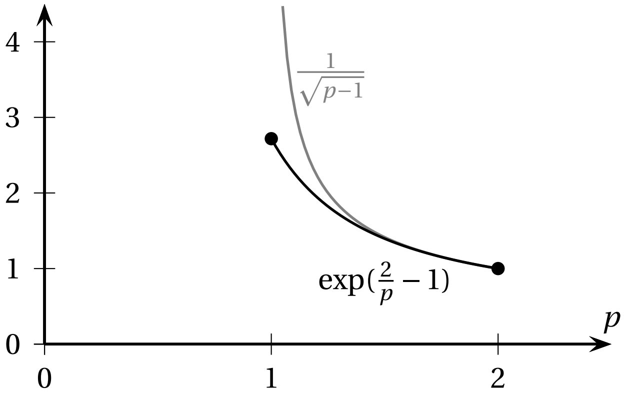
$$\|f\|_{E,2} \overset{\text{Thm } 1.52}{\leq} \|f\|_{E,1}^{\theta} \cdot \|f\|_{2+\varepsilon}^{1-\theta} \overset{\text{Thm } 5.29}{\leq} \|f\|_{E,1}^{\theta} \cdot (1+\varepsilon)^{k(1-\theta)/2} \|f\|_{E,2}^{1-\theta}$$

Then rearranging gives

$$\|f\|_{E,2} \leq \left((1+\varepsilon)^{\frac{1-\theta}{2\theta}}\right)^k \|f\|_{E,1} \overset{\theta=\frac{1}{2}\cdot\frac{\varepsilon}{1+\varepsilon}}{=} \Bigg(\underbrace{(1+\varepsilon)^{\frac{1}{\varepsilon}+\frac{1}{2}}}_{\to e \text{ as } \varepsilon\to 0}\Bigg)^k \|f\|_{E,1} \xrightarrow{\varepsilon\to 0} e^k \|f\|_{E,1}$$

which gives the claim. $\square$

We note that other sources give a base of $\frac{1}{\sqrt{p-1}}$ instead of $\exp(\frac{2}{p}-1)$ but that former factor diverges for $p \to 1$.



5.9 Exponential concentration

We already know from a combination of Lemma 5.2 and the Bonami Lemma (Theorem 5.4) that for every degree-k function $f : \{-1,1\}^n \to \mathbb{R}$ and any $t > 0$,

$$\Pr_{x\sim\{-1,1\}^n}[|f(x)| > t\|f\|_{E,2}] \le \frac{9^k}{t^4}$$

But this only gives an error probability that is *inverse polynomial* in t. For many application it is desirable to have *exponentially* small error bounds. This can be done using the Generalization of Bonami Lemma from Theorem 5.29.

Theorem 5.31. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function of degree at most k. Then for any $t \ge (2e)^{k/2}$ one has*

$$\Pr_{x\sim\{-1,1\}^n}\left[|f(x)| \ge t\|f\|_{E,2}\right] \le \exp\left(-\frac{k}{2e}t^{2/k}\right)$$

Proof. After scaling we may assume that $\|f\|_{E,2} = 1$. Let $q \ge 2$ be a parameter that we determine later. Then for $x \sim \{-1,1\}^n$ one has

$$\begin{aligned}
\Pr[|f(x)| \ge t] &= \Pr[|f(x)|^q \ge t^q] \\
&\overset{\text{Markov}}{\le} \frac{\mathbb{E}[|f(x)|^q]}{t^q} \\
&\overset{\text{Thm 5.29}}{\le} \frac{(q-1)^{qk/2}}{t^q} \cdot \underbrace{\mathbb{E}[f(x)^2]^{q/2}}_{=1} \\
&\le \left(\frac{q^{k/2}}{t}\right)^q \overset{\text{choice of } q}{=} \exp\left(-\frac{k}{2}\cdot \underbrace{t^{2/k}/e}_{=q}\right)
\end{aligned}$$

Here we can see that we should choose q so that $\frac{q^{k/2}}{t} < 1$. We make the choice of $\frac{q^{k/2}}{t} = e^{-k/2}$ which is equivalent to $q = t^{2/k}/e$. Finally we remember that we need $q \geq 2$ for Theorem 5.29 for which we had made the assumption of $t \geq (2e)^{k/2}$. $\square$

Chapter 6

The invariance principle

The goal of this chapter is to prove the *invariance principle* which says that for any low degree multilinear polynomial $F : \mathbb{R}^n \to \mathbb{R}$ without influential coordinates and any "nice" function $\psi : \mathbb{R} \to \mathbb{R}$ one has

$$\mathop{\mathbb{E}}_{X \sim \{-1,1\}^n}[\psi(F(X))] \approx \mathop{\mathbb{E}}_{Y \sim \gamma_n}[\psi(F(Y))]$$

The usefulness of such a statement is that we can prove facts on boolean functions instead for Gaussians where they might be easier to derive. Here γ_n is the n-dimensional (standard) Gaussian distribution with mean $\mathbf{0}$ and covariance matrix I_n. Alternatively we will write $N(\mathbf{0}, \Sigma)$ for the Gaussian distribution with mean $\mathbf{0}$ and covariance matrix Σ; in particular $\gamma_n = N(\mathbf{0}, I_n)$.

6.1 Functions in Gaussian Space

First, we need to leave the realm of functions restricted to the boolean hypercube that we gotten so comfortable with.

Definition 6.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a *multilinear polynomial of degree at most d* if there are coefficients $\alpha_S \in \mathbb{R}$ so that

$$f(x) = \sum_{S \subseteq [n] : |S| \leq d} \alpha_S \cdot \chi_S(x) \quad \forall x \in \mathbb{R}^n$$

Here by a slight abuse of notation we extend $\chi_S(x) = \prod_{i \in S} x_i$ to the whole $\mathbb{R}^n$ and in reverse for a function $f : \mathbb{R}^n \to \mathbb{R}$ we will use notation such as $\hat{f}(S) = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x) \cdot \chi_S(x)]$. With this notation it is clear that the coefficients α_S must be equal to $\hat{f}(S)$ so we can directly write any multilinear polynomial $f : \mathbb{R}^n \to \mathbb{R}$ as

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(x) \quad \forall x \in \mathbb{R}^n$$

If we work with an arbitrary function $f : \mathbb{R}^n \to \mathbb{R}$ then for most operations we need to ensure that integrals are well defined and so we restrict our attention to the class

$$L^2(\mathbb{R}^n, \gamma_n) := \left\{ f : \mathbb{R}^n \to \mathbb{R} \mid f \text{ integrable and } \mathop{\mathbb{E}}_{x \sim \gamma_n}[f(x)^2] < \infty \right\}$$

Note that any multilinear polynomial and also any continuous bounded function is anyway contained in $L^2(\mathbb{R}^n, \gamma_n)$. We can define a *(Gaussian expectation) inner product*

$$\langle f, g \rangle_{\gamma_n} := \mathop{\mathbb{E}}_{x \sim \gamma_n}[f(x) \cdot g(x)]$$

for functions $f, g : \mathbb{R}^n \to \mathbb{R}$. For $p \geq 1$, we also define a norm

$$\|f\|_{\gamma_n, p} := \mathop{\mathbb{E}}_{x \sim \gamma_n}\left[|f(x)|^p\right]^{1/p}$$

In many cases these operations coincide with the boolean case:

Lemma 6.2. *For any multilinear polynomials $f, g : \mathbb{R}^n \to \mathbb{R}$ one has $\langle f, g \rangle_{\gamma_n} = \langle f, g \rangle_E$.*

Proof. By linearity it suffices to consider $f = \chi_S$ and $g = \chi_T$ for $S, T \subseteq [n]$. Then

$$\langle \chi_S, \chi_T \rangle_{\gamma_n} = \mathop{\mathbb{E}}_{x \sim \gamma_n}\left[\prod_{i \in S} x_i \cdot \prod_{i \in T} x_i\right] = \prod_{i \in S \cap T} \underbrace{\mathop{\mathbb{E}}_{x_i \sim \gamma_1}[x_i^2]}_{=1} \prod_{i \in S \Delta T} \underbrace{\mathop{\mathbb{E}}_{x_i \sim \gamma_1}[x_i]}_{=0} = \begin{cases} 1 & \text{if } S = T \\ 0 & \text{if } S \neq T \end{cases}$$

We can see that the only properties of the Gaussian that was relevant here is that (i) coordinates are independent, (ii) the mean of each coordinate is 0 and (iii) the variance of each coordinate is 1. □

Note that Lemma 6.2 is not true for arbitrary functions. For example consider the function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^p$ for $p \in \mathbb{N}$. Then $\langle f, f \rangle_E = \mathbb{E}_{x \sim \{-1,1\}}[x^{2p}] = 1$ while $\langle f, f \rangle_{\gamma_1} = \mathbb{E}_{x \sim \gamma_1}[x^{2p}] = (2p-1)!!$.

Lemma 6.3. *For any multilinear polynomial $f : \mathbb{R}^n \to \mathbb{R}$ one has $\|f\|_{\gamma_n, 2} = \|f\|_{E,2}$.*

Proof. Clear because $\|f\|_{\gamma_n,2}^2 = \langle f, f \rangle_{\gamma_n} = \langle f, f \rangle_E = \|f\|_{E,2}^2$. □

Again, this fails to hold for general p-norms. We make the following observation:

Lemma 6.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial and let $X_1, \dots, X_n \in \mathbb{R}$ be independent random variables so that*

$$\mathbb{E}[X_i] = 0 \quad \forall i \in [n] \quad \textit{and} \quad \mathbb{E}[X_i^2] = 1 \quad \forall i \in [n]$$

Then

$$\mathbb{E}[f(X)] = \hat{f}(\emptyset) \quad \textit{and} \quad \textit{Var}[f(X)] = \sum_{\emptyset \subset S \subseteq [n]} \hat{f}(S)^2$$

These conditions are in particular satisfied for Gaussians and hence this motivates to use the following notation:

Definition 6.5. For a multilinear polynomial $f : \mathbb{R}^n \to \mathbb{R}$ we define

$$\mathrm{Var}[f] := \sum_{\emptyset \subset S \subseteq [n]} \hat{f}(S) \quad \text{and} \quad \mathrm{Inf}_i[f] := \sum_{S \subseteq [n] : i \in S} \hat{f}(S)^2 \quad \forall i \in [n]$$

6.1.1 Stability and noise operators

A particularly important application of the invariance principle will deal with the noise operator and stability of functions which we need to generalize to Gaussian space as well. Recall from Section 1.7 that for any $x \in \{-1,1\}^n$ and $-1 \le \rho \le 1$, $y \sim N_\rho(x)$ is a vector in $\{-1,1\}^n$ with independent coordinates and $\mathbb{E}[x_i y_i] = \rho$. We defined T_ρ as the operator with $(T_\rho f)(x) = \mathbb{E}_{y \sim N_\rho(x)}[f(y)]$, i.e. it provides a noisy version of f. Now, to the Gaussian case.

Definition 6.6. For $x \in \mathbb{R}^n$ we define $N_{\gamma_n,\rho}(x)$ as the distribution over $\rho x + \sqrt{1-\rho^2} g$ where $g \sim \gamma_n$ is an independent Gaussian.

We can see that if $x \sim \gamma_n$ and $y \sim N_{\gamma_n,\rho}(x)$, then $y \sim \gamma_n$ (i.e. y is Gaussian) and the correlation of x and y is $\mathbb{E}[x_i y_i] = \rho$ for all i. For notational convinience, we denote $\mathcal{G}_{n,\rho}$ as the distribution over such *ρ-correlated Gaussian pairs*, i.e. $(x, y) \sim \mathcal{G}_{n,\rho}$ satisfy $x, y \sim \gamma_n$ and $\mathbb{E}[x_i y_i] = \rho$ for all $i \in [n]$. We define the linear operator $T_{\gamma_n,\rho} : L^2(\mathbb{R}^n, \gamma_n) \to L^2(\mathbb{R}^n, \gamma_n)$ with

$$(T_{\gamma_n,\rho} f)(x) := \mathop{\mathbb{E}}_{y \sim N_{\gamma_n,\rho}(x)}[f(y)] \quad \forall x \in \mathbb{R}^n$$

For a function $f \in L^2(\mathbb{R}^n, \gamma_n)$ and $-1 \le \rho \le 1$, we define the *Gaussian stability* as

$$\mathrm{Stab}_{\gamma_n,\rho}[f] := \mathop{\mathbb{E}}_{(x,y) \sim \mathcal{G}_{n,\rho}}[f(x) \cdot f(y)] = \langle T_{\gamma_n,\rho}(f), f \rangle_{\gamma_n}$$

We make an observation that will be useful later:

Lemma 6.7. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial and let $-1 \le \rho \le 1$. Then*

(i) One has $(T_\rho f)(x) = (T_{\gamma_n,\rho} f)(x)$ for all $x \in \{-1,1\}^n$.

(ii) $T_{\gamma_n,\rho} f$ is again a multilinear polynomial.

(iii) One has $Stab_{\gamma_n,\rho}[f] = Stab_\rho[f]$.

Proof. First we verify (i). By linearity of T_ρ and $T_{\gamma_n,\rho}$ it suffices to consider the case that $f(x) = \chi_S(x)$ for some set $S \subseteq [n]$. Fix $x \in \{-1,1\}^n$. Then

$$T_{\gamma_n,\rho}\chi_S(x) = \mathop{\mathbb{E}}_{y \sim N_{\gamma_n,\rho}(x)}[\chi_S(y)] = \prod_{i\in S} \underbrace{\mathop{\mathbb{E}}_{y_i \sim N_{\gamma_1,\rho}(x_i)}[y_i]}_{=\rho x_i} = \rho^{|S|}\chi_S(x)$$

which is exactly the same as in the boolean case. For (ii), since $T_\rho f$ is a multilinear polynomial, the same must hold for $T_{\gamma_n,\rho} f$. For (iii) we use that for multilinear polynomials, $T_{\gamma_n,\rho}$ and $\langle \cdot,\cdot \rangle_{\gamma_n}$ are identical to their boolean counterparts. □

It is not hard to show the the Gaussian noise operator is a *contraction.*

Lemma 6.8. *For any $f : \mathbb{R}^n \to \mathbb{R}$ with $\|f\|_{\gamma_n,1} < \infty$ and $0 \le \rho \le 1$, one has $\|T_{\gamma_n,\rho}(f)\|_{\gamma_n,1} \le \|f\|_{\gamma_n,1}$.*

We will also consider sets $A \subseteq \mathbb{R}^n$ and work with the *Gaussian measure*

$$\gamma_n(A) := \Pr_{x \sim \gamma_n}[x \in A]$$

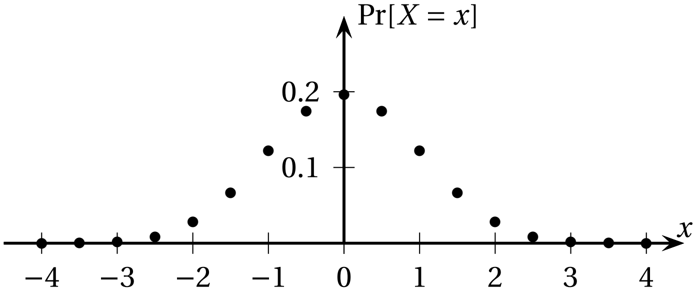
In order for this quantity to be well-defined, A needs to be measurable.

6.2 The Berry-Esseen Theorem

The most classical form of the invariance principle is the *central limit theorem* which says that a sum of many independent random variables converges to a Gaussian with the same mean and variance. A precise quantitative version of this fact is as follows:

Theorem 6.9 (Berry-Esseen Theorem)**.** *Let $X_1, \dots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $\sum_{i=1}^n \sigma_i^2 = 1$ where $\sigma_i^2 :=$ Var$[X_i]$ and let $X := \sum_{i=1}^n X_i$ and $Y \sim \gamma_1$. Then for all $u \in \mathbb{R}$ one has*

$$\left|\Pr[X \le u] - \Pr[Y \le u]\right| \le 0.56 \cdot \sum_{i=1}^n \mathbb{E}[|X_i|^3]$$



Distribution of $X = \sum_{i=1}^n X_i$ where $X_i \sim \{\pm\frac{1}{\sqrt{n}}\}$ for $n = 16$

For example if $|X_i| \le O(\frac{1}{\sqrt{n}})$ for all i, then $\mathbb{E}[|X_i|^3] \le O(n^{-3/2})$ and the right hand side is of the form $O(\frac{1}{\sqrt{n}})$. There are several ways to prove the Berry Esseen Theorem including Fourier analysis. However, we will use a rather flexible technique called the *replacement method* (or *hybrid method*) even though its result will be somewhat suboptimal. One can think of the statement of Berry Esseen as saying that $|\mathbb{E}[\psi(X)] - \mathbb{E}[\psi(Y)]|$ is small, where $\psi(x) := \mathbf{1}_{x \le u}$ is the characteristic function of an interval $(-\infty, u]$. Here we will prove a variant of Berry Esseen using a smooth "test function" ψ instead. In the following statement, ψ''' denotes the 3rd derivative of ψ and $\|\psi'''\|_\infty = \sup_{x\in\mathbb{R}} |\psi'''(x)|$ denotes its largest absolute value.

Theorem 6.10. *Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ be independent random variables so that*

$$\mathbb{E}[X_i] = \mathbb{E}[Y_i] \;\; \forall i \in [n] \quad \textit{and} \quad \mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2] \;\; \forall i \in [n]$$

and set $X := \sum_{i=1}^n X_i$ and $Y := \sum_{i=1}^n Y_i$. Then for any function $\psi : \mathbb{R} \to \mathbb{R}$ with continuous ψ''' one has

$$\left|\mathbb{E}[\psi(X)] - \mathbb{E}[\psi(Y)]\right| \le \frac{1}{6}\|\psi'''\|_\infty \cdot \gamma$$

where $\gamma := \sum_{i=1}^n (\mathbb{E}[|X_i|^3] + \mathbb{E}[|Y_i|^3])$.

Proof. For $t \in \{0, \dots, n\}$ we consider the random variable

$$H_t := Y_1 + \ldots + Y_t + X_{t+1} + \ldots + X_n.$$

One can think of H_t as a hybrid or mixture of the X and the Y-random variables where $H_0 = X$ and $H_n = Y$. It is not hard to see that it suffices to upper bound the error made in any step t.

Claim I. *For each t one has* $|\mathbb{E}[\psi(H_t)] - \mathbb{E}[\psi(H_{t-1})]| \le \frac{1}{6}\|\psi'''\|_\infty \cdot (\mathbb{E}[|X_t|^3] + \mathbb{E}[|Y_t|^3])$.

Proof of Claim I. We abbreviate the random variable

$$U := Y_1 + \ldots + Y_{t-1} + X_{t+1} + \ldots + X_n$$

which leaves out the tth summand. Note that X_t and Y_t are independent from U and $H_{t-1} = U + X_t$ and $H_t = U + Y_t$. Recall that the 2nd order Taylor expansion of ψ at point U is

$$\psi(U+Z) = \psi(U) + \psi'(U) \cdot Z + \frac{1}{2}\psi''(U) \cdot Z^2 + \frac{1}{6}\psi'''(V_{U,Z}) \cdot Z^3$$

where $V_{U,Z}$ is a point in the interval between U and $U+Z$. Applying this twice gives

$$\begin{aligned}
\Big|\mathbb{E}[\psi(H_t)] - \mathbb{E}[\psi(H_{t-1})]\Big| &= \Big|\mathbb{E}[\psi(U+Y_t)] - \mathbb{E}[\psi(U+X_t)]\Big| \\
&\overset{\text{Taylor}}{=} \Big|\mathbb{E}\Big[\psi(U) + \psi'(U)Y_t + \frac{1}{2}\psi''(U)Y_t^2 + \frac{1}{6}\psi'''(V_{U,Y_t})Y_t^3\Big] \\
&\quad - \mathbb{E}\Big[\Big(\psi(U) + \psi'(U)X_t + \frac{1}{2}\psi''(U)X_t^2 + \frac{1}{6}\psi'''(V_{U,X_t})X_t^3\Big)\Big]\Big| \\
&\overset{(*)}{=} \frac{1}{6}\Big|\mathbb{E}\big[\psi'''(V_{U,Y_t}) \cdot Y_t^3 - \psi'''(V_{U,X_t}) \cdot X_t^3\big]\Big| \\
&\leq \frac{1}{6}\|\psi'''\|_\infty \cdot \big(\mathbb{E}[|Y_t^3|] + \mathbb{E}[|X_t^3|]\big)
\end{aligned}$$

In $(*)$ we have used that $\mathbb{E}[X_t] = \mathbb{E}[Y_t]$ and $\mathbb{E}[X_t^2] = \mathbb{E}[Y_t^2]$ which causes the constant, linear and quadratic terms to cancel. □

Now summing over the error terms of all n steps we get

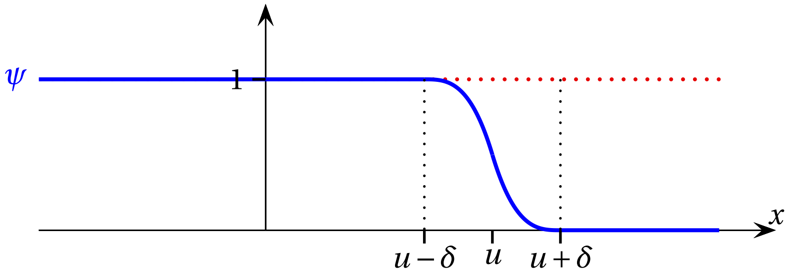
$$\begin{aligned}
\Big|\mathbb{E}[\psi(X)] - \mathbb{E}[\psi(Y)]\Big| &= \Big|\mathbb{E}\Big[\sum_{t=1}^{n}(\psi(H_t) - \psi(H_{t-1}))\Big]\Big| \\
&\leq \sum_{t=1}^{n}\Big|\mathbb{E}[\psi(H_t)] - \mathbb{E}[\psi(H_{t-1})]\Big| \\
&\overset{\text{Claim I}}{\leq} \frac{\|\psi'''\|_\infty}{6} \cdot \sum_{i=1}^{n}\big(\mathbb{E}[|X_i|^3] + \mathbb{E}[|Y_i|^3]\big)
\end{aligned}$$

□

We want to derive at least a weak version of the Berry Esseen Theorem from this result. For that purpose we need a smooth approximation of the indicator function $\mathbf{1}_{x \leq u}$.

Lemma 6.11. *For any* $u \in \mathbb{R}^n$ *and* $\delta > 0$ *there is a function* $\psi : \mathbb{R} \to [0,1]$ *so that* ψ''' *is continuous with* $\|\psi'''\|_\infty \leq O(\frac{1}{\delta^3})$ *so that*

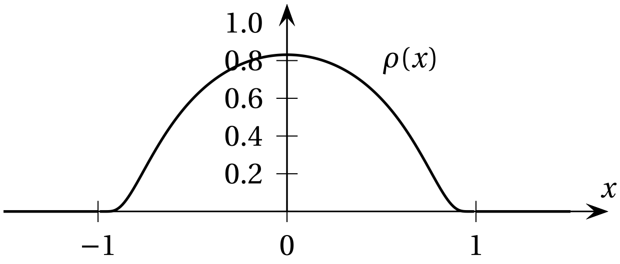
$$\psi(x) = \begin{cases} 1 & \text{if } x \leq u - \delta \\ 0 & \text{if } x \geq u + \delta \end{cases}$$



Proof. Define the symmetric function $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with

$$\rho(x) := \begin{cases} c \exp(-\frac{1}{1-x^2}) & \text{if } -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $c \approx 2.25228$ is the unique scalar so that $\int_{x \in \mathbb{R}} \rho(x) dx = 1$, meaning that ρ is a density function.



One can verify by induction that for any $k \in \mathbb{N}$, the kth derivative of ρ on $(-1,1)$ is of the form $\rho^{(k)}(x) = \frac{p_k(x)}{(1-x^2)^{2k}} \cdot \rho(x)$ where p_k is some polynomial. In particular all derivatives are smooth and $\|\rho^{(k)}\|_\infty \leq C_k$ for some constant k. We write $x \sim \bar{\rho}$ if we sample x according to "squeezed" density function $\bar{\rho}(x) := \frac{1}{\delta}\rho(\frac{x}{\delta})$. Any such sample has $|x| \leq \delta$. The definition of our smooth test function is then given by the convolution of $\mathbf{1}_{\leq u}$ with $\bar{\rho}$, i.e.

$$\psi(x) := \Pr_{g \sim \rho}\left[x + \delta g \in [-\infty, u]\right] = (\mathbf{1}_{\leq u} * \bar{\rho})(x)$$

which indeed has $\|\psi'''\|_\infty \leq O(\frac{1}{\delta^3})$. □

Theorem 6.12 (Weak Berry-Esseen Theorem)**.** *Let $X_1, \dots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $\sum_{i=1}^n \sigma_i^2 = 1$ where $\sigma_i^2 := Var[X_i]$ and let $X := \sum_{i=1}^n X_i$ and $Y \sim \gamma_1$. Then for all $u \in \mathbb{R}$ one has*

$$\left|\Pr[X \leq u] - \Pr[Y \leq u]\right| \leq O(\gamma^{1/4})$$

where $\gamma := \sum_{i=1}^n \mathbb{E}[|X_i|^3]$.

Proof. Let ψ be the smooth approximation to the indicator function $\mathbf{1}_{x \leq u}$ from Lemma 6.11 where the transition from 1 to 0 is between $u - \delta$ and u and we determine the parameter $\delta > 0$ later. For symmetry reasons it suffices to consider the case where $\Pr[Y \leq u]$ is the larger probability. Then

$$\begin{aligned} \Pr[Y \leq u] - \Pr[X \leq u] &\leq \Pr[Y \leq u-\delta] - \Pr[X \leq u] + \underbrace{\Pr[u-\delta \leq Y \leq u]}_{\leq \delta} \\ &\leq |\mathbb{E}[\psi(Y)] - \mathbb{E}[\psi(X)]| + \delta \\ &\overset{\text{Thm } 6.10}{\leq} O\Big(\frac{\gamma}{\delta^3}\Big) + \delta \leq O(\gamma^4) \end{aligned}$$

where the last inequality follows by setting $\delta := \gamma^4$. Note that we have omitted the 3rd moment contribution from the Gaussian part, but splitting $Y = \sum_{i=1}^n Y_i$ with $Y_i \sim \sigma_i \gamma_1$ we may see that $\mathbb{E}[|Y_i|^3] \leq O(\mathbb{E}[|X_i|^3])$ no matter how X_i looks like. □

We want to conclude this section by stating a multi-dimensional Berry-Esseen Theorem without proof for later reference:

Theorem 6.13 (Multidimensional Berry Esseen)**.** *Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be independent random vectors with $\mathbb{E}[X_i] = \mathbf{0}$ for all $i \in [n]$. Set $X := X_1 + \ldots + X_n$ and assume that $\Sigma := Cov[X] = \mathbb{E}[XX^T]$ has full rank and draw $Y \sim N(\mathbf{0}, \Sigma)$. Then for any convex set $U \subseteq \mathbb{R}^d$ one has*

$$\Big|\Pr[X \in U] - \Pr[Y \in U]\Big| \leq O(d^{1/4}\gamma)$$

where $\gamma := \sum_{i=1}^n \mathbb{E}[\|\Sigma^{-1/2} X_i\|_2^3]$.

6.3 The invariance principle

Now we come to the main topic of this chapter, which is the statement and proof of the invariance principle. Recall that a random variable X is B-*reasonable* if $\mathbb{E}[X^4] \leq B \cdot \mathbb{E}[X^2]^2$. If we revisit the statement and proof of Bonami's Lemma (Theorem 5.4) then we can quickly see that there is very little about the hypercube that is being used. In fact, the exact same proof also gives the following more general statement:

Theorem 6.14 (Bonami Lemma on $\mathbb{R}^n$)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial with degree at most d. Then for any independent random variables $X_1, \ldots, X_n \in \mathbb{R}$ with*

$$\mathbb{E}[X_i] = 0, \quad \mathbb{E}[X_i^2] = 1, \quad \mathbb{E}[X_i^3] = 0, \quad \mathbb{E}[X_i^4] \leq 9 \quad \forall i \in [n],$$

the random variable $F(X)$ is 9^d-reasonable.

We give a name to the property that makes Bonami's Lemma work:

Definition 6.15. We say that a vector of independent random variables $X_1, \dots, X_n$ is *nice* if

$$\mathbb{E}[X_i] = 0, \quad \mathbb{E}[X_i^2] = 1, \quad \mathbb{E}[X_i^3] = 0, \quad \mathbb{E}[X_i^4] \leq 9 \quad \forall i \in [n]$$

Theorem 6.16 (Basic Invariance Principle)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial of degree at most d. Let $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ both be random vectors with independent coordinates that are nice (see Def 6.15). Then for any function $\psi : \mathbb{R} \to \mathbb{R}$ with continuous ψ'''' one has*

$$\left| \mathbb{E}[\psi(F(X))] - \mathbb{E}[\psi(F(Y))] \right| \leq \frac{\|\psi''''\|_\infty}{12} \cdot 9^d \cdot \sum_{i=1}^n \mathit{Inf}_i[F]^2$$

Proof. We use a similar hybrid argument as in the proof of Theorem 6.10. For $t \in \{0, \dots, n\}$ we define

$$H_t := F(Y_1, \dots, Y_t, X_{t+1}, \dots, X_n)$$

so that again $H_0 = F(X)$ and $H_n = F(Y)$. Again we account the error made by a single swap:

Claim. *For any t one has $|\mathbb{E}[\psi(H_{t-1})] - \mathbb{E}[\psi(H_t)]| \leq \frac{\|\psi''''\|_\infty}{12} \cdot 9^d \cdot \mathit{Inf}_t[F]^2$.*

Proof of Claim I. We can pull out the t-th variable and write

$$F(x) = \sum_{|S| \leq d} \hat{F}(S)\chi_S(x) = \underbrace{\Big(\sum_{|S| \leq d : t \notin S} \hat{F}(S)\chi_S(x) \Big)}_{=:A(x_1,\dots,x_{t-1},x_{t+1},\dots,x_n)} + x_t \cdot \underbrace{\Big(\sum_{|S| \leq d : t \in S} \hat{F}(S)\chi_{S \setminus \{t\}}(x) \Big)}_{=:B(x_1,\dots,x_{t-1},x_{t+1},\dots,x_n)}$$

Then we define the random variables

$$U := A(X_1, \dots, X_{t-1}, Y_{t+1}, \dots, Y_n) \quad \text{and} \quad D := B(X_1, \dots, X_{t-1}, Y_{t+1}, \dots, Y_n)$$

so that

$$H_{t-1} = U + D \cdot X_t \quad \text{and} \quad H_t = U + D \cdot Y_t$$

Crucially note that (U, D) are independent from X_t and Y_t (but of course U, D themselfs are not necessarily independent). Next, the 3rd degree Taylor approximation of ψ at U is

$$\psi(U + Z) = \psi(U) + \psi'(U) \cdot Z + \frac{1}{2}\psi''(U) \cdot Z^2 + \frac{1}{6}\psi'''(U) \cdot Z^3 + \frac{1}{24}\psi''''(V_{U,Z}) \cdot Z^4$$

where $V_{U,Z}$ is a point on the segment between U and $U+Z$. Now we can bound

$$\begin{aligned}
& \left|\mathbb{E}[\psi(H_{t-1})] - \mathbb{E}[\psi(H_t)]\right| \\
= \quad & \left|\mathbb{E}[\psi(U+D\cdot X_t)] - \mathbb{E}[\psi(U+D\cdot Y_t)]\right| \\
\overset{\text{Taylor}}{=} \quad & \Big|\psi(U)+\psi'(U)\cdot DX_t + \frac{1}{2}\psi''(U)\cdot D^2X_t^2 + \frac{1}{6}\psi'''(U)\cdot D^3X_t^3 + \frac{1}{24}\psi''''(V_{U,DX_t})\cdot D^4X_t^4 \\
& -\Big(\psi(U)+\psi'(U)\cdot DY_t + \frac{1}{2}\psi''(U)\cdot D^2Y_t^2 + \frac{1}{6}\psi'''(U)\cdot D^3Y_t^3 + \frac{1}{24}\psi''''(V_{U,DY_t})\cdot D^4Y_t^4\Big)\Big| \\
= \quad & \frac{1}{24}\left|\mathbb{E}\left[\psi''''(V_{U,DX_t})\cdot D^4X_t^4 - \psi''''(V_{U,DY_t})\cdot D^4Y_t^4\right]\right| \\
\leq \quad & \frac{\|\psi''''\|_\infty}{24}\cdot\mathbb{E}[D^4]\cdot\big(\underbrace{\mathbb{E}[X_t^4]}_{\leq 9}+\underbrace{\mathbb{E}[Y_t^4]}_{\leq 9}\big) \\
\overset{\text{Subclaim I.A}}{\leq} \quad & \frac{\|\psi''''\|_\infty}{12}\cdot 9^d\cdot\mathrm{Inf}_t[F]^2
\end{aligned}$$

Here we use that $\mathbb{E}[DX_t]=\mathbb{E}[D]\,\mathbb{E}[X_t]$ by independence (similar for Y_t and other powers) and we also use that $\mathbb{E}[X_t-Y_t]=0$, $\mathbb{E}[X_t^2-Y_t^2]=0$, $\mathbb{E}[X_t^3-Y_t^3]=0$ so that all except the 4th order terms cancel. In the last step we bound $\mathbb{E}[D^4]$ with the following argument:

Subclaim I.A. *One has* $\mathbb{E}[D^4]\leq 9^{d-1}\cdot \mathit{Inf}_t[F]^2$.

