Intelligent Agents

Fourier Analysis I: Basics

Özgür L. Özçep

Universität zu Lübeck

Institut für Informationssysteme

Todays and next weeks lecture based on

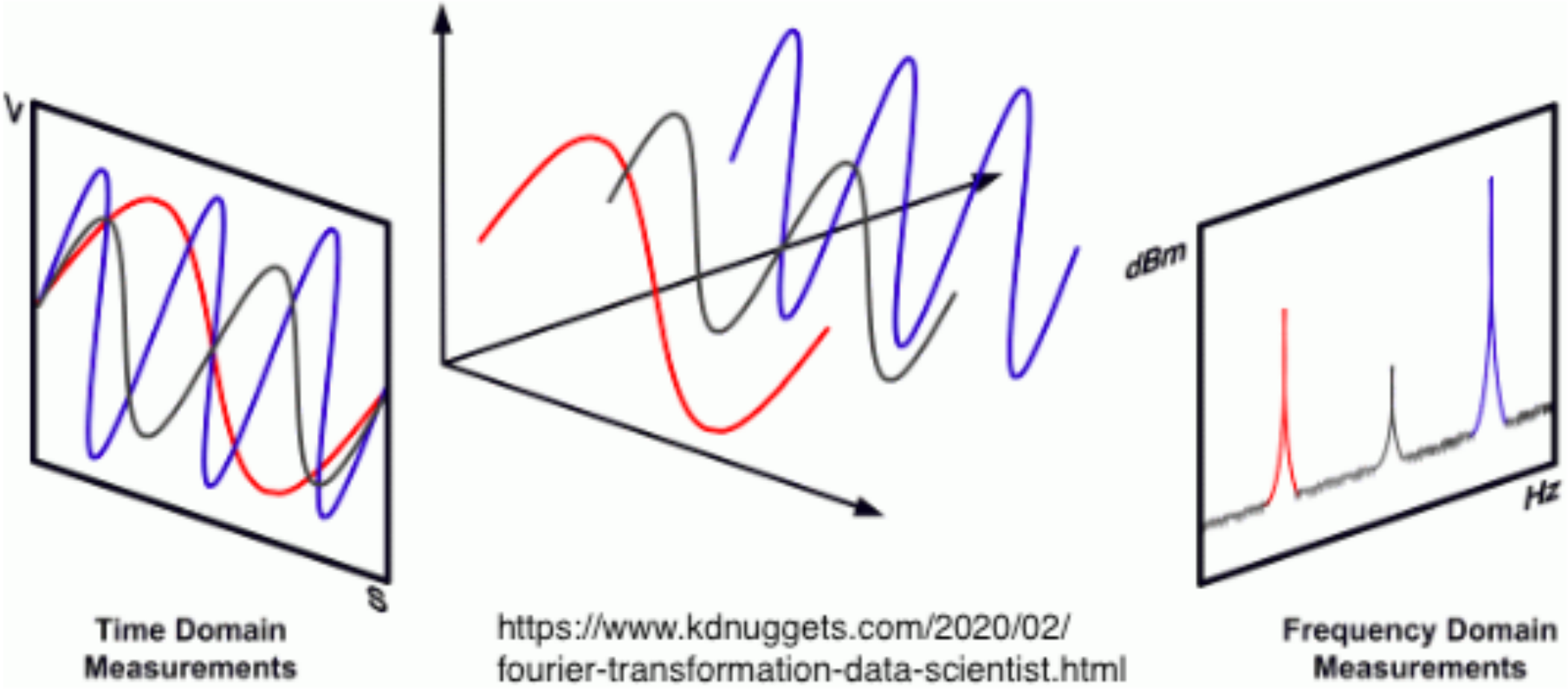
- Lecture notes „Fourier Analysis of Boolean Functions, Witer term 16/17“ M. Schweighofer
  http://www.math.uni-konstanz.de/~schweigh/

- Ryan O‘Donnell: Fourier Analyis of Boolean Functions., CUP 2014.
  Free PDF oavailable at
  https://arxiv.org/pdf/2105.10386.pdf

- Talk of Ronald de Wolf: „Fourier analysis of Boolean functions: Some beautiful examples“ available at
  https://nvti.nl/slides/deWolf.pdf

