http://www.math.uni-konstanz.de/~schweigh/

Fragmentary lecture notes closely following Ryan O'Donnell's book [O'D].

Fourier Analysis of Boolean Functions

Winter Term 2016/2017

Markus Schweighofer

Version of Friday 21st April, 2017, 08:40

Universität Konstanz, Germany

These **incomplete** lecture notes follow very closely Ryan O'Donnell's book [O'D] an electronic version of which can be downloaded freely from:

http://analysisofbooleanfunctions.net

Please report any ambiguities and errors (including typos) to:

markus.schweighofer@uni-konstanz.de

We will always denote by $\mathbb{N} := \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$ the set of positive and nonnegative integers, respectively.

Disclaimer: This document does not claim any originality and is mostly based on the work of other people. We freely reproduce a small part of the ideas presented in the book of Ryan O'Donnell. For the relevant scientific sources, we refer to the bibliography of O'Donnell's book.

Contents

§1 Boolean functions and the Fourier expansion

§1.1 Boolean values and operations

In computer science, the two-element set of Boolean values

$$\mathbb{B} = \{\text{FALSE}, \text{TRUE}\}$$

together with the unary operation NOT and binary operations like AND, OR, XOR (exclusive or) are omnipresent. If one encodes FALSE by 0 and TRUE by 1, as it is common in computer science, the XOR and AND operations become just addition and multiplication in the finite two-element field

$$\mathbb{F}_2 = \{0, 1\}.$$

It is therefore common to work with the field $\mathbb{F}_2$ rather than with $\mathbb{B}$. Perhaps unexpectedly, we will work however most of the time with the field of real numbers $\mathbb{R}$ instead of $\mathbb{F}_2$. At first sight, this seems to complicate things since the addition is no longer the important XOR operation. If we encode FALSE by the *real number* 0 and TRUE by the *real number* 1 and if we have a Boolean-valued function $f\colon D \to \{0,1\} \subseteq \mathbb{R}$ defined on a finite set D, then we can *count* the number of all $x \in D$ mapped to TRUE by $\sum_{x \in D} f(x)$. For such matters, we often work with

$$\{0, 1\} \subseteq \mathbb{R}.$$

In fact, it turns out that most of the time still another model is more advantageous: We will preferably use

$$\{-1, 1\} \subseteq \mathbb{R}$$

with the perhaps counterintuitive convention that

$$-1 \text{ stands for TRUE and } 1 \text{ for FALSE.}$$

The advantages of working with $\{-1, 1\}$ will become clear shortly. For the time being, just note that a Boolean-valued function $f\colon D \to \{-1,1\} \subseteq \mathbb{R}$ defined on a finite set D attains both truth values equally often if and only if $\sum_{x \in D} f(x) = 0$. The reasons for our choice of the counterintuitive convention that -1 stands for TRUE are not very important but note that $-1 = (-1)^1$ and $1 = (-1)^0$. Also note that in this way the multiplication on $\{-1, 1\}$ gets the XOR operation.

Remark 1.1.1. In the following, we will freely move between our different boolean models which are resumed in the following table:

$\mathbb{B}$	$\mathbb{F}_2 = \{0,1\}$	$\{0,1\} \subseteq \mathbb{R}$	$\{-1,1\} \subseteq \mathbb{R}$
FALSE	0	0	1
TRUE	1	1	-1
NOT	$x \mapsto 1+x$	$x \mapsto 1-x$	$x \mapsto -x$
AND	$\cdot$	$\cdot$	$(x,y) \mapsto \frac{1+x+y-xy}{2}$
OR	$(x,y) \mapsto x+y+xy$	$(x,y) \mapsto x+y-xy$	$(x,y) \mapsto \frac{-1+x+y+xy}{2}$
XOR	$+$	$(x,y) \mapsto x+y-2xy$	$\cdot$

Priority will be given to the models from right to left.

§1.2 Real-valued Boolean functions and their Fourier transform

Definition 1.2.1. A *real-valued Boolean function* is a function from $\{-1,1\}^n$ to $\mathbb{R}$ where $n \in \mathbb{N}_0$. The number n is called its *number of input bits*. A *Boolean function* is a function from $\{-1,1\}^n$ to $\{-1,1\}$, in other words a real-valued Boolean function that is Boolean-valued.

The previous definition should be understood in a flexible sense according to Remark 1.1.1. Often, we might use a different model for truth values on the source and the target. For example, a function from $\mathbb{F}_2^6$ to $\{0,1\} \subseteq \mathbb{R}$ is a Boolean function on 6 bits. The main aim of this lecture is to investigate Boolean functions. However, many concepts generalize automatically to *real-valued* Boolean functions and the latter are also an important tool.

Remark 1.2.2. Remember that $\{-1,1\}^0$ is a singleton whose only element is the empty tuple $()$ (which equals the empty map $\emptyset \to \{-1,1\}$ or the empty set $\emptyset$). A real-valued Boolean function on 0 bits is thus given by a real number, and a Boolean function on 0 bits is given by a truth value.

Notation 1.2.3. Let $n \in \mathbb{N}_0$.

(a) We write $[n] := \{1,\dots,n\}$.

(b) For $S \subseteq [n]$, we introduce the *monomial function*

$$\chi_S\colon \{-1,1\}^n \to \mathbb{R},\ x \mapsto x^S := \prod_{i\in S} x_i.$$

(c) For any set A, we denote by

$$\mathscr{P}(A) := \{S \mid S \subseteq A\}$$

its power set, i.e., the set of its subsets.

(d) For sets A and B, we denote by

$$A \triangle B := (A \setminus B) \cup (B \setminus A)$$

their symmetric difference.

(e) For $x \in \{-1,1\}^n$, we introduce

$$\delta_x \colon \{-1,1\}^n \to \mathbb{R},\ y \mapsto \begin{cases} 1 & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases}$$

(f) For $S \subseteq [n]$, we introduce

$$\delta_S \colon \mathscr{P}([n]) \to \mathbb{R},\ T \mapsto \begin{cases} 1 & \text{if } S = T, \\ 0 & \text{otherwise.} \end{cases}$$

(g) We write $\mathbf{x} \sim \{-1,1\}^n$ to denote that $\mathbf{x}$ is a uniformly chosen random element from $\{-1,1\}^n$. Equivalently, the n components of $\mathbf{x}$ are independently chosen to be 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$. We always write random variables in **boldface**. Probabilities $\mathbf{Pr}$ and expectations $\mathbf{E}$ will always be with respect to a uniformly random $\mathbf{x} \sim \{-1,1\}^n$ unless otherwise specified. Thus if $f \colon \{-1,1\} \to \mathbb{R}$, then $\mathbf{E}_{\mathbf{x}}[f(\mathbf{x})]$ stands for $\frac{1}{2^n}\sum_{x\in\{-1,1\}^n} f(x)$. We also often write just $\mathbf{E}[f]$ instead.

Lemma 1.2.4. Let $n \in \mathbb{N}_0$ and $S, T \in \mathscr{P}([n])$. Then $\chi_S\chi_T = \chi_{S\triangle T}$.

Proof. For $x \in \{-1,1\}^n$, we have

$$(\chi_S\chi_T)(x) = \chi_S(x)\chi_T(x) = \prod_{i\in S} x_i \prod_{i\in T} x_i = \prod_{i\in S\triangle T} x_i \prod_{i\in S\cap T} x_i^2 = \prod_{i\in S\triangle T} x_i = \chi_{S\triangle T}(x).$$

□

Lemma 1.2.5. Let $n \in \mathbb{N}_0$ and $S \in \mathscr{P}([n])$. Then

$$\mathbf{E}[\chi_S] = \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}\left[\prod_{i\in S} \mathbf{x}_i\right] = \delta_\emptyset(S) = \begin{cases} 1 & \text{if } S = \emptyset, \\ 0 & \text{if } S \neq \emptyset. \end{cases}$$

Proof. If $S = \emptyset$, then $\mathbf{E}_{\mathbf{x}}[\chi_S(\mathbf{x})] = \mathbf{E}[1] = 1$. Otherwise,

$$\mathop{\mathbf{E}}_{\mathbf{x}}\left[\prod_{i\in S} \mathbf{x}_i\right] = \prod_{i\in S} \mathop{\mathbf{E}}_{\mathbf{x}_i}[\mathbf{x}_i]$$

because the random bits $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independant. But each of the factors $\mathbf{E}_{\mathbf{x}_i}[\mathbf{x}_i]$ in the above *nonempty* product is $\frac{1}{2}1 + \frac{1}{2}(-1) = 0$. □

Definition 1.2.6. Let $n \in \mathbb{N}_0$. We define a scalar product on the finite-dimensional real vector space $\mathbb{R}^{\{-1,1\}^n}$ by

$$\langle f, g\rangle := \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x) = \mathop{\mathbf{E}}_{\mathbf{x} \sim \{-1,1\}^n} [f(\mathbf{x})g(\mathbf{x})] = \mathbf{E}[fg]$$

for $f, g \in \mathbb{R}^{\{-1,1\}^n}$. Equipped with this scalar product, the vector space $\mathbb{R}^{\{-1,1\}^n}$ becomes an euclidean space. It also becomes a normed vector space since the scalar product induces a norm given by

$$\|f\|_2 := \sqrt{\langle f, f\rangle} = \sqrt{\mathbf{E}[f^2]}$$

for $f \in \mathbb{R}^{\{-1,1\}^n}$.

Proposition 1.2.7. *Let $n \in \mathbb{N}_0$. The family $(\chi_S)_{S\subseteq[n]}$ of monomial functions is an orthonormal basis of the euclidean vector space $\mathbb{R}^{\{-1,1\}^n}$.*

Proof. Since $\#\mathscr{P}([n]) = 2^n = \#\{-1,1\}^n = \dim \mathbb{R}^{\{-1,1\}^n}$, it is enough to show that

$$\langle \chi_S, \chi_T\rangle = \delta_S(T)$$

for all $S, T \in \mathscr{P}([n])$. But this follows immediately from Lemmata 1.2.4 and 1.2.5. $\square$

We call $(\chi_S)_{S\subseteq[n]}$ the *Fourier basis* of $\mathbb{R}^{\{-1,1\}^n}$. Our most important tool will be the Fourier expansion of a real-valued Boolean function.

Proposition 1.2.8. *Let $n \in \mathbb{N}_0$. For every $f \colon \{-1,1\}^n \to \mathbb{R}$, there is a unique map*

$$\widehat{f} \colon \mathscr{P}([n]) \to \mathbb{R}$$

such that

$$f(x) = \sum_{S\subseteq[n]} \widehat{f}(S)x^S \tag{1.1}$$

for all $x \in \{-1,1\}^n$. The map

$$\mathscr{F} \colon \mathbb{R}^{\{-1,1\}^n} \to \mathbb{R}^{\mathscr{P}([n])},\ f \mapsto \widehat{f}$$

is a vector space isomorphism.

Proof. Equation (1.1) can be rewritten $f = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S$. The $\widehat{f}(S)$ are therefore just the coefficients of f with respect to the basis $(\chi_S)_{S\subseteq[n]}$ of $\mathbb{R}^{\{-1,1\}^n}$. $\square$

Definition 1.2.9. Let $n \in \mathbb{N}_0$ and $f \colon \{-1,1\}^n \to \mathbb{R}$. The representation (1.1) of f is called the *Fourier expansion* of f. For $S \subseteq [n]$, the real number $\widehat{f}(S)$ is called the *Fourier coefficient* of f on S. The function $\widehat{f}$ is called the *Fourier transform* of f. The map $\mathscr{F}$ is called the *(Boolean) Fourier transform* (on n bits). The degree of f is defined by

$$\deg f := \begin{cases} \max\left\{\#S \mid S \subseteq [n], \widehat{f}(S) \neq 0\right\} & \text{if } f \neq 0 \\ -\infty & \text{if } f = 0. \end{cases}$$

By Proposition 1.2.7 and linear algebra, it is clear that the Fourier coefficient $\widehat{f}(S)$ of f on $S \subseteq [n]$ can be calculated by

$$\widehat{f}(S) = \langle \chi_S, f \rangle = \mathop{\mathbf{E}}_{\mathbf{x} \sim \{-1,1\}^n} \left[\mathbf{x}^S f(\mathbf{x}) \right].$$

Another orthonormal basis of the euclidean space $\{-1,1\}^n$ is the family $(2^{\frac{n}{2}} \delta_x)_{x \in \{-1,1\}^n}$ of *needle functions*. Consider now the Fourier transform $\mathscr{F}$ and its scaled version

$$\mathscr{F}_{\text{scaled}} = 2^{\frac{n}{2}} \mathscr{F} \colon \mathbb{R}^{\{-1,1\}^n} \to \mathbb{R}^{\mathscr{P}([n])},\ g \mapsto 2^{\frac{n}{2}} \widehat{g}.$$

If $f \in \mathbb{R}^{\{-1,1\}^n}$, then the functions $g := 2^{-\frac{n}{2}} f$ and $\widehat{f}$ give the coordinates of f with respect to the basis of needle functions and the Fourier basis, respectively, and $\mathscr{F}_{\text{scaled}}(g) = \widehat{f}$. Hence $\mathscr{F}_{\text{scaled}}$ transforms coordinates with respect to the orthonormal basis $(2^{\frac{n}{2}} \delta_x)_{x \in \{-1,1\}^n}$ into coordinates with respect to the orthonormal basis $(\chi_S)_{S \subseteq [n]}$. In particular, $\mathscr{F}_{\text{scaled}}$ is an *orthogonal* linear map with respect to the *standard scalar products* on $\mathbb{R}^{\{-1,1\}^n}$ and $\mathbb{R}^{\mathscr{P}([n])}$ (which is different on $\mathbb{R}^{\{-1,1\}^n}$ from the one we use). Hence

$$\sum_{x \in \{-1,1\}^n} f(x) g(x) = \sum_{S \subseteq [n]} 2^{\frac{n}{2}} \widehat{f}(S) 2^{\frac{n}{2}} \widehat{g}(S)$$

for all $f, g \in \mathbb{R}^{\{-1,1\}^n}$. This is

Plancherel's Theorem. For any $f, g \colon \{-1,1\}^n \to \mathbb{R}$,

$$\langle f, g \rangle = \mathbf{E}[fg] = \sum_{S \subseteq [n]} \widehat{f}(S) \widehat{g}(S).$$

Of course, it can also be directly verified. A special case of this is

Parseval's Theorem. For any $f \colon \{-1,1\}^n \to \mathbb{R}$,

$$\|f\|_2^2 = \langle f, f \rangle = \mathbf{E}[f^2] = \sum_{S \subseteq [n]} \widehat{f}(S)^2.$$

In particular, if $f \colon \{-1,1\}^n \to \{-1,1\}$ is a Boolean function, then $\|f\|_2 = 1$.