Proof of Subclaim I.A. Let us write $Z:=(X_1,\dots,X_{t-1},Y_{t+1},\dots,Y_n)\in\mathbb{R}^{n-1}$ and recall that $D=B(Z)$ where B is a multilinear polynomial of degree at most $d-1$. Hence we may apply the Bonami Lemma (Theorem 6.14) and get

$$\mathop{\mathbb{E}}_{Z}[B(Z)^4]\leq 9^{d-1}\,\mathbb{E}[B(Z)^2]^2 = 9^{d-1}\Big(\sum_{|S|\leq d:t\in S}\hat{F}(S)^2\Big)^2 = 9^{d-1}\cdot\mathrm{Inf}_t[F]^2 \quad \square$$

Then again summing over all t gives

$$\begin{aligned}
\left|\mathbb{E}[\psi(F(X))]-\mathbb{E}[\psi(F(Y))]\right| \quad & \leq \quad \sum_{t=1}^{n}\left|\mathbb{E}[\psi(H_{t-1})]-\mathbb{E}[\psi(H_t)]\right| \\
& \overset{\text{Claim I}}{\leq} \quad \sum_{t=1}^{n}\frac{\|\psi''''\|_\infty}{12}\cdot 9^d\cdot\mathrm{Inf}_t[F]^2
\end{aligned}$$

$\square$

Recall that a function $\psi:\mathbb{R}\to\mathbb{R}$ is called *c-Lipschitz* if $|\psi(t_1)-\psi(t_2)|\leq c|t_1-t_2|$ for all $t_1,t_2\in\mathbb{R}$. We can also give a guarantee for Lipschitz test functions.

Lemma 6.17. *Let $F:\mathbb{R}^n\to\mathbb{R}$ be a multilinear polynomial of degree at most d and assume additionally that $Var[F]\leq 1$ and $Inf_i[F]\leq\varepsilon$ for all $i\in[n]$. Let $X=$*

$(X_1,\dots,X_n)$ *and* $Y = (Y_1,\dots,Y_n)$ *both be random vectors with independent coordinates that are nice (see Def 6.15). Then for any* c*-Lipschitz function* $\psi : \mathbb{R} \to \mathbb{R}$ *one has*

$$\Big|\mathbb{E}[\psi(F(X))] - \mathbb{E}[\psi(F(Y))]\Big| \leq O(c)\cdot 2^d \cdot \varepsilon^{1/4}$$

Proof. After scaling we may assume that $c = 1$. Let $\eta > 0$ be a parameter that we determine later. As in Lemma 6.11 we can can construct a smooth approximation $\tilde{\psi} : \mathbb{R} \to \mathbb{R}$ so that $\|\tilde{\psi}''''\|_\infty \leq O(\frac{1}{\eta^3})$ while (using Lipschitzness of ψ) one has $|\psi(t) - \tilde{\psi}(t)| \leq \eta$ for all $t \in \mathbb{R}$. From the assumptions we know that

$$\sum_{i=1}^n \mathrm{Inf}_i[F]^2 \leq \varepsilon \sum_{i=1}^n \mathrm{Inf}_i[F] = \varepsilon \underbrace{\sum_{i=1}^n \sum_{|S|\leq d: i\in S} \hat{F}(S)^2}_{\leq \mathrm{Var}[F]\leq 1} \underbrace{|S|}_{\leq d} \leq \varepsilon d$$

Hence we can apply the Invariance Principle (Theorem 6.16) and obtain

$$\Big|\mathbb{E}[\psi(F(X))] - \mathbb{E}[\psi(F(Y))]\Big| \leq \eta + \Big|\mathbb{E}[\tilde{\psi}(F(X))] - \mathbb{E}[\tilde{\psi}(F(Y))]\Big| \leq \eta + O\Big(\frac{\varepsilon d \cdot 9^d}{\eta^3}\Big)$$

Then setting $\eta := (d9^d\varepsilon)^{1/4}$ gives the claim. □

6.4 Comparison inequality between the boolean and the Gaussian case

Now, let us focus on comparing random variables $F(x)$ and $F(y)$ where $x \sim \{-1,1\}^n$ and $y \sim \gamma_n$ as this will be main application of the invariance principle. All statements of the invariance principle that we developed so far have the huge disadvantage that they depend on the degree of the multilinear polynomial F. Dealing with this disadvantage will be the topic of this section. For a multilinear polynomial $F(x) = \sum_{S\subseteq[n]} \hat{F}(S)\chi_S(x)$ and $k \in \mathbb{Z}_{\geq 0}$ we write $F^{\leq k}(x) := \sum_{|S|\leq k} \hat{F}(S)\chi_S(x)$ as the *low degree* part and $F^{>k}(x) := \sum_{|S|>k} \hat{F}(S)\chi_S(x)$ as the *high degree* part. First we derive the (unsurprising) fact that the error in the invariance principle is small if the high degree part has small norm.

Lemma 6.18. *Let* $F : \mathbb{R}^n \to \mathbb{R}$ *be a multilinear polynomial with* Var$[F] \leq 1$ *and let* $k \in \mathbb{N}$ *and* $\varepsilon > 0$ *so that* Inf$_i[F^{\leq k}] \leq \varepsilon$ *for all* $i \in [n]$*. Then for any* c*-Lipschitz function* $\psi : \mathbb{R} \to \mathbb{R}$ *one has*

$$\Big|\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(F(x))] - \mathop{\mathbb{E}}_{y\sim\gamma_n}[\psi(F(y))]\Big| \leq O(c)\cdot\Big(2^k\varepsilon^{1/4} + \|F^{>k}\|_{E,2}\Big)$$

Proof. Recall that $F = F^{\leq k} + F^{>k}$ is the split of F into low degree and high degree parts. Then we can bound

$$\begin{aligned}
& \left| \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(F(x))] - \mathop{\mathbb{E}}_{y\sim\gamma_n}[\psi(F(y))] \right| \\
\overset{\psi \text{ is } c\text{-Lip.}}{\leq} \quad & \underbrace{\left| \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(F^{\leq k}(x))] - \mathop{\mathbb{E}}_{y\sim\gamma_n}[\psi(F^{\leq k}(y))] \right|}_{\leq O(c)\cdot 2^k\varepsilon^{1/4} \text{ by Lem 6.17}} + c \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|F^{>k}(x)|] + c \mathop{\mathbb{E}}_{y\sim\gamma_n}[|F^{>k}(y)|] \\
\leq \quad & O(c)\cdot 2^k\varepsilon^{1/4} + c \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[|F^{>k}(x)|^2\right]^{1/2} + c \mathop{\mathbb{E}}_{y\sim\gamma_n}\left[|F^{>k}(y)|^2\right]^{1/2} \\
\overset{\text{Lem 6.3}}{=} \quad & O(c)2^k\varepsilon^{1/4} + 2c\|F^{>k}\|_{E,2}
\end{aligned}$$

Here we use that ψ is c-Lipschitz. Moreover we use that by Jensen's inequality, for any random variable X one has $\mathbb{E}[|X|] \leq \mathbb{E}[X^2]^{1/2}$. □

Now, we can prove that using a noisy version of F, we can remove the dependence on the degree. Here we remind the reader that by Lemma 6.7, for multilinear polynomials, the operators T_ρ and $T_{\gamma_n,\rho}$ are the same.

Lemma 6.19. *Let* $0 < \varepsilon \leq 1$ *and* $0 < \delta \leq \frac{1}{20}$. *Let* $F : \mathbb{R}^n \to \mathbb{R}$ *be a multilinear polynomial with* $\mathrm{Var}[F] \leq 1$ *and* $\mathrm{Inf}_i[F] \leq \varepsilon$ *for all* $i \in [n]$ *and let* $\psi : \mathbb{R} \to \mathbb{R}$ *be any* c*-Lipschitz function. Then*

$$\left| \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(T_{1-\delta}F(x))] - \mathop{\mathbb{E}}_{y\sim\gamma_n}[\psi(T_{1-\delta}F(y))] \right| \leq O(c)\cdot\varepsilon^{\delta/3}$$

Proof. We define function $H : \mathbb{R}^n \to \mathbb{R}$ with $H(x) := T_{1-\delta}F(x)$ which again is a multilinear polynomial with $\mathrm{Var}[H] \leq \mathrm{Var}[F] \leq 1$ and $\mathrm{Inf}_i[H] \leq \mathrm{Inf}_i[F] \leq \varepsilon$. Let $k \geq 0$ be a parameter that we determine later. The norm of the high degree part of H is

$$\|H^{>k}\|_{E,2}^2 = \sum_{|S|>k} \hat{H}(S)^2 \overset{\text{Prop 1.23}}{=} \underbrace{\sum_{|S|>k} \hat{F}(S)^2}_{\leq \mathrm{Var}[F]\leq 1} \cdot (1-\delta)^{2|S|} \leq \exp(-2\delta k),$$

crucially using the fact that the noise operator $T_{1-\delta}$ dramatically shrinks the high degree Fourier coefficients. Then applying Lemma 6.18 to H we get

$$\begin{aligned}
\left| \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(H(x))] - \mathop{\mathbb{E}}_{y\sim\gamma_n}[\psi(H(y))] \right| \quad & \leq \quad O(c)\cdot\left(2^k\varepsilon^{1/4} + \|H^{>k}\|_{E,2}\right) \\
& \leq \quad O(c)\cdot\left(2^k\varepsilon^{1/4} + \exp(-\delta k)\right) \\
& \overset{k:=\frac{1}{3}\ln(1/\varepsilon)}{\leq} \quad O(c)\cdot\varepsilon^{\delta/3}
\end{aligned}$$

making an appropriate choice for k that balances both error terms. □

Extending a definition from Section 1.8 we can define the ρ*-stable influence* of a multilinear polynomial F as $\mathrm{Inf}_i^{(\rho)}[F] = \sum_{S\subseteq[n]:i\in S} \rho^{|S|-1}\hat{F}(S)^2$. The reader may note that in Lemma 6.19 one could have replaced the assumption of $\mathrm{Inf}_i[F] \le \varepsilon$ by the weaker assumption that $\mathrm{Inf}_i^{(1-\delta)}[F] \le \varepsilon$ as we still would have been able to infer that $\mathrm{Inf}_i[H] \le \mathrm{Inf}_i^{(1-\delta)}[F] \le \varepsilon$. This weaker assumption is usually phrased that F has no (ε,δ)*-notable coordinates.*

Chapter 7

The Majority is Stablest Theorem

7.1 The Majority Function

We want to revisit a topic from the introductory chapter (see Section 1.7 and Section 6.1.1). We remind outselfs that the *stability* of a function $f : \{-1,1\}^n \to \mathbb{R}$ for parameter $-1 \leq \rho \leq 1$ is

$$\mathrm{Stab}_\rho[f] = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n, y\sim N_\rho(x)}[f(x)\cdot f(y)] = \langle T_\rho f, f\rangle_E$$

where $N_\rho(x)$ is the distribution over $\{-1,1\}^n$ with independent coordinates so that $\mathbb{E}_{y\sim N_\rho(x)}[x_i y_i] = \rho$. The question that we want to answer is which functions $f : \{-1,1\}^n \to \{-1,1\}$ maximize the stability $\mathrm{Stab}_\rho[f]$ in the regime $0 < \rho < 1$. Clearly for a constant function the stability is 1, so let us restrict to functions $f : \{-1,1\}^n \to \{-1,1\}$ which are *balanced*, i.e. $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)] = 0$.

Lemma 7.1. *Let $0 < \rho < 1$. The balanced boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ that maximize Stab$_\rho[f]$ are precisely all the (anti-) dictatorship functions $\pm\chi_{\{i\}}$ with $i \in [n]$ which have Stab$_\rho[\pm\chi_{\{i\}}] = \rho$.*

Proof. Fix any balanced function $f : \{-1,1\}^n \to \{-1,1\}$. Then the stability is

$$\mathrm{Stab}_\rho[f] = \langle T_\rho f, f\rangle_E = \sum_{S\subseteq[n]} \rho^{|S|}\hat{f}(S)^2 \leq \rho = \mathrm{Stab}_\rho[\chi_i]$$

where we use that $\hat{f}(\emptyset) = 0$. We note that this inequality can only be tight if all of the Fourier weight is on the 1st level, i.e. $W^1[f] = 1$. By the FKN Theorem (Theorem 5.5) this is true only for (possibly negated) dictatorship functions. □

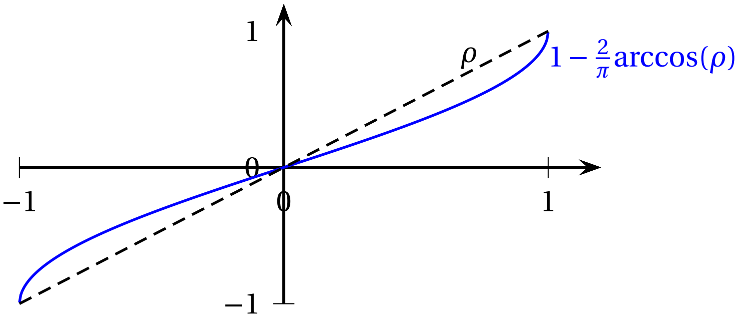
It would be more interesting to require that all coordinates have small influence to rule out dictatorship functions. It turns out that then the problem becomes rather non-trivial. First we discuss one particular function that is of fundamental importance in this context.

For odd n, we consider the *majority function* $\mathrm{Maj}_n : \{-1,1\}^n \to \{-1,1\}$ as

$$\mathrm{Maj}_n(x) := \mathrm{sign}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i\Big) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i > 0 \\ -1 & \text{if } \sum_{i=1}^n x_i < 0 \end{cases}$$

Of course one could drop the scalar $\frac{1}{\sqrt{n}}$ without affecting the definition, but this normalization will be convinient for us later. As we already mentioned in Section 5.3.1, Maj_n is symmetric under permuting the coordinates and $\mathrm{Inf}_i[\mathrm{Maj}_n] = \Theta(\frac{1}{\sqrt{n}})$ for all coordinates i. It will be interesting to determine the stability of the majority function.

Theorem 7.2. *For any* $-1 \leq \rho \leq 1$ *one has* $\mathit{Stab}_\rho[\mathit{Maj}_n] = 1 - \frac{2}{\pi}\mathit{arccos}(\rho) \pm o(1)$.



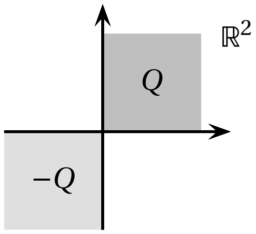
Proof. We will reduce this question on boolean functions to the Gaussian case. Let $\mathcal{D}_\rho$ be the distribution over ρ-correlated pairs $(x,y) \in \{-1,1\}^2$ where $x \sim \{-1,1\}$ and $y \sim N_\rho(x)$. We draw $X_1,\ldots,X_n \sim \frac{1}{\sqrt{n}}\mathcal{D}_\rho$ and set $X := \sum_{i=1}^n X_i \in \mathbb{R}^2$. Then we have the covariance matrices

$$\mathbb{E}[X_i X_i^T] = \frac{1}{n}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \text{and} \quad \mathbb{E}[XX^T] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

To avoid confusion we write the vector as $X = \binom{X(1)}{X(2)}$. We draw

$$Y \sim N\Big(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\Big)$$

which is a 2-dimensional random Gaussian that has the same mean and covariance matrix as X. Let $Q := \mathbb{R}^2_{\geq 0}$ be the positive orthant.



Then

$$\begin{aligned}
\mathrm{Stab}_\rho[\mathrm{Maj}_n] &= \mathbb{E}[\mathrm{sign}(X(1))\cdot\mathrm{sign}(X(2))] \\
&= 2\Pr[\mathrm{sign}(X(1)) = \mathrm{sign}(X(2))] - 1 \\
&= 2\Pr[X \in Q \text{ or } X \in -Q] - 1 \\
&= 4\Pr[X \in Q] - 1 \\
&\overset{\text{Thm 6.13}}{=} 4\Pr[Y \in Q] - 1 \pm o(1) \\
&\overset{\text{Sheppard}}{=} 1 - \frac{2}{\pi}\arccos(\rho) \pm o(1)
\end{aligned}$$

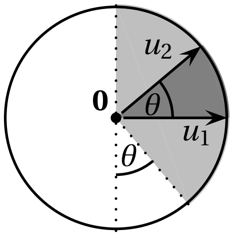
In order to apply the multidimensional Berry Esseen Theorem (Thm 6.13) we use that the orthant Q is convex. In the very last step we use the classic Sheppard's Formula, see Lemma 7.3. ☐

We also prove Sheppard's Formula which we just used:

Lemma 7.3 (Sheppard's Formula)**.** *Let* $-1 \leq \rho \leq 1$. *One has*

$$\Pr\left[Y \in \mathbb{R}^2_{\geq 0}\right] = \frac{1}{2} - \frac{\arccos(\rho)}{2\pi} \quad \textit{where} \quad Y \sim N\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Proof. We generate the Gaussian Y as follows: Let $u_1, u_2 \in S^1$ be two unit vectors in $\mathbb{R}^2$ with inner product $\langle u_1, u_2\rangle = \rho$. Then draw a standard Gaussian $g \sim \gamma_2$ and let $Y_1 = \langle g, u_1\rangle$ and $Y_2 = \langle g, u_1\rangle$. Let θ be the angle between the vectors u_1 and u_2, i.e. $\cos(\theta) = \rho$. By rotational symmetry of the Gaussian we can see that there is an angle of $\pi - \theta$ in which g needs to fall in order to satisfy the event that $Y \geq \mathbf{0}$.



Hence

$$\Pr[Y \geq \mathbf{0}] = \frac{\pi - \theta}{2\pi} = \frac{1}{2} - \frac{\arccos(\rho)}{2\pi}$$

□

We hope that the reader can appreciate how simple the analysis of the stability for majority function was once we transferred the question to the Gaussian setting. We want to prove that the majority function indeed maximizes the stability among all balanced functions with no influencial coordinate. In order to do so we will first prove the analogue in the Gaussian setting and then transfer it back.

7.2 Borell's Isoperimetric Theorem

Recall that in Section 6.1.1 we defined $(x, y) \sim \mathcal{G}_{n,\rho}$ to be ρ-correlated n-dimensional Gaussians. We make the following definition.

Definition 7.4. For $A \subseteq \mathbb{R}^n$ and $\delta \in \mathbb{R}$ we define the *rotational sensitivity of A at δ* as

$$RS_A(\delta) := \Pr_{(x,y)\sim\mathcal{G}_{n,\cos(\delta)}}[\mathbf{1}_A(x) \neq \mathbf{1}_A(y)]$$

We note that

$$\begin{aligned} RS_A(\theta) &= \Pr_{(x,y)\sim\mathcal{G}_{n,\cos(\delta)}}\left[\left(\mathbf{1}_A(x) \text{ and } \mathbf{1}_{\bar{A}}(y)\right) \text{ or } \left(\mathbf{1}_{\bar{A}}(x) \text{ and } \mathbf{1}_A(y)\right)\right] \qquad (7.1)\\ &= 2 \Pr_{(x,y)\sim\mathcal{G}_{n,\cos(\delta)}}[\mathbf{1}_A(x) \text{ and } \mathbf{1}_{\bar{A}}(y)] \end{aligned}$$

If $f : \mathbb{R}^n \to \{-1, 1\}$ is a function with $A := \{x \in \mathbb{R}^n \mid f(x) = 1\}$ and $\bar{A} = \{x \in \mathbb{R}^n \mid f(x) = -1\}$, then we can write $f(x) = 2 \cdot \mathbf{1}_A(x) - 1 = -2 \cdot \mathbf{1}_{\bar{A}}(x) + 1$ and so

$$\begin{aligned} \mathrm{Stab}_{\cos(\theta)}[f] &= \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{G}_{n,\cos(\theta)}}[f(x) \cdot f(y)] \qquad (7.2)\\ &= \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{G}_{n,\cos(\theta)}}[(2 \cdot \mathbf{1}_A(x) - 1) \cdot (-2 \cdot \mathbf{1}_{\bar{A}}(y) + 1)]\\ &= -4 \underbrace{\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{G}_{n,\cos(\theta)}}[\mathbf{1}_A(x) \cdot \mathbf{1}_{\bar{A}}(y)]}_{=RS_A(\theta)/2 \text{ by } (7.1)} + 2 \underbrace{\mathop{\mathbb{E}}_{x\sim\gamma_n}[\mathbf{1}_A(x)]}_{=\gamma_n(A)} + 2 \underbrace{\mathop{\mathbb{E}}_{y\sim\gamma_n}[\mathbf{1}_{\bar{A}}(y)]}_{=1-\gamma_n(A)} - 1\\ &= 1 - 2 \cdot RS_\theta(A) \qquad (7.3) \end{aligned}$$

That means in order to upper bound the stability of a set, it suffices to lower bound the rotational sensitivity of a set.

The following property will be important:

Theorem 7.5 (Subadditivity of Rotational Sensitivity)**.** *For any set* $A \subseteq \mathbb{R}^n$ *and any* $\delta_1, \ldots, \delta_\ell \in \mathbb{R}$ *one has*

$$RS_A\Big(\sum_{i=1}^{\ell} \delta_i\Big) \leq \sum_{i=1}^{\ell} RS_A(\delta_i)$$

Proof. It suffices to prove the result for $\ell = 2$ and then apply induction. Draw two independent Gaussians $g, h \sim \gamma_n$ and for $\theta \in \mathbb{R}$ define the *interpolation*

$$z(\theta) := \cos(\theta) \cdot g + \sin(\theta) \cdot h$$

As $\cos(\theta)^2 + \sin(\theta)^2 = 1$, we have that $z(\theta) \sim \gamma_n$ for all θ. For $\theta, \theta' \in \mathbb{R}$ and any coordinate $i \in [n]$, the Gaussians have a correlation of

$$\mathbb{E}[z(\theta)_i \cdot z(\theta')_i] = \cos(\theta)\cos(\theta') + \sin(\theta)\sin(\theta') = \cos(\theta' - \theta)$$

That means for any δ and θ we have $(z(\theta), z(\theta+\delta)) \sim \mathcal{G}_{n,\cos(\delta)}$ and so $RS_A(\delta) = \Pr[\mathbf{1}_A(z(\theta)) \neq \mathbf{1}_A(z(\theta+\delta))]$. Then using this fact with the union bound we obtain

$$\begin{aligned} RS_A(\delta_1+\delta_2) &= \Pr\left[\mathbf{1}_A(z(\delta_1+\delta_2)) \neq \mathbf{1}_A(z(0))\right] \\ &\leq \Pr\left[\mathbf{1}_A(z(0)) \neq \mathbf{1}_A(z(\delta_1))\right] + \Pr\left[\mathbf{1}_A(z(\delta_1)) \neq \mathbf{1}_A(z(\delta_1+\delta_2))\right] \\ &= RS_A(\delta_1) + RS_A(\delta_2) \end{aligned}$$

□

Now we can prove that in Gaussian space, indeed no balanced function has a higher stability than a majority function (or any indicator function of a halfspace through the origin). We recall that we had defined $\mathrm{Stab}_{\gamma_n,\rho}[f] := \mathbb{E}_{(x,y)\sim\mathcal{G}_{n,\rho}}[f(x) \cdot f(y)]$ in Section 6.1.1.

Theorem 7.6 (Borell's Theorem — Majority is Stablest in Gaussian Space)**.** *Let* $f : \mathbb{R}^n \to [-1,1]$ *and* $\mathbb{E}_{x\sim\gamma_n}[f(x)] = 0$*. Then for any* $0 < \rho < 1$ *one has*

$$Stab_{\gamma_n,\rho}[f] \leq 1 - \frac{2}{\pi} arccos(\rho)$$

Proof. We will only prove the statement for $\cos(\theta) = \rho$ where $\theta = \frac{\pi}{2\ell}$ with $\ell \in \mathbb{N}$. Still these are infinitely many ρ's that cover our later application. First we want to argue that we can restrict our attention to functions $f : \mathbb{R}^n \to \{-1,1\}$. To see this, let $\mathcal{F}_n := \{f \in \mathbb{R}^n \to [-1,1] : \mathbb{E}_{x\sim\gamma_n}[f(x)] = 0\}$ be the space of balanced bounded functions. Note that

$$\mathrm{Stab}_{\gamma_n,\rho}[f] = \langle T_{\gamma_n,\rho} f, f \rangle_{\gamma_n} = \| T_{\gamma_n,\sqrt{\rho}} f \|_{\gamma_n,2}^2$$

As $T_{\gamma_n,\sqrt{\rho}}$ is linear, we know that the map $\Phi : \mathcal{F}_n \to \mathbb{R}_{\geq 0}$ with $\Phi(f) := \mathrm{Stab}_{\gamma_n,\rho}[f]$ is convex. Hence a maximizer should be attained at an extreme point[1] which would be of the form $f : \mathbb{R}^n \to \{-1,1\}$. Then we can set $A := \{x \in \mathbb{R}^n \mid f(x) = 1\}$ and because f was balanced we have $\gamma_n(A) = \frac{1}{2}$ and it suffices to prove the following:
Claim I. *Let $A \subseteq \mathbb{R}^n$ be a set with $\gamma_n(A) = \frac{1}{2}$. Then for $\theta = \frac{\pi}{2\ell}$,*

$$\Pr_{(x,y)\sim\mathcal{G}_{n,\cos(\theta)}}[\mathbf{1}_A(x) \neq \mathbf{1}_A(y)] = RS_A(\theta) \geq \frac{\theta}{\pi}$$

Proof of Claim I. Using the subadditivity of RS_A we have

$$RS_A\Big(\frac{\pi}{2\ell}\Big) \overset{\text{Thm 7.5}}{\geq} \frac{1}{\ell} \cdot RS_A\Big(\frac{\pi}{2}\Big) = \frac{1}{\ell} \cdot \underbrace{\Pr_{\substack{x,y\sim\gamma_n \\ \text{indep.}}}\big[\mathbf{1}_A(x) \neq \mathbf{1}_A(y)\big]}_{=1/2} = \frac{1}{\ell} \cdot \frac{1}{2} = \frac{\theta}{\pi}$$

as for a set A with $\gamma_n(A) = \frac{1}{2}$, one has that $\mathbf{1}_A(x) \neq \mathbf{1}_A(y)$ with probability $1/2$ when x and y are independent Gaussians.

Then

$$\mathrm{Stab}_\rho[f] \overset{(7.2)}{=} 1 - 2 \cdot RS_{\arccos(\rho)}(A) \leq 1 - \frac{2}{\pi}\arccos(\rho)$$

□

A minor modification of the argument (adjusting the balance condition and replacing the probability of 1/2) gives the following:

Corollary 7.7 (Variant of Borell's Theorem)**.** Let $f : \mathbb{R}^n \to [-1,1]$ and $\mu := \mathbb{E}_{x\sim\gamma_n}[f(x)]$. Then for any $0 < \rho < 1$ one has

$$\mathrm{Stab}_{\gamma_n,\rho}[f] \leq 1 - \frac{2}{\pi}\arccos(\rho) + O(\mu)$$

We leave the details as an exercise. See the textbook [O'D21] for the tight bound in terms of μ.

7.3 Majority is stablest

In this section we will finally prove the Majority is Stablest Theorem. But first we show an auxiliary result that bounds the change of $\mathrm{Stab}_\rho[f]$ when we vary ρ.

Lemma 7.8. *For any $f : \{-1,1\}^n \to \mathbb{R}$ and $0 < \rho < 1$ one has $|\frac{d}{d\rho}\mathrm{Stab}_\rho[f]| \leq \frac{1}{1-\rho}\mathrm{Var}[f]$.*

---

[1] We generously skip any compactness issue here.

Proof. Writing out the Fourier representation of stability from Prop 1.24 we have

$$\left|\frac{d}{d\rho}\mathrm{Stab}_\rho[f]\right| \overset{\substack{\text{linearity of}\\ \text{derivative}}}{=} \sum_{S\subseteq[n]} \hat{f}(S)^2 \underbrace{\frac{d}{d\rho}\rho^{|S|}}_{=|S|\cdot\rho^{|S|-1}} = \sum_{S\subseteq[n]} \underbrace{|S|\rho^{|S|-1}}_{\leq\frac{1}{1-\rho}} \hat{f}(S)^2 \leq \frac{1}{1-\rho}\mathrm{Var}[f]$$

Here we use that $k\cdot\rho^{k-1}\leq\frac{1}{1-\rho}$ for all $0<\rho<1$ and $k\in\mathbb{Z}_{\geq 0}$. □

Now to the main result:

Theorem 7.9 (Majority is stablest – Mossel, O'Donnell, Oleszkiewicz [MOO10])**.** *Let* $0\leq\rho<1$*. For any function* $f:\{-1,1\}^n\to[-1,1]$ *with* $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)]=0$ *and* $\mathrm{Inf}_i[f]\leq\varepsilon$ *for all* $i\in[n]$*, one has*

$$Stab_\rho[f]\leq 1-\frac{2}{\pi}arccos(\rho)+O\Big(\frac{\log\log\frac{1}{\varepsilon}}{\log\frac{1}{\varepsilon}}\Big)\cdot\frac{1}{1-\rho}$$

Proof. We appreviate $\phi(\rho):=1-\frac{2}{\pi}\arccos(\rho)$. Let $0<\delta\leq\frac{1}{20}$ be a parameter that we decide later. Eventually we will make use of the invariance principle from Lemma 6.19. For that purpose define an auxiliary function $h:\mathbb{R}^n\to\mathbb{R}$ with

$$h(x):=T_{1-\delta}f(x)\quad\forall x\in\mathbb{R}^n$$

One can think of this as a minimally smoothed version of f. In fact the stability of f and h is very close so that instead we may analyze h.

Claim I. *One has* $|Stab_\rho[h]-Stab_\rho[f]|\leq\frac{2\delta}{1-\rho}$.