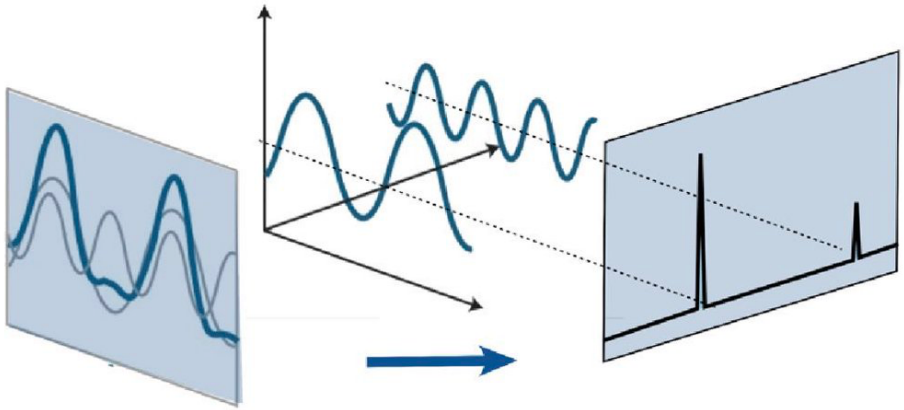
MOTIVATION

The main idea of classical Fourier analysis

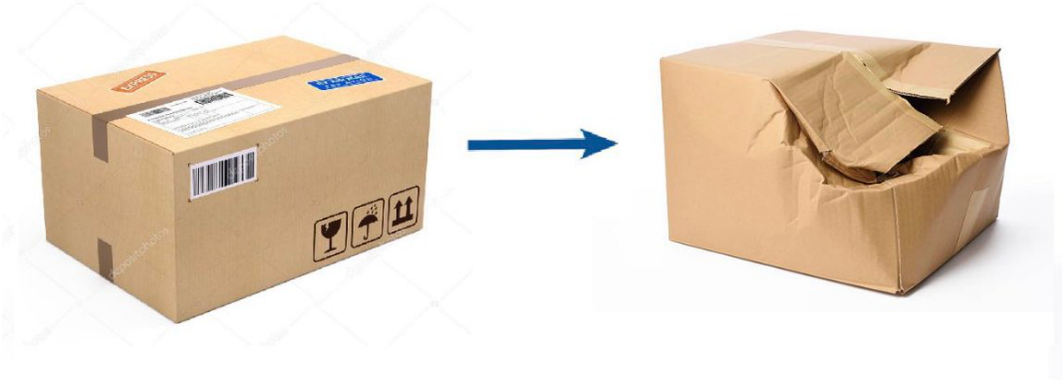


A bad adaptation ...

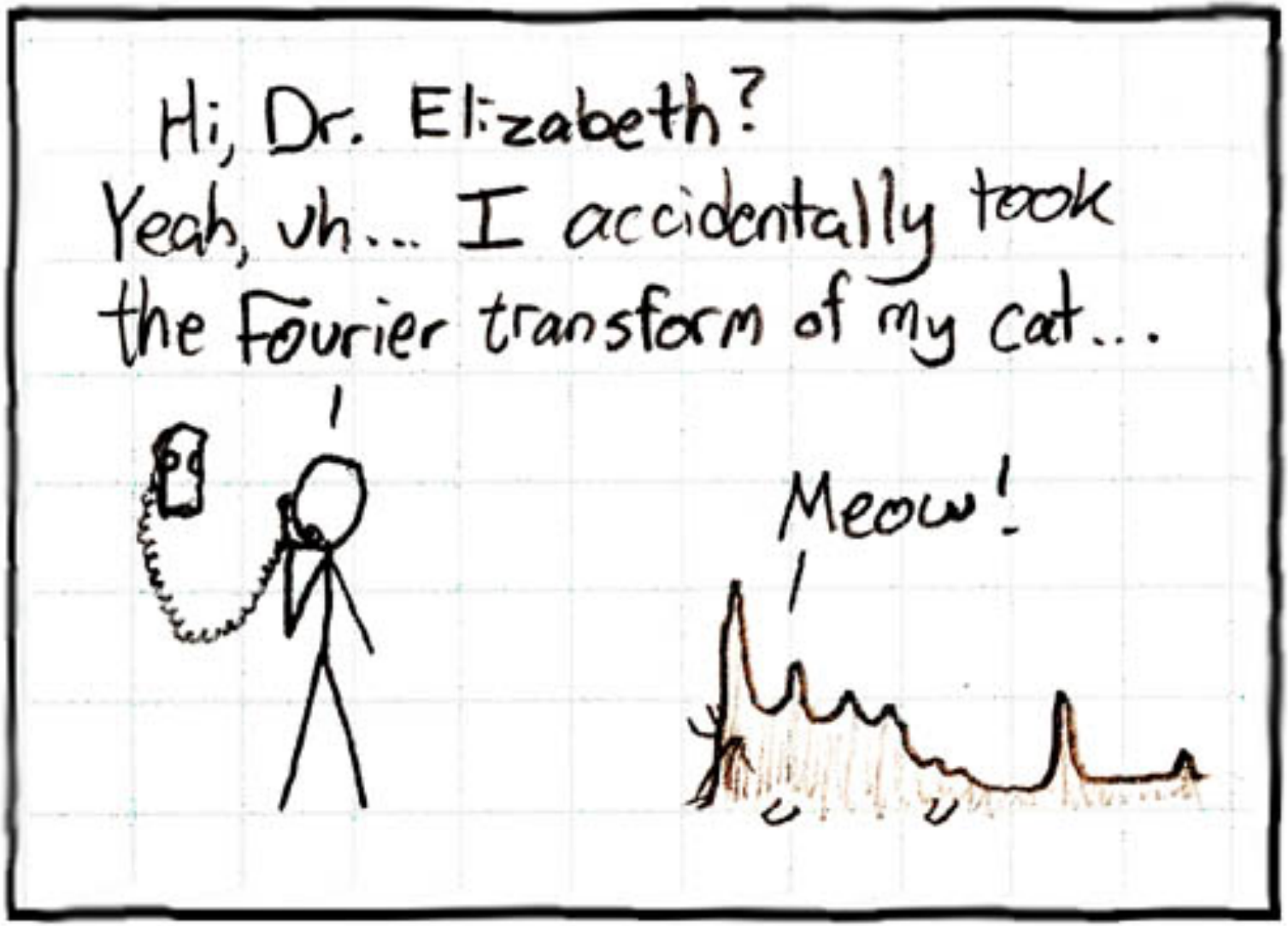
Fourier Transform:



Courier Transform:



A bad application …



Applications

- Many applications in math, physics, engineering, . . . and in computer science:
  - Signal processing
  - Data compression
  - Multiplying two polynomials
- These examples use Fourier analysis over cyclic groups
- We will focus on Fourier analysis over the Boolean cube $= \{-1,1\}^n$

Applications in CS

- Analysis of error-correcting codes
- Learning theory
- Sensitivity of a function to noise on the inputs
- PCPs, NP-hardness of approximation
- Cryptography
- Lower bounds on communication complexity
- Threshold phenomena in random graphs
- Quantum computing
- Notion of influence of variables on a function useful in particular for social theory
- ...

The many faces of Boolean values

- In philosophy: Boolean truth values

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

- In CS this is encoded by field $\mathbb{F}_2 = \{1,0\}$
  - $TRUE \mapsto 1$
  - $FALSE \mapsto 0$

- Sometimes instead work with $\{0, 1\} \subseteq \mathbb{R}$
- In Fourier analysis usually $\{1, -1\} \subseteq \mathbb{R}$ is used where
  - $TRUE \mapsto -1$
  - $FALSE \mapsto 1$

The many faces of Boolean values

$\mathbb{B}$	$\mathbb{F}_2$	$\{0,1\} \subseteq \mathbb{R}$	$\{-1,1\} \subseteq \mathbb{R}$
$\bot$	0	0	1
$\top$	1	1	-1
$\neg$	$1+x$	$1-x$	$-x$
$\wedge$	$\cdot$	$\cdot$	$\frac{1+x+y-xy}{2}$
$\vee$	$x+y+xy$	$x+y-xy$	$\frac{-1+x+y+xy}{2}$
$\oplus$ (XOR)	$+$	$x+y-2xy$	$\cdot$

FOURIER TRANSFORM

Fourier analysis over the Boolean cube

- Real-valued boolean function: $f: \{-1,1\}^n \to \mathbb{R}$
- Boolean function: $f: \{-1,1\}^n \to \{-1,1\}$
- $[n] \coloneqq \{1, \dots, n\}$
- $Pow(A) =$ power set $= \{S \mid S \subseteq A\}$
- Parity functions correspond to the cosines and sines in classical Fourier analysis
- For the $\{1,-1\}$-encoding they are monomials

Definition

For $S \subseteq [n]$ the monomial function $\chi_S(x)$ is defined as

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

Fourier analysis over the boolean cube

- Symmetric difference: $A \Delta B := (A \setminus B) \cup (B \setminus A)$
- Kronecker symbol for x $\in \{1, -1\}$
  $\delta_x : \{1, -1\}^n \to \mathbb{R}; \delta_x(y) = 1$ if $y = x$ else $= 0$
- Kronecker symbol for For $S \subseteq [n]$
  $\delta_S : Pow([n]) \to \mathbb{R}; \delta_S(T) = 1$ if $S =$ T else $= 0$

- We will consider the Boolean cube as probability space
  - $x \sim \{1, -1\}^n$ is a uniformly chosen random element from $\{1, -1\}^n$
  - Expectation value:

$$E_x(f) = E_{x \sim \{1,-1\}^n}(f) = \frac{1}{2^n} \sum_{x \in \{1,-1\}^n} f(x)$$

Fourier analysis over the boolean cube

- Space of functions $\mathbb{R}^{\{1,-1\}^n} = \{ f \mid f: \{1,-1\}^n \to \mathbb{R} \}$ is a vector space.
- It is even an Euclidean space with a normalized inner product $\langle f, g \rangle$

Definition

$$\langle f, g \rangle = E_{x \sim \{1,-1\}^n}\big(f(x)g(x)\big) = \frac{1}{2^n} \sum_{x \in \{1,-1\}^n} f(x)g(x)$$

