LECTURE 4
SELF-REDUCIBILITY OF THE DECODING PROBLEM

Summer School: *Introduction to Quantum-Safe Cryptography*

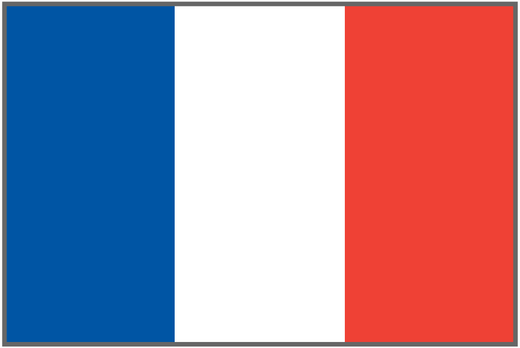
Thomas Debris-Alazard

July 04, 2024

Inria, École Polytechnique

HAPPY INDEPENDENCE DAY!

BUT DON'T FORGET...



Aim of Any Code-Based Cryptosystem:

Security relies on the hardness of the Decoding Problem (DP)

How to trust DP *hardness?*

- Test of time (designing & studying algorithms solving the decoding problem)
- Reduction: prove that decoding is harder than another hard problem

$\longrightarrow$ We will focus on reductions

- A Quick Recap: Decoding Random Codes, an Average Case
- Worst-to-Average-Case Reduction: Framework
- Smoothing Parameter
- Fourier Transform in the Hamming Cube

THE AVERAGE DECODING PROBLEM

Today: focus on binary codes $\big($for the sake of simplicity$\big)$

Linear Codes: Primal Representation

A linear code $\mathcal{C}$ is a subspace of $\mathbb{F}_2^n$

Basis/Generator matrix representation: rows of $\mathbf{A} \in \mathbb{F}_2^{k \times n}$ form a basis,

$$\mathcal{C} = \left\{ \mathbf{s}\mathbf{A} : \ \mathbf{s} \in \mathbb{F}_2^k \right\}$$

The vector/matrix multiplication **sA** is the collection of inner-products

$$\langle \mathbf{s}, \mathbf{a}_1 \rangle, \ldots, \langle \mathbf{s}, \mathbf{a}_n \rangle \ \text{ where } \ \mathbf{a}_i \text{ column of } \mathbf{A} \ \text{ and } \ \langle \mathbf{x}, \mathbf{y} \rangle \stackrel{\text{def}}{=} \textstyle\sum_{i=1}^n x_i y_i \in \mathbb{F}_2$$

Hamming Weight:

$$\forall \mathbf{x} \in \mathbb{F}_2^n, \quad |\mathbf{x}| \stackrel{\text{def}}{=} \left\{ i \in [1, n] : \ x_i \neq 0 \right\}$$

BERNOULLI RANDOM VARIABLE

- $\mathbf{e} \leftarrow \mathrm{Ber}(p)^{\otimes n}$: the e_i's are independent and $\mathbb{P}(e_i = x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}$

Chernoff's Bound: $\mathrm{Ber}(p)^{\otimes n}$ concentrates over words of Hamming weight $\approx np$

Given $\mathbf{e} \leftarrow \mathrm{Ber}(p)^{\otimes n}$,

$$\mathbb{E}\left(|\mathbf{e}|\right) = np \quad \text{and} \quad \mathbb{P}\left(\Big||\mathbf{e}| - np\Big| \geq \varepsilon n\right) \leq 2\, e^{-\varepsilon n^2}$$

First approximation: $\mathrm{Ber}(p)^{\otimes n}$ is a uniform vector of Hamming weight np

Some slight variation of the decoding problem

DP(n, k, t): Average Decoding Problem

- **Input:** $(\mathbf{A}, \mathbf{sA}+\mathbf{t})$ where $\mathbf{A} \in \mathbb{F}_2^{k\times n}$, $\mathbf{s} \in \mathbb{F}_2^k$ are uniform and $\mathbf{t} \leftarrow \mathrm{Ber}(t/n)^{\otimes n}$
- **Output:** recovering $\mathbf{s}$