Proof of Claim I. We use the Fourier representation of stability from Prop 1.24 to write

$$\mathrm{Stab}_\rho[h]=\sum_{S\subseteq[n]}\rho^{|S|}\widehat{T_{1-\delta}f}(S)^2=\sum_{S\subseteq[n]}(\rho\cdot(1-\delta)^2)^{|S|}\hat{f}(S)^2=\mathrm{Stab}_{\rho(1-\delta)^2}[f]$$

Then using the bound on the derivative of stability from Lemma 7.8 we can bound

$$\begin{aligned}|\mathrm{Stab}_\rho(h)-\mathrm{Stab}_\rho(f)| &\leq \int_{\rho(1-\delta)^2}^{\rho}\underbrace{\left|\frac{d}{dt}\mathrm{Stab}_t[f]\right|}_{\leq\frac{1}{1-\rho}\mathrm{Var}[f]}dt\\ &\overset{\text{Lem 7.8}}{\leq} \underbrace{\rho}_{\leq 1}\cdot\underbrace{(1-(1-\delta)^2)}_{\leq 2\delta}\cdot\frac{1}{1-\rho}\underbrace{\mathrm{Var}[f]}_{\leq 1}\leq\frac{2\delta}{1-\rho}\end{aligned}$$

Here we use that $\mathrm{Var}[f]\leq 1$ as $|f(x)|\leq 1$ for all $x\in\{-1,1\}^n$. □

Later we will use the invariance principle in the following form:

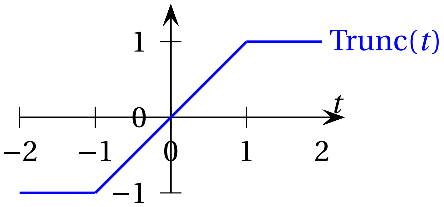
Claim II. *For any $O(1)$-Lipschitz function $\psi : \mathbb{R} \to \mathbb{R}$ one has* $|\mathbb{E}_{y\sim\gamma_n}[\psi(h(y))] - \mathbb{E}_{x\sim\{-1,1\}^n}[\psi(h(x))]| \leq O(\varepsilon^{\delta/3})$.

Proof of Claim II. Simply apply Lemma 6.19. □

If our remaining goal is to upper bound $\mathrm{Stab}_\rho[h]$ then one might be tempted to assume that Borell's Theorem (Theorem 7.6) immediately gives that $\mathrm{Stab}_{\gamma_n,\rho}[h] \leq \phi(\rho)$. But Theorem 7.6 requires that the function is *bounded* between -1 and 1 on the whole $\mathbb{R}^n$ which may not be true (even though indeed $|h(x)| \leq 1$ for all $x \in \{-1,1\}^n$). The way to work around this is to cut the function off outside of the interval $[-1,1]$ and account for the error using the invariance principle.

First we define the *truncation*

$$\mathrm{Trunc}(t) := \begin{cases} t & \text{if } -1 \leq t \leq 1 \\ -1 & \text{if } t < -1 \\ 1 & \text{if } t > 1 \end{cases}$$



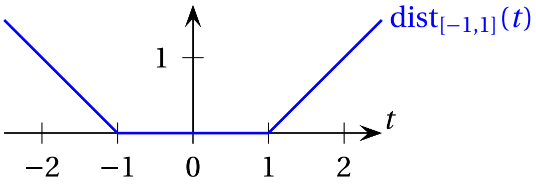
which is a 1-Lipschitz function. Then the overall approach that does work is to bound

$$\begin{aligned}
\mathrm{Stab}_\rho[f] &\overset{\text{Claim I}}{\leq} \mathrm{Stab}_\rho[h] + \frac{2\delta}{1-\rho} \\
&\overset{h\text{ multilin.}}{=} \mathrm{Stab}_{\gamma_n,\rho}[h] + \frac{2\delta}{1-\rho} \\
&\leq \underbrace{\mathrm{Stab}_{\gamma_n,\rho}[\mathrm{Trunc}(h)]}_{\leq \phi(\rho)+O(\varepsilon^{\delta/3}) \quad (*)} + \underbrace{\left|\mathrm{Stab}_{\gamma_n,\rho}[h] - \mathrm{Stab}_{\gamma_n,\rho}[\mathrm{Trunc}(h)]\right|}_{\leq O(\varepsilon^{\delta/3}) \quad (**)} + \frac{2\delta}{1-\rho} \\
&\leq \phi(\rho) + O(\varepsilon^{\delta/3}) + \frac{2\delta}{1-\rho}
\end{aligned}$$

Then choosing a δ that balances the error terms (e.g. $\delta := 3\frac{\log\log(1/\varepsilon)}{\log 1/\varepsilon}$) gives the statement of the theorem. It remains to justify the inequalities claimed in $(*)$ and $(**)$. Towards this goal we prove the following:

Claim III. *One has* $\mathbb{E}_{y\sim\gamma_n}[|h(y) - \mathit{Trunc}(h(y))|] \leq O(\varepsilon^{\delta/3})$.

Proof of Claim III. We define $\mathrm{dist}_A(t)$ as the distance of t to the nearest point in a set A. This always gives a 1-Lipschitz function. In particular we are going to use the function $\mathrm{dist}_{[-1,1]}$.



Then using the invariance principle we can bound

$$\begin{aligned}\mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\left|h(y)-\mathrm{Trunc}(h(y))\right|\right] &= \left|\mathop{\mathbb{E}}_{y\sim\gamma_n}[\mathrm{dist}_{[-1,1]}(h(y))] - \underbrace{\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\mathrm{dist}_{[-1,1]}(h(x))]}_{=0 \text{ since } -1\le h(x)\le 1 \text{ for } x\in\{-1,1\}^n}\right| \\ &\overset{\text{Claim II}}{\le} O(\varepsilon^{\delta/3})\end{aligned}$$

as for any fixed y one has $|h(y)-\mathrm{Trunc}(h(y))| = \mathrm{dist}_{[-1,1]}(h(y))$. □

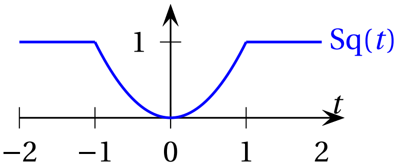
We recall that f is balanced and so h is balanced which (since h is multilinear) also implies that $\mathbb{E}_{y\sim\gamma_n}[h(y)] = 0$. Hence by Claim III, $\mathbb{E}_{y\sim\gamma_n}[\mathrm{Trunc}(h(y))]$ has to be small and we can apply the variant of Borell's Theorem (see Cor 7.7) to obtain

$$\mathrm{Stab}_{\gamma_n,\rho}[\mathrm{Trunc}(h(y))] \overset{\text{Lem 7.7}}{\le} \phi(\rho) + O\left(\left|\mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\mathrm{Trunc}(h(y))\right]\right|\right) \overset{\text{Claim III}}{\le} \phi(\rho) + O(\varepsilon^{\delta/3})$$

Hence we have proven $(*)$.

Next, we want to show $(**)$. We define the *square function* $\mathrm{Sq}:\mathbb{R}\to[0,1]$

$$\mathrm{Sq}(t) := \begin{cases} t^2 & \text{if } -1\le t\le 1 \\ 1 & \text{if } |t|>1\end{cases}$$



It will be useful to note that for any function $g:\mathbb{R}^n\to[-1,1]$ one has

$$\mathrm{Stab}_{\gamma_n,\rho}[g] = \langle T_{\gamma_n,\sqrt{\rho}}g, T_{\gamma_n,\sqrt{\rho}}g\rangle_{\gamma_n} = \mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\mathrm{Sq}(T_{\gamma_n,\sqrt{\rho}}g(y))\right] \tag{7.4}$$

Again, h is not bounded everywhere and so (7.4) does not directly apply, but we can still use the following:

Claim IV. *One has* $\left|Stab_{\gamma_n,\rho}[h] - \mathbb{E}_{y\sim\gamma_n}\left[Sq(T_{\gamma_n,\sqrt{\rho}}h(y))\right]\right| \le O(\varepsilon^{\delta/3})$.

Proof of Claim IV. As h is a multilinear polynomial we have

$$\begin{aligned}\mathrm{Stab}_{\gamma_n,\rho}[h] &= \mathrm{Stab}_\rho[h] \\ &\overset{|h(x)|\le 1\ \forall x\in\{-1,1\}^n}{=} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\mathrm{Sq}(T_{\sqrt{\rho}}h(x))] \\ &= \mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\mathrm{Sq}(T_{\gamma_n,\sqrt{\rho}}h(y))\right] \pm O(\varepsilon^{\delta/3})\end{aligned}$$

using in the last step again Claim II with the fact that Sq is 2-Lipschitz as well as the fact that for the multilinear polynomial h one has $T_{\sqrt{\rho}}h = T_{\gamma_n,\sqrt{\rho}}h$. $\square$

We apply (7.4) for function $g = \text{Trunc}(h)$ directly and use Claim IV to get

$$\begin{aligned}
& \left|\text{Stab}_{\gamma_n,\rho}[h] - \text{Stab}_{\gamma_n,\rho}[\text{Trunc}(h)]\right| \\
\overset{\text{(7.4)+Claim IV}}{\leq} \quad & \left|\mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\text{Sq}\big(T_{\gamma_n,\sqrt{\rho}}(h(y))\big)\right] - \mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\text{Sq}\big(T_{\gamma_n,\sqrt{\rho}}(\text{Trunc}(h(y)))\big)\right]\right| + O(\varepsilon^{\delta/3}) \\
\overset{\text{Sq 2-Lipschitz}}{\leq} \quad & 2\left|\mathop{\mathbb{E}}_{y\sim\gamma_n}\left[T_{\gamma_n,\sqrt{\rho}}h(y)\right] - \mathop{\mathbb{E}}_{y\sim\gamma_n}\left[T_{\gamma_n,\sqrt{\rho}}(\text{Trunc}(h(y)))\right]\right| + O(\varepsilon^{\delta/3}) \\
\overset{\text{Lem 6.8}}{\leq} \quad & 2\left|\mathop{\mathbb{E}}_{y\sim\gamma_n}[h(y)] - \mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\text{Trunc}(h(y))\right]\right| + O(\varepsilon^{\delta/3}) \\
\leq \quad & 2\mathop{\mathbb{E}}_{y\sim\gamma_n}\left[\left|h(y) - \text{Trunc}(h(y))\right|\right] + O(\varepsilon^{\delta/3}) \\
\overset{\text{Claim III}}{\leq} \quad & O(\varepsilon^{\delta/3})
\end{aligned}$$

Here we use that the Gaussian noise operator $T_{\gamma_n,\sqrt{\rho}}$ is linear and a contraction as we know from Lemma 6.8. This shows $(**)$ and concludes the proof. $\square$

The assumption of $\text{Inf}_i[f] \leq \varepsilon$ for all i, can be replaced by the weaker assumption that f has no (ε,δ)-notable coordinates, see the remark after the proof of Lemma 6.19. We want to record two variants for later use.

Theorem 7.10 (Majority is Stablest II)**.** *For any* $0 \leq \rho < 1$ *and* $\eta > 0$ *there are* $\varepsilon > 0$ *and* $d \in \mathbb{N}$ *so that the following holds: For any function* $f : \{-1,1\}^n \to [-1,1]$ *with* $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)] = 0$ *and* $\text{Inf}_i^{\leq d}[f] \leq \varepsilon$ *for all* $i \in [n]$*, one has*

$$Stab_\rho[f] \leq 1 - \frac{2}{\pi}arccos(\rho) + \eta$$

Proof. For a cleaner notation assume that the assumption is that $\text{Inf}_i^{\leq d}[f] \leq \frac{\varepsilon}{2}$ (so we do not need to introduce more constants). Using that assumption we can see that for any $\delta > 0$ one has

$$\text{Inf}_i^{(1-\delta)}[f] = \sum_{S\subseteq[n]:i\in S}(1-\delta)^{|S|}\hat{f}(S)^2 \leq \underbrace{\sum_{|S|\leq d:i\in S}\hat{f}(S)^2}_{\leq \text{Inf}_i^{\leq d}[f]\leq\varepsilon/2} + \underbrace{\sum_{|S|>d}\hat{f}(S)^2}_{\leq 1}\underbrace{\exp(-\delta|S|)}_{\leq\varepsilon/2} \leq \varepsilon$$

if we choose $d := \frac{\ln(\frac{2}{\varepsilon})}{\delta}$. By the remark from above, Theorem 7.9 already applies if $\text{Inf}_i^{(1-\delta)}[f] \leq \varepsilon$ where the choice of δ depends on ε. $\square$

For the regime $-1 < \rho \leq 0$ we can also obtain a *lower* bound on the stability:

Theorem 7.11 (Majority is Stablest III)**.** *For any* $-1 < \rho \leq 0$ *and* $\eta > 0$ *there are* $\varepsilon > 0$ *and* $d \in \mathbb{N}$ *so that the following holds: For any function* $f : \{-1,1\}^n \to [-1,1]$ *with* $\mathrm{Inf}_i^{\leq d}[f] \leq \varepsilon$ *for all* $i \in [n]$*, one has*

$$\mathit{Stab}_\rho[f] \geq \frac{2}{\pi}\mathit{arccos}(-\rho) - 1 - \eta = 1 - \frac{2}{\pi}\mathit{arccos}(\rho) - \eta$$

Proof. Let $f_{\text{odd}} : \{-1,1\}^n \to \mathbb{R}$ be the function with

$$f_{\text{odd}}(x) := \frac{1}{2}(f(x) - f(-x)) \quad \forall x \in \{-1,1\}^n$$

which is also called the *odd part* of the function f. By definition we know that $|f_{\text{odd}}(x)| \leq 1$ for all x because also $|f(x)| \leq 1$ for all x. Moreover, f_{odd} is balanced (even if f was not). We can verify that the Fourier expansion of f_{odd} is simply

$$f_{\text{odd}}(x) = \frac{1}{2}\sum_{S\subseteq[n]} \hat{f}(S)\cdot \underbrace{\left(\chi_S(x) - \chi_S(-x)\right)}_{\substack{=0 \text{ if } |S| \text{ even,} \\ =2\chi_S(x) \text{ if } |S| \text{ odd}}} = \sum_{S\subseteq[n]:|S| \text{ odd}} \hat{f}(S)\cdot\chi_S(x) \qquad (*)$$

from which we also know that $\mathrm{Inf}_i^{\leq d}[f_{\text{odd}}] \leq \mathrm{Inf}_i^{\leq d}[f]$ for all i. Then

$$\begin{aligned} \mathrm{Stab}_\rho[f] &\overset{\text{Prop 1.24}}{=} \sum_{S\subseteq[n]} \rho^{|S|}\hat{f}(S)^2 \\ &\geq \quad - \sum_{S\subseteq[n]:|S| \text{ odd}} (-\rho)^{|S|}\hat{f}(S)^2 \\ &\overset{\text{Prop 1.24}+(*)}{=} -\mathrm{Stab}_{-\rho}[f_{\text{odd}}] \overset{\text{Thm 7.9}}{\geq} -\left(1 - \frac{2}{\pi}\arccos(-\rho) + \eta\right) \end{aligned}$$

where we apply the Majority is Stablest II Theorem to the function f_{odd}. For the alternative representation one can use that $\arccos(-\rho) = \pi - \arccos(\rho)$. $\square$

Chapter 8

Hardness of Approximation II — The Unique Games Conjecture and Hardness for MaxCut

For the MaxCut problem we are given a weighted undirected graph $G = (V, E, w)$ and the goal is to find a cut $S \subseteq V$ that maximizes $w(\delta(S)) := \sum_{e \in E : |e \cap S| = 1} w_e$ which is the weight of the edges with end points in different sides of the cut. We denote $\mathsf{val}(G, w)$ as the value of the optimum solution, i.e. the maximum weight of edges separated by any cut.

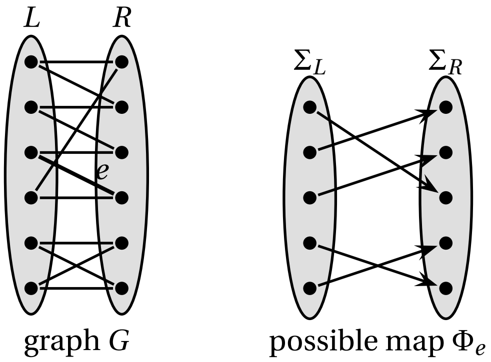
After Chapter 4, this is the second part on hardness of approximation in which we prove an optimum hardness result for MaxCut, assuming the so-called *Unique Games Conjecture (UGC)*. In particular we prove that assuming UGC, there is no polynomial time algorithm that finds a $(\alpha_{GW} + \varepsilon)$-approximation for MaxCut where $\alpha_{GW} \approx 0.878$ is the approximation ratio of the Goemans Williamson SDP rounding algorithm [GW95]. The crucial ingredient for this hardness proof will be the *Majority is Stablest Theorem* that we have just proven in the previous chapter. The presentation here follows the notes of Minzer [Min22] as well as O'Donnell's book [O'D21].

8.1 The Unique Games Conjecture

We have introduced the label cover problem in Section 4.1.3 as an **NP**-hard problem which served as a starting point to derive hardness for 3Lin_2. We recall that a label cover instance is of the form $\Psi = (G, \Sigma_L, \Sigma_R, (\Phi_e)_{e \in E})$ where $G = (L \dot\cup R, E)$ is a bipartite graph, $V := L \cup R$, $\Sigma := \Sigma_L \cup \Sigma_R$ and for every edge $e = (u, v) \in E$ we have a function $\Phi_e : \Sigma_L \to \Sigma_R$ which is satisfied by an assignment $A : V \to \Sigma$ if

$\Phi_e(A(u)) = A(v)$. Using the PCP Theorem and the Parallel Repetition Theorem, we know that for any $\varepsilon > 0$ distinguishing the cases of $\mathsf{val}(\Psi) = 1$ and $\mathsf{val}(\Psi) \leq \varepsilon$ is **NP**-hard. In this chapter we will introduce a special case of label cover where the maps Φ_e are *bijections* rather than arbitrary maps.

Definition 8.1. A *Unique Games* instance is of the form $\Psi = (G, \Sigma_L, \Sigma_R, (\Phi_e)_{e\in E})$ where $G = (V = L\dot{\cup}R, E)$, $|\Sigma_L| = |\Sigma_R|$ and all functions Φ_e are *bijective*. An assignment $A : V \to \Sigma$ satisfies an edge $e = (u, v) \in E$ if $\Phi_e(A(u)) = A(v)$. The goal is to find an assignment A that maximizes the number of satisfied constraints.



graph G possible map Φ_e

If $\mathsf{val}(\Psi) = 1$, then a satisfying assignment A can be found in polynomial time. The reason is that once we know one value $A(u)$, this uniquely determines all assignment values in the connected component of G that contains u. But similar to 3LIN$_2$ this argument does not work if say $\mathsf{val}(\Psi) \leq 1 - \varepsilon$ for some constant $\varepsilon > 0$. So the following is being conjectured:

Conjecture 1 (Unique Games Conjecture; Khot [Kho02])**.** *For all $\varepsilon > 0$, there is a $k \in \mathbb{N}$ so that* UNIQUEGAMES$^{[1-\varepsilon,\varepsilon]}$ *is* **NP***-hard for instances with alphabet size $|\Sigma| \leq k$ where the graph G is regular.*

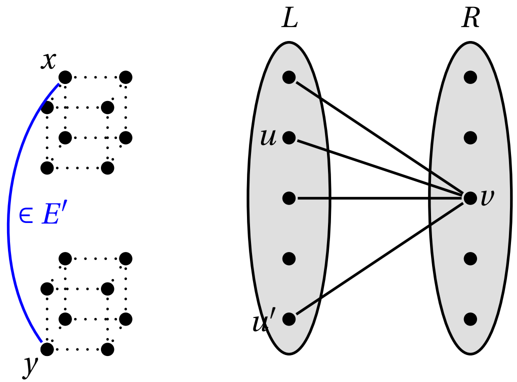
Here *regular* means that all vertices $v \in L \cup R$ have the same degree. In particular this implies that $|L| = |R|$. The assumption that vertices in G need to have regular degree is not actually part of the original Unique Games Conjecture but it would follow from it by standard techniques, hence we add it here for convinience. The Unique Games conjecture has been open for over two decades and we will justify the interest in it by an deriving optimum inapproximability for maxcut from it. We should note that non-trivial algorithms exist for the Unique Games problem. In particular given any instance Ψ with $\mathrm{val}(\Psi) \geq 1 - \varepsilon$ by rounding a semidefinite program one can find an assignment satisfying a $1-\Theta(\sqrt{\varepsilon \log(k)})$ fraction of constraints [CMM06].

8.2 The reduction from Unique Games to MaxCut

We will first describe the reduction from a Unique Games instance to MaxCut and then prove soundness and completeness over the following two sections. Recall that for $-1 \leq \rho \leq 1$ and $x \in \{-1,1\}^n$, $y \sim N_\rho(x)$ provides random vector in $\{-1,1\}^n$ with independent coordinates so that $\mathbb{E}[y_i] = \rho \cdot x_i$. Also recall that for a function $f : \{-1,1\}^n \to \mathbb{R}$, the *stability* is $\mathrm{Stab}_\rho[f] = \mathbb{E}_{x\sim\{-1,1\}^n, y\sim N_\rho(x)}[f(x)\cdot f(y)]$. We refer to Section 1.7 for details.

Now consider a Unique Games instance $\Psi = (G, \Sigma_L, \Sigma_R, (\Phi_e)_{e\in E})$ where we assume the bipartite graph G to be regular and fix a parameter $-1 \leq \rho \leq 0$. We define a maxcut instance $G' = (V', E')$ with weights $w' : E' \to \mathbb{R}_{\geq 0}$ that has vertices $V' := L \times \{-1,1\}^{\Sigma_L}$, corresponding to variables $f_u(x) \in \{-1,1\}$ for $u \in L$ and $x \in \Sigma_L$. We now generate an edge via a random process and the weight of the edge will be the probability/density that the edge is being generated: we pick a uniform node $v \in R$ on the right side. Then we pick two uniform random neighbors $u, u' \in N(v)$. We draw $x \sim \{-1,1\}^{\Sigma_R}$ and $y \sim N_\rho(x)$. We insert the edge

$$\big((u, \Phi_{u,v}(x)), (u', \Phi_{u',v}(y))\big) \in E'$$



part of graph G' graph G

visualization for case that $\Phi_{u,v}, \Phi_{u',v}$ are the identity

Note that the original label cover graph G is bipartite and maxcut has a trivial optimal solution for bipartite graphs. Hence the construction of G' at least passes the sanity check of not being bipartite. It might not yet be obvious what function the noise parameter ρ has. Formally we will prove:

Theorem 8.2 (Analysis of reduction). *For all $-1 \leq \rho \leq 0$ and $\delta > 0$, there is a small enough $\eta > 0$ so that for any Unique Games instance Ψ, the weighted graph (G', w') constructed above satisfies:*

- *Completeness*. $\mathrm{val}(\Psi) \geq 1 - \eta \Longrightarrow \mathrm{val}(G', w') \geq \frac{1}{2}(1-\rho) - \delta$.

- *Soundness*. $\mathsf{val}(\Psi) \leq \eta \Longrightarrow \mathsf{val}(G', w') \leq \frac{1}{\pi} arccos(\rho) + \delta$.

8.3 Completeness

We will first prove completeness.

Lemma 8.3 (Completeness of reduction)**.** *Let* $\eta > 0$*. If* $\mathsf{val}(\Psi) \leq 1 - \eta$*, then* $\mathsf{val}(G', w') \geq \frac{1}{2}(1-\rho) - 2\eta$*.*

Proof. Let $A : L \cup R \to \Sigma$ denote the Unique Games assignment satisfying a $1 - \eta$ fraction of edges. Define functions $\{f_u\}_{u \in L}$ with

$$f_u(x) := x_{A(u)} \quad \forall u \in L \, \forall x \in \{-1,1\}^{\Sigma_L}$$

Note that these functions correspond to the dictatorship functions induced by the labeling. The good cut in G' is given by the set $U := \{(u,x) : f_u(x) = 1\}$. In order to analyze the value of that cut, consider the random process that generates the edges in E'. By regularity, both of the edges (u,v) and (u',v) are uniform random choices from E, hence with probability at least $1-2\eta$, A satisfies both edges $(u,v),(u',v)$. Now condition on this outcome.

W.l.o.g. assume that the maps $\Phi_{u,v}$ and $\Phi_{u',v}$ are bijections. Then there is a single symbol i so that $A(u) = i = A(u')$. Then both vertices (u,x) and (u',y) will end up on different sides of the cut with probability

$$\Pr_{y \sim N_\rho(x)}[x_i \neq y_i] = \frac{1}{2}(1-\rho)$$

which gives the claim. □

8.4 Soundness

Analyzing the soundness however will take a lot more effort and some heavy Fourier analysis machinery. For the soundness direction we need to turn a cut into a good unique games assignment. Recall that the constructed graph G' has vertices $V' := L \times \{-1,1\}^{\Sigma_L}$. But instead of thinking of a cut as the set $U \subseteq V'$ we will rather work with the functions $f_u : \{-1,1\}^{\Sigma_L} \to \{-1,1\}$ so that $U = \{(u,x) \in V' : f_u(x) = 1\}$.

If these functions have significant Fourier coefficients $|\hat{f}_u(S)|$ for small $|S|$ where $S \subseteq \Sigma_L$, then similar to the argument in Prop 4.21 we would be again optimistic that we could extract a good assignment for $A(u)$ by sampling from such a set S. So the difficult case would be where we have no significant Fourier weight on the lower levels and we expect that there cannot be any good unique games solution. Now for the sake of argument imagine all the functions Φ_e are identities and the functions $f_u := f$ are all identical. Then the value of the maxcut solution would be

$$\Pr_{\substack{x\sim\{-1,1\}^{\Sigma_L},\\ y\sim N_\rho(x)}}[f(x)\neq f(y)]$$

Proving that this is less than the value of $\frac{1}{2}(1-\rho)$ obtained for the completeness case is exactly what is done by the Majority is Stablest Theorem.

Proposition 8.4 (Soundness of reduction)**.** *Fix $-1\leq\rho\leq 0$ and Ψ. Suppose there is a cut $U\subseteq V'$ in the constructed graph (G',w') of value $\frac{1}{\pi}\arccos(\rho)+\delta$. Then* $\mathsf{val}(\Psi)\geq c(\delta,\rho)>0$.

Proof. Let $\{f_u\}_{u\in L}$ with $f_u:\{\pm1\}^{\Sigma_L}\to\{\pm1\}$ be the functions representing the cut, i.e. $U=\{(u,x):f_u(x)=1\}$. For a bijective function $\Phi:\Sigma_L\to\Sigma_R$ and $x\in\{-1,1\}^{\Sigma_R}$ we write $\Phi^{-1}(x)=(x_{\Phi^{-1}(i)})_{i\in\Sigma_R}\in\{-1,1\}^{\Sigma_L}$ as the vector with permuted coordinates. Similarly for a set $S\subseteq\Sigma_R$, $\Phi^{-1}(S)=\{\Phi^{-1}(i):i\in S\}$ gives the permutation of the elements in S. Recall that we only have functions f_u defined for vertices u on the left side. However, we want to extend those functions to the right side. We set

$$g_v(x):=\mathop{\mathbb{E}}_{u\sim N(v)}\left[f_u(\Phi_{u,v}^{-1}(x))\right]\quad\forall v\in R\ \forall x\in\Sigma_R$$

Intuitively, $v\in R$ obtains its function values by averaging over the values of its neighbors. Note that $g_v:\{-1,1\}^{\Sigma_R}\to[-1,1]$ is in general not a boolean function. First we can relate the value of the cut to the stability of those functions g_v.
Claim I. *One has $w'(\delta_{G'}(U))=\frac{1}{2}(1-\mathbb{E}_{v\sim R}[Stab_\rho(g_v)])$.*
Proof of Claim I. As in the reduction, let $v\sim R$, then $u,u'\sim N(v)$, $x\sim\{-1,1\}^{\Sigma_R}$ and $y\sim N_\rho(x)$ independently. It will be useful to also draw $\nu\sim N_\rho(\mathbf{1})\in\{-1,1\}^{\Sigma_R}$ independently. We note that y has the same distribution as $x\odot\nu$. We use this fact

to write

$$\begin{aligned}
w'(\delta_{G'}(U)) &= \Pr_{v,u,u',x,y}\left[f_u(\Phi_{u,v}^{-1}(x)) \neq f_{u'}(\Phi_{u,v}^{-1}(y))\right] \\
&= \Pr_{v,u,u',x,\nu}\left[f_u(\Phi_{u,v}^{-1}(x)) \neq f_{u'}(\nu \odot \Phi_{u',v}^{-1}(x))\right] \\
&= \frac{1}{2}\Big(1 - \mathop{\mathbb{E}}_{v,u,u',x,\nu}\left[f_u(\Phi_{u,v}^{-1}(x)) \cdot f_{u'}(\nu \odot \Phi_{u',v}^{-1}(x))\right]\Big) \\
&= \frac{1}{2}\Big(1 - \mathop{\mathbb{E}}_{v,x,\nu}\Big[\underbrace{\mathop{\mathbb{E}}_{u\sim N(v)}\left[f_u(\Phi_{u,v}^{-1}(x))\right]}_{=g_v(x)} \cdot \underbrace{\mathop{\mathbb{E}}_{u'\sim N(v)}\left[f_{u'}(\nu \odot \Phi_{u',v}^{-1}(x))\right]}_{=g_v(\nu\odot x)}\Big]\Big) \\
&= \frac{1}{2}\Big(1 - \mathop{\mathbb{E}}_{v}\Big[\mathop{\mathbb{E}}_{x,\nu}\left[g_v(x) \cdot g_v(\nu \odot x)\right]\Big]\Big) \\
&= \frac{1}{2}\Big(1 - \mathop{\mathbb{E}}_{v}\left[\mathrm{Stab}_\rho(g_v)\right]\Big) \qquad \Box
\end{aligned}$$

Next, we prove that the Fourier coefficients of g_v are simply the averages of the Fourier coefficients of f_u with $u \in N(v)$ (actually this is a simple consequence of the linearity of the Fourier coefficients).

Claim II. *For $v \in R$ and $S \subseteq \Sigma_R$, one has $\hat{g}_v(S) = \mathbb{E}_{u\sim N(v)}[\hat{f}_u(\Phi_{u,v}^{-1}(S))]$.*
Proof of Claim II. We can write

$$\begin{aligned}
\hat{g}_v(S) &= \mathop{\mathbb{E}}_{x\sim\{\pm1\}^{\Sigma_R}}\left[g_v(x)\cdot\chi_S(x)\right] = \mathop{\mathbb{E}}_{u\in N(v)}\Big[\mathop{\mathbb{E}}_{x\sim\{\pm1\}^{\Sigma_R}}\left[f_u(\Phi_{u,v}^{-1}(x))\cdot\chi_S(x)\right]\Big] \\
&= \mathop{\mathbb{E}}_{u\sim N(v)}\Big[\mathop{\mathbb{E}}_{y\sim\{\pm1\}^{\Sigma_L}}\left[f_u(y)\cdot\chi_{\Phi_{u,v}^{-1}(S)}(y)\right]\Big] = \mathop{\mathbb{E}}_{u\sim N(v)}\left[\hat{f}_u(\Phi_{u,v}^{-1}(S))\right]
\end{aligned}$$

From Claim I, we know that

$$\mathop{\mathbb{E}}_{v\sim R}\left[\mathrm{Stab}_\rho(g_v)\right] = 1 - 2w'(\delta_{G'}(U)) \leq 1 - \frac{2}{\pi}\arccos(\rho) - 2\delta$$

We call a vertex $v \in R$ *good* if $\mathrm{Stab}_\rho(g_v) \leq 1 - \frac{2}{\pi}\arccos(\rho) - \delta$ and we denote $R_{\mathrm{good}} \subseteq R$ as the good vertices. By the Reverse Markov inequality (Lem 1.40) and the fact that stability is in $[-1,1]$, we know that $|R_{\mathrm{good}}| \geq \frac{\delta}{2}|R|$ and so it suffices to find an assignment that satisfies a constant fraction of edges incident to good vertices. We fix values for d and τ that in the Majority is Stablest III Theorem (Theorem 7.11) work for parameters ρ and $\eta := \delta/2$. That means for any function $f : \{-1,1\}^n \to [-1,1]$ with $\mathrm{Inf}_i^{\leq d}[f] \leq \tau$ for all $i \in [n]$, one has[1]

$$\mathrm{Stab}_\rho[f] \geq 1 - \frac{2}{\pi}\arccos(\rho) - \frac{\delta}{2}$$

---

[1] Being picky one could say that we actually use that the majority function is *least* stable in the regime $-1 \leq \rho \leq 0$.