- The induced norm is defined as

$$||f||_2 = \sqrt{\langle f, f \rangle} = \sqrt{E(f^2)}$$

Parity functions form an orthormal basis

Theorem

Let $n \in \mathbb{N}_0$. The parity-functions $(\chi_S)_{S\subseteq[n]}$ form an orthonormal basis of the euclidean vector space $\mathbb{R}^{\{1,-1\}^n}$.
We call $(\chi_S)_{S\subseteq[n]}$ the Fourier basis of $\mathbb{R}^{\{1,-1\}^n}$

- Proof
  1. First we show $\chi_S\chi_T = \chi_{S\Delta T}$:
     $\chi_S\chi_T(\mathrm{x}) = \prod_{i\in S} x_i \prod_{i\in T} x_i = \prod_{i\in S\Delta T} x_i \prod_{i\in S\cap T} x_i^2$
     $= \prod_{i\in S\Delta T} x_i = \chi_{S\Delta T}(x)$
  2. Second we show $E_{x\sim\{1,-1^n\}}(\chi_S) = \delta_\emptyset(S)$:
     If $S = \emptyset$, then $E_{x\sim\{1,-1^n\}}(\chi_S) = E_{x\sim\{1,-1^n\}}(1) = 1$. Otherwise $E_{x\sim\{1,-1^n\}}(\prod_{i\in S} x_i) = \prod_{i\in S} E_{x_i\sim\{1,-1\}}(x_i)$ (by independence).
     But $E_{x_i\sim\{1,-1\}}(x_i) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot -1 = 0$)

Parity functions form an orthormal basis

Theorem

Let $n \in \mathbb{N}_0$. The parity-functions $(\chi_S)_{S\subseteq[n]}$ form an orthonormal basis of the euclidean vector space $\mathbb{R}^{\{1,-1\}^n}$.
We call $(\chi_S)_{S\subseteq[n]}$ the Fourier basis of $\mathbb{R}^{\{1,-1\}^n}$

- Proof (continued)
  - It is sufficient to show
$$\langle \chi_S, \chi_T \rangle = \delta_S(T)$$
  - But this follows from 1. 2. above.

Fourier expansion

Theorem

Each real-valued Boolean function f can be uniquely written as

$$f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S \qquad \text{(Fourier expansion)}$$

with unique Fourier coefficients $\hat{f}(S)$ given by:

$$\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{1,-1\}^n} f(x)\chi_S(x)$$

In fact the, (boolean) Fourier transfoem

$$\mathcal{F}\colon \mathbb{R}^{\{1,-1\}^n} \to \mathbb{R}^{Pow(n)}; f \mapsto \hat{f}$$

is a vector space isomorphism

Definition

The degree of f is the largest cardinality of a set S a Fourier coefficient $\hat{f}(S) \neq 0$:

$\deg f = \max\{|S| \mid S \subseteq [n], \hat{f}(S) \neq 0\}$ if $f \neq 0$

$\deg f = -\infty$ else

Examples

Example

1. $f = max^2()$; maximum function on 2 bits (= logical AND: $\wedge$)
   - $max^2(x,y) = \frac{1+x+y-xy}{2}$ (see table from the beginning)
   - $\hat{f}(\{1\}) = \langle f, \chi_{\{1\}} \rangle = \frac{1}{2^2} \sum_{(x,y)\in\{1,-1\}^2} f(x,y)\chi_{\{1\}}(x,y)$

$$= \frac{1}{4} \sum_{\boldsymbol{x}\in\{1,-1\}^2} f(x,y) \cdot x$$

$$= \frac{1}{4} (f(1,1) \cdot 1 + f(-1,1) \cdot (-1) + f(1,-1) \cdot 1 + f(-1,-1) \cdot (-1))$$

$$= \frac{1}{4} (1-1+1+1) = \frac{1}{2}$$

2. Majority function $maj_3(x_1, x_2, x_3)$ which outputs the more frequently bit $1, -1$ occuring in its input

$$maj_3(x_1, x_2, x_3) = \frac{1 + x_1 + x_2 + x_3 - x_1x_2x_3}{2}$$

Intuition on Fourier Expansion

- Fourier expansion: Any real-valued Boolean function can be represented as multilinear polynomial
- Idea: Interpolation with indicator polynomials
  - Let $a = (a_1, \dots, a_n) \in \{1, -1\}^n$ fixed and
    $$x = (x_1, \dots, x_n) \in \{1, -1\}^n$$
  - $1_{\{a\}}(x) = \left(\frac{1+ a_1 x_1}{2}\right) \cdot \dots \cdot \left(\frac{1+ a_n x_n}{2}\right)$
    takes value 1 when $x = a$ and 0 otherwise
  - Hence $f = \sum_{a \in \{1,-1\}^n} f(a) 1_{\{a\}}(x)$
  - Multiplying out indicator polynomials leads to presentation with monomials
  - As inputs are $\{1, -1\}$-bits any x_i^2 reduces to x_i: linearity

Example

$f = max^2()$; maximum function on 2 bits (= logical AND: ∧)