Algorithm $\mathcal{A}$ solving DP in time T and probability ε means

- $\mathcal{A}$ runs in time T,
- Given $\mathbf{A}, \mathbf{s}$ uniform and $\mathbf{t} \leftarrow \mathrm{Ber}(p)^{\otimes n}$,

$$\mathbb{P}_{\mathbf{A},\mathbf{s},\mathbf{t}}\left(\mathcal{A}\left(\mathbf{A}, \mathbf{sA}+\mathbf{t}\right) = \mathbf{s}\right) = \varepsilon$$

- Given $(\mathbf{A}, \mathbf{s}) \in \mathbb{F}_2^{k \times n} \times \mathbb{F}_2^k$ uniform and $\mathbf{t} \leftarrow \mathrm{Ber}(p)^{\otimes n}$,

$$\mathbb{P}_{\mathbf{A},\mathbf{s},\mathbf{t}}\left(\mathcal{A}\left(\mathbf{A}, \mathbf{s}\mathbf{A}+\mathbf{t}\right) = \mathbf{s}\right) = \varepsilon$$

Law of Total Probability:

$$\varepsilon = \frac{1}{2^{k \times n}} \sum_{\mathbf{s}_0, \mathbf{A}_0} \sum_{t} \sum_{\mathbf{t}_0 :\ |\mathbf{t}_0| = t} \mathbb{P}\left(\mathcal{A}\left(\mathbf{A}_0, \mathbf{s}_0\mathbf{A}+\mathbf{t}_0\right) = \mathbf{s}_0\right) \underbrace{p^t(1-p)^{n-t}}_{\mathbb{P}_{\mathbf{t}}(\mathbf{t}=\mathbf{t}_0)}$$

$\longrightarrow \varepsilon$: average success probability of $\mathcal{A}$ over all possible inputs

ε small $\Longrightarrow \mathcal{A}$ fails for almost all instances

Assumption in Code-Based Cryptography:

DP is hard, *i.e.*, for any algorithm, T/ε is large

To Ensure Hardness of DP (Average Hardness):

1. Test of time (designing & studying algorithms solving DP)

2. Reductions: solving the decoding problem on average implies an algorithm which

   (*i*) computes (quantumly) short vectors in the dual code

   (*ii*) solves all instances of another decoding problem (worst-case)

TEST OF TIME, WHAT ELSE?

To Ensure Hardness of DP (Average Hardness):

1. Test of time (designing & studying algorithms solving DP)

2. Reductions: solving the decoding problem on average implies an algorithm which

   (*i*) computes (quantumly) short vectors in the dual code

   (*ii*) solves all instances of another decoding problem (worst-case)

WORST-TO-AVERAGE CASE REDUCTION

Given a fixed instance

$(\mathbf{G}, \mathbf{x}\mathbf{G} + \mathbf{r})$ where Hamming weight of $\mathbf{r}$ is w

we want to recover $\mathbf{r}$

But, we only have an algorithm $\mathcal{A}$ solving DP with probability ε

$$\mathbb{P}_{\mathbf{A},\mathbf{s},\mathbf{t}}\left(\mathcal{A}(\mathbf{A}, \mathbf{s}\mathbf{A} + \mathbf{t}) = \mathbf{t}\right) = \varepsilon$$

Key-idea:

From $(\mathsf{G}, \mathsf{xG} + \mathsf{r})$ build a "uniform decoding" instance being fed to $\mathcal{A}$

1. $\mathbf{e}_i \leftarrow \mathcal{D}$ (distribution)
2. Compute,

$$\langle \mathsf{y}, \mathbf{e}_i \rangle = \langle \mathsf{xG}, \mathbf{e}_i \rangle + \langle \mathsf{r}, \mathbf{e}_i \rangle = \langle \underbrace{\mathsf{x}}_{\text{secret}}, \mathbf{e}_i \mathsf{G}^\top \rangle + \underbrace{\langle \mathsf{r}, \mathbf{e}_i \rangle}_{\text{noise}}$$

Packing Instances Together:

- Build the matrix $\mathbf{A} = (\mathbf{a}_i)$ whose columns are the $\mathbf{e}_i \mathsf{G}^\top$
- Try to decode $(\mathbf{A}, (\langle \mathsf{y}, \mathbf{e}_i \rangle_i)) = (\mathbf{A}, \mathsf{x}\mathbf{A} + \mathbf{t})$ where $\mathbf{t} = (\langle \mathsf{r}, \mathbf{e}_i \rangle)_i$