Claim III. *For each $v \in R_{good}$ there is an $i \in \Sigma_R$ so that $Inf_i^{\le d}[g_v] \ge \tau$.*
Proof of Claim III. If there was no such i, then the Majority is Stablest Theorem would give that $\mathsf{Stab}_\rho(g_v) \ge 1 - \frac{2}{\pi}\arccos(\rho) - \delta/2$ which is a contradiction. □
For $u \in L$ and $v \in R$, let

$$\mathsf{List}_\tau(u) := \left\{ i \in \Sigma_L \mid \mathrm{Inf}_i^{\le d}[f_u] \ge \tau \right\} \quad \text{and} \quad \mathsf{List}_\tau(v) := \left\{ i \in \Sigma_R \mid \mathrm{Inf}_i^{\le d}[g_v] \ge \tau \right\}$$

be the list of influential coordinates. From Claim III and Lemma 1.37 we know that $1 \le |\mathsf{List}_\tau(v)| \le \frac{d}{\tau}$ for each $v \in R_{\mathrm{good}}$. So sampling from $\mathsf{List}_\tau(v)$ will be a good idea to get labellings for vertices in R_{good}. But we still need to find suitable labels for the left hand side L, but fortunately there is some consistency in the influential coordinates between both sides.

Claim IV. *For $v \in R_{good}$ and $i \in \Sigma_R$ with $Inf_i^{\le d}[g_v] \ge \tau$ one has* $\Pr_{u \sim N(v)}[Inf^{\le d}_{\Phi^{-1}_{u,v}(i)}[f_u] \ge \frac{\tau}{2}] \ge \frac{\tau}{2}$.
Proof of Claim IV. In order to simplify notation, let us assume w.l.o.g. that all the maps $\Phi_{u,v}$ for $u \in N(v)$ are identities. Then

$$\begin{aligned} \mathop{\mathbb{E}}_{u \sim N(v)}[\mathrm{Inf}_i^{\le d}[f_u]] &= \mathop{\mathbb{E}}_{u \sim N(v)}\Big[\sum_{|S| \le d, i \in S} \hat{f}_u(S)^2\Big] = \sum_{|S| \le d, i \in S} \mathop{\mathbb{E}}_{u \sim N(v)}\big[\hat{f}_u(S)^2\big] \\ &\overset{\text{Jensen}}{\ge} \sum_{|S| \le d, i \in S} \Big(\underbrace{\mathop{\mathbb{E}}_{u \sim N(v)}[\hat{f}_u(S)]}_{=\hat{g}_u(S)}\Big)^2 = \mathrm{Inf}_i^{\le d}[g_u] \ge \tau \end{aligned}$$

Then by Reverse Markov (Lem 1.40) one has $\Pr_{u \sim N(v)}[\mathrm{Inf}_i^{\le d}[f_u] \ge \frac{\tau}{2}] \ge \tau$ which gives the claim. □

Now we can define partial assignment $A : L \cup R_{\mathrm{good}} \to \Sigma$ that satisfies a constant fraction of edges. For $v \in R_{\mathrm{good}}$, select any $A(v) \in \mathsf{List}_\tau(v)$ (which exists by Claim III). For $u \in L$, draw $A(u) \sim \mathsf{List}_{\tau/2}(u)$ uniformly (or set $A(u)$ arbitrary if $\mathsf{List}_{\tau/2}(u) = \emptyset$).
Claim VI. *For each $v \in R_{good}$, $\Pr_{A, u \sim N(v)}[(u,v)$ satisfied by $A] \ge \frac{\tau}{2} \cdot \frac{\tau}{2d}$.*
Proof of Claim VI. We abbreviate $i := A(v)$. First we draw $u \sim N(v)$, then with probability at least $\frac{\tau}{2}$ we have $\mathrm{Inf}_j^{\le d}[f_u] \ge \frac{\tau}{2}$ for $j := \Phi^{-1}_{u,v}(i)$. We condition on this event. Then $\Pr[A(u) = j] = \frac{1}{|\mathsf{List}_{\tau/2}(f_u)|} \ge \frac{\tau}{d/2}$. Combining both probabilities gives the claim. □

Finally, the assignment A will satisfy a $\frac{\delta}{2} \cdot \frac{\tau}{2} \cdot \frac{\tau}{2d}$ fraction of edges. □

8.5 Conclusion

Now choosing an optimum value for the parameter ρ, we can derive an optimal hardness for MAXCUT.

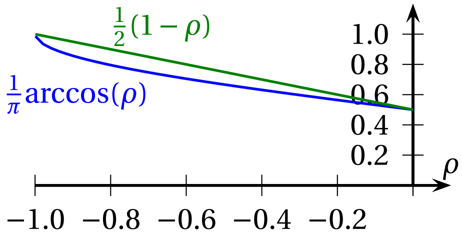
Theorem 8.5. *Assume the Unique Games Conjecture holds. Then for any* $\varepsilon > 0$*, finding a* MAXCUT *approximate solution within* $\alpha_{GW} - \varepsilon$ *is* **NP***-hard where* $\alpha_{GW} \approx 0.878$ *is the same constant as in the Goemans-Williamson SDP rounding algorithm.*

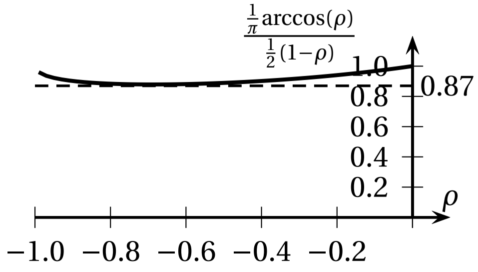
Proof. Combining the result from Lemma 8.3 and Proposition 8.4 we obtain the following. For all $-1 \le \rho \le 0$ and $\delta > 0$, there is a small enough $\eta > 0$ so that the reduction from Section 8.2 satisfies:

(A) $\mathsf{val}(\Psi) \ge 1 - \eta \Longrightarrow \mathsf{val}(G', w') \ge \frac{1}{2}(1-\rho) - \delta$.

(B) $\mathsf{val}(\Psi) \le \eta \Longrightarrow \mathsf{val}(G', w') \le \frac{1}{\pi}\arccos(\rho) + \delta$.

We can make the constant δ as small as desired, hence it remains to maximize the ratio of $\frac{\frac{1}{2}(1-\rho)}{\frac{1}{\pi}\arccos(\rho)}$ over $-1 \le \rho \le 0$. The minimizer is $\rho^* \approx -0.6891$ with a value of approximately 0.8785.



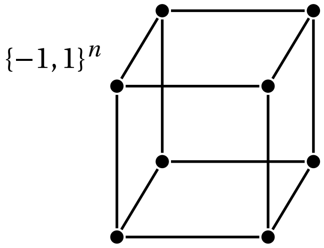


□

Chapter 9

Induced subgraphs of hypercubes

In this chapter we discuss a beautiful result by Huang [Hua19]. For an undirected graph $G = (V, E)$ and a subset of vertices $S \subseteq V$, we write $G[S] := (S, E[S])$ as the *(vertex) induced subgraph* with edgeset $E[S] := \{e \in E \mid e \subseteq S\}$. We also write $\Delta(G)$ as the *maximum degree* of the graph G. This chapter deals with the *hypercube graph* $G_n := (\{-1,1\}^n, E_n)$ where $\{x, y\} \in E_n$ if the *Hamming distance* between x and y is exactly 1.



graph G_n for $n = 3$

We note that the degree of every vertex in G_n is exactly n. We are wondering how small the maximum degree $\Delta(G_n[S])$ of a sizable subset $S \subseteq \{-1,1\}^n$ could be. For example if $S := \{x \in \{-1,1\}^n \mid |\text{ones}(x)| \text{ is even}\}$ then $G[S]$ contains no edge at all and so $\Delta(G[S]) = 0$ and for odd n one has $|S| = 2^{n-1}$. Interestingly as soon as S is one element larger it must contain a lot of edges:

Theorem 9.1. *Let $S \subseteq \{-1,1\}^n$ with $|S| > 2^{n-1}$. Then $\Delta(G_n[S]) \geq \sqrt{n}$.*

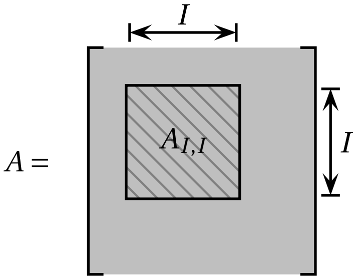
This bound is tight for infinitely many n, see [HKP11, CFGS88][1]. Despite being a claim on the hypercube, the proof uses linear algebraic arguments rather than Fourier analysis.

---

[1] We sketch the construction by [CFGS88]. For that purpose it will be more convinient to use the cube $\{0,1\}^n$. Fix an n so that $\sqrt{n} \in \mathbb{N}$. Partition the variables $[n] = F_1 \dot\cup \ldots \dot\cup F_{\sqrt{n}}$ with $|F_i| = \sqrt{n}$

9.1 Linear algebra background

For a symmetric matrix $A \in \mathbb{R}^{n \times n}$ and $I \subseteq [n]$, $A_{I,I} \in \mathbb{R}^{|I| \times |I|}$ is the *principal submatrix* with entries $(A_{I,I})_{ij} = A_{ij}$ for $i, j \in I$.



Any symmetric $n \times n$ matrix A has only real Eigenvalues which we denote by $\lambda_1(A) \geq \ldots \geq \lambda_n(A)$. We denote the *singular values* (which are the absolute values of the Eigenvalues) by $\sigma_1(A) \geq \ldots \geq \sigma_n(A) \geq 0$. If A has all Eigenvalues in the interval $[a,b]$ then also a principal submatrix must have all Eigenvalues in $[a,b]$. In fact, the following more precise relationship is known:

Lemma 9.2 (Cauchy-Interlacing Theorem)**.** *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix with a principal submatrix $B := A_{I,I}$ where $I \subseteq [n]$. Then*

$$\lambda_i(A) \geq \lambda_i(B) \geq \lambda_{i+n-|I|}(A) \quad \forall i = 1, \ldots, |I|$$

We also need the following:

Lemma 9.3. *For any symmetric matrix $A \in [-1,1]^{n \times n}$ there is a row $i \in [n]$ with $|supp(A_i)| \geq \sigma_1(A)$.*

Proof. Let $\lambda \in \mathbb{R}$ be the Eigenvalue with largest absolute value and let $v \in \mathbb{R}^n$ be the corresponding Eigenvector. Let $i \in [n]$ be a row index with $|v_i| = \|v\|_\infty$. Then

$$|\lambda| \cdot |v_i| \overset{v \text{ Eigenvector}}{=} |(Av)_i| \leq \underbrace{\sum_{j=1}^{n} |A_{ij}|}_{\leq |\text{supp}(A_i)|} \underbrace{|v_j|}_{\leq |v_i|} \leq |\text{supp}(A_i)| \cdot |v_i|$$

Rearranging gives the claim. ☐

---

for all i. Define

$$S := \left\{\mathbf{1}_T : T \subseteq [n] \text{ with } |T| \text{ is even and } \exists i : F_i \subseteq T\right\} \cup \left\{\mathbf{1}_T : T \subseteq [n] \text{ with } |T| \text{ is odd and } \forall i : F_i \not\subseteq T\right\}$$

Then one can prove that $|S| \in \{2^{n-1}+1, 2^{n-1}-1\}$ depending on the parity of $n + \sqrt{n}$. Moreover, $\Delta(G_n[S]), \Delta(G_n[\bar{S}]) \leq \sqrt{n}$. See [CFGS88] for details.

9.2 The adjacency matrix of the hypercube graph

For an undirected graph $G = (V, E)$, the *adjacency matrix* is the symmetric matrix $A \in \{0,1\}^{V\times V}$ with $A_{ij} = 1 \Leftrightarrow \{i, j\} \in E$. We also write A_G if we want to emphasize the graph. For a matrix $A \in \mathbb{R}^{m\times n}$ we write $|A| \in \mathbb{R}^{m\times n}$ as the matrix with entries $|A|_{ij} := |A_{ij}|$.

Proposition 9.4. *Recursively define a matrix* $A_n \in \{-1,0,1\}^{2^n\times 2^n}$ *with*

$$A_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \textit{and} \quad A_n := \begin{pmatrix} A_{n-1} & I_{2^{n-1}} \\ I_{2^{n-1}} & -A_{n-1} \end{pmatrix} \textit{ for } n \geq 2$$

Then

- *(A)* $|A_n|$ *is the adjacency matrix of* G_n*.*
- *(B)* A_n *has* 2^{n-1} *many Eigenvalues* $\sqrt{n}$ *and* 2^{n-1} *many Eigenvalues* $-\sqrt{n}$*.*

Proof. For (A) we can imagine to construct the hybercube graph G_n by taking two copies of G_{n-1} and inserting edges $\{(x,-1),(x,+1)\}$ for $x \in \{-1,1\}^{n-1}$. This is exactly the recursive definition of $|A_n|$. In order to prove (B) we first show the following:
Claim I. *For all* $n \in \mathbb{N}$ *one has* $A_n^2 = nI_{2^n}$*.*
Proof of Claim. We prove the claim by induction over n where the base case $n = 1$ is trivial. For the induction step we have

$$A_n^2 = \begin{pmatrix} A_{n-1}^2 + I_{2^{n-1}} & \mathbf{0} \\ \mathbf{0} & A_{n-1}^2 + I_{2^{n-1}} \end{pmatrix} \overset{\text{induction}}{=} \begin{pmatrix} nI_{2^{n-1}} & \mathbf{0} \\ \mathbf{0} & nI_{2^{n-1}} \end{pmatrix} = nI_{2^n}$$

That means all Eigenvalues of A_n^2 are n and hence the Eigenvalues of A_n must be of the form $\pm\sqrt{n}$. Since $\mathrm{Tr}[A_n] = 0$, there must be an equal number of $+\sqrt{n}$ Eigenvalues and $-\sqrt{n}$ Eigenvalues. $\square$

9.3 The main proof

Now we can give the proof of the main result:

Proof of Theorem 9.1. Let $A := A_n$ be the matrix constructed in Prop 9.4 which has the property that $|A|$ is the adjacency matrix of G_n. Let $S \subseteq \{-1,1\}^n$ be a subset of the hypercube vertices with $S > 2^{n-1}$. Consider the principal submatrix $B := A_{S,S}$ (and note that $|B|$ is the adjacency matrix of the induced subgraph $G_n[S]$). By

Prop 9.4.(B), half the Eigenvalues of A are $+\sqrt{n}$ and so by the Cauchy-Interlacing-Theorem we have $\lambda_1(B) \geq \sqrt{n}$. Then by Lemma 9.3 there is a row $i \in S$ with

$$|\delta_{G[S]}(i)| = |\mathrm{supp}(B_i)| \geq \sqrt{n}.$$

□

9.4 The Sensitivity of Boolean Functions

We also want to show an interesting application of [Hua19]. Recall that for a function $f : \{-1,1\}^n \to \mathbb{R}$ we denote its *degree* as $\deg(f) := \max\{|S| : \hat{f}(S) \neq 0\}$. In other words, $\deg(f)$ is the maximum degree of f, when considering it as a multilinear polynomial. We make some new definitions:

Definition 9.5. For $f : \{-1,1\}^n \to \{-1,1\}$ we define the *sensitivity at* $x \in \{-1,1\}^n$ as

$$S(f,x) := \left|\left\{i \in [n] : f(x) \neq f(x^{\oplus i})\right\}\right|$$

where $x^{\oplus i}$ is the vector x with the ith bit flipped. The *sensitivity* of f is then

$$S(f) := \max_{x \in \{-1,1\}^n} S(f,x)$$

In other words, the sensitivity tells us how many Hamming neighbors may have a different function value. A consequence of Huang's work [Hua19] is the following (where the connection had been known before due to Gotsman and Linial [GL92]).

Theorem 9.6. *For any boolean function $f : \{-1,1\}^n \to \{-1,1\}$ one has $S(f) \geq \sqrt{\deg(f)}$.*

Proof. We first prove the claim for the case that the function has the maximum possible degree of n:
Claim I. *If $f : \{-1,1\}^n \to \{-1,1\}$ has $\deg(f) = n$, then $S(f) \geq \sqrt{n}$.*
Proof of Claim I. The assumption tells us that $\hat{f}([n]) \neq 0$. So suppose w.l.o.g. that $\hat{f}([n]) > 0$. Consider the function $g : \{-1,1\}^n \to \{-1,1\}$ with $g(x) := f(x) \cdot \chi_{[n]}(x)$. Then $\mathbb{E}_{x \sim \{-1,1\}^n}[g(x)] = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x) \cdot \chi_{[n]}(x)] = \hat{f}([n]) > 0$. That means if we define $S := \{x \in \{-1,1\}^n : g(x) = 1\}$ then $|S| > 2^{n-1}$. By Theorem 9.1 we can fix a point $x \in S$ that has at least $\sqrt{n}$ Hamming neighbors in S. Consider one such neighbor $x^{\oplus i} \in S$. Then

$$f(x) \cdot x_i \cdot \chi_{[n]\setminus\{i\}}(x) = g(x) = 1 = g(x^{\oplus i}) = f(x^{\oplus i}) \cdot (-x_i) \cdot \chi_{[n]\setminus\{i\}}(x)$$

which can be rearranged to $f(x) \neq f(x^{\oplus i})$. Hence $S(f,x) \geq \sqrt{n}$. □

Now back to the main claim. Consider a function $f : \{-1,1\}^n \to \{-1,1\}$. Let $S \subseteq [n]$ be one of the sets where f attains the degree, i.e. $|S| = \deg(f)$ and $\hat{f}(S) \neq 0$. Fix an arbitrary $z \in \{-1,1\}^{[n]\setminus S}$ and consider the *restriction* $f_{S|z} : \{-1,1\}^S \to \{-1,1\}$ with $f_{S|z}(x) := f(x,z)$ for $x \in \{-1,1\}^S$. From Prop 1.15 we know that

$$\widehat{f_{S|z}}(S) = \sum_{T \subseteq [n]\setminus S} \underbrace{\hat{f}(S \cup T)}_{=0 \text{ for } T \neq \emptyset} \chi_T(z) = \hat{f}(S) \neq 0$$

That means $\deg(f_{S|z}) = |S| = \deg(f)$. Then by Claim I, $S(f) \geq S(f_{S|z}) \geq \sqrt{\deg(f)}$.

□

This is result is tight in the sense that for infinitely many d and all $n \geq d$ there is a function[2] $f : \{-1,1\}^n \to \{-1,1\}$ with $\deg(f) = d$ and $S(f) \leq \sqrt{d}$. We will discuss more relationships between sensitivity and the degree of boolean functions in the following Chapter 10.

---

[2] Take the set $S \subseteq \{-1,1\}^d$ described in the footnote below Theorem 9.1. Then set $f(x,y) := 1 \Leftrightarrow x \in S$ for all $x \in \{-1,1\}^d$ and $y \in \{-1,1\}^{n-d}$. This function has indeed degree d and $S(f) \leq \sqrt{d}$.

Chapter 10

Bounded low-degree functions and the Aaronson-Ambainis Conjecture

In this chapter we want to study functions of the form $f : \{-1,1\}^n \to [-1,1]$, meaning they are bounded but do not neccessarily have values in $\{-1,1\}$. In particular, we will be interested in functions that additionally have low degree. Aaronson and Ambainis [AA14] came across such functions in the context of the *query complexity* of *quantum computers*. They made the conjecture that any low degree bounded function must have an influential variable:

Conjecture 2 (Aaronson-Ambainis [AA14])**.** *Any function $f : \{-1,1\}^n \to [-1,1]$ of degree d has a coordinate $i \in [n]$ so that*

$$Inf_i[f] \geq poly\left(\frac{Var[f]}{d}\right)$$

We recall from Section 1.9 and Section 1.8 that variance and influence can be expressed as

$$\mathrm{Var}[f] = \sum_{\emptyset \subset S \subseteq [n]} \hat{f}(S)^2 \quad \text{and} \quad \mathrm{Inf}_i[f] = \sum_{S \subseteq [n] : i \in S} \hat{f}(S)^2$$

We also recall that the total influence of a function f is $I[f] := \sum_{i=1}^n \mathrm{Inf}_i[f]$ and we abbreviate the maximum influence by $\mathrm{Inf}_{\max}[f] := \max_{i \in [n]} \mathrm{Inf}_i[f]$. Also recall for any degree-d function $f : \{-1,1\}^n \to [-1,1]$ one has $I[f] \leq d$ (see Lemma 1.33).

Naively, say for $d = O(1)$, just by accounting the Fourier weights it would appear possible that $I[f] = \Theta(1)$ (equivalently $\mathrm{Var}[f] = \Theta(1)$) and $\mathrm{Inf}_i[f] = \Theta(\frac{1}{n})$ for all i. Hence it has to be the boundedness that enforces limitations on how the Fourier weight can be distributed. To illustrate the issue, let us consider the case of $d = 1$ with a linear function $f(x) = \sum_{i=1}^n a_i x_i$ so that $|f(x)| \leq 1$ for all $x \in \{-1,1\}^n$.

The variance of such a function is $\mathrm{Var}[f] = \|a\|_2^2$ and the maximum influence is $\|a\|_\infty^2$. As the function is bounded by 1, we know that $\|a\|_1 \leq 1$. On the other hand, Generalized Cauchy-Schwarz gives that $\|a\|_2^2 \leq \|a\|_1 \|a\|_\infty$ which can be used to derive

$$\mathrm{Inf}_{\max}[f] = \|a\|_\infty^2 \geq \frac{\|a\|_2^4}{\|a_1\|_1^2} \overset{\|a\|_1 \leq 1}{\geq} \mathrm{Var}[f]^2$$

We conclude that Conjecture 2 is indeed true[1] for $d = 1$. On the other hand, we can make the observation that while the maximum function value is $\max_{x \in \{-1,1\}} |f(x)| = \|a\|_1$, the *average* value is rather $\mathbb{E}_{x \sim \{-1,1\}^n}[|f(x)|] \asymp \|a\|_2$. That means, the proof has to necessarily make use of function values that are extremely rare.

At the time of this writing, the Aaronson-Ambainis Conjecture is still open. Inspired by the terrific survey of Backurs [Bac12] we would like to explain the state of the art of what is known towards this conjecture. In particular we discuss the following results:

- We explain the original motivation by Aaronson and Ambainis [AA14] in the area of quantum computing and how their conjecture implies that bounded low degree polynomials have a low average query complexity (see Section 10.1).
- We prove that Conjecture 2 is true for functions $f : \{-1,1\}^n \to \{-1,1\}$. For that purpose we take a detour and prove that low degree functions have low depth *decision trees*. In a second step, we then prove that decision trees have influential variables.
- We prove the currently best known bound for functions $f : \{-1,1\}^n \to [-1,1]$ which are only exponential, rather than polynomial. To be precise, one has $\mathrm{Inf}_{\max}[f] \geq \frac{\sqrt{\mathrm{Var}[f]}}{C^d}$ for a universal constant $C > 0$.
- A result by Lovett and Zhang [LZ23] shows that the *fractional block sensitivity* of a degree-d function $f : \{-1,1\}^n \to [-1,1]$ is small. We reproduce an improved and simplified bound due to Agarwal and Ben-David [ABD24] and show that $FBS(f) \leq O(d^2)$.

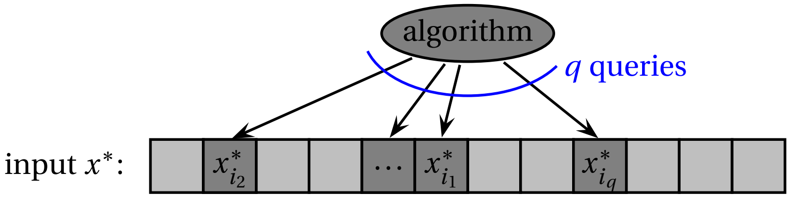
10.1 Average query complexity of bounded functions

We begin by describing a "classical" consequence that an affirmative answer to Conjecture 2 would bring; in the next Section 10.2 we will then explain how this insight can be used in the context of quantum computing.

---

[1] The reader may note that for $a := (\frac{1}{n}, \ldots, \frac{1}{n})$ one has $\mathrm{Inf}_{\max}[f] = \frac{1}{n^2}$ and $\mathrm{Var}[f] = \frac{1}{n}$ and so the exponent of 2 cannot be improved even if the degree is $d = 1$.

Consider a function $f : \{-1,1\}^n \to [-1,1]$. We want to study algorithms that know the function f, but only have *adaptive query access* to the input $x^* \in \{-1,1\}^n$. In other words, the algorithm produces a sequence $i_1, \ldots, i_q \in [n]$ of indices and receives the bits $x^*_{i_1}, \ldots, x^*_{i_q} \in \{-1,1\}$. Here *adaptive* means that the choice of the index i_j may depend on the outcomes of the bits $x^*_{i_1}, \ldots, x^*_{i_{j-1}}$.



At the end the algorithm should output a number that is close to $f(x^*)$ (without having seen all the input x^*). Clearly, some structure is needed for the function f if we want to make sense out of a lot less than n queried bits. And in fact, Aaronson and Ambainis [AA14] have proven that for bounded low-degree functions few queries suffice on average. We would like to emphasize that this is indeed only true for an *average* input.

Theorem 10.1 ([AA14])**.** *Assume the Aaronson-Ambainis Conjecture 2 is true. Then for any degree-d function $f : \{-1,1\}^n \to [-1,1]$ and any $\varepsilon > 0$ there is a randomized algorithm A that makes poly$(d, \frac{1}{\varepsilon})$ many adaptive queries before producing an output so that*

$$\Pr_{x^* \sim \{-1,1\}^n}[|f(x^*) - A(x^*)| > \varepsilon] \le \varepsilon$$

Proof. The algorithm that we will be using is as follows:

```
ALGORITHM A
Input: Degree-$d$ function $f : \{-1,1\}^n \to [-1,1]$. Query access to random input $x^* \sim \{-1,1\}^n$
Output: Estimate on $f(x^*)$
  (1) If $\mathrm{Var}[f] \le \varepsilon^4$ then return $\hat{f}(\emptyset) = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x)]$.
  (2) Select an index $i \in [n]$ with $\mathrm{Inf}_i[f] \ge \delta := \mathrm{poly}(\varepsilon/d)$.
  (3) Query $x^*_i$.
  (4) Recurse on the function $f_{\{i\}|x^*_i} : \{-1,1\}^{n-1} \to [-1,1]$ which is the
      restriction of $f$ on $\{i\}$ using $x^*_i$.
```

We will now prove by induction over n that the algorithm works. First consider the case that $\mathrm{Var}[f] \le \varepsilon^4$ so that we terminate immediately in (1). In that case, as the input is random, its expected error is

$$\mathop{\mathbb{E}}_{x^* \sim \{-1,1\}^n}[|f(x^*) - \hat{f}(\emptyset)|] \le \mathop{\mathbb{E}}_{x^* \sim \{-1,1\}^n}[|f(x^*) - \hat{f}(\emptyset)|^2]^{1/2} = \mathrm{Var}[f]^{1/2} \le \varepsilon^2$$

using Jensen's inequality (Theorem 1.45). Hence $\Pr_{x^*\sim\{-1,1\}^n}[|f(x^*)-\hat{f}(\emptyset)|\geq\varepsilon]\leq\varepsilon$ by Markov's inequality. That means the algorithm is indeed making more than an ε-error on at most an ε-fraction of inputs. Similarly, if the algorithm recurses, then it recurses on the correct "subcube" and the non-queried input is still uniform.

Hence, it only remains to prove that the algorithm indeed terminates after at most $\mathrm{poly}(d,\frac{1}{\varepsilon})$ many recursions. As we only recurse in (2) when $\mathrm{Var}[f]>\varepsilon^4$, by the Aaronson-Ambainis Conjecture 2, there must be some index i so that $\mathrm{Inf}_i[f]\geq\delta$ where $\delta=\mathrm{poly}(d/\varepsilon)$. Now, instead of considering the variance, we analyze how the total influence of the function changes. Recall that the original function has $I[f]\leq\deg(f)\leq d$ by Lemma 1.33. Consider the very first recursion on some coordinate i. As the queried input is assumed to be random, the total influence of the next function is

$$\begin{aligned}
\mathop{\mathbb{E}}_{x_i^*\sim\{-1,1\}}[I[f_{\{i\}|x_i^*}]] &\overset{\text{Thm 1.32.(ii)}}{=} \sum_{S\subseteq[n]}|S|\cdot\mathop{\mathbb{E}}_{x_i^*\sim\{-1,1\}}[\widehat{f_{\{i\}|x_i^*}}(S)^2]\\
&\overset{\text{Prop 1.16.(d)}}{=} \sum_{S\subseteq[n]\setminus\{i\}}|S|\cdot(\hat{f}(S)^2+\hat{f}(S\cup\{i\})^2)\\
&= \sum_{S\subseteq[n]}|S|\cdot\hat{f}(S)^2-\underbrace{\sum_{S\subseteq[n]:i\in S}\hat{f}(S)^2}_{=\mathrm{Inf}_i[f]\geq\delta}\leq I[f]-\delta
\end{aligned}$$

More intuitively, each set $S\subseteq[n]$ with $i\in S$ contributes $|S|\cdot\hat{f}(S)^2$ to the total influence of f but only $(|S|-1)\cdot\hat{f}(S)^2$ to the expected total influence of the next function and the difference when accumulated over all sets is indeed the influence of i. We can conclude that the expected number of iterations until the algorithm terminates is at most[2] $\frac{I[f]}{\delta}\leq\frac{d}{\delta}$ and the probability to not have terminated after $\frac{d}{\varepsilon\delta}$ iterations is at most[3] ε by Markov's inequality. □

---

[2] The right way too see this is as follows: suppose we have a random process $X_0,X_1,\ldots$ where $X_t\in\mathbb{R}_{\geq0}$ for all t and $X_0>0$ is some fixed value. The process is *active* for T iterations (T is also called the *stopping time*) and then becomes *inactive*. The number $T\in\mathbb{Z}_{\geq1}$ is a random variable as well. Let $Y_t\in\{0,1\}$ be the indicator random variable telling if the process is still active in iteration t, i.e. $1=Y_0=\ldots=Y_{T-1}$ and $0=Y_T=Y_{T+1}\ldots$. In particular $T=\sum_{t=0}^{\infty}Y_t$. As long as the random process is active, it decreases in expectation, i.e. $\mathbb{E}[X_{t+1}-X_t\mid Y_t=1]\leq-\delta$ while it freezes afterwards, i.e. $X_{t+1}=X_t$ if $Y_t=0$. Then by linearity of expectation

$$X_0\geq\sum_{t=0}^{\infty}\mathbb{E}[X_t-X_{t+1}]=\sum_{t=0}^{\infty}\Pr[Y_t=1]\cdot\mathbb{E}[X_t-X_{t+1}\mid Y_t=1]\geq\delta\sum_{t=0}^{\infty}\mathbb{E}[Y_t]=\delta\,\mathbb{E}[T]$$

which can be rearranged to $\mathbb{E}[T]\leq\frac{X_0}{\delta}$.

[3] We can modify the algorithm and agree to return 0 if the number of recursions have exceeded our limit which lets us incur another ε.

We should note that the algorithm is deterministic as long as we assume that variances, influences and expectations can be computed for all restrictions. We also note that Theorem 10.1 could be rephrased as the statement that any function $f : \{-1,1\}^n \to [-1,1]$ has a *decision tree* of depth at most $\mathrm{poly}(d/\varepsilon)$ so that the expected error (on average over the inputs) is at most ε.

10.2 Query complexity for quantum computers

Now we come to the application to *quantum computing* that was the motivation of Aaronson and Ambainis [AA14]. Again, there is some input $x \in \{-1,1\}^n$ but now we have a quantum computer that has query access to the input. It is beyond the scope of these notes to explain the query model of quantum computing, but a very readable introduction can be found in the survey of Buhrman and de Wolf [Bd02]. For a more extensive introduction to quantum computing in general we recommend the popular textbook of Nielsen and Chuang [NC00]. Quantum algorithms are inherently randomized and after making some number q of queries to the input x, the algorithm accepts with some probability $Q(x)$. This gives us a function $Q : \{-1,1\}^n \to [0,1]$ that represents the acceptance probability of the quantum algorithm. The only fact on quantum computers that we then need is the following result by Beals at al.