- $max^2(x, y) =$

$$(1)\ \left(\frac{1+ x_1}{2}\right)\left(\frac{1+ x_2}{2}\right)$$
$$+ (1)\ \left(\frac{1- x_1}{2}\right)\left(\frac{1+ x_2}{2}\right)$$
$$+ (1)\ \left(\frac{1+ x_1}{2}\right)\left(\frac{1- x_2}{2}\right)$$
$$+ (-1)\left(\frac{1- x_1}{2}\right)\left(\frac{1- x_2}{2}\right) = \frac{1 + x + y - xy}{2}$$

Frequently used insight: Plancherel

Theorem (Plancherel)

For any real-valued Boolean function $f: \{1, -1\}^n \rightarrow \mathbb{R}$:

$$\langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \hat{g}(S)$$

Proof idea

- Use the Fourier expansions of f, g , the linearity of the scalar product and the fact that the parity functions are orthonormal.

Special Case: Parseval

Theorem (Parseval)

For any real-valued Boolean function $f: \{1, -1\}^n \rightarrow \mathbb{R}$:

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

- In particular, if f: $\{1, -1\}^n \rightarrow \{1, -1\}$ then: $||f||_2^2 = 1$.
- So the squares of fourer coefficients $\hat{f}(S)^2$, called the „Fourier weights“, can be interpreted as a probability mass function on $Pow([n])$.

Examples

$\mathbb{B}$	$\mathbb{F}_2$	$\{0,1\} \subseteq \mathbb{R}$	$\{-1,1\} \subseteq \mathbb{R}$
$\bot$	0	0	1
$\top$	1	1	-1
$\neg$	$1+x$	$1-x$	$-x$
$\wedge$	$\cdot$	$\cdot$	$\frac{1+x+y-xy}{2}$
$\vee$	$x+y+xy$	$x+y-xy$	$\frac{-1+x+y+xy}{2}$
$\oplus$ (XOR)	$+$	$x+y-2xy$	$\cdot$

- $||\top||_2 = \sqrt{(-1)^2} = 1$
- $||\wedge||_2 = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2} = 1$

Hamming distance

Definition

For two Boolean functions f, g the relative Hamming distance is defined as the fraction of inputs they disagree:

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

- It is easily shown that $dist(,)$ is a metric on the set of boolean functions.
- $dist(,)$ gives a nice interpretation of the scalar product as a measure of similarity.

Theorem

For two Boolean functions f, g:

$$\langle f,g \rangle = \Pr_x\big(f(x) = g(x)\big) - \Pr_x\big(f(x) \neq g(x)\big) = 1 - 2dist(f,g)$$

Wake-Up Exercise

Theorem

For two Boolean functions f, g:

$$\langle f, g \rangle = \Pr_x\big(f(x) = g(x)\big) - \Pr_x\big(f(x) \neq g(x)\big) = 1 - 2dist(f, g)$$

Prove the theorem above on the representation of the scalar product with the Hamming distance.

Answer to wake up exercise

Theorem

For two Boolean functions f, g:

$$\langle f, g \rangle = \Pr_x\big(f(x) = g(x)\big) - \Pr_x\big(f(x) \neq g(x)\big) = 1 - 2dist(f, g)$$

Proof

$$\langle f, g \rangle = E_{x \sim \{1,-1\}^n}(f(x)g(x))$$
$$= \Pr_x(f(x) = g(x)) - \Pr_x\big(f(x) \neq g(x)\big)$$
$$= \Big(1 - \Pr_x\big(f(x) \neq g(x)\big)\Big) - \Pr_x\big(f(x) \neq g(x)\big)$$
$$= 1 - 2dist(f, g)$$

Moments of (real-valued) boolean functions f

- The mean of f: $E(f)$
  - If $E(f) = 0$, then f is called unbiased
  - In particular, if $f: \{1, -1\}^n \to \{1, -1\}$, then $E(f) = 0$ means: f attains each truth value on exactly half of the input bit vectors.

Theorem

For Boolean function f its mean is given by the Fourier coeefficient for the empty set: $E(f) = \hat{f}(\emptyset)$

- $\hat{f}(\emptyset)$ exemplifies general idea in Boolean Fourier analysis: Each Fourier coefficient $\hat{f}(S)$ gives global beahiour of f
  - $\hat{f}(S)$ for small S describe rough global behaviour of f
  - $\hat{f}(S)$ for large S describe fine-tuned global behaviour of f

  Hence: Study f by considering Fourier coefficients for small S

Wake-Up Exercise

Theorem

For Boolean function f its mean is given by the Fourier coeefficient for the empty set: $E(f) = \hat{f}(\emptyset)$

Prove the theorem above on the mean of a function.