Example 1.2.10. Reconsidering the rightmost column of the table in Remark 1.1.1, we see that indeed

$$\begin{aligned}
\|\,\text{FALSE}\,\|_2 &= \sqrt{1^2} = 1, \\
\|\,\text{TRUE}\,\|_2 &= \sqrt{(-1)^2} = 1, \\
\|\,\text{NOT}\,\|_2 &= \sqrt{(-1)^2} = 1, \\
\|\,\text{AND}\,\|_2 &= \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2} = 1, \\
\|\,\text{OR}\,\|_2 &= \sqrt{\left(-\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 1 \text{ and} \\
\|\,\text{XOR}\,\|_2 &= \sqrt{1^2} = 1
\end{aligned}$$

Of course, if $n \geq 1$ is fixed, not every $f\colon \{-1,1\}^n \to \mathbb{R}$ lying on the unit sphere is a Boolean function since the unit sphere contains infinitely many points.

Example 1.2.11. Fix $x \in \{-1,1\}^n$. Since $\widehat{\delta_x}(S) = \langle \chi_S, \delta_x \rangle = \mathbf{E}_{\mathbf{y}\sim\{-1,1\}^n}[\mathbf{y}^S \delta_x(\mathbf{y})] = \frac{x^S}{2^n}$, the Fourier expansion of $2^n \delta_x$ is

$$2^n \delta_x = \sum_{S \subseteq [n]} x^S \chi_S.$$

Definition 1.2.12. Given two Boolean functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$, we define their *(relative Hamming) distance* to be

$$\mathrm{dist}(f,g) := \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})]$$

the fraction of inputs on which they disagree.

Proposition 1.2.13. *The relative Hamming distance is a metric on the set of boolean functions on n bits.*

Proof. Let $f, g, h\colon \{-1,1\}^n \to \{-1,1\}$. Then it is clear that $\mathrm{dist}(f,g) \geq 0$, $\mathrm{dist}(f,g) = 0 \iff f = g$ and $\mathrm{dist}(f,g) = \mathrm{dist}(g,f)$. Finally, $\mathrm{dist}(f,g) \leq \mathrm{dist}(f,h) + \mathrm{dist}(h,g)$ since

$$\begin{aligned}
\mathrm{dist}(f,g) = \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})] &\leq \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq h(\mathbf{x}) \text{ or } h(\mathbf{x}) \neq g(\mathbf{x})] \\
&\leq \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq h(\mathbf{x})] + \Pr_{\mathbf{x}}[h(\mathbf{x}) \neq g(\mathbf{x})] = \mathrm{dist}(f,h) + \mathrm{dist}(h,g)
\end{aligned}$$

□

The relative Hamming distance gives a nice interpretation of the scalar product between two Boolean functions, namely as a measure of how similar they are.

Proposition 1.2.14. *If $f, g\colon \{-1,1\}^n \to \{-1,1\}$,*

$$\langle f, g \rangle = \Pr_{\mathbf{x}}[f(\mathbf{x}) = g(\mathbf{x})] - \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})] = 1 - 2\,\mathrm{dist}(f,g).$$

Proof.

$$\begin{aligned}
\langle f, g \rangle &= \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x})g(\mathbf{x})] \\
&= \Pr_{\mathbf{x}}[f(\mathbf{x})g(\mathbf{x}) = 1] - \Pr_{\mathbf{x}}[f(\mathbf{x})g(\mathbf{x}) = -1] \\
&= \Pr_{\mathbf{x}}[f(\mathbf{x}) = g(\mathbf{x})] - \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})] \\
&= (1 - \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})]) - \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})] = 1 - 2\,\mathrm{dist}(f,g)
\end{aligned}$$

□

The *mean* of $f\colon \{-1,1\}^n \to \mathbb{R}$ is $\mathbf{E}[f]$. When f has mean 0, we say that it is *unbiased*, or *balanced*. In the particular case where $f\colon \{-1,1\}^n \to \{-1,1\}$ is a Boolean function, its mean is $\mathbf{E}[f] = \mathbf{Pr}[f=1] - \mathbf{Pr}[f=-1]$ and thus f is unbiased if and only if it attains each truth value at exactly half of the points of $\{-1,1\}^n$. The next proposition shows that a real-valued Boolean function f is unbiased if and only if its empty-set Fourier coefficient is 0.

Proposition 1.2.15. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$. Then $\mathbf{E}[f] = \widehat{f}(\emptyset)$.*

Proof. $\mathbf{E}[f] = \mathbf{E}[1f] = \langle 1, f\rangle = \langle \chi_\emptyset, f\rangle = \widehat{f}(\emptyset)$ □

The Fourier coefficient $\widehat{f}(\emptyset)$ thus yields already an important global information on f. This is an instance of the following more general idea behind the Fourier basis: Given $f\colon \{-1,1\}^n \to \{-1,1\}$, each Fourier coefficient of f gives information on the *global behavior* of f. The coefficients $\widehat{f}(S)$ for small sets S give the rough global behavior, and the coefficients $\widehat{f}(S)$ for big S are responsible for the global fine tuning. The overall hope is that the fine tuning is not so important for many Boolean functions appearing in practice and one can therefore get a good idea of the global behavior of f by just studying $\widehat{f}(S)$ for small sets S. In contrast to this, the coefficients with respect to the basis $(2^{\frac{n}{2}}\delta_x)_{x\in\{-1,1\}^n}$ of needle functions give only information about the *local behavior* of f, namely about the values of f at individual points. The Fourier transform $\mathscr{F}$ converts all the local information to global information since its scaled version $\mathscr{F}_{\text{scaled}}$ performs the base change from the basis of needle functions to the Fourier basis.

Next we obtain formulas for the *variance* of a real-valued Boolean function (with the same conventions as for the expectation and the probability).

Proposition 1.2.16. (a) *The* variance *of $f\colon \{-1,1\}^n \to \mathbb{R}$ is*

$$\mathbf{Var}[f] = \mathbf{E}[(f-\mathbf{E}[f])^2] = \mathbf{E}[f^2] - \mathbf{E}[f]^2 = \|f-\mathbf{E}[f]\|_2^2 \overset{\text{Parseval}}{=} \sum_{\emptyset\neq S\subseteq[n]} \widehat{f}(S)^2.$$

(b) *For $f\colon \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbf{Var}[f] = 1 - \mathbf{E}[f]^2 = 4\,\mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x})=1]\,\mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x})=-1] \in [0,1]$$

and this is 0 *if and only if f is constant, and* 1 *if and only if f is unbiased.*

Proof. (a) $\mathbf{Var}[f] = \mathbf{E}[(f-\mathbf{E}[f])^2]$ equals on the one hand

$$\mathbf{E}[f^2 - 2f\,\mathbf{E}[f] + \mathbf{E}[f]^2] = \mathbf{E}[f^2] - 2\,\mathbf{E}[f]^2 + \mathbf{E}[f]^2 = \mathbf{E}[f^2] - \mathbf{E}[f^2]$$

and on the other $\|f-\mathbf{E}[f]\|_2^2$. From Proposition 1.2.15 and Proposition 1.2.8, we get

$$f - \mathbf{E}[f] = \sum_{\emptyset\neq S\subseteq[n]} \widehat{f}(S)\chi_S$$

and therefore by Parseval

$$\|f - \mathbf{E}[f]\|_2^2 = \sum_{\emptyset \neq S \subseteq [n]} \widehat{f}(S)^2.$$

(b)

$$\begin{aligned}
\mathbf{Var}[f] &\overset{(a)}{=} \mathbf{E}[f^2] - \mathbf{E}[f]^2 = \mathbf{E}[1] - \mathbf{E}[f]^2 = 1 - \mathbf{E}[f]^2 \\
&= (\Pr_{\mathbf{x}}[f(\mathbf{x}) = 1] + \Pr_{\mathbf{x}}[f(\mathbf{x}) = -1])^2 - (\Pr_{\mathbf{x}}[f(\mathbf{x}) = 1] - \Pr_{\mathbf{x}}[f(\mathbf{x}) = -1])^2 \\
&= 4 \Pr_{\mathbf{x}}[f(\mathbf{x}) = 1] \Pr_{\mathbf{x}}[f(\mathbf{x}) = -1] = 4a(1-a)
\end{aligned}$$

with $a := \mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x}) = 1] \in [0,1]$. The rest follows by discussing the graph of

$$[0,1] \to \mathbb{R},\ b \mapsto 4b(1-b).$$

$\square$

Proposition 1.2.17. *Let* $f\colon \{-1,1\}^n \to \{-1,1\}$*. Then*

$$2\varepsilon \le \mathbf{Var}[f] \le 4\varepsilon$$

where $\varepsilon := \min\{\mathrm{dist}(f,1), \mathrm{dist}(f,-1)\}$.

Proof. Set $a := \mathrm{dist}(f,1)$. Then $\mathrm{dist}(f,-1) = \mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x}) \neq -1] = 1 - \mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x}) \neq 1] = 1 - \mathrm{dist}(f,1) = 1 - a$ and by Proposition 1.2.16(b) $\mathbf{Var}[f] = 4(1-a)a$. If $a \le \frac{1}{2}$, then $1 - a \ge \frac{1}{2}$ and $2\varepsilon = 2a = 4\frac{1}{2}a \le 4(1-a)a = \mathbf{Var}[f] \le 4a = 4\varepsilon$. If $a \ge \frac{1}{2}$, then $1 - a \le \frac{1}{2}$ and $2\varepsilon = 2(1-a) = 4(1-a)\frac{1}{2} \le 4(1-a)a = \mathbf{Var}[f] \le 4(1-a)$. $\square$

By using Plancherel in place of Parseval, we get a generalization of Proposition 1.2.16(a) to covariance:

Proposition 1.2.18. *The* covariance *of* $f, g\colon \{-1,1\}^n \to \mathbb{R}$ *is*

$$\mathbf{Cov}[f,g] = \mathbf{E}[(f - \mathbf{E}[f])(g - \mathbf{E}[g])] = \mathbf{E}[fg] - \mathbf{E}[f]\,\mathbf{E}[g] = \sum_{\emptyset \neq S \subseteq [n]} \widehat{f}(S)\widehat{g}(S).$$

§1.3 Probability densities and convolution

Definition 1.3.1. Let $D \neq \emptyset$ be a finite set. A function $f\colon D \to \mathbb{R}_{\ge 0}$ is called a *(probability)* $\left\{\begin{matrix} \textit{mass} \\ \textit{density} \end{matrix}\right\}$ *function* on D if $\sum_{x \in D} f(x) = \left\{\begin{matrix} 1 \\ \#D \end{matrix}\right\}$. We write $\mathbf{x} \sim f$ to denote that a random element $\mathbf{x}$ from D is drawn from the associated probability distribution which is defined by $\mathbf{Pr}_{\mathbf{x} \sim f}[\mathbf{x} = y] = \left\{\begin{matrix} f(y) \\ \frac{f(y)}{\#D} \end{matrix}\right\}$ for $y \in D$.

Definition 1.3.2. The *(Fourier) weight* of $f\colon \{-1,1\}^n \to \mathbb{R}$ on the set $S \subseteq [n]$ is defined to be the squared Fourier coefficient $\widehat{f}(S)^2$.

The Fourier weights of a Boolean function sum up to 1 by Parseval's Theorem. If f is a Boolean function, then $\widehat{f}^{\,2}$ is thus a probability mass function on $\mathscr{P}([n])$.

Definition 1.3.3. Given $f\colon \{-1,1\}^n \to \{-1,1\}$, the *spectral sample* for f is the probability distribution in $\mathscr{P}([n])$ with probability mass function $\widehat{f}^{\,2}$.

The spectral samples of AND and OR are uniformly distributed on $\{\emptyset, \{1\}, \{2\}, \{1,2\}\}$ as can be seen from the table in Remark 1.1.1.

Definition 1.3.4. Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $0 \le k \le n$. The *degree k part of f* is

$$f_{=k} := \sum_{\substack{S\subseteq[n]\\ \#S=k}} \widehat{f}(S)\chi_S$$

and we call

$$\|f_{=k}\|_2^2 = \sum_{\substack{S\subseteq[n]\\ \#S=k}} \widehat{f}(S)^2$$

the *(Fourier) weight of f at degree k*. If f is a Boolean function, then

$$\|f_{=k}\|_2^2 = \Pr_{\mathbf{S}\sim\widehat{f}^{\,2}}[\#\mathbf{S} = k].$$

We will also sometimes use notation like

$$f_{\le k} := \sum_{\substack{S\subseteq[n]\\ \#S\le k}} \widehat{f}(S)\chi_S$$

and call

$$\|f_{\le k}\|_2^2 = \sum_{\substack{S\subseteq[n]\\ \#S\le k}} \widehat{f}(S)^2$$

the *weight of f in degree $\le k$*.

Remark 1.3.5. If φ is a density function on $\{-1,1\}^n$ and $g\colon \{-1,1\}^n \to \mathbb{R}$, then

$$\mathop{\mathbf{E}}_{\mathbf{y}\sim\varphi}[g(\mathbf{y})] = \langle \varphi, g\rangle = \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}[\varphi(\mathbf{x})g(\mathbf{x})].$$

Definition 1.3.6. If $\emptyset \ne A \subseteq \{-1,1\}^n$, we write φ_A for the density function associated to the uniform distribution on A, i.e., $\varphi_A(x) = \begin{cases} \frac{2^n}{\#A} & \text{if } x \in A \\ 0 & \text{if } x \in \{-1,1\}^n \setminus A \end{cases}$. We typically write $\mathbf{y} \sim A$ rather than $\mathbf{y} \sim \varphi_A$ [$\to$ 1.2.3(g)].

Example 1.3.7. By Example 1.2.11, every Fourier coefficient of $\varphi_{\{(1,\ldots,1)\}}$ is 1,

Reminder 1.3.8. A *commutative real algebra* is a real vector space $(V,+,\cdot)$ together with a binary operation $\circ$ such that

- $(V,+,\circ)$ is a commutative ring (with one) and
- $(\lambda\cdot x)\circ y=\lambda\cdot(x\circ y)$ for all $\lambda\in\mathbb{R}$ and $x,y\in V$, i.e., the scalar multiplication and the ring multiplication are compatible.