Theorem 10.2 ([BBC$^+$01])**.** *Suppose a quantum algorithm makes q many queries to an input $x \in \{-1,1\}^n$. Then the acceptance probability $Q : \{-1,1\}^n \to [0,1]$ is a multi-linear real polynomial with $\deg(Q) \leq 2q$.*

Then combining this fact with Theorem 10.1 we can conclude:

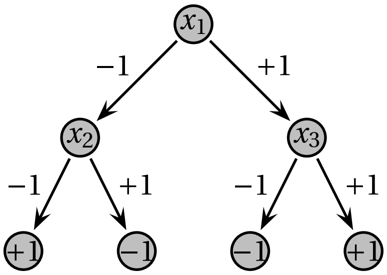
Theorem 10.3. *Assume the Aaronson-Ambainis Conjecture 2 is true. Suppose a quantum algorithm makes q queries to an input $x \in \{-1,1\}^n$ and let $Q(x) \in [0,1]$ be the acceptance probability on input x. Then there is a classical randomized algorithm A that makes $\mathrm{poly}(q, \frac{1}{\varepsilon})$ many queries and satisfies*

$$\Pr_{x \sim \{-1,1\}^n}[|Q(x) - A(x)| \geq \varepsilon] \leq \varepsilon$$

We would like to emphasize that the classical algorithm is only able to approximate the answer on average over the inputs. Also, the algorithm would need to have access to the polynomial Q and be able to compute variances and influences for restrictions.

10.3 Decision trees

The material from this section is mainly taken from the survey of Buhrman and de Wolf [Bd02]. A *decision tree* is a binary tree with a distinguished root in which each interior node is labeled with a variable from $x_1, \dots, x_n$ and each leaf is labeled with an output from $\{-1,1\}$. Moreover, each edge is labeled with a number -1 or $+1$. Given an input $x^* \in \{-1,1\}^n$, we can follow the unique path from the root to a leaf where at an interior node labeled with x_i we take the -1 arc if $x_i^* = -1$ and otherwise the $+1$ arc. The decision tree then defines a function $f : \{-1,1\}^n \to \{-1,1\}$ where the function value $f(x^*)$ corresponds to the label of the leaf that we reach on input x^*. We are free to query the variables in any order.



decision tree of depth 2

The *depth* of a decision tree is the maximum length of a root-leaf path (in terms of number of edges). Note that in a minimal decision tree, we would never query the same variable twice. Typically one is interested in either minimizing the depth or the size of a decision tree. For our purposes here, it is the depth that matters:

Definition 10.4. For a function $f : \{-1,1\}^n \to \{-1,1\}$, the *decision tree complexity* is

$$D(f) := \min\left\{\text{depth}(T) \mid T \text{ is a decision tree computing } f\right\}$$

One can interpret $D(f)$ as the number of variables that need to be queried in order to determine the function value $f(x)$. Clearly, some functions need decision trees of high depth. For example for the parity function $f(x) = \prod_{i=1}^n x_i$ we always need to query all variables and so $D(f) = n$. Also by a counting argument one can easily estimate that a random function f would have $D(f) \geq n(1 - o(1))$.

10.3.1 Sensitivity and block sensitivity

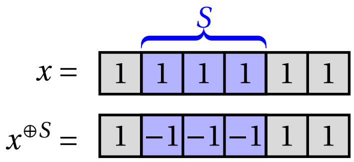
We want to introduce other complexity measures for a function f that turn out to be closely related to $D(f)$. Recall that for point $x \in \{-1,1\}^n$ and $i \in [n]$, we denote the vector with the ith bit flipped by $x^{\oplus i} = (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n)$. We recall a definition that we had earlier given in Sec 9.4.

Definition 10.5. For a function $f : \{-1,1\}^n \to \{-1,1\}$, the *sensitivity at* $x \in \{-1,1\}^n$ is the number of hamming neighbors with different function values, i.e.

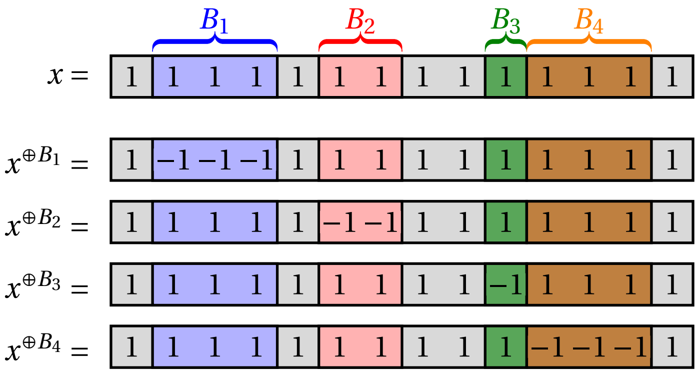
$$S(f,x) := |\{i \in [n] : f(x) \neq f(x^{\oplus i})\}|$$

The *sensitity* of f itself is $S(f) := \max_{x \in \{-1,1\}^n} S(f,x)$.

Next, we introduce a generalization of this quantity. For a set $S \subseteq [n]$, we define $x^{\oplus S}$ as the vector x where precisely the signs of the entries in S are flipped. In particular x and $x^{\oplus S}$ differ in exactly $|S|$ many coordinates.



Definition 10.6. For a function $f : \{-1,1\}^n \to \{-1,1\}$, the *block sensitivity at* $x \in \{-1,1\}^n$, denoted by $BS(f,x)$ is the maximum number b so that there are disjoint sets $B_1, \dots, B_b \subseteq [n]$ so that $f(x) \neq f(x^{\oplus B_i})$ for all $i = 1, \dots, b$. The *block sensitivity* of f itself is again $BS(f) := \max_{x \in \{-1,1\}^n} BS(f,x)$.



It is not hard to see that $S(f) \leq BS(f)$. In the following chapters we want to elaborate that also $BS(f)$ can be bounded by a polynomial in $S(f)$. First, we need to take a detour and discuss a method to lower bound the degree of a polynomial.

10.4 Lower bounds on the degree of a polynomial

In this section we want to discuss methods to prove lower bounds on the degrees of polynomials. For this chapter it will be notationally convinient to work with functions $g : \{0,1\}^n \to \mathbb{R}$ instead of the usual functions on the $\{\pm 1\}^n$-cube. As we are working towards degree lower bound and by Prop 1.9 a shift does not change the degree, this is indeed only a change in notation.

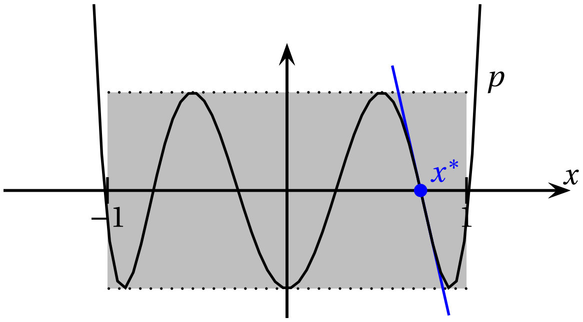
10.4.1 Univariate polynomials

Even though the functions $g : \{0,1\}^n \to \mathbb{R}$ that we are interested in are multivariate polynomials, we begin the discussion with the univariate case. A very classic result is the following which relates the derivative, range and degree of a polynomial:

Theorem 10.7 (Markov brothers' inequality (1890s))**.** *For a univariate polynomial* $p : \mathbb{R} \to \mathbb{R}$ *of degree* d*, one has*

$$\max_{-1\le x\le 1} |p'(x)| \le d^2 \cdot \max_{-1\le x\le 1} |p(x)|$$

This means that if a polynomial stays in some range over a longer interval and we have a lower bound on the derivative, then this implies a lower bound on the degree.



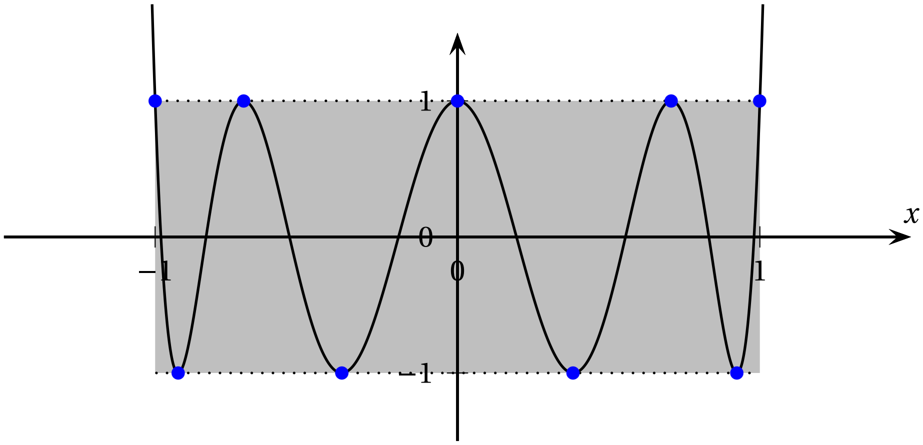
polynomial p with $x^* \in [-1,1]$ maximizing $|p'(x^*)|$

In fact, an inequality can be proven more generally for the k-th derivative in which case it states that

$$\max_{-1\le x\le 1} |p^{(k)}(x)| \le d^{2k} \cdot \max_{-1\le x\le 1} |p(x)|$$

(here we slightly simplified the dependence compared to the tight actual inequality).

Remark 2. The bound from Theorem 10.7 is tight. Let T_d be the d-th Chebychev polynomial which is a univariate degree-d polynomial. Then $|T_d(x)| \le 1$ for all $-1 \le x \le 1$ while $|T_d'(x)| \le T_d'(1) = d^2$ for $-1 \le x \le 1$.



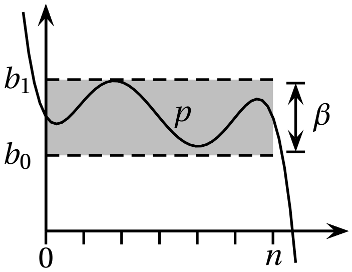
8th Chebychev polynomial $T_8(x)$ with extrema

However, for our purpose it will be more convinient to use a variant where only the function values on a discrete set of points matter. Also, we stretch the interval of consideration from $[-1,1]$ to $[0,n]$.

Theorem 10.8 (Ehlich and Zeller [EZ64], Rivlin and Cheney [RC66])**.** *Let* $p:\mathbb{R}\to\mathbb{R}$ *be a univariate polynomial and let* $b_0 \le p(i) \le b_1$ *for all* $i \in \{0,\ldots,n\}$*,* $\beta := b_1 - b_0$ *and* $\gamma := \max_{x\in[0,n]} |p'(x)|$*. Then*

$$\deg(p) \ge \sqrt{n \cdot \frac{1}{1+\frac{\beta}{\gamma}}}$$

If $0 \le \frac{\gamma}{\beta} \le 1$*, then* $\deg(p) \ge \sqrt{\frac{\gamma n}{2\beta}}$*.*



10.4.2 Symmetrization of multi-variate polynomials

The idea is to apply the degree lower bound for univariate polynomials to the multi-variate case. In order to do so we need to be able to turn a multivariate polynomial into a univariate one. The first step is to symmetrize the polynomial:

Definition 10.9. Given a function $g : \mathbb{R}^n \to \mathbb{R}$, the *symmetrization* is the function $g_{\text{sym}} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$g_{\text{sym}}(x) := \mathop{\mathbb{E}}_{\pi:[n]\to[n]}[g(x_{\pi(1)}, \dots, x_{\pi(n)})]$$

where the expectation is over a uniform random permutation π.

For example the function $g(x_1, x_2, x_3) = x_1 - x_2 x_3 + 1$ has the symmetrization

$$g_{\text{sym}}(x_1, x_2, x_3) = \frac{x_1 + x_2 + x_3}{3} - \frac{x_1 x_2 + x_2 x_3 + x_1 x_3}{3} + 1$$

We summarize a few properties of the symmetrization:

Lemma 10.10. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial. Then*

(a) One has $\deg(g_{sym}) \leq \deg(g)$.

(b) If $|g(x)| \leq 1$ for all $x \in \{0,1\}^n$, then also $|g_{sym}(x)| \leq 1$ for all $x \in \{0,1\}^n$.

Proof. For (a) we observe that

$$\widehat{g_{\text{sym}}}(S) = \mathop{\mathbb{E}}_{T\sim\binom{[n]}{|S|}}[\hat{g}(T)]$$

and so the degree cannot increase when symmetrizing. The symmetrization is obtained by averaging which implies (b).

$\square$

Note that (b) also holds if $\{0,1\}^n$ is replaced by $\{-1,1\}^n$ (or any other set that is closed under taking permutations).

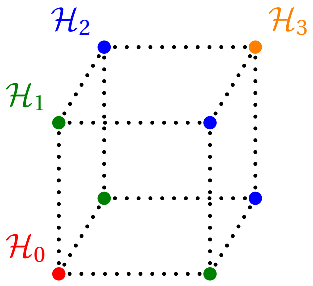
Once we have symmetrized a function it is easy to express it as a univariate polynomial:

Lemma 10.11 (Minsky and Papert [MP69]). *For any function $g : \{0,1\}^n \to \mathbb{R}$, there is a univariate polynomial $p : \mathbb{R} \to \mathbb{R}$ with $p(\sum_{i=1}^n x_i) = g_{sym}(x)$ for all $x \in \{0,1\}^n$ and $\deg(p) = \deg(g_{sym}) \leq \deg(g)$.*

See the survey of Buhrman and de Wolf [Bd02] for a proof.

10.4.3 Degree lower bounds for polynomials on the hypercube

Next, we want to explain how to use symmetrization to make the degree lower bound from Theorem 10.8 work for functions on the hypercube. While this is in preparation of the Theorem of Nisan and Szegedy, we keep it rather general. For $\ell \in \{0, \dots, n\}$, let us define $\mathcal{H}_\ell := \{x \in \{0,1\}^n \mid \sum_{i=1}^n x_i = \ell\}$ as the ℓth *Hamming level* of the hypercube.



Theorem 10.12. *Let* $g:\{0,1\}^n \to [-1,1]$ *and let* $a,b \in \{0,\dots,n\}$ *with* $a \neq b$ *be two Hamming levels. Then*

$$\deg(g) \geq \sqrt{\gamma n/4}$$

where

$$\gamma := \frac{|\mathbb{E}_{x\sim\mathcal{H}_a}[g(x)] - \mathbb{E}_{x\sim\mathcal{H}_b}[g(x)]|}{|a-b|}$$

Proof. Let p be the corresponding univariate polynomial from Lemma 10.11, so that

$$p\Big(\sum_{i=1}^{n} x_i\Big) = g_{\text{sym}}(x) \quad \forall x \in \{0,1\}^n$$

We also note that $|p(\ell)| \leq 1$ for all $\ell = 0,\dots,n$ since $|g(x)| \leq 1$ for all $x \in \{0,1\}^n$, making use of Lemma 10.10.(b). Note that for a particular Hamming level $\ell \in \{0,\dots,n\}$ we have

$$p(\ell) = \mathop{\mathbb{E}}_{x\sim\mathcal{H}_\ell}[g(x)]$$

Then there is a point z between a and b where the derivative of p is at least its average, i.e.

$$|p'(z)| \geq \frac{|p(a)-p(b)|}{|a-b|} = \gamma.$$

Applying Theorem 10.8 with parameters γ and $\beta := 2$, we obtain a degree lower bound of

$$\deg(g) \geq \deg(p) \overset{\text{Thm 10.8}}{\geq} \sqrt{\frac{\gamma n}{4}}$$

□

10.5 The Theorem of Nisan and Szegedy

Now we have everything in place in order to prove the result of Nisan and Szegedy [NS92] which says that any low degree boolean function has low block sensitivity.

Theorem 10.13 (Nisan and Szegedy [NS92]). *For any $f : \{-1,1\}^n \to \{-1,1\}$ one has $BS(f) \le 2\deg(f)^2$.*

Proof. Let $b := BS(f)$ be the block sensitivity of f and assume by symmetry reasons that it is attained at point $x := \mathbf{1}$. Let $B_1, \dots, B_b \subseteq [n]$ be the corresponding disjoint subsets. Again, by symmetry we may assume that the coordinates are sorted in the order of $B_1, \dots, B_b$ followed by all remaining coordinates. We write $\mathbf{1}_{B_i}$ as the all-ones vector with $|B_i|$ many entries. We define a new function $g : \{0,1\}^b \to \{-1,1\}$ by letting

$$g(y_1, \dots, y_b) := f(\mathbf{1} - 2y_1 \cdot \mathbf{1}_{B_1}, \dots, \mathbf{1} - 2y_b \cdot \mathbf{1}_{B_b}, 1, \dots, 1).$$

We note that $g(\mathbf{0}) = f(\mathbf{1})$ and because of the definition of block sensitivity one has $g(e_i) = f(\mathbf{1}^{\oplus B_i}) \neq f(\mathbf{1}) = g(\mathbf{0})$ for all i. That means the function g has a value of $f(\mathbf{1})$ on Hamming level 0 and a value of $-f(\mathbf{1})$ on every point of Hamming level 1. Then applying Theorem 10.12 with $\gamma := 2$ we have

$$\deg(f) \ge \deg(g) \ge \sqrt{b/2}$$

□

This closes the last part in the proof that indeed both notions of sensitivity as well as the degree are polynomially related.

Corollary 10.14. For any $f : \{-1,1\}^n \to \{-1,1\}$ one has

$$\sqrt{\deg(f)} \overset{(1)}{\le} S(f) \overset{(2)}{\le} BS(f) \overset{(3)}{\le} 2\deg(f)^2$$

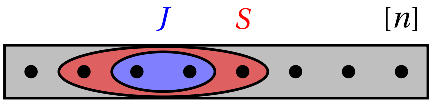
Proof. We have proven (1) in Theorem 9.6 by combining the breakthrough of Huang [Hua19] with a reduction of Gotsman and Linial [GL92]. For (2) one can use that even pointwise $S(f,x) \le BS(f,x)$ by defining $B_1, \dots, B_b$ as the singletons i where $f(x) \neq f(x^{\oplus i})$. Finally, (3) is Theorem 10.13 due to Nisan and Szegedy [NS92].

□

10.6 Low degree boolean functions have low depth decision trees

In this section, we want to prove that any function $f : \{-1,1\}^n \to \{-1,1\}$ satisfies that $D(f) \le O(\deg(f)^4)$. In other words, any boolean function has a decision tree of depth at most $O(\deg(f)^4)$. Again, we rely on the survey of Buhrman and de Wolf [Bd02]. First we obtain a simple lemma that will allow is to find function value "flips".

Lemma 10.15. *Let $f : \{-1,1\}^n \to \{-1,1\}$, let $S \subseteq [n]$ be any inclusion-wise maximal set so that $\hat{f}(S) \neq 0$ and let $x \in \{-1,1\}^n$. Then there is a set $J \subseteq S$ so that $f(x^{\oplus J}) \neq f(x)$.*



$$x = (\;1\;\;1\;\;1\;\;1\;\;1\;\;1\;\;1\;\;1\;)$$

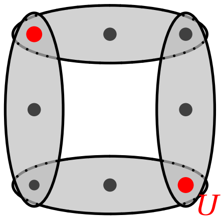
$$x^{\oplus J} = (\;1\;\;1\;\;-1\;\;-1\;\;1\;\;1\;\;1\;\;1\;)$$

Proof. Consider the function $g : \{-1,1\}^S \to \{-1,1\}$ with $g := f_{S|x_{[n]\setminus S}}$, i.e. g is obtained by restricting f to S using $x_{[n]\setminus S}$. Then by Prop 1.15, the Fourier coefficient of the restriction for the "top-level" set S itself is

$$\hat{g}(S) = \sum_{T\subseteq[n]\setminus S} \underbrace{\hat{f}(S\cup T)}_{=0 \text{ for } T\neq\emptyset} \cdot \chi_T(x) = \hat{f}(S)$$

using the maximality of S. In particular $\hat{g}(S) \neq 0$ and so the function g cannot be constant. Hence the function value of g cannot be equal to $g(x_S)$ everywhere. Then denote $x_S^{\oplus J}$ as any such point so that $g(x_S) \neq g(x_S^{\oplus J})$. That settles the claim. $\square$

Consider a set family $\mathcal{F} \subseteq 2^{[n]}$. A standard notion in combinatorics is the one of a *transversal* or *hitting set* for $\mathcal{F}$, which is a subset $U \subseteq [n]$ so that $U \cap S \neq \emptyset$ for all $S \in \mathcal{F}$.



Set family $\mathcal{F}$ with hitting set U

Crucially we can prove that for a boolean function the maximum cardinality Fourier support has a small hitting set.

Lemma 10.16. *For $f : \{-1,1\}^n \to \{-1,1\}$ with $d := \deg(f)$, let $\mathcal{F} := \{S \subseteq [n] \mid \hat{f}(S) \neq 0 \text{ and } |S| = d\}$ be the maximum cardinality Fourier support. Then $\mathcal{F}$ admits a hitting set of size $O(d^3)$.*

Proof. Let $\mathcal{M} := \{S_1, \ldots, S_k\} \subseteq \mathcal{F}$ be any maximal hypergraph matching in $\mathcal{F}$, i.e. the sets $S_1, \ldots, S_k$ are disjoint and no other set from $\mathcal{F}$ could be added without destroying that property. We fix $x \in \{-1,1\}^n$ arbitrarily. By Lemma 10.15 there are subsets $J_i \subseteq S_i$ for all $i = 1, \ldots, k$ so that $f(x^{\oplus J_i}) \neq f(x)$. Then by the definition of block sensitivity and Lemma 10.13 we know that $k \leq BS(f) \leq O(d^2)$. Hence, the set $U := S_1 \cup \ldots \cup S_k$ is a hitting set for $\mathcal{F}$ of size $|U| \leq O(d^3)$. □

Finally we can prove that any degree-d boolean function has a decision tree of depth at most $O(d^4)$.

Theorem 10.17. *For any $f : \{-1,1\}^n \to \{-1,1\}$ one has $D(f) \leq O(\deg(f)^4)$.*

Proof. Let $x \in \{-1,1\}^n$ be an unknown input so that we have to determine $f(x)$ by querying at most $O(d^4)$ many variables.

Let $U \subseteq [n]$ be the set from Lemma 10.16 which is a hitting set for the size-d sets in the Fourier support of f. Recall that $|U| \leq O(d^3)$. We can query all values of x in U. Consider the restriction $f_{d-1} : \{-1,1\}^n \to \{-1,1\}$ obtained by fixing all variables in U accordingly. Then $\deg(f_{d-1}) \leq d-1$ (see Prop 1.15). We repeat the argument with f_{d-1} until we reach a function of degree 0. This requires a total number of $O(d^4)$ queries. □

Clearly a function $f : \{-1,1\}^n \to \{-1,1\}$ depends on at most $2^{D(f)} \leq 2^{O(\deg(f)^4)}$ many variables. We would like to point out that there are functions whose number of non-redundant variables is exponential in $D(f)$. One such example is the *address function*. For $x \in \{-1,1\}^k$, let $\text{bin}(x) \in \{1, \ldots, 2^k\}$ be the number represented by the bits in x. Define $f : \{-1,1\}^{k+2^k} \to \{-1,1\}$ with $f(x,y) := y_{\text{bin}(x)}$. In other words, the function returns the entry of y that is indexed by x. One can observe that f depends on all variables while $D(f) \leq k+1$.

10.7 Every decision tree has an influential variable

As mentioned above, any degree-d function $f : \{-1,1\}^n \to \{-1,1\}$ without redundant variables satisfies that $n \leq 2^{O(d^4)}$. Moreover, we know that $\text{Inf}_{\max}[f] \geq \Omega(\frac{\log(n)}{n}) \cdot \text{Var}[f]$ by the KKL Theorem 5.10, and so we can already conclude an exponential bound for the Aaronson-Ambainis problem for boolean functions. Goal of this section is to prove a polynomial bound instead (again, only for boolean functions). The result that we will be discussion is due to O'Donnell, Saks, Schramm and Servedio [OSSS05].

First, we provide alternative expressions for variance and influence for boolean functions that will come in handy.

Lemma 10.18. *For a boolean function* $f : \{-1,1\}^n \to \{-1,1\}$ *one has*

$$\mathit{Var}[f] = \mathop{\mathbb{E}}_{x,y\sim\{-1,1\}^n}[|f(x)-f(y)|] \quad \textit{and} \quad \mathit{Inf}_i[f] = \mathop{\mathbb{E}}_{(x,y)\sim\Omega_i}[|f(x)-f(y)|]$$

where Ω_i *is the distribution over pairs* (x,y) *where* $x \sim \{-1,1\}^n$ *is uniform and* $y_j = x_j$ *for all* $j \neq i$ *and* $y_i \sim \{-1,1\}$ *independently.*

Proof. We draw $x, y \sim \{-1,1\}^n$ independently. Then the variance is

$$\begin{aligned}\mathrm{Var}[f] &\overset{\text{Def 1.38}}{=} \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \\ &= \frac{1}{2}\left(\mathbb{E}[f(x)^2] - 2\,\mathbb{E}[f(x)]\,\mathbb{E}[f(y)] + \mathbb{E}[f(y)]^2\right) \\ &= \frac{1}{2}\,\mathbb{E}[\underbrace{(f(x)-f(y))^2}_{\in\{0,4\}}] = \mathbb{E}[\underbrace{|f(x)-f(y)|}_{\in\{0,2\}}]\end{aligned}$$

Moreover, drawing $(x,y) \sim \Omega_i$, the influence of variable i is

$$\mathrm{Inf}_i[f] \overset{\text{Lem 1.30}}{=} \Pr[f(x) \neq f(x^{\oplus i})] = 2\Pr[f(x) \neq f(y)] = \mathbb{E}[\underbrace{|f(x)-f(y)|}_{\in\{0,2\}}]$$

□

For a decision tree T, and variable i we define

$$p_i(T) := \Pr_{x^*\sim\{-1,1\}^n}\left[\text{variable } x_i \text{ is queried when evaluating } x^*\right]$$

In particular if T has depth D then the expected number of queried variables is $\sum_{i=1}^n p_i(T) \leq D$. We prove the following inequality that relates variance and influences:

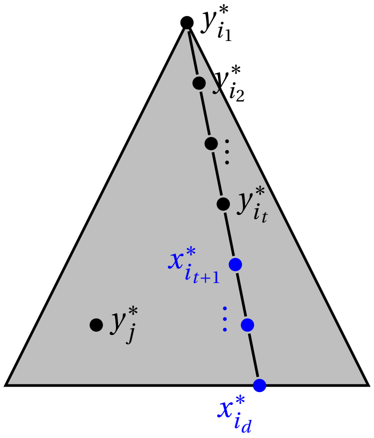
Theorem 10.19. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be a boolean function that is computed by decision tree* T*. Then*

$$\mathit{Var}[f] \leq \sum_{i=1}^n p_i(T) \cdot \mathit{Inf}_i[f]$$

Proof. Consider two independent random inputs $x^*, y^* \sim \{-1,1\}^n$. On input x^*, the computation of $f(x^*)$ follows a random path in the tree T. Let $x_{i_1}, \ldots, x_{i_d}$ be the variables that are being queried on this path. Note that the indices $i_1, \ldots, i_d$ as well as length d of the path are random variables that depend on x^*. For $t \geq 0$, let $u^{(t)} \in \{-1,1\}^n$ be the vector with

$$u_i^{(t)} = \begin{cases} x_i^* & \text{if } i \in \{i_{t+1}, i_{t+2}, \ldots, i_d\} \\ y_i^* & \text{otherwise} \end{cases}$$

meaning that the variables on the evaluation path after step t are taken from x^*; everything else is from y^*. In particular for any $t \geq d$ one has $u^{(d)} = y^*$ and while in general $u^{(0)} \neq x^*$, one still has $f(u^{(0)}) = f(x^*)$ because all values on the computation path are taken from x^* in this case. So, one can think of $u^{(t)}$ as an interpolation between x^* and y^*.



Visualization of $u^{(t)}$. Note that the

decision tree eval. path is w.r.t. x^* instead

Then we can rewrite the variance using the triangle inequality and the interpolation from above as

$$\begin{aligned}
\mathrm{Var}[f] &\overset{\text{Lem 10.18}}{=} \mathop{\mathbb{E}}_{x^*,y^* \sim \{-1,1\}^n}[|f(x^*) - f(y^*)|] \\
&\overset{\text{triangle ineq.}}{\leq} \sum_{t\geq 1} \mathbb{E}[|f(u^{(t)}) - f(u^{(t-1)})|] \\
&= \sum_{t\geq 1}\sum_{i=1}^{n} \Pr[i_t = i] \cdot \underbrace{\mathbb{E}[|f(u^{(t)}) - f(u^{(t-1)})| \mid i_t = i]}_{=\mathrm{Inf}_i[f] \text{ by Claim I}} \\
&= \sum_{i=1}^{n} \mathrm{Inf}_i[f] \cdot \underbrace{\sum_{t\geq 1} \Pr[i_t = i]}_{=p_i(T)}
\end{aligned}$$

Hence it remains to prove the following:
Claim I. *Fix* $i \in [n]$ *and* $t \geq 1$. *Then* $\mathbb{E}[|f(u^{(t)}) - f(u^{(t-1)})| \mid i_t = i] = \mathit{Inf}_i[f]$.
Proof of Claim I. We fix outcomes $X := (x^*_1, \ldots, x^*_{i_{t-1}})$ so that in iteration t the decision tree (on input of x^*) queries the ith variable, i.e. indeed $i_t = i$. It suffices to prove that then $\mathbb{E}[|f(u^{(t)}) - f(u^{(t-1)})| \mid X] = \mathrm{Inf}_i[f]$. We note that the vector $u^{(t-1)}$ contains the variables $y^*_{i_1}, \ldots, y^*_{i_{t-1}}$ instead, which are independent from X. In

particular, the vector $u^{(t-1)}$ is still uniformly from $\{-1,1\}^n$ even under conditioning on X. The vector $u^{(t)}$ differs from $u^{(t-1)}$ only in coordinate $i_t = i$. Moreover we make the observation that $u^{(t-1)}$ contains x_i^* and the vector $u^{(t-1)}$ contains y_i^*. Hence we may conclude that[4] $(u^{(t-1)}, u^{(t)})_{|X} \sim \Omega_i$. Then

$$\mathbb{E}[|f(u^{(t)}) - f(u^{(t-1)})| \mid X] = \mathop{\mathbb{E}}_{(x,y)\sim\Omega_i}[|f(x)-f(y)|] \overset{\text{Lem 10.18}}{=} \text{Inf}_i[f]$$

□

Now, as a consequence, each low degree function with low decision tree complexity must have an influential variable.

Theorem 10.20. *Any function $f : \{-1,1\}^n \to \{-1,1\}$ with degree $d := \deg(f)$ has a variable i with*

$$\mathit{Inf}_i[f] \geq \frac{\mathit{Var}[f]}{D(f)} \geq \Omega\Big(\frac{\mathit{Var}[f]}{d^4}\Big)$$

Proof. We denote $\text{Inf}_{\max}[f] := \max_{i\in[n]} \text{Inf}_i[f]$ as the maximum influence of any variable. Let T be the decision tree that has depth $D(f) \leq O(d^4)$, according to Theorem 10.17. As never more than $D(f)$ many variables are being queried, we have

$$\text{Var}[f] \leq \underbrace{\sum_{i=1}^n p_i(T)}_{\leq D(f)} \cdot \underbrace{\text{Inf}_i[f]}_{\leq \text{Inf}_{\max}[f]} \leq D(f)\cdot \text{Inf}_{\max}[f]$$

Rearranging gives the claim. □

10.8 Maximum values of functions with significant linear part

We will now switch gears and focus our attention on functions of the form $f : \{-1,1\}^n \to \mathbb{R}$ which are much less understood and less structured than their boolean special cases. Much of what we know is from the work of Dinur, Friedgut, Kindler and O'Donnell [DFKO06]. Most of their work deals with probability estimates for anti-concentration of degree-d functions. Instead we will here focus on simply

---

[4] One subtle aspect that might lead to confusion is the following: the indices $i_1, \ldots, i_d$ are defined dependent on the decision tree path for input x^*. On the other hand, in this claim we account the change arising from the vectors $u^{(t)}$ whose decision tree paths are not even considered.

proving that $\max_{x\in\{-1,1\}^n}|f(x)|$ is large depending on variance, maximum influence and degree. This loss of generality will allow us a much simpler proof where we deviate quite a bit from the original.