From the fixed decoding instance $\mathbf{G}, \mathbf{x}\mathbf{G} + \mathbf{r}$, we build

$$\langle \mathbf{y}, \mathbf{e}_i \rangle = \langle \mathbf{x}\mathbf{G}, \mathbf{e} \rangle + \langle \mathbf{r}, \mathbf{e} \rangle = \langle \underbrace{\mathbf{x}}_{\text{secret}}, \mathbf{e}_i \mathbf{G}^\top \rangle + \underbrace{\langle \mathbf{r}, \mathbf{e}_i \rangle}_{\text{noise}}$$

Packing Instances Together:

- Build the matrix $\mathbf{A} = (\mathbf{a}_i)$ whose columns are the $\mathbf{e}_i \mathbf{G}^\top$
- Try to decode $(\mathbf{A}, (\langle \mathbf{y}, \mathbf{e}_i \rangle)_i)) = (\mathbf{A}, \mathbf{x}\mathbf{A} + \mathbf{t})$ where $\mathbf{t} = (\langle \mathbf{r}, \mathbf{e}_i \rangle)_i$

$\longrightarrow$ Feed $(\mathbf{A}, (\langle \mathbf{y}, \mathbf{e}_i \rangle)_i))$ to the average decoding algorithm $\mathcal{A}$. But what happens?

- Columns of $\mathbf{A}$, *i.e.*, $\mathbf{e}_i \mathbf{G}^\top$, are not uniform
- Noise $\langle \mathbf{r}, \mathbf{e}_i \rangle$ and $\mathbf{e}_i \mathbf{G}^\top$ are correlated
- How does $\langle \mathbf{r}, \mathbf{e}_i \rangle$ behave?

Our Goal:

Estimate success probability of $\mathcal{A}$ being fed with the biased instance $(\mathbf{A}, (\langle \mathbf{y}, \mathbf{e}_i \rangle)_i))$

Statistical Distance:

Given two random variables X, Y,

$$\Delta(X, Y) = \Delta(f, g) = \frac{1}{2}\sum_a |\mathbb{P}(X = a) - \mathbb{P}(Y = a)|$$

$\longrightarrow$ It captures the differences between two random variables

- **Data processing inequality:** for any function/algorithm h

$$\Delta(h(X), h(Y)) \leq \Delta(X, Y)$$

- For any event $\mathcal{E}$,

$$|\mathbb{P}(X \in \mathcal{E}) - \mathbb{P}(Y \in \mathcal{E})| \leq \Delta(X, Y)$$

If an algorithm succeeds with inputs X *and probability* ε*, then it succeeds given* Y *with probability* $\varepsilon + \Delta(X, Y)$

True average decoding instance

1. We want the following to be small:

$$\alpha \overset{\text{def}}{=} \Delta\Big((\mathbf{e}_j\mathbf{G}^\top, \langle \mathbf{x}, \mathbf{e}_j\mathbf{G}^\top\rangle + \langle \mathbf{r}, \mathbf{e}_j\rangle), (\underbrace{\mathbf{a}}_{\text{uniform}}, \langle \mathbf{x}, \mathbf{a}\rangle + \underbrace{e}_{\text{same distrib as } \langle \mathbf{r}, \mathbf{e}_j\rangle})\Big)$$

2. We feed $\left(\mathbf{e}_j\mathbf{G}^\top, \langle \mathbf{x}, \mathbf{e}_j\mathbf{G}^\top\rangle + \langle \mathbf{r}, \mathbf{e}_j\rangle\right)$ to the decoding-solver $\mathcal{A}$ with success probability ε

3. If we give n samples to $\mathcal{A}$, it will recover $\mathbf{x}$ with probability $\varepsilon + n\alpha$

Simplification:

Target: $\Delta\left(\mathbf{e}_j\mathbf{G}^\top, \underbrace{\mathbf{a}}_{\text{uniform}}\right)$ small when $\mathbf{G}$ is fixed but $\mathbf{e}_j$ random variable.