Answer to wake up exercise

Theorem

For Boolean function f its mean is given by the Fourier coeefficient for the empty set: $E(f) = \hat{f}(\emptyset)$

Proof

$$E(f) = E(f \cdot 1) = \langle f, 1 \rangle = \langle f, \chi_\emptyset \rangle = \hat{f}(\emptyset)$$

Moments of (real-valued) boolean functions f

- The variance of f: $\mathrm{Var(f)} = E(\,(f - E(f))^2\,)$

Theorem

1. For real-valued boolean function f ist variance is the sum of all squared Fourier coefficients except that for ∅:
   $Var(f) = E(f^2) - E(f)^2 = \left|\left|f - E(f)\right|\right|_2^2 = \sum_{\emptyset \neq S \subseteq [n]} \hat{f}(S)^2$
2. For boolean function f one even has:
   $Var(f) = 1 - E(f)^2 = 4 \Pr_x(f(x) = 1) \Pr_x(f(x) = -1) \in [0,1]$

Theorem

For a Boolean function f then following bounds hold:

$$2\,\epsilon \leq Var(f) \leq 4\,\epsilon$$

where $\epsilon = \min\{dist(f,1), dist(f,-1))\}$

Proof of first theorem

Proof

1. $Var(f)$

$$= E\big((f - E(f)^2)\big)$$
$$= E(f^2 - 2fE(f) + E(f)^2)$$
$$= E(f^2) - 2E(f)^2 + E(f)^2$$
$$= E(f^2) - E(f)^2$$

On the other hand $E\big((f - E(f)^2)\big) = \big||f - E(f)|\big|_2^2$ by definition.

But $f - E(f) = \sum_{\emptyset \neq S \subseteq [n]} \hat{f}(S)\chi_S$. Now can apply Parseval.

2. Using 1. we have $Var(f) = E(f^2) - E(f)^2$. But $E(f^2) = E(1) = 1$.

But $1 - E(f)^2 = (\Pr_x(f(x) = 1) + \Pr_x(f(x) = -1))^2$
$$- (\Pr_x(f(x) = 1) - \Pr_x(f(x) = -1))^2$$
$$= 4\Pr_x(f(x) = 1)\Pr_x(f(x) = -1)$$

Moments of (real-valued) Boolean functions f,g

- The covariance of f, g:

$$\mathrm{Cov(f)} = E(\,(f - E(f)) \cdot (g - E(g))\,)$$

Theorem

For real-valued Boolean function f, g the covariance is the sum of all componentwise products of Fourier coefficients except those for ∅:

$$Cov(f,g) = \sum_{\emptyset \neq S \subseteq [n]} \hat{f}(S)\hat{g}(S)$$

PROBABILITY DENSITIES AND CONVOLUTION

Mass / density function

- We follow the probabilistic perspective on $\{1, -1\}^n$ and on the Fourier coefficients
- The operation of convolution has a special role in this setting

Definition

Let $D \neq \emptyset$ be a finite set. $f: D \to \mathbb{R}_{\geq 0}$ is a probability mass function [density function on D] iff $\sum_{x \in D} f(x) = 1$ [iff $\sum_{x \in D} f(x) = |D|$]

- $x \sim f$ means that x is drawn w.r.t. probability distribution associated with f defined as:
  - $\Pr_{x \sim f}(x = y) = f(y)$
  - $[\ \Pr_{x \sim f}(x = y) = \frac{f(y)}{|D|}\ ]$

Fourier weights

Definition

Let f be n-ary real-valued function and $0 \leq k \leq n$

- $f_{=k} \coloneqq \sum_{S\subseteq[n],|S|=k} \hat{f}(S)\chi_S$ is the degree k part o f
- $||f_{=k}||_2^2 = \sum_{S\subseteq[n],|S|=k} \hat{f}(S)^2$ is the Fourier weight of f at degree k
- $f_{\leq k} \coloneqq \sum_{S\subseteq[n],|S|\leq k} \hat{f}(S)\chi_S$
- $||f_{\leq k}||_2^2 = \sum_{S\subseteq[n],|S|\leq k} \hat{f}(S)^2$ weight of f in degree $\leq k$

Fact

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

$$E_{y\sim\phi}(g(y)) = \langle \phi, g\rangle = E_{x\sim\{1,-1\}^n}(\phi(x)g(x))$$

Density for uniform distribution

Definition

For $\emptyset \neq A \subseteq \{1,-1\}^n$, function ϕ_A is the density function associated with the uniform distribution on A, i.e.:

$$\phi_A(a) = \frac{2^n}{|A|} \quad \text{if } x \in A \text{ else } \phi_A(a) = 0$$

We write $y \sim A$ instead of $y \sim \phi_A$

Example

Every Fourier coefficient of $\phi_{\{(1,1,\dots,1)\}}$ is 1 as for any $x \in \{1,-1\}^n$:

- $\widehat{\delta_x}(S) = \langle \chi_S, \delta_x \rangle = E_{y\sim\{1,-1\}^n}\left(y^S, \delta_x(y)\right) = \frac{x^S}{2^n}$
- So $2^n \delta_x = \sum_{S\subseteq[n]} x^S \chi_S$
- in particular for $x = (1,1,\dots,1)$

$$\phi_{\{(1,1,\dots,1)\}} = 2^n \delta_{(1,1,\dots,1)} = \sum_{S\subseteq[n]} 1 \cdot \chi_S$$

Convolution

Definition

Let $f, g: \{1, -1\}^n \to \mathbb{R}$ be n-ary real-valued functions.