As mentioned in Proposition 1.2.8, the Fourier transform $\mathscr{F}$ is a vector space isomorphism from $\mathbb{R}^{\{-1,1\}^n}$ to $\mathbb{R}^{\mathscr{P}([n])}$. Now both $\mathbb{R}^{\{-1,1\}^n}$ and $\mathbb{R}^{\mathscr{P}([n])}$ are not just real vector spaces but even commutative real algebras with the *pointwise multiplication* we have already used in the case of $\mathbb{R}^{\{-1,1\}^n}$. We now introduce on either of $\mathbb{R}^{\{-1,1\}^n}$ and $\mathbb{R}^{\mathscr{P}([n])}$ an alternative multiplication called *convolution* and denoted by $*$ that corresponds to pointwise multiplication "on the other side of the Fourier transform".

Definition 1.3.9. (a) The *convolution* of $f,g\colon\{-1,1\}^n\to\mathbb{R}$ is defined by

$$f*g:=\mathscr{F}^{-1}((\mathscr{F}(f))(\mathscr{F}(g))).$$

(b) The *convolution* of $F,G\colon\mathscr{P}([n])\to\mathbb{R}$ is defined by

$$F*G:=\mathscr{F}((\mathscr{F}^{-1}(F))(\mathscr{F}^{-1}(G))).$$

By construction, we have $\mathscr{F}(f*g)=\mathscr{F}(f)\mathscr{F}(g)$ for all $f,g\colon\{-1,1\}^n\to\mathbb{R}$ and $\mathscr{F}^{-1}(F*G)=(\mathscr{F}^{-1}(F))(\mathscr{F}^{-1}(G))$ for all $F,G\colon\mathscr{P}([n])\to\mathbb{R}$. Therefore the convolution makes each of $\mathbb{R}^{\{-1,1\}^n}$ and $\mathbb{R}^{\mathscr{P}([n])}$ into a commutative algebra and $\mathscr{F}$ is not just a vector space isomorphism but even an algebra isomorphism (i.e., in addition a ring homomorphism) if one takes the pointwise multiplication on one of $\mathbb{R}^{\{-1,1\}^n}$ and $\mathbb{R}^{\mathscr{P}([n])}$ and the convolution on the other. As this will be extremely important, we formulate part of this observation in the following proposition:

Proposition 1.3.10. *For all $f,g,h\colon\{-1,1\}^n\to\mathbb{R}$,*

(a) $\widehat{f*g}=\widehat{f}\,\widehat{g}$

(b) $\widehat{f}*\widehat{g}=\widehat{fg}$

(c) $(f*g)*h=f*(g*h)$

(d) $(\widehat{f}*\widehat{g})*\widehat{h}=\widehat{f}*(\widehat{g}*\widehat{h})$

(e) $f*g=g*f$

(f) $\widehat{f}*\widehat{g}=\widehat{g}*\widehat{f}$.

Theorem 1.3.11. *Consider the abelian groups $\{-1,1\}^n$ with pointwise multiplication and $\mathscr{P}([n])$ with the symmetric difference [$\to$ 1.2.3(d)]. Consider the group isomorphism*

$$\iota\colon\{-1,1\}^n\to\mathscr{P}([n]),\ x\mapsto\{i\mid x_i=-1\}$$

and the natural algebra isomorphism

$$\iota^*\colon \mathbb{R}^{\mathscr{P}([n])} \to \mathbb{R}^{\{-1,1\}^n},\ F \mapsto F \circ \iota.$$

Up to this isomorphism, the scaled version of the Fourier transform $\mathscr{F}_{scaled} = 2^{\frac{n}{2}}\mathscr{F}$ is its own inverse, i.e.,

$$\iota^* \circ \mathscr{F}_{scaled} \circ \iota^* \circ \mathscr{F}_{scaled} = \mathrm{id}\,.$$

Proof. Let $f\colon \{-1,1\}^n \to \mathbb{R}$. We have to show $2^n\iota^*(\mathscr{F}(\iota^*(\widehat{f}))) = f$, i.e.,

$$2^n \left(\widehat{\widehat{f} \circ \iota} \circ \iota\right) = f.$$

Evaluate this in a fixed $x \in \{-1,1\}^n$. Then the claim becomes

$$2^n \widehat{\widehat{f} \circ \iota}(\{i \mid x_i = -1\}) = f(x).$$

We rewrite this as

$$2^n \langle \chi_{\{i|x_i=-1\}}, \widehat{f} \circ \iota \rangle = f(x).$$

This becomes

$$2^n \mathop{\mathbf{E}}_{\mathbf{y}\sim\{-1,1\}^n}[\chi_{\{i|x_i=-1\}}(\mathbf{y})\widehat{f}(\{i \mid \mathbf{y}_i = -1\})] = f(x).$$

The crucial trick is now to observe

$$\chi_{\{i|x_i=-1\}}(y) = (-1)^{\#\{i|x_i=-1, y_i=-1\}} = \chi_{\{i|y_i=-1\}}(x)$$

for all $y \in \{-1,1\}^n$ from which we get the equivalent formulation

$$2^n \mathop{\mathbf{E}}_{\mathbf{y}\sim\{-1,1\}^n}[\chi_{\{i|\mathbf{y}_i=-1\}}(x)\widehat{f}(\{i \mid \mathbf{y}_i = -1\})] = f(x).$$

This is equivalent to

$$\sum_{S\subseteq[n]} \chi_S(x)\widehat{f}(S) = f(x)$$

which is clearly fulfilled due to

$$f(x) = \left(\sum_{S\subseteq[n]} \widehat{f}(S)\chi_S\right)(x) = \sum_{S\subseteq[n]} \chi_S(x)\widehat{f}(S).$$

□

Corollary 1.3.12. *The scaled version $\iota^*_{scaled} := 2^n\iota^*\colon \mathbb{R}^{\mathscr{P}([n])} \to \mathbb{R}^{\{-1,1\}^n}$ of the natural algebra isomorphism ι^* [→ 1.3.11] remains an algebra isomorphism if one takes the convolution (instead of pointwise multiplication) on both sides.*

Proof. Let $f, g\colon \{-1,1\}^n \to \mathbb{R}$. We have to show $2^n(\iota^*)^{-1}(f * g) = (\iota^*)^{-1}(f) * (\iota^*)^{-1}(g)$. By Definition 1.3.9, this means

$$2^n(\iota^*)^{-1}(\mathscr{F}^{-1}((\mathscr{F}(f))(\mathscr{F}(g)))) = \mathscr{F}((\mathscr{F}^{-1}((\iota^*)^{-1}(f)))(\mathscr{F}^{-1}((\iota^*)^{-1}(g)))),$$

which becomes immediately

$$2^n(\mathscr{F} \circ \iota^*)^{-1}((\mathscr{F}(f))(\mathscr{F}(g))) = \mathscr{F}(((\iota^* \circ \mathscr{F})^{-1}(f))((\iota^* \circ \mathscr{F})^{-1}(g))).$$

In terms of $\mathscr{F}_{\text{scaled}} = 2^{\frac{n}{2}}\mathscr{F}$, this means

$$(\mathscr{F}_{\text{scaled}} \circ \iota^*)^{-1}((\mathscr{F}_{\text{scaled}}(f))(\mathscr{F}_{\text{scaled}}(g))) = \\ \mathscr{F}_{\text{scaled}}(((\iota^* \circ \mathscr{F}_{\text{scaled}})^{-1}(f))((\iota^* \circ \mathscr{F}_{\text{scaled}})^{-1}(g))).$$

But $\mathscr{F}_{\text{scaled}} \circ \iota^*$ and $\iota^* \circ \mathscr{F}_{\text{scaled}}$ are both self-inverse by Theorem 1.3.11 so that this reduces to

$$(\mathscr{F}_{\text{scaled}} \circ \iota^*)((\mathscr{F}_{\text{scaled}}(f))(\mathscr{F}_{\text{scaled}}(g))) = \mathscr{F}_{\text{scaled}}(((\iota^* \circ \mathscr{F}_{\text{scaled}})(f))((\iota^* \circ \mathscr{F}_{\text{scaled}})(g)))$$

which is clear. $\square$

Theorem 1.3.13. *Let $f, g\colon \{-1,1\}^n \to \mathbb{R}$. Then*

(a) $(f * g)(x) = \mathbf{E}_{\mathbf{y} \sim \{-1,1\}^n}[f(\mathbf{y})g(x\mathbf{y})] = \mathbf{E}_{\mathbf{y} \sim \{-1,1\}^n}[f(x\mathbf{y})g(\mathbf{y})]$ *for all $x \in \{-1,1\}^n$ and*

(b) $(\widehat{f} * \widehat{g})(S) = \sum_{T \subseteq [n]} \widehat{f}(T)\widehat{g}(S \triangle T) = \sum_{T \subseteq [n]} \widehat{f}(S \triangle T)\widehat{g}(T)$ *for all $S \subseteq [n]$.*

Proof. (b) follows easily from the Fourier expansion since

$$(\widehat{f} * \widehat{g})(S) = \widehat{fg}(S) = \sum_{\substack{T,U \subseteq [n] \\ T \triangle U = S}} \widehat{f}(T)\widehat{g}(U)$$

for $S \subseteq [n]$ because

$$fg = \left(\sum_{T \subseteq [n]} \widehat{f}(T)\chi_T\right)\left(\sum_{U \subseteq [n]} \widehat{g}(U)\chi_U\right) = \sum_{S \subseteq [n]} \left(\sum_{\substack{T,U \subseteq [n] \\ T \triangle U = S}} \widehat{f}(T)\widehat{g}(U)\right)\chi_S.$$

To prove (a), we start with a reformulation of (b), namely that

$$(F * G)(S) = \sum_{T \subseteq [n]} F(T)G(S \triangle T)$$

for all $F, G \in \mathbb{R}^{\mathscr{P}([n])}$ and $S \subseteq [n]$. Using the group isomorphism $\iota\colon \{-1,1\}^n \to \mathscr{P}([n])$ from Theorem 1.3.11, this means

$$(F * G)(\iota(x)) = \sum_{y \in \{-1,1\}^n} F(\iota(y))G(\iota(xy)),$$

or in other words

$$(\iota^*(F * G))(x) = \sum_{y \in \{-1,1\}^n} (\iota^*(F))(y)(\iota^*(G))(xy),$$

for all $F, G \in \mathbb{R}^{\mathscr{P}([n])}$ and $x \in \{-1,1\}^n$. Using ι^*_{scaled} from Corollary 1.3.12, this can be rewritten as

$$((\iota^*_{\text{scaled}}(F)) * (\iota^*_{\text{scaled}}(G)))(x) = \mathop{\mathbf{E}}_{\mathbf{y} \in \{-1,1\}^n} [(\iota^*_{\text{scaled}}(F))(\mathbf{y})(\iota^*_{\text{scaled}}(G))(x\mathbf{y})]$$

for all $F, G \in \mathbb{R}^{\mathscr{P}([n])}$ and $x \in \{-1,1\}^n$. □

Corollary 1.3.14. *Let φ be a probability density function on $\{-1,1\}^n$.*

(a) *If $f\colon \{-1,1\}^n \to \mathbb{R}$, then $(\varphi * f)(x) = \mathbf{E}_{\mathbf{y} \sim \varphi}[f(x\mathbf{y})]$ for all $x \in \{-1,1\}^n$.*

(b) *If ψ is also a probability density function on $\{-1,1\}^n$, then so is $\varphi * \psi$, and if $\mathbf{x} \sim \varphi$ and $\mathbf{y} \sim \psi$ are drawn independently, then $\mathbf{xy} \sim \varphi * \psi$.*

§1.4 Application: The test of Blum, Luby and Rubinfeld

Proposition 1.4.1. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a boolean function. Then the following are equivalent:*

(a) $\forall x, y \in \{-1,1\}^n : f(xy) = f(x)f(y)$

(b) $\exists S \subseteq [n] : f = \chi_S$

Proof. (b) $\Longrightarrow$ (a) If $S \subseteq [n]$ and $f = \chi_S$, then $f(x)f(y) = x^S y^S = (xy)^S = f(xy)$ for all $x, y \in \{-1,1\}^n$.

(a) $\Longrightarrow$ (b) Suppose (a) holds. For $i \in \{1, \dots, n\}$, let $e^{(i)} \in \{-1,1\}^n$ be defined by $e^{(i)}_j := 1$ if $j \neq i$ and $e^{(i)}_i := -1$. Set $S := \{i \mid f(e^{(i)}) = -1\}$. Then for all $x \in \{-1,1\}^n$,

$$f(x) = f\left(\prod_{\substack{i=1 \\ x_i=-1}}^{n} e^{(i)}\right) \overset{(a)}{=} \prod_{\substack{i=1 \\ x_i=-1}}^{n} f(e^{(i)}) = \prod_{\substack{i \in S \\ x_i=-1}} (-1) = x^S.$$

□

Remark 1.4.2. In the preceding proposition, consider f as a function $\mathbb{F}_2^n \to \mathbb{F}_2$ according to Remark 1.1.1. Then condition (a) becomes $\forall x, y \in \mathbb{F}_2^n : f(x+y) = f(x) + f(y)$. This implies $f(0) = 0$ and therefore $\forall \lambda \in \mathbb{F}_2 : f(\lambda x) = \lambda f(x)$. So condition (a) is equivalent to f being $\mathbb{F}_2$-linear. Condition (b) becomes $\exists S \subseteq [n] : \forall x \in \mathbb{F}_2^n : f(x) = \sum_{i \in S} x_i$ which means that f has a matrix representation. Indeed, Proposition 1.4.1 and its proof are known from linear algebra.

Lemma 1.4.3. Consider the function $h\colon [0,\frac{1}{4}] \to \mathbb{R},\ x \mapsto 3x - 10x^2 + 8x^3$.

(a) h is strictly monotonically increasing on $\left[0, \frac{2-\sqrt{2}}{4}\right]$ with $h(0) = 0$ and $h\left(\frac{2-\sqrt{2}}{4}\right) = \frac{1}{4}$.