First, we need another result that deals with univariate polynomials.

Lemma 10.21. *For any odd $d \in \mathbb{Z}_{\geq 0}$ there is a set $P \subseteq [-\frac{1}{2}, \frac{1}{2}]$ of size $|P| \leq d+2$ so that the following holds: let $p(x) := \sum_{i=0}^{d} a_i x^i$ be a degree-d polynomial. Then there is an $x^* \in P$ so that $|p(x^*)| \geq \frac{|a_1|}{2d+2}$.*

For details, we again refer to [DFKO06]. We just would like to point out that this lemma can be derived from extremal properties of the Chebychev polynomials.

For a linear function $f(x) = \sum_{i=1}^n a_i x_i$ we can maximize $|f(x)|$ by simply picking $x_i := \text{sign}(a_i)$ and obtain a function value of $f(x) = \|a\|_1$. Quite surprisingly, we obtain almost the same bound if arbitrary other Fourier coefficients are present.

Theorem 10.22. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-d function with linear coefficients $a_i := \hat{f}(\{i\})$ for $i = 1,\ldots,n$. Then*

$$\max_{x\in\{-1,1\}^n}|f(x)| \geq \frac{\|a\|_1}{Cd}$$

where $C > 0$ is a universal constant.

Proof. For symmetry reasons we may assume that $a_i \geq 0$ for all $i = 1,\ldots,n$. For $-1 \leq \rho \leq 1$, recall that $x \sim N_\rho(\mathbf{1})$ gives a random vector $x \in \{-1,1\}^n$ with independent coordinates so that $\mathbb{E}_{x\sim N_\rho(\mathbf{1})}[x_i] = \rho$ for all i. In other words, x is a biased random vector. Consider

$$g(\rho) := \mathop{\mathbb{E}}_{x\sim N_\rho(\mathbf{1})}[f(x)] = \sum_{S\subseteq[n]} \hat{f}(S)\cdot\prod_{i\in S}\underbrace{\mathop{\mathbb{E}}_{x\sim N_\rho(\mathbf{1})}[x_i]}_{=\rho} = \sum_{k=0}^{d}\rho^k\cdot\Big(\sum_{|S|=k}\hat{f}(S)\Big)$$

We note that g is a univariate polynomial with variable ρ and its linear coefficient is $\sum_{i=1}^n \hat{f}(\{i\}) = \|a\|_1$. Hence by Lemma 10.21, there exists a value ρ^* so that $|g(\rho^*)| \geq \Theta(\frac{\|a\|_1}{d})$. Then there has to be at least one outcome $x^* \in \{-1,1\}^n$ so that $|f(x^*)| \geq |g(\rho^*)| \geq \Theta(\frac{\|a\|_1}{d})$. That settles the claim. ☐

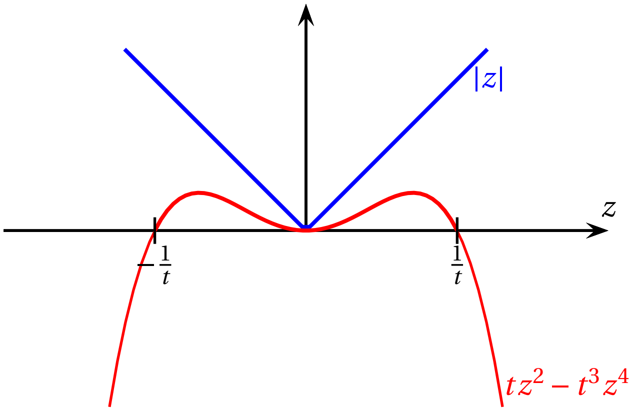
10.9 Maximum values of arbitrary functions

We continue our discussion of [DFKO06]. The next goal is to be able to lower bound $\max_{x\in\{-1,1\}^n}|f(x)|$ for an arbitrary function f that may not even have any linear part. First we prove a lemma that will be useful:

Lemma 10.23. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function of degree at most d. Then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|f(x)|] \geq 2^{-\Theta(d)} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[f(x)^2]^{1/2}$$

Proof. First we prove the following useful fact:
Claim I. *For all $t > 0$ and $z \in \mathbb{R}$ one has $|z| \geq tz^2 - t^3z^4$.*
Proof of Claim I. If $|z| \leq \frac{1}{t}$ then $|z| \geq tz^2$ and the claim is true. If $|z| \geq \frac{1}{t}$, then $tz^2 - t^3z^4 \leq 0$. ☐



The claim is invariant under scaling f, hence we may assume that $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)^2] = 1$. We abbreviate $X := f(x)$ where $x \sim \{-1,1\}^n$. Then we know that X is 9^d-reasonable. For $t > 0$ we have

$$\mathbb{E}[|X|] \overset{\text{Claim I}}{\geq} t\underbrace{\mathbb{E}[X^2]}_{=1} - t^3\underbrace{\mathbb{E}[X^4]}_{\leq 9^d} \geq t - t^3 9^d \overset{t:=2^{-\Theta(d)}}{\geq} 2^{-\Theta(d)}$$

☐

It will also be useful to understand how the Fourier weight of random restrictions behaves. Recall that for any k, we abbreviate $W_k[f] = \sum_{S\subseteq[n]:|S|=k} \hat{f}(S)^2$ as the Fourier weight on level k. More generally, for an interval I, we write $W_I[f] := \sum_{S:|S|\in I} \hat{f}(S)^2$.

Lemma 10.24 (Properties of random restrictions)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and let $k \in \mathbb{N}$. Sample $J \subseteq [n]$ by including each index independently with probability $\frac{1}{k}$, then sample $z \sim \{-1,1\}^{\bar{J}}$ and let $g := f_{|J,z} : \{-1,1\}^J \to \mathbb{R}$ be the restriction of f to J using z.*

(i) One has

$$\mathop{\mathbb{E}}_{J,z}[W_1[g]] \gtrsim W_{[k,2k)}[f].$$

(ii) For fixed J and $S \subseteq J$, the map $z \mapsto \hat{g}(S)$ has degree at most deg(f).

(iii) For each fixed $J \subseteq [n]$ and $i \in J$ one has $\mathbb{E}_z[\hat{g}(\{i\})^2] \leq$ Inf$_i[f]$.

Proof. **For (i).** We recall from Prop 1.15 that the Fourier expansion of such a restriction is

$$g(y) = \sum_{S \subseteq J} \chi_S(y) \underbrace{\sum_{T \subseteq [n] \setminus J} \hat{f}(S \cup T) \chi_T(z)}_{=\hat{g}(S)} \tag{10.1}$$

for $y \in \{-1,1\}^J$. We need to prove that g has a significant linear part in expectation. Using Prop 1.16.(d) we know that

$$\begin{aligned} \mathop{\mathbb{E}}_J\Big[\sum_{i \in J} \mathop{\mathbb{E}}_z[\hat{g}(\{i\})^2]\Big] &= \mathop{\mathbb{E}}_J\Big[\sum_{i \in J} \sum_{T \subseteq [n] \setminus J} \hat{f}(\{i\} \cup T)^2\Big] \\ &= \sum_{S \subseteq [n]} \hat{f}(S)^2 \cdot \mathop{\mathbb{E}}_J[\mathbf{1}_{|S \cap J| = 1}] \\ &\geq \sum_{k \leq |S| < 2k} \hat{f}(S)^2 \cdot \Pr_J[|S \cap J| = 1] \\ &\geq \frac{1}{10} W_{[k,2k)}[f] \end{aligned}$$

Here we use that a term $\hat{f}(S)^2$ appears in the expression on the first line if and only if $|S \cap J| = 1$. Additionally, we used that for a set $S \subseteq [n]$ of size $k \leq |S| < 2k$ we have

$$\Pr_J[|S \cap J| = 1] = \sum_{i \in S} \Pr_J[S \cap J = \{i\}] = \underbrace{|S| \cdot \frac{1}{k}}_{\geq 1} \cdot \underbrace{\Big(1 - \frac{1}{k}\Big)^{|S|-1}}_{\geq 0.1} \geq \frac{1}{10}$$

For (ii). For fixed $J \subseteq [n]$ and $S \subseteq [n]$, reinspecting (10.1) we know that

$$\hat{g}(S) = \sum_{T \subseteq [n] \setminus J} \hat{f}(S \cup T) \cdot \chi_T(z)$$

Considering this as a function in z, clearly the degree is at most as high as the degree of f itself.

For (iii). We fix $J \subseteq [n]$ and $i \in J$. Then

$$\mathop{\mathbb{E}}_z[\hat{g}(\{i\})^2] \overset{(10.1)}{=} \sum_{T \subseteq [n] \setminus J} \hat{f}(\{i\} \cup T)^2 \leq \sum_{S \subseteq [n] : i \in S} \hat{f}(S)^2 = \text{Inf}_i[f]$$

which finishes the proof. □

Now to the main result whose remaining proof is short as we have already done much of the legwork:

Theorem 10.25. *For any degree-d function $f : \{-1,1\}^n \to \mathbb{R}$ one has*

$$\max_{x\in\{-1,1\}^n} |f(x)| \geq \frac{\mathit{Var}[f]}{C^d\sqrt{\mathit{Inf}_{\max}[f]}}$$

where $C > 0$ is a large enough universal constant.

The proof strategy is as follows: a suitable random restriction of f will have a significant linear part. Then by Theorem 10.22 the restriction must have a high maximum value.

Proof of Theorem 10.25. By a bucketing argument, there must be some value k which is a power of 2 in $\{1,\dots,d\}$ so that $W_{[k,2k)}[f] \geq \frac{\mathrm{Var}[f]}{2\log(d)}$. We pick a set $J \subseteq [n]$ at random by independently including every coordinate with probability $\frac{1}{d}$. Then we draw $z \sim \{-1,1\}^{[n]\setminus J}$ and consider the random restriction $g := f_{|J,z} : \{\pm 1\}^J \to [-1,1]$. By Lemma 10.24.(i) we know that $\mathbb{E}_{J,z}[W_1[g]] \gtrsim W_{[k,2k)}[f]$. We fix a set J so that this expectation is attained, i.e. $\mathbb{E}_z[W_1[g]] = \mathbb{E}_z[\sum_{i\in J}\hat{g}(\{i\})^2] \gtrsim W_{[k,2k)}[f]$. We abbreviate $\mu_i := \mathbb{E}_z[\hat{g}(\{i\})^2]$ and $\mu_{\max} := \max_{i\in J}\mu_i$. We note that by Lemma 10.24.(iii) we have $\mu_{\max} \leq \mathrm{Inf}_{\max}[f]$. Then as for each i, the function $z \mapsto \hat{g}(\{i\})$ is a function of degree at most d, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_z\Big[\sum_{i=1}^n |\hat{g}(\{i\})|\Big] &\overset{\text{Lem } 10.23}{\geq} 2^{-\Theta(d)}\sum_{i=1}^n \sqrt{\mu_i} \\
&\geq 2^{-\Theta(d)}\sum_{i=1}^n \frac{\mu_i}{\sqrt{\mu_{\max}}} \\
&\geq \frac{2^{-\Theta(d)}}{\Theta(\log(d))}\frac{\mathrm{Var}[f]}{\sqrt{\mu_{\max}}} \geq \frac{\mathrm{Var}[f]}{2^{\Theta(d)}\sqrt{\mathrm{Inf}_{\max}[f]}}
\end{aligned}$$

Now fix any outcome of z attaining this expectation. We abbreviate the linear coefficients as $a_i := \hat{g}(\{i\})$. Then applying Theorem 10.22 gives

$$\max_{x\in\{-1,1\}^n}|f(x)| \overset{g \text{ restriction of } f}{\geq} \max_{y\in\{-1,1\}^J}|g(y)| \overset{\text{Theorem } 10.22}{\geq} \frac{\|a\|_1}{Cd} \geq \frac{\mathrm{Var}[f]}{2^{O(d)}\sqrt{\mathrm{Inf}_{\max}[f]}}$$

as claimed. □

One can rearrange the statement of Theorem 10.25 to provide an exponential (rather than polynomial) bound for the Aaronson-Ambainis problem:

Corollary 10.26. For any degree-d function $f : \{-1,1\}^n \to [-1,1]$ one has $\mathrm{Inf}_{\max}[f] \geq \frac{\mathrm{Var}[f]^2}{C^d}$ where $C > 0$ is a universal constant.

Proof. Rearranging

$$1 \geq \max_{x \in \{-1,1\}^n} |f(x)| \overset{\text{Thm } 10.25}{\geq} \frac{\mathrm{Var}[f]}{C^d \sqrt{\mathrm{Inf}_{\max}[f]}}$$

gives the claim. □

The reader might already suspect that the Aaronson-Ambainis conjecture appears to be equivalent to finding large function values depending on variance and influence. We can make that explicit:

Conjecture 3. *There are small enough constants $C_0, \delta > 0$ and a large enough constant $C_1 > 0$ so that for any degree-d function $f : \{-1,1\}^n \to \mathbb{R}$ one has*

$$\max_{x \in \{-1,1\}^n} |f(x)| \geq C_0 \frac{Var[f]^{1/2+\delta}}{d^{C_1} \cdot Inf_{\max}[f]^{\delta}}$$

Every function f has an $x \in \{-1,1\}^n$ with $|f(x)| \geq \mathrm{Var}[f]^{1/2}$, hence the goal is to beat this trivial bound.

Lemma 10.27. *Conj 2 ⇔ Conj 3.*

10.10 Bounded low degree functions are close to juntas

The proof of Theorem 10.25 is somewhat flexible and would allow to modify the function f. In particular it can provide the following:

Theorem 10.28. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a function of degree at most d and for a set $I \subseteq [n]$ of variables, we let $h(x) := \sum_{S \subseteq [n] : S \cap I \neq \emptyset} \hat{f}(S) \cdot \chi_S(x)$. Then*

$$\max_{x \in \{-1,1\}^n} |f(x)| \geq \frac{Var[h]}{C^d \sqrt{Inf_{\max}[h]}}$$

where $C > 0$ is a large enough universal constant.

We note that the way that h is defined one has $\mathrm{Inf}_i[h] = \mathrm{Inf}_i[f]$ for all $i \in I$ and $0 \leq \mathrm{Inf}_i[h] \leq \mathrm{Inf}_i[f]$ for $i \notin I$. The proof of Theorem 10.25 can be modified by changing the definition of $\mathcal{F}$ to $\mathcal{F} := \{S \subseteq [n] : 2^{s-1} \leq |S \cap I| < 2^s\}$ and sampling coordinates $U \subseteq I$ independently with probability 2^{-s}. Then set $\mu_{\max} := \max_{i \in I} \mu_i \leq \mathrm{Inf}_{\max}[h]$. We leave further details of the modification to the interested reader.

We recall that a function f that only depends on at most k coordinates is called a *k-junta* (see Def 5.27). We also recall that $\mathrm{dist}(f,g) = \mathbb{E}_{x\sim\{-1,1\}^n}[(f(x)-g(x))^2]$ denotes the distance between two functions. The work of Dinur, Friedgut, Kindler and O'Donnell [DFKO06] also contains the following result:

Theorem 10.29. *Let $f : \{-1,1\}^n \to [-1,1]$ be a function so that*

$$\sum_{|S|>k} \hat{f}(S)^2 \leq \exp\Big(-\Theta\Big(\frac{k^2\log(k)}{\varepsilon}\Big)\Big)$$

for some $k \in \mathbb{N}$ and some $\varepsilon > 0$. Then there is a $2^{O(k)}/\varepsilon^2$ junta $g : \{-1,1\}^n \to \mathbb{R}$ so that dist$(f,h) \leq \varepsilon$.

We will not prove this result in full generality here, but we prove the special case where the function f has low degree (rather than very low Fourier weight above some level).

Theorem 10.30. *Let $f : \{-1,1\}^n \to [-1,1]$ be a function of degree at most d. Then for any $\varepsilon > 0$, there is a $2^{O(d)}/\varepsilon^2$-junta $h : \{-1,1\}^n \to \mathbb{R}$ so that dist$(f,h) \leq \varepsilon$. In particular there is a set $J \subseteq [n]$ of size $|J| \leq 2^{O(d)}/\varepsilon^2$ so that $\sum_{S\subseteq[n]:S\not\subseteq J} \hat{f}(S)^2 \leq \varepsilon$.*

Proof. We abbreviate the influential coordinates of f as

$$J := \Big\{i \in [n] \mid \mathrm{Inf}_i[f] \geq \frac{\varepsilon^2}{C^{2d}}\Big\}$$

where $C > 0$ is the same constant as in Theorem 10.28. Since $\sum_{i=1}^n \mathrm{Inf}_i[f] \leq d$ we know that $|J| \leq \frac{dC^{2d}}{\varepsilon^2}$ (see Lemma 1.37). We set $g(x) := \sum_{S\subseteq J} \hat{f}(S)\chi_S(x)$ which by definition is a $|J|$-junta and let $h := f - g = \sum_{S:S\not\subseteq J} \hat{f}(S)\chi_S$ be the error that we are making by this approximation. Then $\mathrm{dist}(f,g) = \sum_{S:S\not\subseteq J} \hat{f}(S)^2 = \mathrm{Var}[h]$. Hence it remains to prove that indeed $\mathrm{Var}[h] \leq \varepsilon$. By construction, h has degree at most d and $\mathrm{Inf}_{\max}[h] < \frac{\varepsilon^2}{C^{2d}}$. Then applying Theorem 10.28 to f and h (with $I := [n]\setminus J$) we have

$$1 \geq \max_{x\in\{-1,1\}^n} |f(x)| \overset{\text{Thm 10.28}}{\geq} \frac{\mathrm{Var}[h]}{C^d\sqrt{\mathrm{Inf}_{\max}[h]}} \geq \frac{\mathrm{Var}[h]}{C^d\sqrt{\frac{\varepsilon^2}{C^{2d}}}} = \frac{\mathrm{Var}[h]}{\varepsilon}$$

Then rearranging gives the claim. □

The reader may note the similarity of Theorem 10.30 with Friedgut's Junta Theorem (Theorem 5.28).

10.11 Sensitivity, block sensitivity and fractional block sensitivity for bounded functions

The notions of sensitivity and block sensitivity can be extended from boolean functions to arbitrary functions on the hypercube.

Definition 10.31. Let $f : \{-1,1\}^n \to \mathbb{R}$. The *sensitivity of f at a point $x \in \{-1,1\}^n$* is

$$S(f,x) := \frac{1}{2}\sum_{i=1}^{n} |f(x) - f(x^{\oplus i})|$$

Moreover, the *block sensitivity of f at a point $x \in \{-1,1\}^n$* is

$$BS(f,x) := \frac{1}{2} \max_{\substack{\text{disjoint} \\ B_1,\ldots,B_k \subseteq [n]}} \sum_{i=1}^{k} |f(x) - f(x^{\oplus B_i})|$$

Again, $S(f) := \max_{x\in\{-1,1\}^n} S(f,x)$ and $BS(f) := \max_{x\in\{-1,1\}^n} BS(f,x)$ are the *sensitivity* and *block sensitivity* of the function itself.

Here we have added the factor $1/2$ so that the definition is consistent with the boolean case as in Def 10.6. Of course we still have $S(f,x) \leq BS(f,x)$ and $S(f) \leq BS(f)$ for all f and x.

Backurs and Bavarian [BB14] were the first to prove that $BS(f) \leq \text{poly}(d)$ for a degree-d function $f : \{-1,1\}^n \to [-1,1]$. Filmus, Hatami, Keller and Lifshitz [FHKL15] improved and simplified the bound to $BS(f) \leq O(d^2)$. We somewhat deviate from their exposition and prove a auxiliary result first that will turn out to be useful. Again using the $\{0,1\}^n$-cube is notationally convinient:

Proposition 10.32. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial of degree at most d so that $|h(x)| \leq 1$ for all $x \in \{0,1\}^n$. Then for any $t > 0$ one has*

$$\sum_{i=1}^{n} |h(\mathbf{0}) - h(te_i)| \leq O(d^2 t)$$

Proof. First we want to get rid of the absolute value. Let $I_+ := \{i \in [n] : h(te_i) > h(\mathbf{0})\}$ and $I_- := \{i \in [n] : h(te_i) < h(\mathbf{0})\}$. Then we can write

$$\sum_{i=1}^{n} |h(\mathbf{0}) - h(te_i)| = \sum_{i\in I_+} (h(te_i) - h(\mathbf{0})) + \sum_{i\in I_-} (h(\mathbf{0}) - h(te_i))$$

We note that the restriction $h_{|I_+} : \mathbb{R}^{I_+} \to \mathbb{R}$ with $h_{|I_+}(y) = h(y,\mathbf{0})$ is again a multilinear polynomial of degree at most d which is bounded on the hypercube (same for

$h_{|I_-}$). Hence, resetting the notation, it suffices to prove that $\sum_{i=1}^{n}(h(te_i)-h(\mathbf{0}))\leq O(d^2t)$ for any degree-d function $h:\mathbb{R}^n\to\mathbb{R}$ with $|h(x)|\leq 1$ for $x\in\{0,1\}^n$ (i.e. proving the original claim without the absolute values).

We use Lemma 10.11 to construct a univariate polynomial $p:\mathbb{R}\to\mathbb{R}$ of degree at most d so that

$$p(\ell)=\mathop{\mathbb{E}}_{x\sim\mathcal{H}_\ell}[h(x)]\quad\forall\ell=0,\dots,n$$

where $\mathcal{H}_\ell$ is again the ℓth Hamming level. Since $|h(x)|\leq 1$ for $x\in\{0,1\}^n$, we know that $|p(\ell)|\leq 1$ for all $\ell\in\{0,\dots,n\}$. Then using the degree lower bound from Theorem 10.12 with levels $a:=0$ and $b:=1$ we obtain

$$d^2\geq\frac{n}{4}\cdot\left|\mathop{\mathbb{E}}_{x\sim\mathcal{H}_1}[h(x)]-\mathop{\mathbb{E}}_{x\sim\mathcal{H}_0}[h(x)]\right|=\frac{1}{4}\sum_{i=1}^{n}(h(e_i)-h(\mathbf{0}))$$

This only proves our claim if $t=1$. But h is a *multilinear* function. That means for any x and i, the function $s\mapsto h(x+se_i)$ is an affine linear function. In particular

$$\sum_{i=1}^{n}(h(te_i)-h(\mathbf{0}))=t\sum_{i=1}^{n}(h(e_i)-h(\mathbf{0}))\leq 4td^2$$

This finishes the claim. □

Theorem 10.33. *For any function $f:\{-1,1\}^n\to[-1,1]$ of degree d one has $BS(f)\leq O(d^2)$.*

Proof. Fix $x\in\{-1,1\}^n$ and let $B_1,\dots,B_k\subseteq[n]$ be any disjoint blocks. For symmetry reasons we may assume that $x=\mathbf{1}$. Define $g:\{-1,1\}^k\to\mathbb{R}$ by letting

$$g(y_1,\dots,y_k):=f\big(\underbrace{(1-2y_1)\mathbf{1}_{B_1},\dots,(1-2y_k)\cdot\mathbf{1}_{B_k},\mathbf{1}}_{\in[-1,1]^n\text{ for }y\in[0,1]^k}\big).$$

Then $g(\mathbf{0})=f(\mathbf{1})$ and $g(e_i)=f(\mathbf{1}^{\oplus B_i})$. Moreover $|g(y)|\leq 1$ for $y\in[0,1]^k$ because f is bounded on the $[-1,1]^n$ cube. Also we can see that $\deg(g)\leq\deg(f)\leq d$. Then

$$\sum_{i=1}^{k}|f(\mathbf{1})-f(\mathbf{1}^{\oplus B_i})|=\sum_{i=1}^{k}|g(\mathbf{0})-g(e_i)|\leq O(d^2)$$

by applying Prop 10.32 with $t:=1$. □

Next, we can actually make the argument work also when the blocks are not disjoint! We need the following fact:

Lemma 10.34. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial of degree at most d and let $h : \mathbb{R}^m \to \mathbb{R}^n$ with $h(y) := Ay + b$ be affine linear with $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$. Then also the function $y \mapsto f(h(y))$ is multilinear with degree at most d.*

Proof. By assumption $f(x) = \sum_{|S|\le d} \hat{f}(S) \cdot \prod_{i\in S} x_i$. Then we can write

$$f(h(y)) = \sum_{|S|\le d} \hat{f}(S) \cdot \prod_{i\in S} \Big(b_i + \sum_{j=1}^{m} A_{ij} y_j\Big)$$

Multiplying out the product we can see that only multilinear products $\chi_T(y)$ for $|T| \le d$ appear which gives the claim. □

Now we can prove the following:

Proposition 10.35. *Let $f : \{-1,1\}^n \to [-1,1]$ be a function of degree d and let $x \in \{-1,1\}^n$. Let $S_1,\dots,S_m \subseteq [n]$ be a set family that contains each element at most t times. Then*

$$\sum_{j=1}^{m} \big|f(x) - f(x^{\oplus S_j})\big| \le O(d^2 t)$$

Proof. Let $t \ge 1$. Again assume w.l.o.g. that $x = \mathbf{1}$. Define the function $g : \mathbb{R}^m \to \mathbb{R}$ as

$$g(y) := f\Big(\mathbf{1} - \frac{2}{t}\sum_{j=1}^{m} y_j \mathbf{1}_{S_j}\Big)$$

where $\mathbf{1}_S \in \{0,1\}^n$ is the characteristic vector of a set $S \subseteq [n]$ and we think of f also as its extension to $\mathbb{R}^n$. Then by Lemma 10.34, g is again a multilinear polynomial of degree at most d. Moreover $g(\mathbf{0}) = f(\mathbf{1})$ and $g(te_j) = f(\mathbf{1} - t\frac{2}{t}\cdot \mathbf{1}_{S_j}) = f(\mathbf{1}^{\oplus S_j})$. For any $y \in [0,1]^m$ one has $\mathbf{1} - \frac{2}{t}\sum_{j=1}^{m} y_j \mathbf{1}_{S_j} \in [-1,1]^n$ using the assumption that each element i appears in at most t sets. Then $|g(y)| \le 1$ for $y \in \{0,1\}^m$. Applying Prop 10.32 we know that

$$\sum_{j=1}^{m} |f(\mathbf{1}) - f(\mathbf{1}^{\oplus S_j})| = \sum_{j=1}^{m} |g(\mathbf{0}) - g(te_j)| \le O(d^2 t)$$

□

It will be convinient to also allow to take blocks fractionally:

Definition 10.36. Let $f : \{-1,1\}^n \to \mathbb{R}$. The *fractional block sensitivity of f at a point $x \in \{-1,1\}^n$* is the optimum value of the linear program

$$FBS(f,x) := \max\Big\{ \sum_{S\subseteq[n]} y_S \cdot |f(x) - f(x^{\oplus S})| : \sum_{S:i\in S} y_S \le 1 \ \forall i \in [n];\ y_S \ge 0 \ \forall S \subseteq [n] \Big\}$$

We set $FBS(f) := \max_{x\in\{-1,1\}^n} FBS(f,x)$ as the maximum fractional block sensitivity of f itself.

We note that integral solutions to this LP correspond to disjoint blocks and so $S(f,x) \leq BS(f,x) \leq FBS(f,x)$ for any f and x. Lovett and Zhang [LZ23] were the first to prove some upper bound on the fractional block sensitivity (though they used a slightly different definition). Here we reprove a bound that is due to Agarwal and Ben-David [ABD24].

Theorem 10.37. *For any function $f : \{-1,1\}^n \to [-1,1]$ of degree at most d one has $FBS(f) \leq O(d^2)$.*

Proof. Fix $x \in \{-1,1\}^n$. Let $\{y_S\}_{S\subseteq[n]}$ be a solution for the LP from Def 10.36 whose value attains $FBS(f,x)$. There is always such an optimum solution which is rational and hence we may assume that for some $k \in \mathbb{N}$ one has $y_S \in \frac{\mathbb{Z}_{\geq 0}}{k}$ for all $S \subseteq [n]$. Let $S_1,\dots,S_m$ be the set system that contains each set S exactly $ky_S \in \mathbb{Z}_{\geq 0}$ times. Because of the packing constraint of the LP, each element $i \in [m]$ will be in at most k of the sets in $S_1,\dots,S_m$ (of course counted with multiplicity). Then

$$k\cdot FBS(f,x) = \sum_{S\subseteq[n]} ky_S\cdot|f(x)-f(x^{\oplus S})| = \sum_{j=1}^{m}|f(x)-f(x^{\oplus S_j})| \overset{\text{Prop 10.35}}{\leq} O(d^2k)$$

Dividing by k gives the claim. ☐

10.12 Geometric properties of bounded low degree functions

Finally, we want to discuss some results of Lovett and Zhang [LZ23] on the geometry of bounded low degree functions.

Distances in the hypercube

We begin by discussing general aspects about distances. For $p \geq 1$ and a compact set $A \subseteq \mathbb{R}^n$ we write

$$d_p(x,A) := \min\left\{\|x-y\|_p : y \in A\right\}$$

as the minimum L_p-distance of a point x to A. For two compact sets $A,B \subseteq \mathbb{R}^n$ we also write

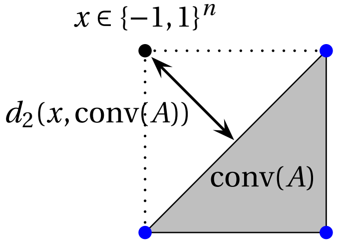
$$d_p(A,B) := \min\left\{\|x-y\|_p : x \in A, y \in B\right\}.$$

Here, we will be exclusively consider the cases $p \in \{2,\infty\}$.

A rather powerful result to bound distances on the hypercube is due to Talagrand. For a set $A \subseteq \{-1,1\}^n$, we write $\mu_n(A) := \frac{|A|}{2^n}$ as the uniform measure w.r.t. to the hypercube points. Moreover $\text{conv}(A)$ denotes the *convex hull* of A, i.e. the unique smallest convex set containing A. One can prove that most points will be very close to the convex hull of a not too tiny set A.

Theorem 10.38 (Talagrand [Tal95]). *Let $A \subseteq \{-1,1\}^n$ be non-empty. Then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\exp\left(\frac{1}{16}\cdot d_2(x,conv(A))^2\right)\right] \le \frac{1}{\mu_n(A)}.$$



Arguably, we are not doing justice to this inequality, which holds in much more generality for arbitrary product spaces. A very readable account can be found in the textbook by Alon and Spencer [AS16]. In particular by Jensen's inequality, this implies that $\mathbb{E}_{x\sim\{-1,1\}^n}[d_2(x,\text{conv}(A))] \le O(\sqrt{\ln\frac{1}{\mu_n(A)}})$. Another variation of Talagrand's Theorem is as follows:

Corollary 10.39. *Let $X, Y \subseteq \{-1,1\}^n$. Then*

$$\mu_n(X)\cdot\mu_n(Y) \le \exp\left(-\frac{1}{16}d_2(X,conv(Y))^2\right)$$

A fractional certificate

In the following we interpret a distribution π over $[n]$ also as a probability vector $\pi \in \mathbb{R}^n_{\ge 0}$ with $\sum_{i=1}^n \pi_i = 1$. For $S \subseteq [n]$ we write $\pi(S) := \Pr_{i\sim\pi}[i \in S] = \sum_{i\in S}\pi_i$. The fractional block sensitivity bound from Theorem 10.33 also has a dual interpretation:

Proposition 10.40. *Let $f : \{-1,1\}^n \to [-1,1]$ be a function of degree at most d. Then for any $x \in \{-1,1\}^n$ there is a distribution π_x over $[n]$ so that*

$$|f(x) - f(x^{\oplus S})| \le O(d^2)\cdot\pi_x(S) \quad \forall S \subseteq [n]$$

Proof. As we mentioned before, $FBS(f,x)$ is the optimum value of a linear program — in fact, the LP in Def 10.36 is a *packing LP*. Let $c \in \mathbb{R}^{2^{[n]}}$ be the vector with entries $c_S := |f(x) - f(x^{\oplus S})|$ and let $A \in \mathbb{R}^{n \times 2^{[n]}}$ be the matrix with entries

$$A_{i,S} := \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in [n] \ \forall \ S \subseteq [n]$$

Then the LP from Def 10.36 is

$$FBS(f,x) = \max\left\{c^T y \mid Ay \leq \mathbf{1};\ y \geq \mathbf{0}\right\}$$

Such a linear program must have a *dual* with the same optimum value, which will be the *covering LP*

$$\min\left\{\mathbf{1}^T z \mid z^T A \geq c^T;\ z \geq \mathbf{0}\right\}$$

Writing this out gives

$$\begin{aligned} \sum_{i=1}^{n} z_i &= FBS(f,x) \\ \sum_{i \in S} z_i &\geq |f(x) - f(x^{\oplus S})| \quad \forall S \subseteq [n] \\ z &\in \mathbb{R}^n_{\geq 0} \end{aligned}$$

Now setting $\pi_i := \frac{z_i}{FBS(f,x)}$ and using the estimate of $FBS(f,x) \leq O(d^2)$ from Theorem 10.37 gives the claim. □

We also include a simple lemma for later:

Lemma 10.41. *Let* $X \in [-1,1]$ *be a random variable with* $\Pr[X \geq \mathbb{E}[X] + \varepsilon] \leq \varepsilon$. *Then* $\mathrm{Var}[X] \leq 12\varepsilon$.