Their convolution $f * g: \{1, -1\}^n \to \mathbb{R}$ is

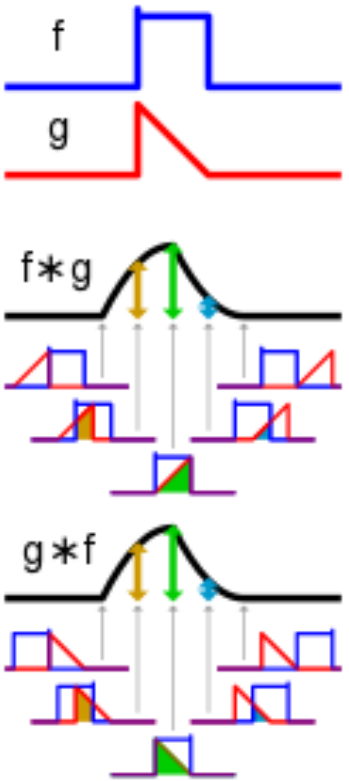
$$(f * g)(x) = E_{y \sim \{1,-1\}^n}\big(f(y)g(x \circ y)\big)$$
$$= E_{y \sim \{1,-1\}^n}\big(f(x \circ y)g(x)\big)$$

where $\circ$ is bitwise multiplication

Proposition

The convolution operator is associative and commutative:

$$f * (g * h) = (f * g) * h; \quad f * g = g * f$$



Convolution in classical Furier analysis

Convolution and densities

Proposition

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

1. $\phi * g(x) = E_{y\sim\phi}\big(g(x \circ y)\big)$
   In particular $E_{y\sim\phi}\big(g(y)\big) = \phi * g((1,1,\dots,1))$
2. If $g = \psi$ is itself a probability density then so is $\phi * \psi$
   It represents the distribution on $x \in \{1,-1\}^n$ by
   - choosing $y \sim \phi$ and $z \sim \psi$ independently and
   - setting $x = y \circ z$.

Note:

- if we use the encoding $\mathbb{F}_2$ then $\circ$ bcomes addition. $x = y + z$.
- So convolution gives a means to calculate the probability of a sum $x = y + z$ of two random variables y, z

Main theorem on convolution

Theorem

Let $f, g: \{1,-1\}^n \to \mathbb{R}$ be n-ary real-valued functions.. Then for all $S \subseteq [n]$:

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

Proof

- $\widehat{f * g}(S)$
  $= E_{x\sim\{1,-1\}^n}\big((f * g)(x)\chi_S(x)\big)$ (Fourier formula)
  $= E_{x\sim\{1,-1\}^n}\big(\, E_{y\sim\{1,-1\}^n}(\, f(y)g(y \circ x)\,)\chi_S(x)\,\big)$ (by definition)
  $= E_{\substack{x,z\sim\{1,-1\}^n \\ independently}}\,(\, f(y)g(z)\,)\chi_S(y \circ z)$ (as $x \circ y$ is uniform on $\{1,-1\}^n$ for all x)
  $= E_{x,z\sim\{1,-1\}^n}(f(y)\chi_S(y)g(z)\chi_S(z))$ (because $\chi_S(x \circ y) = \chi_S(x)\chi_S(y)$)
  $= \hat{f}(S)\hat{g}(S)$ (Fourier formula and independence)

HIGHLIGHT APPLICATION: BLR TEST

Almost linear functions and the BLR test

Definition

A Boolean function $f: \{1,-1\}^n \to \{1,-1\}$ is called linear if either of the following equivalent conditions hold:

1. $\forall x,y \in \{1,-1\}^n: f(x \circ y) = f(x)f(y)$
2. There is a set $S \subseteq [n]$ such that: $f = \chi_S$

Proof of equivalence

- 2. → 1.: $f(x)f(y) = x^S y^S = (x \circ y)^S = f(x \circ y)$
- 1. → 2.: $e^i \in \{1,-1\}^n$ defined by: $e^i_j = -1$ if $i = j$ else $e^i_j = 1$.
  $S := \{i \mid f(e^i) = -1\}$. Then for all $x \in \{1,-1\}^n$:

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

(Note: $\prod$ stands for bitwise multiplication)

Linearity

- If we use the encoding wih $\mathbb{F}_2$ then the two conditions become

1. f$(x + y) = f(x) + f(y)$
   (and this already entails $f(0) = 0$ and $f(\lambda x) = \lambda f(x)$ for all $x \in \mathbb{F}_2$)
2. $f(x) = \sum_{i \in S} x_i$, i.e., there is $a \in \mathbb{F}_2$ such that
   $$f(x) = a \cdot x$$

- So the equivalence amounts to the fact (known from linear algebra) that linear functions are representable as matrix multiplication.