(b) $h(x) \geq \frac{1}{4}$ for all $x \in \left[\frac{2-\sqrt{2}}{4}, \frac{1}{4}\right]$

Proof. (a) $h'(x) = 3 - 20x + 24x^2 = 24\left(x - \frac{5-\sqrt{7}}{12}\right)\left(x - \frac{5+\sqrt{7}}{12}\right)$ for all $x \in [0, \frac{1}{4}]$, $h'(0) = 3 > 0$, $\frac{2-\sqrt{2}}{4} < \frac{5-\sqrt{7}}{12}$

(b) $h(x) - \frac{1}{4} = 8\left(x - \frac{2-\sqrt{2}}{4}\right)(x - \frac{1}{4})\left(x - \frac{2+\sqrt{2}}{4}\right)$ for all $x \in [0, \frac{1}{4}]$

$\square$

Lemma 1.4.4. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a boolean function. Then

$$\Pr_{\substack{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n \\ \text{independent}}}[f(\mathbf{x}\mathbf{y}) = f(\mathbf{x})f(\mathbf{y})] = \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \widehat{f}(S)^3.$$

Proof.

$$\begin{aligned}
\Pr_{\substack{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n \\ \text{independent}}}[f(\mathbf{x}\mathbf{y}) = f(\mathbf{x})f(\mathbf{y})] &= \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{x})f(\mathbf{y})f(\mathbf{x}\mathbf{y}) = 1] \\
&= \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y}}\left[\frac{1}{2} + \frac{1}{2}f(\mathbf{x})f(\mathbf{y})f(\mathbf{x}\mathbf{y})\right] = \frac{1}{2} + \frac{1}{2}\mathop{\mathbf{E}}_{\mathbf{x}}[f(\mathbf{x})\mathop{\mathbf{E}}_{\mathbf{y}}[f(\mathbf{y})f(\mathbf{x}\mathbf{y})]] \\
&\overset{1.3.13}{=} \frac{1}{2} + \frac{1}{2}\mathop{\mathbf{E}}_{\mathbf{x}}[f(\mathbf{x})(f*f)(\mathbf{x})] \overset{\text{Plancherel}}{=} \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \widehat{f}(S)\widehat{f*f}(S) \\
&\overset{1.3.10}{=} \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \widehat{f}(S)^3
\end{aligned}$$

$\square$

Theorem 1.4.5 (robust version of 1.4.1)**.** *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a boolean function. Consider the following conditions for $\varepsilon \in \mathbb{R}_{\geq 0}$:*

(a_ε) $\Pr_{\substack{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n \\ \textit{independent}}}[f(\mathbf{x}\mathbf{y}) = f(\mathbf{x})f(\mathbf{y})] \geq 1 - \varepsilon$

(b_ε) $\exists S \subseteq [n] : \mathrm{dist}(f, \chi_S) \leq \varepsilon$

Then (b_ε) *implies* ($\mathrm{a}_{3\varepsilon}$)*, and conversely*

- (a_ε) *implies* (b_ε) *for all* $\varepsilon \geq 0$*, and*
- ($\mathrm{a}_{3\varepsilon-10\varepsilon^2+8\varepsilon^3}$) *implies* ($\mathrm{b}_\varepsilon$) *if* $\varepsilon < \frac{2-\sqrt{2}}{4} \approx 0.146$.

Proof. The proof of (b_ε) $\implies$ ($a_{3\varepsilon}$) is a robust version of the corresponding part of the proof of 1.4.1: Choose $S \subseteq [n]$ with $\operatorname{dist}(f, \chi_S) \leq \varepsilon$. Then

$$\begin{aligned}
\Pr_{\substack{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n \\ \text{independent}}}[f(\mathbf{xy}) = f(\mathbf{x})f(\mathbf{y})] &\geq \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{xy}) = \chi_S(\mathbf{xy}) \,\&\, f(\mathbf{x}) = \chi_S(\mathbf{x}) \,\&\, f(\mathbf{y}) = \chi_S(\mathbf{y})] \\
&= 1 - \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{xy}) \neq \chi_S(\mathbf{xy}) \text{ or } f(\mathbf{x}) \neq \chi_S(\mathbf{x}) \text{ or } f(\mathbf{y}) \neq \chi_S(\mathbf{y})] \\
&\geq 1 - \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{xy}) \neq \chi_S(\mathbf{xy})] - \Pr_{\mathbf{x}}[f(\mathbf{x}) \neq \chi_S(\mathbf{x})] - \Pr_{\mathbf{y}}[f(\mathbf{y}) \neq \chi_S(\mathbf{y})] \\
&\overset{1.3.14(b)}{=} 1 - 3\Pr_{\mathbf{x}}[f(\mathbf{x}) \neq \chi_S(\mathbf{x})] \overset{1.2.12}{=} 1 - 3\operatorname{dist}(f, \chi_S) \geq 1 - 3\varepsilon.
\end{aligned}$$

For the proof of ($a_{3\varepsilon-10\varepsilon^2+8\varepsilon^3}$) $\implies$ (b_ε), the ideas of the proof of 1.4.1 do not help. We have to use Lemma 1.4.4. Choose $T \subseteq [n]$ minimizing $\delta := \operatorname{dist}(f, \chi_T)$. For all $S \subseteq [n]$ we have $\operatorname{dist}(f, \chi_S) \geq \operatorname{dist}(f, \chi_T)$ and therefore $\widehat{f}(S) \leq \widehat{f}(T) = 1 - 2\delta$ by 1.2.14. Hence we have by Lemma 1.4.4 that

$$(*) \qquad \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{xy}) = f(\mathbf{x})f(\mathbf{y})] \leq \frac{1}{2} + \frac{1}{2}(1-2\delta)\sum_{S\subseteq[n]} \widehat{f}(S)^2 = \frac{1}{2} + \frac{1}{2}(1-2\delta) = 1 - \delta.$$

We first suppose that $\varepsilon \geq 0$ satisfies (a_ε) and we will show that $\delta \leq \varepsilon$. Now

$$1 - \varepsilon \overset{(a_\varepsilon)}{\leq} \Pr_{\substack{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n \\ \text{independent}}}[f(\mathbf{xy}) = f(\mathbf{x})f(\mathbf{y})] \overset{(*)}{\leq} 1 - \delta$$

from which $\delta \leq \varepsilon$ indeed follows.

Now suppose that $\varepsilon \in \left[0, \frac{2-\sqrt{2}}{4}\right)$ fulfills ($a_{3\varepsilon-10\varepsilon^2+8\varepsilon^3}$). We have to show again $\delta \leq \varepsilon$. Using the function h from Lemma 1.4.3, we have then

$$1 - h(\varepsilon) \leq \Pr_{\mathbf{x},\mathbf{y}}[f(\mathbf{xy}) = f(\mathbf{x})f(\mathbf{y})] \overset{(*)}{\leq} 1 - \delta,$$

i.e.,

$$\delta \leq h(\varepsilon) \underset{1.4.3(a)}{\overset{\varepsilon<\frac{2-\sqrt{2}}{4}}{<}} h\left(\frac{2-\sqrt{2}}{4}\right) \overset{1.4.3(a)}{=} \frac{1}{4}.$$

For $S \subseteq [n]$ with $S \neq T$, we have $\langle \chi_S, \chi_T \rangle = 0$ and therefore $\operatorname{dist}(\chi_S, \chi_T) = \frac{1}{2}$ from which it follows that

$$1 - 2\operatorname{dist}(f, \chi_S) = 2(\operatorname{dist}(\chi_S, \chi_T) - \operatorname{dist}(f, \chi_S)) \overset{1.2.13}{\leq} 2\operatorname{dist}(f, \chi_T) = 2\delta.$$

and therefore $\widehat{f}(S) \leq 2\delta$. We now get

$$\begin{aligned}
1 - h(\varepsilon) &\leq \mathop{\mathbf{Pr}}_{\substack{\mathbf{x},\mathbf{y}\sim\{-1,1\}^n \\ \text{independent}}} [f(\mathbf{xy}) = f(\mathbf{x})f(\mathbf{y})] \\
&\overset{1.4.4}{=} \frac{1}{2} + \frac{1}{2} \sum_{S\subseteq[n]} \widehat{f}(S)^3 \\
&\leq \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + \frac{1}{2} \sum_{\substack{S\subseteq[n] \\ S\neq T}} 2\delta\widehat{f}(S)^2 \\
&= \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + \delta \sum_{\substack{S\subseteq[n] \\ S\neq T}} \widehat{f}(S)^2 \\
&= \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + \delta(1 - \widehat{f}(T)^2) \\
&= \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + \delta(1 - (1-2\delta)^2) \\
&= \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + 4\delta^2 - 4\delta^3 \\
&= 1 - 3\delta + 10\delta^2 - 8\delta^3 \\
&= 1 - h(\delta),
\end{aligned}$$

i.e., $h(\delta) \leq h(\varepsilon) < \frac{1}{4}$ and so by Lemma 1.4.3(b) we have $\delta < \frac{2-\sqrt{2}}{4}$ and therefore by Lemma 1.4.3(a) $\delta \leq \varepsilon$. ☐

Remark 1.4.6. (a) Proposition 1.4.1 is the special case of Theorem 1.4.5 where $\varepsilon = 0$.

(b) In Theorem 1.4.5, for small $\varepsilon \geq 0$, the condition $(\mathrm{a}_{3\varepsilon-10\varepsilon^2+8\varepsilon^3})$ is just a little stronger than $(\mathrm{a}_{3\varepsilon})$, i.e., the shown implication $(\mathrm{a}_{3\varepsilon-10\varepsilon^2+8\varepsilon^3}) \implies (\mathrm{b}_\varepsilon)$ is just a little weaker than $(\mathrm{a}_{3\varepsilon}) \implies (\mathrm{b}_\varepsilon)$. For small ε, Theorem 1.4.4 proves thus almost equivalence of $(\mathrm{a}_{3\varepsilon})$ and (b_ε).

§2 Basic concepts and social choice

§2.1 Social choice functions

A Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ can be thought of as a *voting rule* or a *social choice function* for an election with 2 candidates and n voters. The most familiar voting rule is the majority function:

Definition 2.1.1. Let $n \in \mathbb{N}_0$.

(a) For odd n, the *majority* function $\mathrm{Maj}_n\colon \{-1,1\}^n \to \{-1,1\}$ is defined by $\mathrm{Maj}_n(x) := \mathrm{sgn}(x_1 + \cdots + x_n)$ for all $x \in \{-1,1\}^n$.

(b) A function $f\colon \{-1,1\}^n \to \{-1,1\}$ is called a *weighted majority* or *(linear) threshold function* if there are $a_0, \ldots, a_n \in \mathbb{R}$ such that $f(x) = \mathrm{sgn}(a_0 + a_1x_1 + \ldots + a_nx_n)$ for all $x \in \{-1,1\}^n$.

(c) $\mathrm{AND}_n, \mathrm{OR}_n\colon \{-1,1\}^n \to \{-1,1\}$ are defined by

$$\mathrm{AND}_n(x) = -1 :\iff x_1 = \ldots = x_n = -1 \qquad \text{and}$$
$$\mathrm{OR}_n(x) = 1 :\iff x_1 = \ldots = x_n = 1.$$

for all $x \in \{-1,1\}^n$ [$\to$ 1.1.1].

(d) For $i \in [n]$, $\chi_i := \chi_{\{i\}}$ is called the i-th *dictator* function.

(e) For $k \in \mathbb{N}_0$, a function $f\colon \{-1,1\}^n \to \{-1,1\}$ is called a *k-junta* if it depends on at most k of its input coordinates, i.e., there is $\ell \in \{0, \ldots, k\}$, pairwise different $i_1, \ldots, i_\ell \in [n]$ and $g\colon \{-1,1\}^\ell \to \{-1,1\}$ such that $f(x) = g(x_{i_1}, \ldots, x_{i_\ell})$ for all $x \in \{-1,1\}^\ell$.

(f) For odd n and for $d \in \mathbb{N}_0$, the *depth-d recursive majority of n* function, denoted $\mathrm{Maj}_n^{\otimes d}$, is the Boolean function of n^d bits defined inductively by $\mathrm{Maj}_n^{\otimes 0}(x) := x$ and

$$\mathrm{Maj}_n^{\otimes d+1}(x^{(1)}, \ldots, x^{(n)}) := \mathrm{Maj}_n(\mathrm{Maj}_n^{\otimes d}(x^{(1)}), \ldots, \mathrm{Maj}_n^{\otimes d}(x^{(n)}))$$

for all $x^{(1)}, \ldots, x^{(n)} \in \{-1,1\}^{n^d}$. In particular, $\mathrm{Maj}_n^{\otimes 1} = \mathrm{Maj}_n$ for all odd n.

(g) For $w, s \in \mathbb{N}_0$, the *tribes* function of width w and size s, $\mathrm{Tribes}_{w,s}\colon \{-1,1\}^{ws} \to \{-1,1\}$, is defined by

$$\mathrm{Tribes}_{w,s}(x^{(1)}, \ldots, x^{(s)}) := \mathrm{OR}_s(\mathrm{AND}_w(x^{(1)}), \ldots, \mathrm{AND}_w(x^{(s)}))$$

for all $x^{(1)}, \ldots, x^{(n)} \in \{-1,1\}^w$.

While the tribes function seems implausible, it appears in practice: Consider s nuclear states (tribes) each having w military commanders. A nuclear war is started if and only if in at least one of the nuclear states the military commanders are unanimously in favor of throwing a nuclear bomb. Here are some natural properties of 2-candidate social choice functions which may be considered desirable:

Definition 2.1.2. We say that a function $f\colon \{-1,1\}^n \to \mathbb{R}$ is

- *monotone* if $f(x) \leq f(y)$ whenever $x, y \in \{-1,1\}^n$ with $x_i \leq y_i$ for all $i \in \{-1,1\}^n$,
- *odd* if $f(x) = -f(-x)$ for all $x \in \{-1,1\}^n$,
- *unanimous* if $f(1,\ldots,1) = 1$ and $f(-1,\ldots,-1) = -1$,
- *symmetric* if $f(x_{\sigma(1)},\ldots,x_{\sigma(n)}) = f(x)$ for all $x \in \{-1,1\}^n$ and all $\sigma \in S_n$,
- *transitive-symmetric* if for all $i, j \in [n]$ there is some $\sigma \in S_n$ such that $\sigma(i) = j$ and $f(x_{\sigma(1)},\ldots,x_{\sigma(n)}) = f(x)$ for all $x \in \{-1,1\}^n$.

Example 2.1.3. Let $n \in \mathbb{N}_0$. Consider the properties from Definition 2.1.2.

(a) For odd n, Maj_n has all properties and it is the the only monotone odd symmetric Boolean function on n bits.

(b) Maj_n (for odd n), AND_n, OR_n and χ_i (for $i \in [n]$) are Boolean linear threshold functions.

(c) AND_n and OR_n satisfy all properties except oddness for $n \neq 1$ and unanimity for $n = 0$.

(d) The dictator functions satisfy the first three properties but for $n \geq 2$ they do not satisfy the last two.

(e) There are exactly $2n + 2$ 1-juntas on n bits, namely the n dictators, the n negated dictators and the two constant Boolean functions.