Proof. After shifting we may assume that $-2 \leq X \leq 2$ with $\mathbb{E}[X] = 0$. Then

$$\mathrm{Var}[X] = \mathbb{E}[X^2] \leq 2\mathbb{E}[|X|] = 4\mathbb{E}[\max\{X,0\}] \leq 4 \cdot (\varepsilon \cdot 2 + 1 \cdot \varepsilon) \leq 12\varepsilon.$$

□

Distances to the convex hull of ε-separated points

Now, back to functions $f : \{-1,1\}^n \to [-1,1]$ of degree at most d. Suppose that the variance of such function is $\mathrm{Var}[f] = \Theta(1)$ and $\mathbb{E}_{x \sim \{-1,1\}^n}[f(x)] = 0$. By Lemma 10.41, for some constant $\varepsilon > 0$ one has that

$$X := \left\{x \in \{-1,1\}^n \mid f(x) \geq \varepsilon\right\} \quad \text{and} \quad Y := \left\{x \in \{-1,1\}^n \mid f(x) \leq -\varepsilon\right\}$$

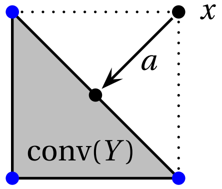
have a measure of $\mu_n(X), \mu_n(Y) \geq \varepsilon$. Then by Talagrand's inequality (Theorem 10.38), the average Euclidean distance is $\mathbb{E}_{x \sim X}[d_2(x, \text{conv}(Y))] \leq O(1)$. The Euclidean ball B_2^n has the same volume as the scaled cube $\Theta(\frac{1}{\sqrt{n}}) \cdot B_\infty^n$. So intuitively, one might then think that $\mathbb{E}_{x \sim X}[d_\infty(x, \text{conv}(Y))] \leq O(\frac{1}{\sqrt{n}})$ should hold as well. But this is very wrong and the opposite is true — the L_∞-distance is surprisingly large! For a function f and a point $x \in \{-1,1\}^n$ we say that $S \subseteq [n]$ is an (r,ε)-*improvement* if $f(x^{\oplus S}) \geq f(x) + \varepsilon$ and $|S| \leq r$. We can also give a condition for the existence of such improvement steps with small r. For a vector $a \in \mathbb{R}^n$ we write $|a| \in \mathbb{R}^n$ as the vector with entries $|a|_i := |a_i|$ for $i \in [n]$.

Proposition 10.42. *Let $C > 0$ be the implicit constant from Proposition 10.40. Let $f : \{\pm 1\}^n \to [-1,1]$ be a function of degree at most d and let $x \in \{-1,1\}^n$ and $Y \subseteq \{y \in \{\pm 1\}^n \mid f(y) \leq f(x) - \varepsilon\}$ for some $\varepsilon > 0$. Then*

- *(i) For any $a \in \mathbb{R}^n$ with $x + a \in \text{conv}(Y)$ one has $\langle \pi, |a| \rangle = \mathbb{E}_{i \sim \pi}[|a_i|] \geq \frac{2\varepsilon}{Cd^2}$ where $\pi := \pi_x$ is the distribution from Proposition 10.40.*
- *(ii) One has $d_\infty(x, \text{conv}(Y)) \geq \frac{2\varepsilon}{Cd^2}$.*
- *(iii) At least one point in Y admits an $(r, \frac{\varepsilon}{2})$-improvement where $r := \text{poly}(d, \frac{1}{\varepsilon}) \cdot d_2(x, \text{conv}(Y))^2$.*

Proof. For (i). Since $x + a \in \text{conv}(Y)$, there is a distribution σ over sets so that $x + a = \mathbb{E}_{S \sim \sigma}[x^{\oplus S}]$ and $x^{\oplus S} \in Y$. We note that $|a| = 2\mathbb{E}_{S \sim \sigma}[\mathbf{1}_S]$. Recall that $|f(x^{\oplus S}) - f(x)| \geq \varepsilon$ for all $S \in \text{supp}(\sigma)$ and by Proposition 10.40 we have $\pi(S) \geq \frac{|f(x) - f(x^{\oplus S})|}{Cd^2}$ for all $S \subseteq [n]$. This implies that

$$\langle \pi, |a| \rangle = \mathop{\mathbb{E}}_{i \sim \pi}[|a_i|] = 2 \mathop{\mathbb{E}}_{S \sim \sigma}\Big[\underbrace{\mathop{\mathbb{E}}_{i \sim \pi}[(\mathbf{1}_S)_i]}_{= \pi(S)} \Big] \geq \frac{2\varepsilon}{Cd^2}$$



For (ii), let a be the vector with $x + a \in \text{conv}(Y)$ that minimizes $\|a\|_\infty$. As the maximum entry is at least the average entry (w.r.t. distribution π) we have $\|a\|_\infty \geq \frac{2\varepsilon}{Cd^2}$ by (i).

Finally we prove (iii). Now, let $a \in \mathbb{R}^n$ be the vector with $x + a \in \text{conv}(Y)$ that minimizes $\|a\|_2$. Let σ be the distribution over sets so that $x + a = \mathbb{E}_{S \sim \sigma}[x^{\oplus S}]$. For some parameter $\delta > 0$ that we determine later, let $J := \{i \in [n] \mid |a_i| \geq \delta\}$ be the large coordinates of a. Clearly $\|a\|_2^2 \geq |J| \cdot \delta^2$ which can be rearranged to $|J| \leq r := \frac{\|a\|_2^2}{\delta^2}$. We define a set Y' and a vector a' so that

$$Y' := \left\{(x_J, y_{[n]\setminus J}) \mid y \in Y\right\} \quad \text{and} \quad x + a' = \mathop{\mathbb{E}}_{S\sim\sigma}\left[x^{\oplus(S\setminus J)}\right] \in \text{conv}(Y')$$

In other words, for all $y \in Y$, we overwrite the coordinates in J with the corresponding entries in x. For each point this changes at most r coordinates. Moreover $a'_J = \mathbf{0}$ and $a'_{[n]\setminus J} = a_{[n]\setminus J}$. Now suppose for the sake of contradiction that $Y' \subseteq \{y' \in \{\pm 1\}^n \mid f(y') \leq f(x) - \frac{\varepsilon}{2}\}$. Then

$$\delta \overset{\text{choice of } J}{\geq} d_\infty(x, \text{conv}(Y')) \overset{(ii)}{\geq} \frac{\varepsilon}{Cd^2}$$

which is a contradiction if we choose say $\delta := \frac{\varepsilon}{2Cd^2}$. From that contradiction we learn that there is indeed a $y' \in Y'$ with $f(y') > f(x) - \frac{\varepsilon}{2}$. That point is of the form $y' = (x_J, y_{[n]\setminus J})$ for some $y \in Y$. That means if from that y we overwrite the J coordinates with x_J, we increase the function value by at least $\frac{\varepsilon}{2}$. Hence y is our $(r, \frac{\varepsilon}{2})$-sensitive point where $r = \frac{\|a\|_2^2}{\delta^2} = \Theta(\frac{d^4}{\varepsilon^2}) \cdot d_2(x, \text{conv}(Y))^2$. □

Many points have small improvement steps

We can now prove that a large fraction of points must have a significant improvement by flipping only few bits.

Theorem 10.43. *For a large enough constant $C > 0$ the following holds: Let $f : \{-1,1\}^n \to [-1,1]$ be a function of degree at most d with $\text{Var}[f] \geq C\varepsilon$. Then there is a set $Y_{imp} \subseteq \{-1,1\}^n$ of points with $\mu_n(Y_{imp}) \geq \varepsilon$ so that all points in Y_{imp} have an (r,ε)-improvement for $r := \text{poly}(d, \frac{1}{\varepsilon})$.*

Proof. To simplify notation we may shift the function so that $\mathbb{E}_{x\sim\{-1,1\}^n}[f(x)] = 0$. We define the sets

$$\begin{aligned} X &:= \{x \in \{-1,1\}^n \mid f(x) \geq \varepsilon\} \\ Y &:= \{x \in \{-1,1\}^n \mid f(x) \leq -\varepsilon\} \\ Y_{\text{imp}} &:= \{x \in Y \mid x \text{ has an } (r,\varepsilon)\text{-improvement}\} \\ Y_{\text{no-imp}} &:= \{x \in Y \mid x \text{ has } \underline{\text{no}}\ (r,\varepsilon)\text{-improvement}\} \end{aligned}$$

We note that $Y = Y_{\text{imp}} \dot\cup Y_{\text{no-imp}}$. By Lemma 10.41 we know that $\mu_n(X), \mu_n(Y) \geq 2\varepsilon$ if we choose $C > 0$ large enough. Now assume for the sake of contradiction that

$\mu(Y_{\text{imp}}) \leq \varepsilon$. Then $\mu_n(Y_{\text{no-imp}}) = \mu_n(Y) - \mu_n(Y_{\text{imp}}) \geq \varepsilon$. By Talagrand's bound from Cor 10.39 we have

$$\underbrace{\mu_n(X)}_{\geq 2\varepsilon} \cdot \underbrace{\mu_n(Y_{\text{no-imp}})}_{\geq \varepsilon} \leq \exp\left(-\frac{1}{16} d_2(X, \mathrm{conv}(Y_{\text{no-imp}}))^2\right)$$

from which we know that $d_2(X, \mathrm{conv}(Y_{\text{no-imp}})) \leq O(\sqrt{\ln(1/\varepsilon)})$. Let $x \in X$ be a point attaining this bound. By Prop 10.42.(iii) there is a $y \in Y_{\text{no-imp}}$ that admits an (r, ε)-improvement for $r := \mathrm{poly}(d, \frac{1}{\varepsilon}) \cdot d_2(x, \mathrm{conv}(Y))^2 \leq \mathrm{poly}(d, \frac{1}{\varepsilon})$. This is a contradiction.

$\square$

Chapter 11

The Bohnenblust-Hille Inequality

Littlewood's 4/3 inequality and its generalization, the Bohnenblust-Hille inequality [BH31] are important results in mathematical analysis. We will discuss their proofs and show applications to the Aaronson-Ambainis conjecture and learning of low-degree functions. To keep the notation simple, we will only cover those inequalities for functions $f : \{-1,1\}^n \to \mathbb{R}$ on the hypercube even though they apply in much more generality.

11.1 Preliminaries

First we review some notation. As before, for any $p \geq 1$ and any function $f : \{-1,1\}^n \to \mathbb{R}$ we define a norm $\|f\|_{E,p} = (\mathbb{E}_{x \sim \{-1,1\}^n}[|f(x)|^p])^{1/p}$. On the other hand, without the expectation the norm is just a sum, i.e. $\|f\|_p = (\sum_{x \in \{-1,1\}^n} |f(x)|^p)^{1/p}$. For us important will be the maximum function value of $\|f\|_\infty := \max_{x \in \{-1,1\}^n} |f(x)|$. We also use the notation $\hat{\|} f \hat{\|}_p := (\sum_{S \subseteq [n]} |\hat{f}(S)|^p)^{1/p}$ for the corresponding norm of the Fourier coefficients. We know that the balls B_p^n are getting larger as p increases, e.g. $B_1^n \subseteq B_2^n \subseteq B_\infty^n$. In reverse, this implies the following:

Lemma 11.1. *For any vector $x \in \mathbb{R}^n$, the function $p \mapsto \|x\|_p$ is decreasing in p.*

We will also use a version of the Generalized Bonami Lemma from Chapter 5 which we restate for convinience.

Theorem (Theorem 5.30 — Generalized Bonami Lemma II)**.** *For any function $f : \{-1,1\}^n \to \mathbb{R}$ of degree at most k and any $1 \leq p \leq 2$ one has*

$$\|f\|_{E,2} \leq (e^{\frac{2}{p}-1})^k \cdot \|f\|_{E,p}$$

Next, we want to work towards a matrix sum inequality that will be crucial for our proof of the Bohnenblust-Hille Inequality. First we prove a helper lemma that allows us to swap the summation order as long as we move the larger exponent inside.

Lemma 11.2. *For any $B \in \mathbb{R}_{\geq 0}^{m\times n}$ and $1 \leq p \leq s$ one has*

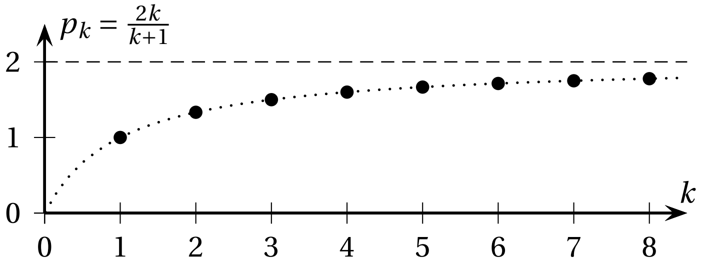
$$\left(\sum_{i=1}^{m}\left(\sum_{j=1}^{n} B_{ij}^{p}\right)^{\frac{s}{p}}\right)^{\frac{p}{s}} \leq \sum_{j=1}^{n}\left(\sum_{i=1}^{m} B_{ij}^{s}\right)^{\frac{p}{s}}$$

Proof. We can write

$$\begin{aligned}\left(\sum_{i=1}^{m}\left(\sum_{j=1}^{n} B_{ij}^{p}\right)^{\frac{s}{p}}\right)^{\frac{p}{s}} &\overset{\frac{s}{p}\geq 1}{=} \left\|\sum_{j=1}^{n}\left(B_{ij}^{p}\right)_{i\in[m]}\right\|_{\frac{s}{p}} \\ &\overset{\text{triangle inequ.}}{\leq} \sum_{j=1}^{n}\left\|\left(B_{ij}^{p}\right)_{i\in[m]}\right\|_{\frac{s}{p}} = \sum_{j=1}^{n}\sum_{i=1}^{m}\left(B_{ij}^{p\cdot\frac{s}{p}}\right)^{\frac{p}{s}}\end{aligned}$$

where we use that $\|\cdot\|_{\frac{s}{p}}$ is a norm so that the triangle inequality can be used. □

Now we come to the crucial matrix sum lemma. For a parameter $k \geq 1$ we define exponents $p_k := \frac{2k}{k+1}$. We can define those for any real $k \geq 1$, though later we only need the values p_k for integer k. The p_k's give an increasing sequence of exponents that approaches 2 as $k \to \infty$. But the power of the Bohnenblust-Hille Inequality will lie in the fact that $p_k < 2$. It will also be useful to keep the identity $p_{k/2} = \frac{2(k/2)}{k/2+1} = \frac{2k}{k+2}$ in mind.



Now we can prove the matrix sum inequality which due to Defant, Popa and Schwarting [DPS10], extending work of Blei [Ble01].

Lemma 11.3. *Let $A \in \mathbb{R}^{m\times n}$ be a matrix and let $k > 0$. Then*

$$\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^{p_k}\right)^{\frac{1}{p_k}} \leq \left(\sum_{i=1}^{m}\|A_i\|_2^{p_{k/2}}\right)^{\frac{1}{2p_{k/2}}}\left(\sum_{j=1}^{n}\|A^j\|_2^{p_{k/2}}\right)^{\frac{1}{2p_{k/2}}}$$

Proof. W.l.o.g. we assume that $A_{ij} \geq 0$ for all i, j. We abbreviate $\alpha := \frac{1}{2} p_k$. For any $p > 1$, we define $p^* := \frac{p}{p-1} > 1$ as the *Hölder conjugate* which is the unique value so that $\frac{1}{p} + \frac{1}{p^*} = 1$, see again Section 5.5.1.

Now, let $p, s > 1$ be two parameters that we determine later. Then

$$\begin{aligned}
\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^{p_k} &= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{ij}^{\alpha} \cdot A_{ij}^{\alpha} \Big) & (11.1)\\
&\overset{\text{Hölder}}{\leq} \sum_{i=1}^{m} \Big(\Big(\sum_{j=1}^{n} A_{ij}^{\alpha p} \Big)^{\frac{1}{p}} \cdot \Big(\sum_{j=1}^{n} A_{ij}^{\alpha p^*} \Big)^{\frac{1}{p^*}} \Big) & (11.2)\\
&\overset{\text{Hölder}}{\leq} \Big(\sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{ij}^{\alpha p} \Big)^{\frac{s}{p}} \Big)^{\frac{1}{s}} \cdot \Big(\sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{ij}^{\alpha p^*} \Big)^{\frac{s^*}{p^*}} \Big)^{\frac{1}{s^*}} & (11.3)\\
&\overset{\text{Lem 11.2}}{\leq} \Big(\sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{ij}^{\alpha p} \Big)^{\frac{s}{p}} \Big)^{\frac{1}{s}} \cdot \Big(\sum_{j=1}^{n} \Big(\sum_{i=1}^{m} A_{ij}^{\alpha s^*} \Big)^{\frac{p^*}{s^*}} \Big)^{\frac{1}{p^*}} & (11.4)
\end{aligned}$$

Here in line (11.2) we apply Hölder's inequality with exponents (p, p^*) to the inner terms. Then in (11.3) we apply Hölder's inequality again but with exponents (s, s^*) applied to the outer terms. Finally in line (11.3) we bound the right hand side factor using Lemma 11.2 keeping in mind that we will need that $s^* \geq p^* \Leftrightarrow s \leq p$.

It then remains to pick the parameters p, p^*, s, s^* so that the outcome in (11.4) matches the expression in our claim. In fact, by symmetry we can see that we will need to set $p = s^*$ and $p^* = s$. Next, we want an inner exponent of $\frac{p_k}{2} \cdot p = \alpha \cdot p = 2$ which means that we should set $p = \frac{4}{p_k} = s^*$. Then by definition of p_k, the conjugate is $s := p^* = \frac{p}{p-1} = \frac{2(k+1)}{k+2} = \frac{2p_{k/2}}{p_k}$. We note that conviniently, $\frac{s}{p} = \frac{p^*}{s^*} = \frac{2p_{k/2}}{p_k} \cdot \frac{p_k}{4} = \frac{p_{k/2}}{2} < 1$. Hence we can rewrite (11.1)+(11.4) to

$$\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^{p_k} \leq \Big(\sum_{i=1}^{m} \Big(\sum_{j=1}^{n} A_{ij}^{2} \Big)^{\frac{p_{k/2}}{2}} \Big)^{\frac{p_k}{2p_{k/2}}} \cdot \Big(\sum_{j=1}^{n} \Big(\sum_{i=1}^{m} A_{ij}^{2} \Big)^{\frac{p_{k/2}}{2}} \Big)^{\frac{p_k}{2p_{k/2}}}$$

which was exactly the claim (after taking the p_k-th root). □

11.2 Littlewood's 4/3 Inequality

To warm up, we prove *Littlewood's 4/3 Inequality*. Luckily, the bulk of the technical work has already been taken care of in Lemma 11.3. We note that functions of the form $f(x, y) = x^T A y$ are also called *bilinear forms*. In our usual Fourier analytic notation this means that each set S with $\hat{f}(S) \neq 0$ has (i) $|S| = 2$ and (ii) S contains exactly one of the x-variables and one of the y-variables.

Theorem 11.4 (Littlewood's 4/3 inequality [LIT30])**.** *For $A \in \mathbb{R}^{m\times n}$, let $f : \{-1,1\}^{m+n} \to \mathbb{R}$ be the function with $f(x,y) = x^T Ay$. Then*

$$\Big(\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^{4/3}\Big)^{3/4} \leq C \cdot \|f\|_\infty$$

where $C > 0$ is a universal constant.

Proof. After scaling we may assume that $\|f\|_\infty = 1$. Then applying Lemma 11.3 with $k = 2$ one has $p_2 = \frac{4}{3}$ and $p_1 = 1$ so that

$$\Big(\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^{4/3}\Big)^{3/4} \leq \Big(\sum_{i=1}^{m}\|A_i\|_2\Big)^{1/2}\Big(\sum_{j=1}^{n}\|A^j\|_2\Big)^{1/2}$$

By symmetry, it remains to prove the following:
Claim I. *One has $\sum_{i=1}^m \|A_i\|_2 \leq O(1)$.*
Proof of Claim I. We can write

$$\begin{aligned}\sum_{i=1}^{m}\|A_i\|_2 &\asymp \sum_{i=1}^{m}\mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}[|\langle A_i,y\rangle|] \\ &= \mathop{\mathbb{E}}_{y\in\{-1,1\}^n}\Big[\sum_{i=1}^{m}\max_{x_i\in\{-1,1\}} x_i\langle A_i,y\rangle\Big] = \mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}\Big[\max_{x\in\{-1,1\}^m} f(x,y)\Big] \leq 1\end{aligned}$$

using Khintchine's inequality and the fact that f is bounded. □

11.3 The Bohnenblust-Hille Inequality

Now we come to the central part of this chapter, the Bohnenblust-Hille Inequality which generalizes Littlewood's 4/3-Inequality. A function $f : \{-1,1\}^V \to \mathbb{R}$ is called k-*multilinear* if there is a partition of $V = V_1\dot\cup\ldots\dot\cup V_k$ so that

$$|V_i \cap S| = 1 \quad \forall i \in [k] \;\; \forall S \subseteq V \text{ with } \hat{f}(S) \neq 0$$

In other words, for each block i, the function is linear in the variables x_{V_i}.

For our purpose it will be notationally cleaner to write such a k-multilinear function as $f : \{-1,1\}^{nk} \to \mathbb{R}$ with k blocks of exactly n variables. Then each monomial can be indexed by a vector $i = (i_1,\ldots,i_k) \in [n]^k$. For $x = (x^{(1)},\ldots,x^{(k)}) \in \{-1,1\}^{nk}$ (i.e. $x^{(j)} \in \{-1,1\}^n$ for each $j \in [k]$) we then write the characters as

$$\chi_i(x) = x^{(1)}_{i_1}\cdot\ldots\cdot x^{(k)}_{i_k}$$

In particular a k-multilinear function $f : \{-1,1\}^{nk} \to \mathbb{R}$ can be written as its Fourier expansion

$$f(x) = \sum_{i_1,\dots,i_k \in [n]} \hat{f}(i_1,\dots,i_k) \cdot x^{(1)}_{i_1} \cdot \ldots \cdot x^{(k)}_{i_k} = \sum_{i \in [n]^k} \hat{f}(i) \cdot \chi_i(x)$$

for $x = (x^{(1)},\dots,x^{(k)}) \in \{-1,1\}^{kn}$.

Then the Bohnenblust-Hille Inequality is as follows:

Theorem 11.5 (Bohnenblust-Hille Inequality [LIT30])**.** *For any k-multilinear function $f : \{-1,1\}^{nk} \to \mathbb{R}$ and any $p \geq \frac{2k}{k+1}$ one has*

$$\hat{\|} f \hat{\|}_p = \Big(\sum_{i_1,\dots,i_k \in [n]} |\hat{f}(i_1,\dots,i_k)|^p \Big)^{1/p} \leq C_k \cdot \|f\|_\infty$$

Here one can pick $C_k \lesssim k^{\log_2(e)} \leq k^{3/2}$.

We should mention that the polynomial bound on C_k is due to Pellegrino and Seoane-Sepúlveda [PSS12]. For the proof, we will follow the exposition of Montanaro [Mon12]. From Lemma 11.1 we know that $\hat{\|} f \hat{\|}_p$ is decreasing in p and so it suffices to prove the claim for $p = \frac{2k}{k+1}$. One can replace a given k by the nearest larger power of 2 which only changes C_k by a constant. We can also scale f so that $\|f\|_\infty \leq 1$. Then the exact statement that we prove is then the following:

Proposition 11.6 (Bohnenblust-Hille Inequality II [LIT30])**.** *For any k-multilinear function $f : \{-1,1\}^{nk} \to [-1,1]$ and any $k \geq 1$ that is a power of 2 one has*

$$\hat{\|} f \hat{\|}_{p_k} \leq C_k$$

where $C_k \leq k^{\log_2(e)} \leq k^{1.45}$ and $p_k := \frac{2k}{k+1}$.

Proof. We prove the claim by induction. For $k = 1$ we have $p_1 = 1$ and the inequality is of the form $\sum_{i \in [n]} |\hat{f}(i)| \leq C_1$ which is true for $C_1 = 1$ as $\|f\|_\infty = 1$ (see e.g. the argument at the beginning of Chapter 10).

Now suppose k is a power of 2 with $k \geq 2$. We split the block indices $[k]$ into two parts $A := \{1,\dots,\frac{k}{2}\}$ and $B := \{\frac{k}{2}+1,\dots,k\}$ so that $|A| = |B| = \frac{k}{2}$. For a tuple $i \in [n]^k$ we write $i_A = (i_1,\dots,i_{k/2})$; similar for i_B. Moreover, we split $x = (x_A, x_B) \in \{-1,1\}^{nk}$ into the variables belonging to the corresponding blocks.

We define a matrix M that has row indices $i_A \in [n]^{k/2}$ and column indices $i_B \in [n]^{k/2}$ where the entries are the Fourier coefficients of f, i.e.

$$M_{i_A,i_B} := \hat{f}(i_A, i_B)$$

Then applying Lemma 11.3 to the matrix M gives

$$\begin{aligned}
\hat{\|} f \hat{\|}_{p_k} &= \Big(\sum_{i_A \in [n]^{k/2}} \sum_{i_B \in [n]^{k/2}} |M_{i_A, i_B}|^{p_k} \Big)^{1/p_k} \\
&\overset{\text{Lem 11.3}}{\leq} \underbrace{\Big(\sum_{i_A \in [n]^{k/2}} \|M_{i_A}\|_2^{p_{k/2}} \Big)}_{(*)}{}^{1/(2p_{k/2})} \cdot \underbrace{\Big(\sum_{i_B \in [n]^{k/2}} \|M^{(i_B)}\|_2^{p_{k/2}} \Big)}_{(**)}{}^{1/(2p_{k/2})}
\end{aligned}$$

Now we have broken the sum into two parts and we can bound both by the same quantity.

Claim I. *One has* $(*), (**) \leq e^{p_{k/2}} \cdot C_{k/2}^{p_{k/2}}$.

Proof of Claim I. For symmetry reasons it suffices to upper bound the column sum $(**)$. For each $x_A \in \{-1,1\}^{nk/2}$ we consider the function $h_{x_A} : \{-1,1\}^{nk/2} \to \mathbb{R}$ which is the restriction of f when fixing x_A. We know that we can write

$$h_{x_A}(x_B) = f(x_A, x_B) = \sum_{i_B \in [n]^{k/2}} \underbrace{\Big(\sum_{i_A \in [n]^{k/2}} \hat{f}(i_A, i_B) \cdot \chi_{i_A}(x_A) \Big)}_{=: g_{i_B}(x_A)} \cdot \chi_{i_B}(x_B) \quad (*)$$

where we have abbreviated $g_{i_B}(x_A)$ as the arising Fourier coefficients of that restriction.

For each i_B we can bound the length of each column by

$$\begin{aligned}
\|M^{(i_B)}\|_2 &= \Big(\sum_{i_A \in [n]^{k/2}} \hat{f}(i_A, i_B)^2 \Big)^{1/2} \qquad (11.5) \\
&\overset{\text{Parsival}}{=} \|g_{i_B}\|_{E,2} \\
&\overset{\text{hypercontr.}}{\leq} \big(e^{\frac{2}{p_{k/2}} - 1} \big)^{k/2} \cdot \|g_{i_B}\|_{p_{k/2}} \\
&= \exp\Big(\underbrace{\Big(\frac{k+2}{k} - 1 \Big) \cdot \frac{k}{2}}_{=1} \Big) \cdot \|g_{i_B}\|_{p_{k/2}} \\
&= e \cdot \|g_{i_B}\|_{p_{k/2}}
\end{aligned}$$

where we use hypercontractivity from Lemma 5.30 with parameter $p_{k/2} = \frac{2k}{k+2} \in [1,2)$ as. We observe that for any $r \geq 1$ one has

$$\begin{aligned}
\sum_{i_B \in [n]^{k/2}} \|g_{i_B}\|_{E,r}^r &\overset{\text{Def } \|\cdot\|_{E,r}}{=} \sum_{i_B \in [n]^{k/2}} \mathop{\mathbb{E}}_{x_A \sim \{-1,1\}^{nk/2}} [|g_{i_B}(x_A)|^r] \qquad (11.6) \\
&= \mathop{\mathbb{E}}_{x_A \sim \{-1,1\}^{nk/2}} \Big[\sum_{i_B \in [n]^{k/2}} |g_{i_B}(x_A)|^r \Big] \\
&\overset{\text{Def } \hat{\|} \cdot \hat{\|}_r}{=} \mathop{\mathbb{E}}_{x_A \sim \{-1,1\}^{nk/2}} [\hat{\|} h_{x_A} \hat{\|}_r^r]
\end{aligned}$$

using that $g_{i_B}(x_A)$ are the Fourier coefficients of the restriction h_{x_A}. Then we can bound

$$\begin{aligned}(**) &= \sum_{i_B \in [n]^{k/2}} \|M^{(i_B)}\|_2^{p_{k/2}} \\ &\overset{(11.5)}{\leq} e^{p_{k/2}} \sum_{i_B \in [n]^{k/2}} \|g_{i_B}\|_{p_{k/2}}^{p_{k/2}} \\ &\overset{(11.6)}{=} e^{p_{k/2}} \cdot \mathop{\mathbb{E}}_{x_A \sim \{-1,1\}^{nk/2}} [\hat{\|} h_{x_A} \hat{\|}_{p_{k/2}}^{p_{k/2}}] \\ &\overset{\text{induction}}{\leq} e^{p_{k/2}} \cdot C_{k/2}^{p_{k/2}}\end{aligned}$$

using the inductive hypothesis with the fact that the restriction is also bounded, i.e. $\|h_{x_A}\|_\infty \leq 1$. That concludes the proof of Claim I. $\square$

We continue the main claim. We can bound

$$\hat{\|} f \hat{\|}_{p_{k/2}} \leq (*)^{1/(2p_{k/2})} \cdot (**)^{1/(2p_{k/2})} \overset{\text{Claim I}}{\leq} \left(e^{p_{k/2}} \cdot C_{k/2}^{p_{k/2}}\right)^{1/p_{k/2}} = e \cdot C_{k/2}$$

Hence we obtain the recursion of $C_k \leq e \cdot C_{k/2}$ which can be resolved to $C_k \leq e^{\log_2(k)} = k^{\log_2(e)} \leq k^{1.45}$. $\square$

11.4 An application to the Aaronson-Ambainis Conjecture

We will now demonstrate that the Bohnenblust-Hille inequality implies the Aaronson-Ambainis conjecture for a special class of functions:

Theorem 11.7. *Let $f : \{-1,1\}^{nk} \to [-1,1]$ be a bounded k-multilinear form where for some $\alpha > 0$ one has $\hat{f}(i) \in \{-\alpha, \alpha\}$ for all $i \in [n]^k$. Then*

$$Inf_{\max}[f] \geq \frac{Var[f]^2}{\Theta(k^3)}$$

Proof. One has $\mathrm{Var}[f] = n^k \alpha^2$ and $\mathrm{Inf}_j[f] = n^{k-1}\alpha^2$ for each variable j as one can easily see. Using the Bohnenblust-Hille inequality (Theorem 11.6) with $p := \frac{2k}{k+1}$ we can bound the p-norm of the Fourier coefficients by

$$n^{(k+1)/2} \cdot \alpha = (n^k \alpha^p)^{1/p} = \hat{\|} f \hat{\|}_p \leq C_k \leq O(k^{3/2})$$

We rearrange to obtain an upper bound of $\alpha \leq O(\frac{k^{3/2}}{n^{(k+1)/2}})$. This gives us exactly the saving that we need and

$$\frac{\mathrm{Var}[f]^2}{\mathrm{Inf}_{\max}[f]} = \frac{n^{2k}\alpha^4}{n^{k-1}\alpha^2} = n^{k+1}\alpha^2 \leq O(k^3)$$

□

11.5 A generalization and an application to learning low degree functions

In Theorem 11.6 we stated the Bohnenblust-Hille inequality for k-multilinear functions with a bound of $C_k \leq \mathrm{poly}(k)$. But of course, the same inequality makes sense for arbitrary functions on the hypercube. Here is what is known:

Theorem 11.8 (Bohnenblust-Hille Inequality III [DMoP19])**.** *For any function* $f : \{-1,1\}^n \to \mathbb{R}$ *of degree at most* k *and any* $p \geq \frac{2k}{k+1}$ *one has*

$$\hat{\|} f \hat{\|}_p = \Big(\sum_{S \subseteq [n]} |\hat{f}(S)|^p \Big)^{1/p} \leq \tilde{C}_k \cdot \|f\|_\infty$$

Here one can pick $\tilde{C}_k \leq \exp(\Theta(\sqrt{k}\log(k)))$.