Robust linearity

Does the equivalence hold in a robust version?

(1) f$(x \circ y) = f(x)f(y)$ for almost all $x \in \{1,-1\}^n$

(2) There is a set $S \subseteq [n]$ such that: $f(x) = \chi_S(x)$ for almost all $x \in \{1,-1\}^n$

- The proof for 2 -> 1 is robust: translates directly to proof (2)-> (1).
- The part (1)-> (2) is not. Needs a theorem $\Longrightarrow$ BLR Test (Blum, Luby, Rubinfeld 93)

Property testing

- For large data want to test properties approximately
  $\Longrightarrow$ Field of Property testing. (Oded Goldreich 17)
  - f is black box
    - Can query on any input bit vector of your choosing
    - Want to verify some property with few queries accurately (error less than ϵ)
- In particualr: f is ϵ-close to being linear if for some truly linear $g(x) = \chi_S$

Property testing

- For large data want to test properties aproximately
  $\Longrightarrow$ Field of Property testing. (Oded Goldreich 17)
  - f is black box
  - Can query on any input bit vector of your choosing
  - Want to verify some property with few queries accurately (error less than ϵ)

Definition

g, $f: \{1,-1\}^n \to \{1,-1\}$ are ϵ-close iff $dist(f,g) \leq \epsilon$.
For a property P (i.e. a subset) of Boolean functions let
$dist(f,P) = \min_{g \in P}(dist(f,g))$. f is ϵ-close to P iff $dist(f,P) \leq \epsilon$.

- In particular: f is ϵ-close to being linear if for some truly linear $g(x) = \chi_S$

BLR Test

- BLR Test shows that indeed (1)-> (2) holds.

Algorithm (BLR Test)

Given query access to $f: \{1,-1\}^n \to \{1,-1\}$:

- Choose $x \sim \{1,-1\}^n$ and $y \sim \{1,-1\}^n$ independently
- Query f at x, y and $x \circ y$
- Accept if $f(x)f(y) = f(x \circ y)$

Theorem

Suppose the BLR Test accepts $f: \{1,-1\} \to \{1,-1\}$ with probability $1-\epsilon$. Then f is ϵ- close to being linear

BLR Test

Proof

- We define a characteristic:

$$\frac{1}{2} + \frac{1}{2}\mathrm{f(x)f(y)f(x} \circ y) = \ 1 \text{ if } f\,(x)f(y) = f(x \circ y)$$

$$= \ 0 \text{ if } f\,(x)f(y) \neq f(x \circ y)$$

- $1 - \epsilon = \Pr(\mathrm{BLR}\ accepts\ f) = E_{x,y}\left(\frac{1}{2} + \frac{1}{2}\mathrm{f(x)f(y)f(x} \circ y)\right)$
  - $= \frac{1}{2} + \frac{1}{2}E_x\left(f(x) \cdot E_y(f(y)f(x \circ y))\right)$
  - $= \frac{1}{2} + \frac{1}{2}E_x(f(x) \cdot (f\ * f)(x))$ (by definition)
  - $= \frac{1}{2} + \frac{1}{2}\sum_{s\subseteq[n]} \hat{f}(S)\ \widehat{f * f}(S)$ (Plancherel)
  - $= \frac{1}{2} + \frac{1}{2}\sum_{s\subseteq[n]} \hat{f}(S)^3$ (main theorem on convolution)

BLR Test

Proof (continued)

- Rearranging the equality gives
  - $1 - 2\epsilon = \sum_{s \subseteq [n]} \hat{f}(S)^3$
  - $\leq \max_{S \subseteq [n]} \{\hat{f}(S)\} \cdot \sum_{s \subseteq [n]} \hat{f}(S)^2$
  - $= \max_{S \subseteq [n]} \{\hat{f}(S)\}$ (by Parseval)
- But $\hat{f}(S) = \langle f, \chi_S \rangle = 1 - 2dist(f, \chi_S)$
- Hence there exists some $S^* \subseteq [n]$ such that $1 - 2\,\epsilon \leq 1 - 2\,dist(f, \chi_{S^*})$
- That is, f is ϵ −close to the linear function χ_{S^*}

Uhhh, a lecture with a hopefully useful

APPENDIX

References

- (Oded Goldreich 17)
  Oded Goldreich: Introduction to property testing, freely available draft of a book, 2017.
  https://www.wisdom.weizmann.ac.il/~oded/PDF/pt-v3.pdf
- (Blum, Luby, Rubinfeld 93)
  M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. Journal of Computer and System Sciences, 47(3):549–595, 1993.

Color Convention in this course

- Formulae, when occurring inline
- Newly introduced terminology and definitions
- Important results (observations, theorems) as well as emphasizing some aspects
- Examples are given with standard orange with possibly light orange frame
- Comments and notes
- Algorithms