(f) For $d \in \mathbb{N}_0$ and for odd n, $\mathrm{Maj}_n^{\otimes d}$ satisfies all properties except, if $n \geq 3$ and $d \geq 2$, symmetry.

(g) For $w, s \in \mathbb{N}_{\geq 2}$, $\mathrm{Tribes}_{w,s}$ is monotone, not odd, unanimous, not symmetric but transitive-symmetric.

§2.2 Influences and derivatives

Definition 2.2.1. For $x \in \{-1,1\}^n$ and $i \in [n]$, we set

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

We say that the coordinate $i \in [n]$ is *pivotal* for $f\colon \{-1,1\}^n \to \{-1,1\}$ on input x if $f(x) \neq f(x^{\oplus i})$. The *influence* of coordinate $i \in [n]$ on $f\colon \{-1,1\}^n \to \{-1,1\}$ is defined to be the probability that i is pivotal on a random input:

$$\mathbf{Inf}_i[f] := \Pr_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})].$$

Example 2.2.2. Let $n \in \mathbb{N}$. Then $\mathbf{Inf}_i[\mathrm{OR}_n] = \mathbf{Inf}_i[\mathrm{AND}_n] = 2^{1-n}$. For odd n,

$$\mathbf{Inf}_i[\mathrm{Maj}_n] = \binom{n-1}{\frac{n-1}{2}} 2^{1-n}.$$

Definition 2.2.3. For $x \in \{-1,1\}^n$, $i \in [n]$ and $b \in \{-1,1\}$, we set

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

For $i \in [n]$, we introduce the *i-th (discrete) derivative operator*

$$D_i\colon \mathbb{R}^{\{-1,1\}^n} \to \mathbb{R}^{\{-1,1\}^n},\ f \mapsto \left(x \mapsto \frac{f(x^{(i\mapsto 1)}) - f(x^{(i\mapsto -1)})}{2} \right).$$

The derivative operators just introduced are vector space endomorphisms. If

$$f\colon \{-1,1\}^n \to \{-1,1\}$$

is a Boolean function, then $x \mapsto D_i f(x)^2$ is the 0-1-indicator for whether i is pivotal for f on $x \in \{-1,1\}^n$ and we conclude that $\mathbf{Inf}_i[f] = \mathbf{E}_{\mathbf{x}\sim\{-1,1\}^n}[D_i f(\mathbf{x})^2]$. We take this formula as a *definition* for the influences of real valued Boolean functions.

Definition 2.2.4. We generalize Definition 2.2.1 to functions $f\colon \{-1,1\}^n \to \mathbb{R}$ by defining the *influence* of coordinate $i \in [n]$ on f to be

$$\mathbf{Inf}_i[f] := \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}[D_i f(\mathbf{x})^2] = \|D_i f\|_2^2.$$

Definition 2.2.5. We say that coordinate $i \in [n]$ is *relevant* for $f\colon \{-1,1\}^n \to \mathbb{R}$ if $\mathbf{Inf}_i[f] > 0$, i.e., $f(x^{(i\mapsto 1)}) \neq f(x^{(i\mapsto -1)})$ for at least one $x \in \{-1,1\}^n$.

The discrete derivative operators are quite analogous to the usual partial derivatives.

Remark 2.2.6. Let $i \in [n]$ and $f\colon \{-1,1\}^n \to \mathbb{R}$.

(a) $D_i f = \sum_{\substack{S\subseteq[n]\\ i\in S}} \widehat{f}(S)\chi_{S\setminus\{i\}}$

(b) For $i \in [n]$, $\mathbf{Inf}_i[f] = \sum_{\substack{S\subseteq[n]\\ i\in S}} \widehat{f}(S)^2$.

Proposition 2.2.7. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be monotone and $i \in [n]$. Then*

$$\mathbf{Inf}_i[f] = \widehat{f}(\{i\}).$$

Proof.

$$\begin{aligned}\mathbf{Inf}_i[f] &\overset{2.2.1}{=} \Pr_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x}) \neq f(\mathbf{x}^{\oplus i})] = \Pr_{\mathbf{x}\sim\{-1,1\}^n}[f(\mathbf{x}^{(i\mapsto 1)}) \neq f(\mathbf{x}^{(i\mapsto -1)})] \\ &\overset{f \text{ monotone}}{=} \mathop{\mathbf{E}}_{\mathbf{x}\sim\{-1,1\}^n}\left[\frac{f(\mathbf{x}^{(i\mapsto 1)}) - f(\mathbf{x}^{(i\mapsto -1)})}{2}\right] \overset{2.2.3}{=} \mathbf{E}[D_i f] = \widehat{D_i f}(\emptyset) \overset{2.2.6(a)}{=} \widehat{f}(\{i\})\end{aligned}$$

□

Proposition 2.2.8. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be transitive-symmetric and monotone and $i \in [n]$. Then*

$$\mathbf{Inf}_i[f] \leq \frac{1}{\sqrt{n}}.$$

Proof. $1 = \sum_{S\subseteq[n]} \widehat{f}(S)^2 \geq \sum_{i=1}^n \widehat{f}(\{i\})^2 \overset{2.2.7}{=} \sum_{i=1}^n \mathbf{Inf}_i[f]^2 \overset{f \text{ transitive-symmetric}}{=} n\,\mathbf{Inf}_i[f]^2.$ □

Definition 2.2.9. Let $i \in [n]$. The *i-th* $\begin{Bmatrix}\text{expectation}\\ \text{Laplacian}\end{Bmatrix}$ operator is the vector space endomorphism $\begin{Bmatrix}E_i\\ L_i\end{Bmatrix}$ of $\mathbb{R}^{\{-1,1\}^n}$ defined by

$$E_i f(x) := \mathop{\mathbf{E}}_{\mathbf{y}\sim\{-1,1\}}[f(x_1,\dots,x_{i-1},\mathbf{y},x_{i+1},\dots,x_n)]$$

for all $f\colon \{-1,1\}^n \to \mathbb{R}$ and $x \in \{-1,1\}^n$ and

$$L_i f := f - E_i f$$

for all $f\colon \{-1,1\}^n \to \mathbb{R}$.

Remark 2.2.10. Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $x \in \{-1,1\}^n$.

(a) $E_i f(x) = \frac{f(x)+f(x^{\oplus i})}{2}$

(b) $f(x) = E_i f(x) + x_i D_i f(x) = E_i f(x) + L_i f(x)$

(c) $E_i f(x) = \sum_{\substack{S\subseteq[n]\\ i\notin S}} \widehat{f}(S) x^S$

(d) $L_i f(x) = \sum_{\substack{S\subseteq[n]\\ i\in S}} \widehat{f}(S) x^S$

(e) $\langle f, L_i f\rangle = \langle L_i f, L_i f\rangle = \mathbf{Inf}_i[f]$

Definition 2.2.11. [→ 2.2.4] The *total influence* of $f\colon \{-1,1\}^n \to \mathbb{R}$ is defined to be

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

Definition 2.2.12. Let $f\colon \{-1,1\}^n \to \{-1,1\}$. The *sensitivity* $\mathrm{sens}_f(x)$ of f at x is defined to be the number of pivotal coordinates for f on input x [$\to$ 2.2.1].

Remark 2.2.13. For $f\colon \{-1,1\}^n \to \{-1,1\}$, $\mathbf{I}[f] = \mathbf{E}_{\mathbf{x}}[\mathrm{sens}_f(\mathbf{x})]$. [$\to$ 2.2.1]

Theorem 2.2.14. *Fix $n \in \mathbb{N}_0$. For $f\colon \{-1,1\}^n \to \{-1,1\}$,*

$$\mathop{\mathbf{E}}_{\mathbf{x}}[\#\{i \in [n] \mid \mathbf{x}_i = f(\mathbf{x})\}] = \frac{n}{2} + \frac{1}{2}\sum_{i=1}^{n} \widehat{f}(\{i\})$$

and, for odd n, this is maximized if and only if $f = \mathrm{Maj}_n$.

Proof. Let $f\colon \{-1,1\}^n \to \{-1,1\}$. The left hand side equals

$$\sum_{i=1}^{n} \frac{1 + \mathbf{E}_{\mathbf{x}}[f(\mathbf{x})\mathbf{x}_i]}{2} = \frac{n}{2} + \frac{1}{2}\sum_{i=1}^{n} \langle f, \chi_{\{i\}} \rangle$$

which equals the right hand side. Moreover,

$$\frac{1}{2}\sum_{i=1}^{n} \widehat{f}(\{i\}) = \mathop{\mathbf{E}}_{\mathbf{x}}[f(\mathbf{x})(\mathbf{x}_1 + \cdots + \mathbf{x}_n)] \le \mathop{\mathbf{E}}_{\mathbf{x}}[|\mathbf{x}_1 + \cdots + \mathbf{x}_n|]$$

with equality if and only if $f(x) = \mathrm{sgn}(x_1 + \cdots + x_n)$ for all $x \in \{-1,1\}^n$ with $x_1 + \cdots + x_n \neq 0$. If n is odd, then $x_1 + \cdots + x_n \neq 0$ for all $x \in \{-1,1\}^n$. $\square$

Corollary 2.2.15. [$\to$ 2.2.7] *Let $n \in \mathbb{N}$ be odd. Among all monotone $f\colon \{-1,1\}^n \to \{-1,1\}$, Maj_n is the only one with maximal total influence.*

Definition 2.2.16. [$\to$ 2.2.3, 2.2.9] We introduce the *(discrete) gradient* operator

$$\nabla\colon \mathbb{R}^{\{-1,1\}^n} \to (\mathbb{R}^n)^{\{-1,1\}^n},\ f \mapsto \left(x \mapsto \begin{pmatrix} D_1 f(x) \\ \vdots \\ D_n f(x) \end{pmatrix} \right)$$

and the *Laplacian operator* $L := \sum_{i=1}^{n} L_i\colon \mathbb{R}^{\{-1,1\}^n} \to \mathbb{R}^{\{-1,1\}^n}$.

These are of course linear maps.

Remark 2.2.17. [$\to$ 2.2.10] Let $f\colon \{-1,1\}^n \to \mathbb{R}$.

(a) $Lf = \sum_{S \subseteq [n]} (\#S)\widehat{f}(S)\chi_S$

(b) $\langle f, Lf \rangle = \mathbf{I}[f]$

(c) $\mathbf{I}[f] = \sum_{S \subseteq [n]} (\#S)\widehat{f}(S)^2 = \sum_{k=0}^{n} k\|f_{=k}\|_2^2$ [$\to$ 1.3.4]

Remark 2.2.18. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ and $x \in \{-1,1\}^n$.

(a) $\|\nabla f(x)\|^2 = \mathrm{sens}_f(x)$

(b) $Lf(x) = f(x)\,\mathrm{sens}_f(x)$

(c) $\mathbf{I}[f] = \mathbf{E}_{\mathbf{S}\sim\widehat{f}^{\,2}}[\#S]$ [→ 1.3.3]

Proposition 2.2.19 (Poincaré inequality)**.** *Let* $f\colon \{-1,1\}^n \to \mathbb{R}$*. Then* $\mathbf{Var}[f] \leq \mathbf{I}[f]$*.*

Proof. $\mathbf{Var}[f] \overset{1.2.16(a)}{=} \sum_{\substack{S\subseteq[n]\\ S\neq\emptyset}} \widehat{f}(S)^2 \leq \sum_{S\subseteq[n]} (\#S)\widehat{f}(S)^2 \overset{2.2.17(c)}{=} \mathbf{I}[f]$ ☐

§2.3 Noise stability

Reminder 2.3.1. Consider random bits $\mathbf{x} \sim \{-1,1\}$ and $\mathbf{y} \sim \{-1,1\}$ both drawn uniformly from $\{-1,1\}$, but this time not necessarily independently (e.g., one could draw $\mathbf{x} \sim \{-1,1\}$ and then set $\mathbf{y} := \mathbf{x}$). The joint probability distribution of $\mathbf{x}$ and $\mathbf{y}$ (i.e., the distribution of $(\mathbf{x},\mathbf{y})$) is given by $\lambda := \mathbf{Pr}[\mathbf{x} = -1 = \mathbf{y}] \in [0, \frac{1}{2}]$ since

$$\lambda + \mathbf{Pr}[\mathbf{x} = -1 \,\&\, \mathbf{y} = 1] = \mathbf{Pr}[\mathbf{x} = -1] = \frac{1}{2}$$

and analogously $\lambda + \mathbf{Pr}[\mathbf{x} = 1 \,\&\, \mathbf{y} = -1] = \frac{1}{2}$. Actually, one sees immediately that each $\lambda := \mathbf{Pr}[\mathbf{x} = -1 = \mathbf{y}] \in [0, \frac{1}{2}]$ is realized by a unique probability distribution on $\{-1,1\}^2$. We have $\mathbf{E}[\mathbf{x}] = 0 = \mathbf{E}[\mathbf{y}]$ and therefore

$$\mathbf{Cov}[\mathbf{x},\mathbf{y}] = \mathbf{E}\left[(\mathbf{x} - \mathbf{E}[\mathbf{x}])\,(\mathbf{y} - \mathbf{E}[\mathbf{y}])\right] = \mathbf{E}[\mathbf{x}\mathbf{y}],$$

$\mathbf{Var}[\mathbf{x}] = \mathbf{E}[(\mathbf{x} - \mathbf{E}[\mathbf{x}])^2] = 1$ and similarly $\mathbf{Var}[\mathbf{y}] = 1$. Therefore the *correlation* of $\mathbf{x}$ and $\mathbf{y}$ is

$$\begin{aligned}\mathbf{Corr}[\mathbf{x},\mathbf{y}] &= \frac{\mathbf{Cov}[\mathbf{x},\mathbf{y}]}{\sqrt{\mathbf{Var}[\mathbf{x}]\,\mathbf{Var}[\mathbf{y}]}} = \mathbf{Cov}[\mathbf{x},\mathbf{y}] = \mathbf{E}[\mathbf{x}\mathbf{y}] \\ &= \mathbf{Pr}[f(\mathbf{x}) = f(\mathbf{y})] - \mathbf{Pr}[f(\mathbf{x}) \neq f(\mathbf{y})] = 2\lambda - (1-2\lambda) = 4\lambda - 1 \in [-1,1].\end{aligned}$$

The joint probability distribution of $\mathbf{x}$ and $\mathbf{y}$ is thus uniquely determined by the correlation $\mathbf{Corr}[x,y]$ of $\mathbf{x}$ and $\mathbf{y}$ which can take arbitrary values between -1 and 1.