It is unknown if the bound of $\tilde{C}_k$ can be improved, possibly to a $\mathrm{poly}(k)$. In a talk on the work [EI22], Ivanisvili even mentions that there is no known construction proving that $\tilde{C}_k$ needs to grow with k[1]. But just the fact that there is such a constant independent of n already has interesting consequences. We will show case this with an application to learning low degree functions which is due to Eskenazis and Ivanisvili [EI22].

Suppose we are given random query access to a bounded function $f : \{-1,1\}^n \to [-1,1]$ of degree at most d and from the queries, we want to learn f up to an error of ε. More precisely, for some number N, we may draw uniform independent points $x^{(1)}, \ldots, x^{(N)} \sim \{-1,1\}^n$ and are then being informed of the function values $f(x^{(1)}), \ldots, f(x^{(N)}) \in [-1,1]$. From those values we have to construct a function $h : \{-1,1\}^n \to \mathbb{R}$ so that $\|f - h\|_{E,2} \leq \varepsilon$. We note that this is a variant of the question that we studied in Chapter 3.

To warm up, we discuss a classical result:

Theorem 11.9 (Linial, Mansour, Nisan [LMN93])**.** *Let* $\varepsilon > 0$. *Given* $N := \Theta(\frac{d}{\varepsilon^2} n^d \log(n))$ *many random samples from a degree-d function* $f : \{-1,1\}^n \to [-1,1]$*, with high probability one can construct a function* h *so that* $\|f - h\|_{E,2} \leq \varepsilon$.

Proof. Let $\delta > 0$ and $N \in \mathbb{N}$ be parameters that we determine later. We draw samples $x^{(1)}, \ldots, x^{(N)} \sim \{-1,1\}^n$ and use those samples to create an estimate $\alpha_S := \frac{1}{N}\sum_{i=1}^N f(x^{(i)}) \cdot \chi_S(x^{(i)})$ for the Fourier coefficient $\hat{f}(S)$.

---

[1] See https://www.ipam.ucla.edu/abstract/?tid=17275&pcode=CV2022

We have proven in Lemma 3.1 that for a choice of $N = \Theta(\frac{s}{\delta^2})$ one has

$$\Pr[|\hat{f}(S) - \alpha_S| \leq \delta] \geq 1 - e^{-s}$$

for each fixed set S. Then taking the union bound over the $O(n^d)$ many sets S with $|S| \leq d$ and letting $s := \Theta(\log(n^d))$, we know that

$$\Pr[|\hat{f}(S) - \alpha_S| \leq \delta\ \forall |S| \leq d] \geq 1 - \frac{1}{\mathrm{poly}(n)}$$

for $N := \Theta(\frac{d}{\delta^2}\log(n))$. Condition on this event to happen. We set $h := \sum_{|S| \leq d} \alpha_S \chi_S$ as the approximation to f. Then the error satisfies

$$\|f - h\|_{E,2}^2 = \sum_{|S| \leq d} \underbrace{(\hat{f}(S) - \alpha_S)}_{\leq \delta}{}^2 \leq O(n^d) \cdot \delta^2 \leq \varepsilon^2$$

if we make a choice of $\delta := \Theta(\frac{\varepsilon}{n^{d/2}})$. □

If we think of ε and d as constants then this bound is of the form $O_{\varepsilon,d}(n^d \log(n))$. Surprisingly, one can reduce the number of samples down to only $\Theta_{\varepsilon,d}(\log n)$ with almost the same choice of h.

Theorem 11.10 (Eskenazis, Ivanisvili [EI22]). *For any degree-d function $f : \{-1,1\}^n \to [-1,1]$, $N := \frac{e^{\mathrm{poly}(d)}}{\varepsilon^{\Theta(d)}}\log(n) = \Theta_{\varepsilon,d}(\log n)$ many random samples suffice to construct a function h with $\|f - h\|_{E,2} \leq \varepsilon$.*

Proof. As before, we may assume to know estimates α_S so that $|\hat{f}(S) - \alpha_S| \leq \delta$ for all $|S| \leq d$. Again, we know that $N := \Theta(\frac{d}{\delta^2}\log(n))$ samples suffice. So the surprising part is to argue that we can choose $\delta > 0$ independent of n. Let $\mathcal{L} := \{S \subseteq [n] \mid |S| \leq d \text{ and } |\alpha_S| \geq 2\delta\}$ be the *large* Fourier coefficients. Then we define a function h according to our estimates — but only on the large Fourier coefficients. That means we set

$$h(x) := \sum_{S \in \mathcal{L}} \alpha_S \cdot \chi_S(x)$$

The error in the approximation is

$$\|f - h\|_{E,2}^2 = \underbrace{\sum_{S \in \mathcal{L}} (\hat{f}(S) - \alpha_S)^2}_{(*)} + \underbrace{\sum_{|S| \leq d : S \notin \mathcal{L}} \hat{f}(S)^2}_{(**)}$$

We will analyze the parts $(*)$ and $(**)$ separately.

- *Error* $(*)$ *on the large Fourier coefficients.* The number of large Fourier coefficients is
$$|\mathcal{L}| \leq \sum_{S \in \mathcal{L}} \Big(\frac{|\hat{f}(S)|}{\delta}\Big)^{\frac{2d}{d+1}} \overset{\text{Thm } 11.8}{\leq} \Big(\frac{\tilde{C}_d}{\delta}\Big)^{\frac{2d}{d+1}}$$
using the Bohnenblust-Hille inequality III (Theorem 11.8). Here we use that for each $S \in \mathcal{L}$ one has $\frac{|\hat{f}(S)|}{\delta} \geq 1$. The error coming from the large Fourier coefficients can then be bounded by
$$(*) = \sum_{S \in \mathcal{L}} (\hat{f}(S) - \alpha_S)^2 \leq |\mathcal{L}|\delta^2 \leq \Big(\frac{\tilde{C}_d}{\delta}\Big)^{2d/(d+1)} \delta^2 = \delta^{2/(d+1)} \cdot C_d^{2d/(d+1)} \leq \frac{\varepsilon}{2}$$
for a choice of $\delta \leq \frac{\varepsilon^{\Theta(d)}}{e^{\text{poly}(d)}}$.
- *Error* $(**)$ *on the small Fourier coefficients.* We have
$$(**) = \sum_{S \notin \mathcal{L}} \hat{f}(S)^2 \leq (3\delta)^{2/(d+1)} \sum_{S \notin \mathcal{L}} \hat{f}(S)^{2d/(d+1)} \overset{\text{Thm } 11.8}{\leq} (3\delta)^{2/(d+1)} \tilde{C}_d^{2d/(d+1)} \leq \frac{\varepsilon}{2}$$
for the same choice of δ. Here we use that $\hat{f}(S)^2 \leq (3\delta)^{2/(d+1)} \hat{f}(S)^{2d/(d+1)}$ since for each $S \notin \mathcal{L}$ we know that $|\hat{f}(S)| \leq 3\delta$.

This concludes the claim. $\square$

Chapter 12

Isoperimetric inequalities

Generally speaking, an *isoperimetric inequality* considers a set with a given volume or measure and lower bounds the volume or measure of its surface. Of course it is natural to ask such questions for the hypercube as well. We will discuss one particular variant due to Talagrand [Tal93] where our exposition follows the simpler proof presented by Eldan, Kindler, Lifshitz and Minzer [EKLM25].

12.1 Talagrand's Isoperimetric Inequality

First we introduce the necessary notation. Throughout this chapter we will restrict our attention to *boolean functions* $f : \{-1,1\}^n \to \{-1,1\}$. Recall that in Section 9.4, we defined the *sensitivity*

$$S(f,x) := |\{i \in [n] : f(x) \neq f(x^{\oplus i})\}|$$

of the function f at point $x \in \{-1,1\}^n$ as the number of coordinates that flip the value.

We will also revisit the *variance* of a function (see also Section 1.9) which is

$$\mathrm{Var}[f] = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\left(f(x) - \mathop{\mathbb{E}}_{y\sim\{-1,1\}^n}[f(y)]\right)^2\right] = \sum_{\emptyset\subset S\subseteq[n]} \hat{f}(S)^2$$

Recall that for a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ with $p := \Pr_{x\in\{-1,1\}^n}[f(x) = 1]$ as the fraction of points with value 1, one can conveniently write

$$\mathrm{Var}[f] = 4p(1-p) \in [0,1]$$

The main result of this chapter will be:

Theorem 12.1 (Talagrand [Tal93])**.** *For any non-constant $f : \{-1,1\}^n \to \{-1,1\}$ one has*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{S(f,x)}\right] \geq \Omega\Big(\mathrm{Var}[f]\cdot\sqrt{\log\Big(\frac{2}{\mathrm{Var}[f]}\Big)}\Big) \tag{12.1}$$

It will be instructive to study this inequality on two natural functions:

- *Majority.* For n odd, let $f : \{-1,1\}^n \to \{-1,1\}$ be the *majority function* (see Section 5.3.1). Here we have a $\Theta(\frac{1}{\sqrt{n}})$ fraction of points x where $S(f,x) = \Theta(n)$ and $S(f,x) = 0$ for all other points. On the other hand, $\mathrm{Var}[f] = 1$ and so both sides of inequality (12.1) are $\Theta(1)$ implying that Talagrand's inequality is tight here.

- *Subcubes.* For some $k \in \{1,\dots,n\}$, consider the function

  $$f(x) = \begin{cases} 1 & \text{if } x_1 = \ldots = x_k = 1 \\ -1 & \text{otherwise} \end{cases}$$

  Then $\mathrm{Var}[f] = \Theta(2^{-k})$ and so the right hand side of (12.1) is $\Theta(\sqrt{k}\cdot 2^{-k})$. For the left hand side, points x in the subcube (which are a 2^{-k} fraction of points) have $S(f,x) = k$. Points that are Hamming neighbors of the subcube — which is a $\Theta(k\cdot 2^{-k})$ fraction of points — have $S(f,x) = 1$. Hence the left hand side is $\Theta(\frac{\sqrt{k}+k}{2^k})$, meaning that Talagrand's inequality is only tight for constant k but not for superconstant k. But there is an asymmetry to the situation to which we will get back later.

12.2 The proof of Talagrand's inequality

In this section, we will present a proof of Theorem 12.1. As indicated earlier we will not use Talagrand's original proof but the more recent one by [EKLM25]. Recall that the *derivative* of f at a point $x \in \{-1,1\}^n$ in direction i is

$$D_i f(x) = \frac{1}{2}\cdot(f(x^{i\to 1}) - f(x^{i\to -i})),$$

see Section 1.8. We use this to define the *gradient* $\nabla f(x) \in \mathbb{R}^n$ as

$$\nabla f(x) := \big(D_1 f(x),\dots,D_n f(x)\big)$$

For hypercube vertices $x \in \{-1,1\}^n$, our definition $\nabla f(x)$ indeed coincides with the gradient of the multilinear function f — and we will never consider it on other points anyway. Most of the time we are only interested in the absolute value of the derivative which simplifies to

$$|D_i f(x)| = \frac{1}{2}|f(x) - f(x^{\oplus i})|.$$

We also write $|\nabla f(x)| \in \mathbb{R}^n$ for the vector with entries corresponding to the absolute values of the gradient. We note that for any boolean function f and $x \in \{-1,1\}^n$ one has $\sqrt{|S(f,x)|} = \|\nabla f(x)\|_2$, hence Talagrand's inequality actually lower bounds the average Euclidean length of the gradient of f.

Gradient length vs. Level-1 weight

Recall that $W_k[f] = \sum_{|S|=k} \hat{f}(S)^2$ is the *Level-k weight* and for an interval I we write $W_I[f] = \sum_{|S|\in I} \hat{f}(S)^2$. First we prove a lower bound on the length of the gradient in terms of the level-1 weight. That lower bound by itself might not be very strong; in fact any character function $f := \chi_S$ for $|S| \geq 2$ has variance 1 and $W_1[f] = 0$.

Lemma 12.2. *For any $f : \{-1,1\}^n \to \{-1,1\}$ one has*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2] \geq \sqrt{W_1[f]}$$

Proof. For a boolean function f we have $D_i f(x) \in \{-1,0,1\}$ and so

$$\begin{aligned} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|\nabla f(x)|_i] &= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|D_i(x)|] \qquad (12.2) \\ &= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|D_i(x)|^2] \overset{\text{Def Inf}}{=} \mathrm{Inf}_i[f] \geq \hat{f}(\{i\})^2 \end{aligned}$$

Then

$$\begin{aligned} \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2] &\overset{\text{Jensen}}{\geq} \left\| \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[|\nabla f(x)|] \right\|_2 \\ &\overset{(12.2)}{\geq} \left| \sum_{i=1}^n \hat{f}(\{i\})^2 \right|^{1/2} \\ &= \sqrt{W_1[f]} \end{aligned}$$

where we use Jensen's inequality[1] with the convexity of $\|\cdot\|_2$. □

---

[1] See Theorem 1.45 or to be more precise its extension to convex function $F : \mathbb{R}^n \to \mathbb{R}$.

Gradient length vs. approximate level d weight

Next, we want to compare to average length of the gradient to level d weights. We will do this by considering a random restriction of f and then applying Lemma 12.2 to the random restriction.

Lemma 12.3. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and let* $d \in \mathbb{N}$. *Then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2] \gtrsim \sqrt{d}\cdot W_{[d,2d)}[f]$$

Proof. First fix a point $x \in \{-1,1\}^n$. We consider the following random experiment: we sample $J \subseteq [n]$ so that each index $i \in [n]$ is included independently with probability $\frac{1}{d}$ and let $g := f_{|J,x_{\bar{J}}}$ be the restriction of f to variables J. Then

$$\mathop{\mathbb{E}}_{J}[\|\nabla g(x_J)\|_2] \overset{\text{Jensen}}{\leq} \Big(\mathop{\mathbb{E}}_{J}[\|\nabla g(x_J)\|_2^2]\Big)^{1/2} = \Big(\sum_{i=1}^{n} \underbrace{\mathop{\mathbb{E}}_{J}[\mathbf{1}_{i\in J}]}_{=1/d} \cdot D_i f(x)^2\Big)^{1/2} = \frac{\|\nabla f(x)\|_2}{\sqrt{d}} \tag{12.3}$$

where we use Jensen's inequality with the concavity of $z \mapsto \sqrt{z}$. Then averaging over $x \sim \{-1,1\}^n$ gives

$$\begin{aligned}
\frac{\mathbb{E}_x[\|\nabla f(x)\|_2]}{\sqrt{d}} &\overset{(12.3)}{\geq} \mathop{\mathbb{E}}_{x}\Big[\mathop{\mathbb{E}}_{J}[\|\nabla g(x_J)\|_2]\Big] \\
&= \mathop{\mathbb{E}}_{J,x_{\bar{J}}}\Big[\mathop{\mathbb{E}}_{x_J}\big[\|\nabla g(x_J)\|_2\big]\Big] \\
&\overset{\text{Lem 12.2}}{\gtrsim} \mathop{\mathbb{E}}_{J,x_{\bar{J}}}\Big[\sqrt{W_1[g]}\Big] \\
&\geq \mathop{\mathbb{E}}_{J,x_{\bar{J}}}[W_1[g]] \\
&\overset{\text{Lem 10.24}}{\gtrsim} W_{[d,2d)}[f]
\end{aligned}$$

using that $0 \leq W_1[g] \leq 1$ as g is always a boolean function. Moreover, we used in the last step the properties of random restrictions from Lemma 10.24. □

We can bootstrap this inequality to get a slightly stronger bound:

Lemma 12.4. *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *and let* $d \in \mathbb{N}$. *Then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2] \gtrsim \sqrt{d}\cdot W_{\geq d}[f]$$

Proof. One has $\frac{1}{2+\sqrt{2}}\sum_{j=0}^{\infty}2^{-j/2}=1$. Applying the previous Lemma to $d' := 2^j d$ for $j=0,1,2,\ldots$ gives

$$\frac{\mathbb{E}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2]}{\sqrt{d}} = \frac{1}{2+\sqrt{2}}\sum_{j=0}^{\infty}\frac{\mathbb{E}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2]}{\sqrt{2^j d}} \overset{\text{Lem 12.3}}{\gtrsim} \sum_{j=0}^{\infty} W_{[2^j d,2^{j+1}d)}[f] = W_{\geq d}[f]$$

□

The main proof

One last technical ingredient that we need is that low variance boolean functions must have most of their Fourier weight on higher levels. This is not a new insight; we have seen similar argument e.g. in the proof of the KKL Theorem (see Prop 5.9). It will be more convenient to switch to $\{0,1\}$-values here.

Lemma 12.5. *There is a constant $\varepsilon > 0$ so that for any function $h : \{-1,1\}^n \to \{0,1\}$ with $Var[h] \leq \varepsilon$ one has $W_{>d}[h] \geq 0.99 \cdot Var[h]$ where $d := \frac{1}{8}\log_2(\frac{1}{Var[h]})$.*

Proof. Let $\alpha := \Pr_{x\sim\{-1,1\}^n}[h(x)=1]$. After possibly replacing h by $1-h$ which does not affect the variance or high level weight we may assume that $0 \leq \alpha \leq \frac{1}{2}$. As $\text{Var}[h] = \alpha(1-\alpha)$ we know that the assumption means that we can make α as small as we want. We recall that for any $p \geq 1$, the function h has a norm of $\|h\|_{E,p} = \alpha^{1/p}$. Using Hölder's Inequality and the Bonami Lemma we have

$$\begin{aligned} W_{[0,d]}[h] &= \|h^{\leq d}\|_{E,2}^2 \\ &\leq \langle h, h^{\leq d}\rangle_E \\ &\overset{\text{Thm 1.49}}{\leq} \|h\|_{E,4/3}\|h^{\leq d}\|_{E,4} \\ &\overset{\text{Thm 5.4}}{\leq} \|h\|_{E,4/3}\sqrt{3}^d\|h\|_{E,2} \leq \alpha^{5/4}\cdot\left(\frac{1}{\text{Var}[h]}\right)^{1/8} \leq 0.01\text{Var}[h] \end{aligned}$$

for α small enough. □

Now we have everything together for the proof of the main result of this chapter, which we restate for convinience:

Theorem (Talagrand [Tal93])**.** *For any non-constant $f : \{-1,1\}^n \to \{-1,1\}$ one has*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{S(f,x)}\right] \geq \Omega\left(Var[f]\cdot\sqrt{\log\frac{2}{Var[f]}}\right)$$

Proof. Applying Lemma 12.3 with $d = 1$ gives

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{S(f,x)}\right] = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\|\nabla f(x)\|_2] \gtrsim W_{\geq 1}[f] = \mathrm{Var}[f]$$

In particular this settles Talagrand's inequality whenever $\mathrm{Var}[f] \geq \Omega(1)$. Now consider the case where $\mathrm{Var}[f] \leq \varepsilon$ for a small enough constant $\varepsilon > 0$. Given f, define $h : \{-1,1\}^n \to \{0,1\}$ so that $f(x) = 1 - 2h(x)$. Then $|\hat{h}(S)| = \frac{1}{2}|\hat{f}(S)|$ for $|S| \geq 1$ and $\mathrm{Var}[h] = \frac{1}{4}\mathrm{Var}[f]$. Hence setting $d := \frac{1}{8}(\log_2(\frac{1}{\mathrm{Var}[f]}) + 2)$, we can apply Lemma 12.5 and conclude that

$$W_{\geq d}[f] \geq 0.99\mathrm{Var}[f]$$

We now apply Lemma 12.3 to obtain

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{S(f,x)}\right] \gtrsim \sqrt{d}\cdot \underbrace{W_{\geq d}[f]}_{\geq \Omega(\mathrm{Var}[f])} \geq \Theta\left(\sqrt{\log\frac{2}{\mathrm{Var}[f]}}\cdot \mathrm{Var}[f]\right)$$

which concludes the proof of the theorem. □

We recall our discussion from Section 12.1 that Talagrand's inequality was not quite tight for subcubes of codimension $k \gg 1$ as we had two classes of edges and one contributed more than the other. But Talagrand offered an extension that can prove a tight bound for this case as well. For a boolean function f and $x \in \{-1,1\}^n$, let $E_f(x) := \{(x, x^{\oplus i}) \mid f(x) \neq f(x^{\oplus i})$ where $i \in [n]\}$ be the sensitive edges incident to x. Consider a 2-coloring of the hypercube edges and call the color classes `RED` and `BLUE`. Define

$$\begin{aligned} S_{\mathrm{blue}}(f,x) &:= \begin{cases} |E_f(x) \cap \mathtt{BLUE}| & \text{if } f(x) = 1 \\ 0 & \text{if } f(x) = -1 \end{cases} \\ S_{\mathrm{red}}(f,x) &:= \begin{cases} 0 & \text{if } f(x) = 1 \\ |E_f(x) \cap \mathtt{RED}| & \text{if } f(x) = -1 \end{cases} \end{aligned}$$

In other words, $+1$-valued points only pay for the incident blue edges, -1-valued points only pay for the incident red edges. One can prove:

Theorem 12.6 (Talagrand [Tal93])**.** *For any non-constant $f : \{-1,1\}^n \to \{-1,1\}$ and any 2-coloring of the hypercube edges into RED and BLUE edges one has*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{S_{red}(f,x)} + \sqrt{S_{blue}(f,x)}\right] \geq \Omega\left(\mathrm{Var}[f]\cdot\sqrt{\log\frac{2}{\mathrm{Var}[f]}}\right) \tag{12.4}$$

Now, let us go back to the subcube function

$$f(x) = \begin{cases} 1 & \text{if } x_1 = \ldots = x_k = 1 \\ -1 & \text{otherwise} \end{cases}$$

from the Section 12.1. We color all edges BLUE, so that we only pay for incidences of +1-points. And indeed the left hand side of (12.4) is $\Theta(\sqrt{k} \cdot 2^{-k})$ which is the same as the right hand side.

12.3 The vertex and edge boundaries

We also want to discuss an isoperimetric inequality due to Margulis [Mar74] which can be derived from Theorem 12.1.

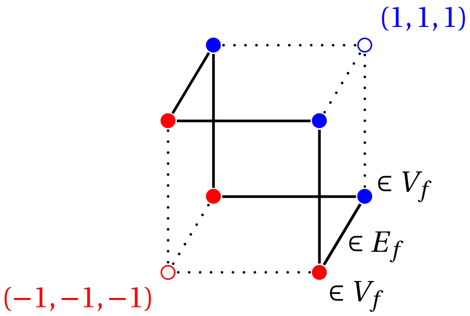
Definition 12.7. For a boolean function $f : \{-1,1\}^n \to \{-1,1\}$, the *vertex boundary*

$$V_f := \left\{ x \in \{-1,1\}^n \mid S(f,x) > 0 \right\}$$

are all the vertices that have a neighbor with different value. Moreover the *edge boundary*

$$E_f := \left\{ (x, x^{\oplus i}) \mid x \in \{-1,1\}^n,\ i \in [n] \text{ and } f(x) \neq f(x^{\oplus i}) \right\}$$

are all the edges between Hamming neighbors with different signs.



Vertex/edge boundaries for Maj_3

$f(x) = 1$ for blue points, $f(x) = -1$ for red points.

For a function with $\text{Var}[f] = \Theta(1)$ it is possible that $\frac{|V_f|}{2^n} = \Theta(\frac{1}{\sqrt{n}})$ (for example for the Majority function) and it is also possible that $|E_f| = \Theta(1)$ — see e.g. dictatorship functions. But it is not possible that both are true for the same function.

Theorem 12.8 (Margulis [Mar74])**.** *For any* $f : \{-1,1\}^n \to \{-1,1\}$ *one has*

$$\frac{|V_f|}{2^n} \cdot \frac{|E_f|}{2^n} \geq \Omega(\text{Var}[f]^2)$$

We note that Margulis' Theorem preceeds the one of Talagrand by two decades. But as we have proven Talagrand's Theorem already, we can now use it to derive Theorem 12.8.

Proof of Theorem 12.8. We can write

$$\begin{aligned}
\frac{|V_f|}{2^n}\cdot\frac{|E_f|}{2^n} &= \frac{1}{2}\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\mathbf{1}_{S(f,x)>0}\right]\cdot\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[S(f,x)] \\
&\overset{(*)}{\geq} \frac{1}{2}\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{\mathbf{1}_{S(f,x)>0}}\cdot\sqrt{S(f,x)}\right]^2 \\
&= \frac{1}{2}\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\sqrt{S(f,x)}\right]^2 \\
&\gtrsim \mathrm{Var}[f]^2
\end{aligned}$$

using Cauchy-Schwarz in $(*)$ and the basic version of Talagrand's Theorem in the form of $\mathbb{E}_{x\sim\{-1,1\}^n}[\sqrt{S(f,x)}]\gtrsim\mathrm{Var}[f]$ in $(**)$. □

Bibliography

[AA14] Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *Theory Comput.*, 10:133–166, 2014.

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach.* Cambridge University Press, 2009.

[ABD24] Avantika Agarwal and Shalev Ben-David. Oracle separations for the quantum-classical polynomial hierarchy, 2024.

[AS16] Noga Alon and Joel H. Spencer. *The probabilistic method.* Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[Bac12] Arturs Backurs. Influences in low-degree polynomials. 2012.

[BB14] Arturs Backurs and Mohammad Bavarian. On the sum of L1 influences. In *CCC*, pages 132–143. IEEE Computer Society, 2014.

[BBC$^+$01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[Bd02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002. Complexity and Logic.

[BH31] H. F. Bohnenblust and Einar Hille. On the absolute convergence of dirichlet series. *Annals of Mathematics*, 32(3):600–622, 1931.

[Ble01] Ron Blei. *Analysis in Integer and Fractional Dimensions.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

[BLR90] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *STOC*, pages 73–83. ACM, 1990.

[CFGS88] F.R.K Chung, ZoltÃ¡n FÃ¼redi, R.L Graham, and P Seymour. On induced subgraphs of the cube. *Journal of Combinatorial Theory, Series A*, 49(1):180–187, 1988.

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 205–214, New York, NY, USA, 2006. Association for Computing Machinery.

[DFKO06] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the fourier tails of bounded functions over the discrete cube. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 437–446, New York, NY, USA, 2006. Association for Computing Machinery.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007.

[DMoP19] Andreas Defant, Mieczysł aw Mastył o, and Antonio Pérez. On the Fourier spectrum of functions on Boolean cubes. *Math. Ann.*, 374(1-2):653–680, 2019.

[DPS10] Andreas Defant, Dumitru Popa, and Ursula Schwarting. Coordinatewise multiple summing operators in banach spaces. *Journal of Functional Analysis*, 259(1):220–242, 2010.

[EI22] Alexandros Eskenazis and Paata Ivanisvili. Learning low-degree functions from a logarithmic number of random queries. In *STOC*, pages 203–207. ACM, 2022.

[EKLM25] Ronen Eldan, Guy Kindler, Noam Lifshitz, and Dor Minzer. Isoperimetric Inequalities Made Simpler. *Discrete Analysis*, jul 24 2025.

[EZ64] H. Ehlich and K. Zeller. Schwankung von polynomen zwischen gitterpunkten. *Mathematische Zeitschrift*, 86(1):41–44, 1964.

[FHKL15] Yuval Filmus, Hamed Hatami, Nathan Keller, and Noam Lifshitz. On the sum of the l1 influences of bounded functions, 2015.

[GL92] C Gotsman and N Linial. The equivalence of two problems on the cube. *Journal of Combinatorial Theory, Series A*, 61(1):142–146, 1992.

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995.

[HKP11] Pooya Hatami, Raghav Kulkarni, and Denis Pankratov. *Variations on the Sensitivity Conjecture*. Number 4 in Graduate Surveys. Theory of Computing Library, 2011.

[Hol07] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, pages 411–419, New York, NY, USA, 2007. Association for Computing Machinery.

[H01̊] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, July 2001.

[Hua19] Hao Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Annals of Mathematics*, 190(3):949–955, 11 2019.

[JOW15] Jacek Jendrej, Krzysztof Oleszkiewicz, and Jakub O. Wojtaszczyk. On some extensions of the fkn theorem. *Theory of Computing*, 11(18):445–469, 2015.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, New York, NY, USA, 2002. Association for Computing Machinery.

[LIT30] J. E. LITTLEWOOD. On bounded bilinear forms in an infinite number of variables. *The Quarterly Journal of Mathematics*, os-1(1):164–174, 01 1930.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, July 1993.

[LZ23] Shachar Lovett and Jiapeng Zhang. Fractional certificates for bounded functions. In *ITCS*, volume 251 of *LIPIcs*, pages 84:1–84:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

[Mar74] G. A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy Peredači Informacii*, 10(2):101–108, 1974.

[Min21] Dor Minzer. Topics in combinatorics: Analysis of boolean functions, 2021.

[Min22] Dor Minzer. Topic in tcs / probabilistically checkable proofs, 2022.

[Mon12] Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *J. Math. Phys.*, 53(12):122206, 15, 2012.

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.

[Mos14] Dana Moshkovitz. Parallel repetition from fortification. In *FOCS*, pages 414–423. IEEE Computer Society, 2014.

[MP69] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, USA, 1969.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[NS92] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. In *STOC*, pages 462–467. ACM, 1992.

[O'D21] Ryan O'Donnell. Analysis of boolean functions. *CoRR*, abs/2105.10386, 2021.

[OSSS05] Ryan O'Donnell, Michael E. Saks, Oded Schramm, and Rocco A. Servedio. Every decision tree has an influential variable. In *FOCS*, pages 31–39. IEEE Computer Society, 2005.

[PSS12] Daniel Pellegrino and Juan B. Seoane-Sepúlveda. New upper bounds for the constants in the Bohnenblust-Hille inequality. *J. Math. Anal. Appl.*, 386(1):300–307, 2012.

[Rao08] Anup Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 1–10, New York, NY, USA, 2008. Association for Computing Machinery.

[Raz95] Ran Raz. A parallel repetition theorem. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '95, pages 447–456, New York, NY, USA, 1995. Association for Computing Machinery.

[RC66] T. J. Rivlin and E. W. Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on Numerical Analysis*, 3(2):311–320, 1966.

[Tal93] M. Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric & Functional Analysis GAFA*, 3(3):295–314, 1993.

[Tal95] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995.