Definition 2.3.2. Fix $\varrho \in [-1,1]$. If two random $\mathbf{x},\mathbf{y} \in \{-1,1\}$ are chosen in such a way that the n pairs $(\mathbf{x}_1,\mathbf{y}_1),\ldots,(\mathbf{x}_n,\mathbf{y}_n)$ are independant, both $\mathbf{x}$ and $\mathbf{y}$ are distributed uniformly on $\{-1,1\}^n$ and for each $i \in [n]$, the joint distribution of $\mathbf{x}_i$ and $\mathbf{y}_i$ is the one with correlation ϱ [→ 2.3.1], then we call $\mathbf{x}$ and $\mathbf{y}$ *ϱ-correlated*.

Definition 2.3.3. Let $x \in \{-1,1\}^n$. For all $\varrho \in [0,1]$, we write $\mathbf{y} \sim N_\varrho(x)$ to denote that $\mathbf{y} \in \{-1,1\}^n$ is chosen at random as follows: For $i \in [n]$ independently,

$$\mathbf{y}_i := \begin{cases} x_i & \text{with probability } \varrho \\ \text{uniformly random} & \text{with probability } 1-\varrho. \end{cases}$$

For all $\varrho \in [-1,0]$, we set $N_\varrho(x) := N_{-\varrho}(-x)$. If $\varrho \in [-1,1]$ such that $\mathbf{y} \sim N_\varrho(x)$, we say that $\mathbf{y}$ is a *ϱ-reliable version of* $\mathbf{x}$. Thus, if $\varrho \in [-1,1]$, we get a $\mathbf{y} \sim N_\varrho(x)$ by setting

$$\mathbf{y}_i := \begin{cases} x_i & \text{with probability } \frac{1}{2} + \frac{1}{2}\varrho \\ -x_i & \text{with probability } \frac{1}{2} - \frac{1}{2}\varrho \end{cases}$$

for each $i \in [n]$ independently.

Lemma 2.3.4. Let $\varrho \in [-1,1]$. Suppose that $\mathbf{x} \sim \{-1,1\}^n$ is drawn uniformly at random and then $\mathbf{y} \sim N_\varrho(\mathbf{x})$ is chosen as a random ϱ-reliable version of $\mathbf{x}$. Then $\mathbf{x}$ and $\mathbf{y}$ are ϱ-*correlated.*

Proof. Fix $i \in [n]$. We have to show $\mathbf{E}[\mathbf{x}_i] = 0 = \mathbf{E}[\mathbf{y}_i]$ and $\mathbf{E}[\mathbf{x}_i\mathbf{y}_i] = \varrho$. It is trivial that $\mathbf{E}[\mathbf{x}_i] = 0$. From Definition 2.3.3, we get $\mathbf{E}[y_i] = (\frac{1}{2} + \frac{1}{2}\varrho)\,\mathbf{E}[\mathbf{x}_i] + (\frac{1}{2} - \frac{1}{2}\varrho)\,\mathbf{E}[-\mathbf{x}_i] = 0 + 0 = 0$ and $\mathbf{E}[\mathbf{x}_i\mathbf{y}_i] = (\frac{1}{2} + \frac{1}{2}\varrho)\,\mathbf{E}[\mathbf{x}_i^2] + (\frac{1}{2} - \frac{1}{2}\varrho)\,\mathbf{E}[-\mathbf{x}_i^2]) = (\frac{1}{2} + \frac{1}{2}\varrho) + (\frac{1}{2} - \frac{1}{2}\varrho)(-1) = \varrho$. □

Definition 2.3.5. For $f\colon \{-1,1\}^n \to \mathbb{R}$ and $\varrho \in [-1,1]$, the *noise stability of f at ϱ* is

$$\mathbf{Stab}_\varrho[f] := \mathop{\mathbf{E}}_{\substack{\mathbf{x},\mathbf{y} \\ \varrho\text{-correlated}}} [f(\mathbf{x})f(\mathbf{y})].$$

If $f\colon \{-1,1\}^n \to \{-1,1\}$, we have

$$\mathbf{Stab}_\varrho[f] = \mathop{\mathbf{Pr}}_{\substack{\mathbf{x},\mathbf{y} \\ \varrho\text{-corr.}}} [f(\mathbf{x}) = f(\mathbf{y})] - \mathop{\mathbf{Pr}}_{\substack{\mathbf{x},\mathbf{y} \\ \varrho\text{-corr.}}} [f(\mathbf{x}) \neq f(\mathbf{y})] = 2 \mathop{\mathbf{Pr}}_{\substack{\mathbf{x},\mathbf{y} \\ \varrho\text{-corr.}}} [f(\mathbf{x}) = f(\mathbf{y})] - 1.$$

Definition 2.3.6. For $f\colon \{-1,1\}^n \to \{-1,1\}$ and $\delta \in [0,1]$, we write $\mathbf{NS}_\delta[f]$ for *noise sensitivity of f at δ*, defined to be the probability that $f(\mathbf{x}) \neq f(\mathbf{y})$ when $\mathbf{x} \sim \{-1,1\}^n$ is uniformly random and $\mathbf{y}$ is formed from $\mathbf{x}$ by reversing each bit independently with probability δ (so that $\mathbf{E}[\mathbf{x}\mathbf{y}] = (1-\delta)\,\mathbf{E}[\mathbf{x}\mathbf{x}] + \delta\,\mathbf{E}[\mathbf{x}(-\mathbf{x})] = 1 - \delta - \delta = 1 - 2\delta$). In other words,

$$\mathbf{NS}_\delta[f] = \frac{1}{2} - \frac{1}{2}\mathbf{Stab}_{1-2\delta}[f].$$

Example 2.3.7. The constant functions ± 1 have noise stability 1 at each $\varrho \in [-1,1]$. The dictator functions $\chi_{\{i\}}$ satisfy $\mathbf{Stab}_\varrho[\chi_{\{i\}}] = \varrho$ for all $\varrho \in [-1,1]$ and $i \in [n]$ (equivalently $\mathbf{NS}_\delta[\chi_{\{i\}}] = \delta$ for all $\delta \in [0,1]$ and $i \in [n]$). More generally,

$$\mathbf{Stab}_\varrho[\chi_S] = \mathop{\mathbf{E}}_{\substack{\mathbf{x},\mathbf{y} \\ \varrho\text{-corr.}}} [\mathbf{x}^S\mathbf{y}^S] = \mathop{\mathbf{E}}_{\substack{\mathbf{x},\mathbf{y} \\ \varrho\text{-corr.}}} \left[\prod_{i\in S}(\mathbf{x}_i\mathbf{y}_i)\right] = \prod_{i\in S} \mathop{\mathbf{E}}_{\mathbf{x}_i,\mathbf{y}_i}[\mathbf{x}_i\mathbf{y}_i] = \prod_{i\in S}\varrho = \varrho^{\#S},$$

where we used the fact that the bit pairs $(\mathbf{x}_i, \mathbf{y}_i)$ are independent across i to convert the expectation of a product into a product of expectations.

Definition 2.3.8. Let $\varrho \in [-1,1]$. The *noise operator with parameter* ϱ is the vector space endomorphism T_ϱ of $\mathbb{R}^{\{-1,1\}^n}$ defined by

$$T_\varrho f(x) := \mathop{\mathbf{E}}_{\mathbf{y}\sim N_\varrho(x)}[f(\mathbf{y})]$$

for all $f\colon \{-1,1\}^n \to \mathbb{R}$ and $x \in \{-1,1\}^n$.

Proposition 2.3.9. *For $\varrho \in [-1,1]$ and $f\colon \{-1,1\}^n \to \mathbb{R}$,*

$$T_\varrho f = \sum_{S\subseteq[n]} \varrho^{\#S}\widehat{f}(S)\chi_S = \sum_{k=0}^{n} \varrho^k f_{=k}.$$

Proof. By linearity, it is enough to show $T_\varrho\chi_S = \varrho^{\#S}\chi_S$ for all $\varrho \in [-1,1]$ and $S \subseteq [n]$ which follows from

$$T_\varrho\chi_S(x) = \mathop{\mathbf{E}}_{\mathbf{y}\sim N_\varrho(x)}[\mathbf{y}^S] = \prod_{i\in S}\mathop{\mathbf{E}}_{\mathbf{y}\sim N_\varrho(x)}[\mathbf{y}_i] = \prod_{i\in S}(\varrho x_i) = \varrho^{\#S}\chi_S(x).$$

Here we used the fact that for $\mathbf{y} \sim N_\varrho(x)$ the bits $\mathbf{y}_i$ are independent and satisfy $\mathbf{E}[\mathbf{y}_i] = (\frac{1}{2}+\frac{1}{2}\varrho)x_i + (\frac{1}{2}-\frac{1}{2}\varrho)(-x_i) = \varrho x_i$. □

Proposition 2.3.10. *Let $\varrho \in [-1,1]$ and $f\colon \{-1,1\}^n \to \mathbb{R}$. Then*

$$\mathbf{Stab}_\varrho[f] = \langle f, T_\varrho f\rangle.$$

Proof.

$$\mathbf{Stab}_\varrho[f] \overset{2.3.5}{\underset{2.3.4}{=}} \mathop{\mathbf{E}}_{\substack{\mathbf{x}\sim\{-1,1\}^n\\ \mathbf{y}\sim N_\varrho(\mathbf{x})}}[f(\mathbf{x})f(\mathbf{y})] = \mathop{\mathbf{E}}_{\mathbf{x}}\left[f(\mathbf{x})\mathop{\mathbf{E}}_{\mathbf{y}\sim N_\varrho(x)}[f(\mathbf{y})]\right]$$

□

Corollary 2.3.11. *Let $\varrho \in [-1,1]$. For all $f\colon \{-1,1\}^n \to \mathbb{R}$, we have*

$$\mathbf{Stab}_\varrho[f] = \sum_{S\subseteq[n]} \varrho^{\#S}\widehat{f}(S)^2 = \sum_{k=0}^{n} \varrho^k \|f_{=k}\|_2^2.$$

In particular,

$$\mathbf{Stab}_\varrho[f] = \mathop{\mathbf{E}}_{\mathbf{S}\sim\widehat{f}^2}[\varrho^{\#\mathbf{S}}]$$

for all $f\colon \{-1,1\}^n \to \{-1,1\}$.

Proposition 2.3.12. *Suppose that $\varrho \in (0,1)$ and $f\colon \{-1,1\}^n \to \{-1,1\}$ is unbiased. Then*

$$\mathbf{Stab}_\varrho[f] \le \varrho,$$

with equality if and only if $f \in \{\chi_{\{i\}} \mid i \in [n]\} \cup \{-\chi_{\{i\}} \mid i \in [n]\}$.

Proof. Since f is an unbiased Boolean function, we have $\sum_{k=1}^{n} \|f_{=k}\|_2^2 = 1$ and

$$\mathbf{Stab}_\varrho[f] = \sum_{k=1}^{n} \|f_{=k}\|_2^2 \varrho^k \le \sum_{k=1}^{n} \|f_{=k}\|_2^2 \varrho = \varrho,$$

by 2.3.11 with equality if and only if $\|f_{=k}\|_2^2 = 0$ for all $k \in \{2, \dots, n\}$.[1] □

Proposition 2.3.13. *For* $f\colon \{-1,1\}^n \to \mathbb{R}$,

$$\frac{d}{d\varrho} \mathbf{Stab}_\varrho[f]\Big|_{\varrho=0} = \|f_{=1}\|_2^2,$$
$$\frac{d}{d\varrho} \mathbf{Stab}_\varrho[f]\Big|_{\varrho=1} = \mathbf{I}[f].$$

Proof. Use 2.3.11, in case of the second claim together with 2.2.17(c). □

Definition 2.3.14. [→ 2.2.4, 2.2.11] For $f\colon \{-1,1\}^n \to \mathbb{R}$, $\varrho \in [0,1]$ and $i \in [n]$, we define the *ϱ-stable influence* of coordinate i on f

$$\mathbf{Inf}_i^{(\varrho)}[f] := \mathbf{Stab}_\varrho[D_i f] \overset{2.3.10}{=} \langle D_i f, T_\varrho D_i f\rangle \underset{2.2.6(a)}{\overset{2.3.9}{=}} \sum_{\substack{S \subseteq [n] \\ i \in S}} \varrho^{\#S-1} \widehat{f}(S)^2$$

and the *ϱ-stable total influence* of f

$$\mathbf{I}^{(\varrho)}[f] := \sum_{i=1}^{n} \mathbf{Inf}_i^{(\varrho)}[f] = \sum_{k=1}^{n} k\varrho^{k-1} \|f_{=k}\|_2^2 \overset{2.3.11}{=} \frac{d}{d\varrho} \mathbf{Stab}_\varrho[f].$$

As ϱ increases from 0 to 1, $\mathbf{Inf}_i^{(\varrho)}[f]$ increases from $\mathbf{Inf}_i^{(0)}[f] = \widehat{f}(\{i\})^2$ to $\mathbf{Inf}_i^{(1)}[f] \overset{2.3.8}{=} \|D_i f\|_2^2 \overset{2.2.4}{=} \mathbf{Inf}_i[f]$ for any $f\colon \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$. Consequently, at the same time $\mathbf{I}^{(\varrho)}[f]$ increases from $\mathbf{I}^{(0)}[f] = \|f_{=1}\|_2^2$ to $\mathbf{I}^{(1)}[f] \overset{2.2.11}{=} \mathbf{I}[f]$. For $\varrho \in (0,1)$, we are not aware of an especially natural combinatorial meaning of the ϱ-stable influence. However, we will see later that the stable influences are technically very useful. One reason for this is that every function $f\colon \{-1,1\}^n \to \{-1,1\}$ has only a few "stably-influental" coodinates.

Proposition 2.3.15. *Suppose* $f\colon \{-1,1\}^n \to \mathbb{R}$ *has* $\mathbf{Var}[f] \le 1$ *(e.g., if f is Boolean). Let* $\delta, \varepsilon \in (0,1]$*. Then* $I := \{i \in [n] \mid \mathbf{Inf}_i^{(1-\delta)}[f] \ge \varepsilon\}$ *has at most* $\frac{1}{\delta\varepsilon}$ *elements.*

Proof. We have $\mathbf{I}^{(1-\delta)}[f] \ge \varepsilon \# I$. It is therefore enough to show that $\mathbf{I}^{(1-\delta)}[f] \le \frac{1}{\delta}$. We show in fact that $\delta\, \mathbf{I}^{(1-\delta)}[f] \le \mathbf{Var}[f]$. Comparing 1.2.16 with 2.3.14, it suffices to show that $k\delta(1-\delta)^{k-1} \le 1$ for all $k \in \{1, \dots, n\}$. This follows by discussing the graph of

$$[0,1] \to \mathbb{R},\ t \mapsto kt(1-t)^{k-1}.$$

□

[1] Now the reader concludes with a little argument also used as at the end of the proof of 2.4.3. In future versions of this script, one might want to formulate a suitable lemma.

It is good to think of the elements of the set I in this proposition as the "notable" coordinates for the function f. The monomial function $\chi_{[n]}$ has n coordinates with influence 1 [$\rightarrow$ 2.2.6] but has no "notable" coordinate in this sense as soon as $\mathbf{Inf}_i^{(1-\delta)}[\chi_{[n]}] = (1-\delta)^{n-1} < \varepsilon$, i.e., as soon as n is large. The intuition becomes quite clear when one discusses $\chi_{[n]}$ as a voting rule. This voting rule maximizes theoretically the influence of each voter [$\rightarrow$ 2.2.1] and yet in practice the voters would consider that they have no influence, at least if n is big, for a couple of reasons one of which is that small noise would lead to a quite random outcome of the election.

§2.4 Application: Arrow's Theorem

When there are just 2 candidates, the majority rule seems to possess about all of the mathematical properties that one ideally could expect from a voting rule [$\rightarrow$ 2.1.2, 2.1.3, 2.2.15]. Unfortunately, as soon as there are 3 or more candidates, the problem of social choice becomes much more difficult. For example, suppose we have three candidates A, B and C and each of the n voters has a ranking of them. How should we aggregate these preferences to produce a winning candidate? Condorcet proposed to conduct the three pairwise elections A versus B, B versus C and C versus A and to hope that this produces a *Condorcet winner*, i.e., a candidate winning both pairwise elections in which he participates. Suppose 1 stands for a preference for the first candidate and -1 for a preference of the second candidate in such a pairwise election. Then we encode a ranking of an individual voter by a 3-tuple of consistent preferences, i.e., by an element of the set

$$R := \{(1,1,-1),(1,-1,-1),(-1,1,-1),(-1,1,1),(1,-1,1),(-1,-1,1)\}$$

having $3! = 6$ elements. For example $(1,1,-1)$ stands for the ranking where A is first, B second and C third.

Theorem 2.4.1. *Consider a 3-candidate Condorcet election using the same voting rule $f\colon \{-1,1\}^n \rightarrow \{-1,1\}$ for each pairwise election. If each of the n voters chooses uniformly and independently one of the $3! = 6$ candidate rankings, then the probability of a Condorcet winner is precisely $\frac{3}{4} - \frac{3}{4}\,\mathbf{Stab}_{-\frac{1}{3}}[f]$.*

Proof. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \{-1,1\}^n$ be the votes for the elections A versus B, B versus C and C versus A, i.e., the $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i)$ are chosen uniformly from R and independently across i. The function

$$g\colon \{-1,1\}^3 \rightarrow \{0,1\},\ w \mapsto \frac{3}{4} - \frac{1}{4}w_1w_2 - \frac{1}{4}w_1w_3 - \frac{1}{4}w_2w_3$$

is the 0-1-indicator function for R. The probability that there is a Condorcet winner is thus

$$\mathbf{E}[g(f(\mathbf{x}), f(\mathbf{y}), f(\mathbf{z}))] = \frac{3}{4} - \frac{1}{4}\,\mathbf{E}[f(\mathbf{x})f(\mathbf{y})] - \frac{1}{4}\,\mathbf{E}[f(\mathbf{x})f(\mathbf{z})] - \frac{1}{4}\,\mathbf{E}[f(\mathbf{y})f(\mathbf{z})].$$

Since $\mathbf{E}[\mathbf{x}_i] = 0 = \mathbf{E}[\mathbf{y}_i]$ and $\mathbf{E}[\mathbf{x}_i\mathbf{y}_i] = \frac{2}{6} - \frac{4}{6} = -\frac{1}{3}$ for each i, we see that $\mathbf{x}$ and $\mathbf{y}$ are $(-\frac{1}{3})$-correlated in the sense of 2.3.2 so that $\mathbf{E}[f(\mathbf{x})f(\mathbf{y})] = \mathbf{Stab}_{-\frac{1}{3}}[f]$ by Definition 2.3.5. Similarly, $\mathbf{E}[f(\mathbf{x})f(\mathbf{z})] = \mathbf{E}[f(\mathbf{y})f(\mathbf{z})] = \mathbf{Stab}_{-\frac{1}{3}}[f]$ and the proof is complete. □

Corollary 2.4.2. *In a 3-candidate Condorcet election using $f\colon \{-1,1\}^n \to \{-1,1\}$, the probability of a Condorcet winner is at most $\frac{7}{9} + \frac{2}{9}\|f_{=1}\|_2^2$.*

Proof. From Theorem 2.3.11, we have that the probability in question is

$$\begin{aligned}
\frac{3}{4} - \frac{3}{4}\mathbf{Stab}_{-\frac{1}{3}}[f] &= \frac{3}{4} - \frac{3}{4}\left(\|f_{=0}\|_2^2 - \frac{1}{3}\|f_{=1}\|_2^2 + \frac{1}{9}\|f_{=2}\|_2^2 - \frac{1}{27}\|f_{=3}\|_2^2 + \dots\right) \\
&\leq \frac{3}{4}\left(1 + \frac{1}{3}\|f_{=1}\|_2^2 + \frac{1}{27}\|f_{=3}\|_2^2 + \frac{1}{243}\|f_{=5}\|_2^2 + \dots\right) \\
&\leq \frac{3}{4}\left(1 + \frac{1}{3}\|f_{=1}\|_2^2 + \frac{1}{27}(\|f_{=3}\|_2^2 + \|f_{=5}\|_2^2 + \dots)\right) \\
&\leq \frac{3}{4}\left(1 + \frac{1}{3}\|f_{=1}\|_2^2 + \frac{1}{27}(1 - \|f_{=1}\|_2^2)\right) = \frac{7}{9} + \frac{2}{9}\|f_{=1}\|_2^2.
\end{aligned}$$

□

Corollary 2.4.3 (Arrow's Theorem)**.** *Suppose $f\colon \{-1,1\}^n \to \{-1,1\}$ is a unanimous voting rule used in a 3-candidate Condorcet election. If there is always a Condorcet winner, then f must be a dictatorship.*

Proof. If there is always a Condorcet winner, then $1 \leq \frac{7}{9} + \frac{2}{9}\|f_{=1}\|_2^2 \leq \frac{7}{9} + \frac{2}{9}\|f\|_2^2 = \frac{7}{9} + \frac{2}{9} = 1$ and thus $\|f_{=1}\|_2^2 = 1 = \|f\|_2^2$. Hence $f = f_{=1}$. We leave it to the reader to show that f must then be a dictator or a negated dictator. Since f is unanimous, f is a dictator.[2] □

[2] This is the same argument as at the end of the proof of 2.3.12. In future versions of this script, one might want to formulate a suitable lemma.

§3 Spectral structure and learning

§3.1 Low degree spectral concentration

Definition 3.1.1. We say that the Fourier spectrum of $f\colon \{-1,1\}^n \to \mathbb{R}$ is ε-*concentrated on degree up to* k if $\|f_{>k}\|_2^2 \leq \varepsilon$ [$\to$ 1.3.4]. For $f\colon \{-1,1\}^n \to \{-1,1\}$ we can express this condition using the spectral sample [$\to$ 1.3.3]: $\mathbf{Pr}_{\mathbf{S}\sim\widehat{f}^2}[\#\mathbf{S} > k] \leq \varepsilon$.

Proposition 3.1.2. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and suppose $\varepsilon > 0$. Then the Fourier spectrum of f is ε-concentrated on degree up to $\frac{\mathbf{I}[f]}{\varepsilon}$.*

Proof. First consider the special case where $\mathbf{I}[f] = 0$. Then $\mathbf{Var}[f] = 0$ by the Poincaré inequality 2.2.19. This implies that f is constant by 1.2.16. Then $\|f_{>0}\|_2^2 = 0 \leq \varepsilon$. Now consider the case where $\mathbf{I}[f] > 0$. From 2.2.17(c) it follows that $\frac{\mathbf{I}[f]}{\varepsilon}\left\|f_{\geq\frac{\mathbf{I}[f]}{\varepsilon}}\right\|_2^2 \leq \mathbf{I}[f]$. Thus $\|f_{>\frac{\mathbf{I}[f]}{\varepsilon}}\|_2^2 \leq \|f_{\geq\frac{\mathbf{I}[f]}{\varepsilon}}\|_2^2 \leq \varepsilon$. □

Proposition 3.1.3. *For any $f\colon \{-1,1\}^n \to \{-1,1\}$ and $\delta \in (0,\frac{1}{2}]$, the Fourier spectrum of f is ε-concentrated on degree up to $\frac{1}{\delta}$ for $\varepsilon := \frac{2}{1-e^{-2}}\,\mathbf{NS}_\delta[f] \leq 3\,\mathbf{NS}_\delta[f]$.*

Proof. Consider the function $\varphi\colon\ (0,\frac{1}{2}] \to \mathbb{R},\ t \mapsto 1-(1-2t)^{\frac{1}{t}}$. We have

$$\varphi(t) = 1 - e^{\frac{\log(1-2t)}{t}}$$

and therefore

$$\varphi'(t) = e^{\frac{\log(1-2t)}{t}}\,\frac{t\frac{1}{1-2t}(-2) - \log(1-2t)}{t^2}$$

for all $t \in (0,\frac{1}{2})$. Introducing $g\colon\ [0,\frac{1}{2}) \to \mathbb{R},\ t \mapsto -\frac{2t}{1-2t} - \log(1-2t)$, it follows that

$$\operatorname{sgn}\varphi'(t) = -\operatorname{sgn} g(t)$$

for all $t \in (0,\frac{1}{2})$. Because of $g'(t) = \frac{-2(1-2t)-4t}{(1-2t)^2} + \frac{2}{1-2t} = -\frac{4t}{(1-2t)^2} \leq 0$ for $t \in [0,\frac{1}{2})$ and $g(0) = 0$, g is pointwise positive. Hence φ is monotonically increasing on $(0,\frac{1}{2})$. Because of

$$\lim_{t\to\frac{1}{2}}\varphi(t) = \lim_{t\to\frac{1}{2}}\left(1-e^{\frac{\log(1-2t)}{t}}\right) = 1 = 1-0^2 = \varphi\left(\frac{1}{2}\right),$$

φ is moreover continuous. Therefore φ is monotonically decreasing on its whole domain $(0, \frac{1}{2}]$. Now

$$2\,\mathbf{NS}_\delta[f] \overset{2.3.6}{=} 1 - \mathbf{Stab}_{1-2\delta}[f] \overset{2.3.11}{=} 1 - \mathop{\mathbf{E}}_{\mathbf{S}\sim\widehat{f}^2}[(1-2\delta)^{\#\mathbf{S}}] = \mathop{\mathbf{E}}_{\mathbf{S}\sim\widehat{f}^2}[1-(1-2\delta)^{\#\mathbf{S}}]$$
$$\geq (1-(1-2\delta)^{\frac{1}{\delta}}) \mathop{\mathbf{Pr}}_{\mathbf{S}\sim\widehat{f}^2}\left[\#\mathbf{S} \geq \frac{1}{\delta}\right] \geq (1-e^{-2}) \mathop{\mathbf{Pr}}_{\mathbf{S}\sim\widehat{f}^2}\left[\#\mathbf{S} \geq \frac{1}{\delta}\right]$$

where for the last inequality we use that φ is monotonically increasing and that

$$\lim_{t\to 0} \varphi(t) = 1 - e^{-2}.$$

$\square$

Lemma 3.1.4. For all $f\colon \{-1,1\}^n \to \mathbb{R}$ with $f \neq 0$, we have [$\to$ 1.2.9]

$$\mathop{\mathbf{Pr}}_{\mathbf{x}}[f(\mathbf{x}) \neq 0] \geq \frac{1}{2^{\deg f}}.$$

Proof. We prove by induction on $d \in \mathbb{N}_0$ that for all $n \in \mathbb{N}_0$ and $f\colon \{-1,1\}^n \to \mathbb{R}$ of degree d, we have $\mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x}) \neq 0] \geq \frac{1}{2^d}$.

$\underline{d=0}$ For all $n \in \mathbb{N}_0$ and $f\colon \{-1,1\}^n \to \mathbb{R}$ of degree 0, we have $\mathbf{Pr}_{\mathbf{x}}[f(\mathbf{x}) \neq 0] = 1 = \frac{1}{2^0}$.

$\underline{d-1 \to d \quad (d \in \mathbb{N})}$ Let $n \in \mathbb{N}_0$ and let $f\colon \{-1,1\}^n \to \mathbb{R}$ be of degree d. Choose $i \in [n]$ and $g, h\colon \{-1,1\}^{n-1} \to \mathbb{R}$ such that

$$f(x) = g(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{n-1}) + h(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{n-1})x_i$$

for all $x \in \{-1,1\}^n$ and $\deg h = d-1$. WLOG $i = n$. Now we have

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\mathbf{x}}[f(\mathbf{x}) \neq 0] &= \frac{1}{2} \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^{n-1}}[f(\mathbf{x},-1) \neq 0] + \frac{1}{2} \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^{n-1}}[f(\mathbf{x},1) \neq 0] \\
&= \frac{1}{2} \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^{n-1}}[g(\mathbf{x}) - h(\mathbf{x}) \neq 0] + \frac{1}{2} \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^{n-1}}[g(\mathbf{x}) + h(\mathbf{x}) \neq 0] \\
&\geq \frac{1}{2} \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^{n-1}}[g(\mathbf{x}) - h(\mathbf{x}) \neq 0 \text{ or } g(\mathbf{x}) + h(\mathbf{x}) \neq 0] \\
&\geq \frac{1}{2} \mathop{\mathbf{Pr}}_{\mathbf{x}\sim\{-1,1\}^{n-1}}[h(\mathbf{x}) \neq 0] \underset{\text{hypothesis}}{\overset{\text{induction}}{\geq}} \frac{1}{2} \cdot \frac{1}{2^{d-1}} = \frac{1}{2^d}.
\end{aligned}$$

$\square$

Proposition 3.1.5. *Let $d \in \mathbb{N}_0$, $f\colon \{-1,1\}^n \to \{-1,1\}$ be of degree $\leq d$ and $i \in [n]$. Then $\mathbf{Inf}_i[f]$ is either 0 or at least 2^{1-d}.*

Proof. $\mathbf{Inf}_i[f] \underset{2.2.3}{\overset{2.2.1}{=}} \mathbf{Pr}_{\mathbf{x}}[D_i f(\mathbf{x}) \neq 0]$ is zero if $D_i f = 0$ and $\geq \frac{1}{2^{\deg(D_i f)}}$ otherwise. By 2.2.6, $\deg(D_i f) \leq (\deg f) - 1 \leq d - 1$. $\square$

Lemma 3.1.6. Let $f\colon \{-1,1\}^n \to \{-1,1\}$. Then $\mathbf{I}[f] \leq \deg f$.

Proof.

$$\mathbf{I}[f] \overset{2.2.17(c)}{=} \sum_{k=0}^{n} k\|f_{=k}\|_2^2 = \sum_{k=0}^{\deg f} k\|f_{=k}\|_2^2 \leq (\deg f) \sum_{k=0}^{\deg f} \|f_{=k}\|_2^2$$
$$= (\deg f) \sum_{k=0}^{n} \|f_{=k}\|_2^2 = (\deg f)\|f\|_2^2 = \deg f$$

□

Theorem 3.1.7. *Let $d \in \mathbb{N}_0$ and suppose $f\colon \{-1,1\}^n \to \{-1,1\}$ has degree $\leq d$. Then f is a $d2^{d-1}$-junta* [→ 2.1.1(e)].

Proof. Let $k \in \{0,\dots,n\}$ denote the number of relevant coordinates for f [→ 2.2.5]. By Remark 2.2.6(b), we have to show $k \leq d2^{d-1}$. This follows from

$$k2^{1-d} \overset{3.1.5}{\leq} \sum_{i=1}^{n} \mathbf{Inf}_i[f] \overset{2.2.11}{=} \mathbf{I}[f] \overset{3.1.6}{\leq} \deg f \leq d.$$

□

§3.2 Subspaces and decision trees

Definition 3.2.1. [→ 1.2.6] We will use the norms defined on $\mathbb{R}^{\{-1,1\}^n}$ and $\mathbb{R}^{\mathscr{P}([n])}$ by

$$\|f\|_\infty := \max\{|f(x)| \mid x \in \{-1,1\}^n\},$$
$$\|\widehat{f}\|_\infty := \max\{|\widehat{f}(S)| \mid S \subseteq [n]\},$$
$$\|f\|_1 := \sum_{x\in\{-1,1\}^n} |f(x)| \qquad \text{and}$$
$$\|\widehat{f}\|_1 := \sum_{S\subseteq[n]} |\widehat{f}(S)|$$

for all $f\colon \{-1,1\}^n \to \mathbb{R}$.

Definition 3.2.2. Let $\varepsilon \in \mathbb{R}$ and $f\colon \{-1,1\}^n \to \mathbb{R}$. We say that $\widehat{f}$ is ε-granular if $\widehat{f}(S)$ is an integer multiple of ε for all $S \subseteq [n]$.

Definition and Proposition 3.2.3. Let $n \in \mathbb{N}_0$ and let K be a field. For each subspace U of the K-vector space K^n, we define the subspace

$$U^\perp := \{x \in K^n \mid \forall y \in U : \sum_{i=1}^{n} x_i y_i = 0\}$$

of K^n. Then for any subspace U of K^n, we have $\dim(U) + \dim(U^\perp) = n$ and

$$U^{\perp\perp} := (U^\perp)^\perp = U.$$

Proof. Let U be a subspace of K^n and choose an $m \in \mathbb{N}_0$ and a matrix $A \in K^{m\times n}$ whose rows form a basis of U. Then $U^\perp = \ker A$ and thus

$$\dim(U) + \dim(U^\perp) = m + \dim\ker A = \operatorname{rk} A + \dim\ker A = n.$$

For each $x \in U$, we have $\sum_{i=1}^n x_i y_i = \sum_{i=1}^n y_i x_i = 0$ for all $y \in U^\perp$ and thus $x \in U^{\perp\perp}$. This shows $U \subseteq U^{\perp\perp}$. But U and $U^{\perp\perp}$ have the same dimension and are therefore equal since $\dim(U) + \dim(U^\perp) = n = \dim(U^\perp) + \dim(U^{\perp\perp})$. □

Proposition 3.2.4. *Let $n \in \mathbb{N}_0$. Consider the vector space $\mathbb{F}_2^n$ over the field $\mathbb{F}_2 = \{0,1\}$. Consider the natural maps $\{-1,1\}^n \xleftarrow{\iota} \mathbb{F}_2^n \xrightarrow{\iota'} \mathscr{P}([n])$ defined by*

$$\text{the } i\text{-th component of } \iota(x) \text{ equals } -1 \iff x_i = 1 \iff i \in \iota'(x)$$

for all $x \in \mathbb{F}_2^n$ and $i \in [n]$. These maps are group isomorphisms [→ 1.3.11]. Let U be a subspace of $\mathbb{F}_2^n$ of codimension $k := n - \dim U$ and let $v \in \mathbb{F}_2^n$. For the the 0-1-indicator function

$$1_{\iota(v+U)} \colon \{-1,1\}^n \to \{0,1\} \subseteq \mathbb{R}$$

of $\iota(v+U) \subseteq \{-1,1\}^n$, we have for all $S \subseteq [n]$

$$\widehat{1_{\iota(v+U)}}(S) = \begin{cases} (\iota(v))^S 2^{-k} & \text{if } S \in (\iota'(U))^\perp, \\ 0 & \text{otherwise.} \end{cases}$$

Hence $\varphi_{\iota(v+U)} = \sum_{S\in\iota'(U^\perp)} (\iota(v))^S \chi_S$ where $\varphi_{\iota(v+U)}$ is the density function associated to the uniform distribution on $\iota(v+U)$ from 1.3.6. We have $\{S \subseteq [n] \mid \widehat{1_{\iota(v+U)}}(S) \neq \emptyset\} = 2^k$, $\widehat{1_{\iota(v+U)}}$ is 2^{-k}-granular, $\|\widehat{1_{\iota(v+U)}}\|_\infty = 2^{-k}$ and $\|\widehat{1_{\iota(v+U)}}\|_1 = 1$.

Proof. Choose a basis $u_1, \dots, u_k$ of $U^\perp$. Then

$$\begin{aligned} w \in v + U \iff w - v \in U \iff w + v \in U &\overset{3.2.3}{\iff} v + w \in U^{\perp\perp} \\ &\iff \forall i \in [k] : (\iota(v+w))^{\iota'(u_i)} = 1 \end{aligned}$$

for all $w \in \mathbb{F}_2^n$. Hence

$$\begin{aligned} 2^k 1_{\iota(v+U)}(x) &= 2^k \prod_{i=1}^k \left(\tfrac{1}{2} + \tfrac{1}{2}(\iota(v + \iota^{-1}(x)))^{\iota'(u_i)}\right) \\ &= \sum_{I\subseteq[n]} \prod_{i\in I} (\iota(v+\iota^{-1}(x)))^{\iota'(u_i)} \\ &\overset{\iota' \text{ group hom.}}{=} \sum_{I\subseteq[n]} (\iota(v+\iota^{-1}(x)))^{\iota'(\sum_{i\in I} u_i)} \\ &= \sum_{\lambda_1,\dots,\lambda_k\in\mathbb{F}_2} (\iota(v+\iota^{-1}(x)))^{\iota'(\lambda_1 u_1 + \dots + \lambda_k u_k)} \\ &\overset{u_1,\dots,u_k \text{ basis}}{=} \sum_{y\in U^\perp} (\iota(v+\iota^{-1}(x)))^{\iota'(y)} \\ &= \sum_{S\in\iota'(U^\perp)} (\iota(v+\iota^{-1}(x)))^S \\ &\overset{\iota \text{ group hom.}}{=} \sum_{S\in\iota'(U^\perp)} (\iota(v)\iota(\iota^{-1}(x)))^S \\ &= \sum_{S\in\iota'(U^\perp)} (\iota(v)x)^S \\ &= \sum_{S\in\iota'(U^\perp)} (\iota(v))^S x^S \end{aligned}$$

for all $x \in \{-1,1\}^n$. The rest is easy. □

Definition 3.2.5. A *decision tree* T is a representation of a real-valued Boolean function $f\colon \{-1,1\}^n \to \mathbb{R}$. It consists of a rooted binary tree in which the internal nodes are labeled by coordinates $i \in [n]$, the outgoing edges of each internal node are labeled -1 and 1, and the leaves are labeled by real numbers. We insist that no coordinate $i \in [n]$ appears more than once on any root-to-leaf path.

On input $x \in \mathbb{F}_2^n$, the tree T constructs a *computation path* from the root node to a leaf. Specifically, when the computation path reaches an internal node labeled by coordinate $i \in [n]$ we say that T *queries* x_i. The computation path then follows the outgoing edge labeled by x_i. The output of T (and hence f) on input x is the label of the leaf reached by the computation path.

The *size* s of a decision tree is the total number of leaves. The *depth* k of T is the maximum length of any root-to-leaf path (where you count its number of edges or, equivalently, the number of its internal nodes). Given $f\colon \{-1,1\}^n \to \mathbb{R}$, we write $\mathrm{DT}(f)$ (respectively, $\mathrm{DT}_{\mathrm{size}}(f)$) for the least depth (respectively, size) of a decision tree computing f.

Proposition 3.2.6. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ be computable by a decision tree of size s and depth k. Then*

(a) $\deg f \le k$,

(b) $\#\{S \subseteq [n] \mid \widehat{f}(S) \neq 0\} \le s2^k \le 4^k$,

(c) $\|\widehat{f}\|_1 \le s\|f\|_\infty \le 2^k\|f\|_\infty$ *and*

(d) $\widehat{f}$ *is 2^{-k}-granular if $f(\{-1,1\}^n) \subseteq \mathbb{Z}$.*

Proof. Easy. □

Lemma 3.2.7. If $k, n \in \mathbb{N}_0$ and $f\colon \{-1,1\}^n \to \{-1,1\}$ is a Boolean function computed by a decision tree T with exactly s leafs whose connecting path to the root hat at least k edges, then $2^k\|f_{\ge k}\|_2^2 \le s$.

Proof. Induction on k. For $k = 0$, we have $2^k\|f_{\ge k}\|_2^2 = \|f_{\ge k}\|_2^2 = \|f\|_2^2 = 1 \le s$. Now consider $k \in \mathbb{N}$ and suppose the claim is already proven for $k-1$ instead of k. If T consists only of its root, then f is constant so that we have $2^k\|f_{\ge k}\|_2^2 = 0 = s$. From now on suppose that the root of T is an internal node and denote by T_b the decision subtree of T whose root is connected to the root of T by the edge labeled with $b \in \{-1,1\}$. In particular $n \ge 1$ and WLOG the root of T is labeled by coordinate 1. Then T_b computes $f_b\colon \{-1,1\}^{n-1} \to \{-1,1\}$, $(x_2,\ldots,x_n) \mapsto f(b,x_2,\ldots,x_n)$ for each $b \in \{-1,1\}$. Denote by s_b the number of leafs of T_b whose connecting path to the root of T_b has at least $k-1$ edges. By induction hypothesis, we have $2^{k-1}\|(f_b)_{\ge k-1}\|_2^2 \le s_b$ for each $b \in \{-1,1\}$. Now

$$f(x) = \left(\frac{1-x_1}{2}\right) f_{-1}(x_2,\ldots,x_n) + \left(\frac{1+x_1}{2}\right) f_1(x_2,\ldots,x_n)$$

for all $x \in \mathbb{R}^n$. Defining

$$g\colon \{-1,1\}^n \to \mathbb{R},$$
$$x \mapsto \left(\frac{1-x_1}{2}\right)(f_{-1})_{\geq k-1}(x_2,\ldots,x_n) + \left(\frac{1+x_1}{2}\right)(f_1)_{\geq k-1}(x_2,\ldots,x_n),$$

we obviously have that $f_{\geq k} = g_{\geq k}$ and thus $\|f_{\geq k}\|_2^2 = \|g_{\geq k}\|_2^2 \leq \|g\|_2^2$. Moreover,

$$\begin{aligned}
2^k\|f_{\geq k}\|_2^2 = 2^k\|g_{\geq k}\|_2^2 \leq 2^k\|g\|_2^2 = \\
2^k \mathop{\mathbf{E}}_{\mathbf{x}} \left[\begin{array}{l} \left(\frac{1-\mathbf{x}_1}{2}\right)^2 (f_{-1})_{\geq k-1}(\mathbf{x}_2,\ldots,\mathbf{x}_n))^2 + \\ 2\left(\frac{1-\mathbf{x}_1}{2}\right)\left(\frac{1+\mathbf{x}_1}{2}\right)(f_{-1})_{\geq k-1}(\mathbf{x}_2,\ldots,\mathbf{x}_n))^2(f_1)_{\geq k-1}(\mathbf{x}_2,\ldots,\mathbf{x}_n))^2 + \\ \left(\frac{1+\mathbf{x}_1}{2}\right)^2 (f_1)_{\geq k-1}(\mathbf{x}_2,\ldots,\mathbf{x}_n))^2] \end{array}\right] \\
= 2^k\left(\frac{1}{2}\|(f_{-1})_{\geq k-1}\|_2^2 + \frac{1}{2}\|(f_1)_{\geq k-1}\|_2^2\right) \\
= 2^{k-1}\|(f_{-1})_{\geq k-1}\|_2^2 + 2^{k-1}\|(f_1)_{\geq k-1}\|_2^2 \leq s_{-1} + s_1 = s.
\end{aligned}$$

□

Theorem 3.2.8. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be computable by a decision tree of size s and let $\varepsilon \in (0,1]$. Then the spectrum of f is ε-concentrated on degree up to $\log_2\left(\frac{s}{\varepsilon}\right)$* [$\to$ 3.1.1].

Proof. Setting $k := \lfloor \log_2\left(\frac{s}{\varepsilon}\right)\rfloor \in \mathbb{N}_0$, we have to show that $\|f_{\geq k+1}\|_2^2 = \|f_{>k}\|_2^2 \leq \varepsilon$. From $\log_2\left(\frac{s}{\varepsilon}\right) \leq k+1$, we obtain $\frac{s}{\varepsilon} \leq 2^{k+1}$. It therefore suffices to show that

$$\|f_{\geq k+1}\|_2^2 \leq \frac{s}{2^{k+1}}.$$

This follows from 3.2.7. □

Bibliography

[O'D] R. O'Donnell: Analysis of Boolean functions, Cambridge University Press, New York, 2014, available online at analysisofbooleanfunctions.net 1, 2