Aim: $\Delta\left(\mathbf{e}\mathbf{G}^\top, \underbrace{\mathbf{a}}_{\text{uniform}}\right)$ small

Which object is $\mathbf{e}\mathbf{G}^\top$*?*

Take the code $\mathcal{C} \subseteq \mathbb{F}_2^n$ point of view

$$\mathcal{C} = \left\{\mathbf{c} : \mathbf{c}\mathbf{G}^\top = \mathbf{0}\right\}$$

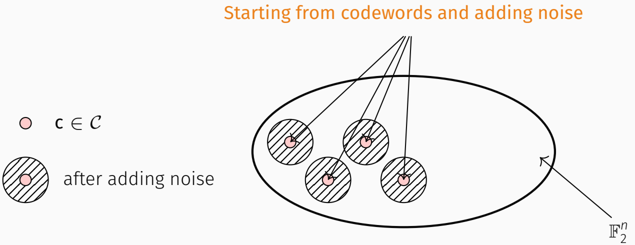
$\longrightarrow$ $\mathbf{e}\mathbf{G}^\top$ *defines a coset of* $\mathcal{C}$

Primal Representation:

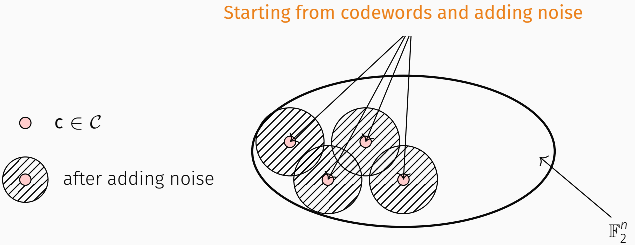
$\mathbf{e}\mathbf{G}^\top$ uniform $\iff$ uniform in $\mathbb{F}_2^n/\mathcal{C}$, *i.e.* uniform modulo $\mathcal{C}$

$\mathbf{e}\mathbf{G}^\top$ uniform for $\mathbf{e} \longleftarrow \mathcal{D}$ $\iff$ $\mathbf{c} + \mathbf{e}$ uniform in $\mathbb{F}_2^n$ where $\mathbf{c} \overset{\text{unif}}{\longleftarrow} \mathcal{C}$ and $\mathbf{e} \longleftarrow \mathcal{D}$

$\mathbf{c} + \mathbf{e}$ uniform in $\mathbb{F}_2^n$ where $\mathbf{c} \xleftarrow{unif} \mathcal{C}$ and $\mathbf{e} \longleftarrow \mathcal{D}$



$\mathbf{c} + \mathbf{e}$ uniform in $\mathbb{F}_2^n$ where $\mathbf{c} \xleftarrow{unif} \mathcal{C}$ and $\mathbf{e} \longleftarrow \mathcal{D}$



$\longrightarrow$ To be uniform: necessary to cover the whole space after adding noise!

$\mathbf{c} + \mathbf{e}$ uniform in $\mathbb{F}_2^n$ where $\mathbf{c} \xleftarrow{unif} \mathcal{C}$ and $\mathbf{e} \longleftarrow \mathcal{D}$

If $\mathbf{e}$ concentrates over words of Hamming weight $\leq t$, it is necessary that

t is such that: $\sharp\mathcal{C} \cdot \binom{n}{t} \geq 2^n$

$\mathbf{c} + \mathbf{e}$ uniform in $\mathbb{F}_2^n$ where $\mathbf{c} \xleftarrow{unif} \mathcal{C}$ and $\mathbf{e} \longleftarrow \mathcal{D}$

If $\mathbf{e}$ concentrates over words of Hamming weight $\leq t$, it is necessary that

t is such that: $\sharp\mathcal{C} \cdot \binom{n}{t} \geq 2^n$

Gilbert-Varshamov Radius of $\mathcal{C}$:

t_{GV}: smallest radius t_0 such that $\sharp\mathcal{C} \cdot \binom{n}{t_0} \geq 2^n$

If one targets $\mathbf{c} + \mathbf{e}$ uniform with $\mathbf{e}$ concentrating over words of Hamming weight t,

then one wants t as small as possible which is t_{GV}

But why?

An algorithm solving the average decoding problem with noise

$$e_i = \langle \mathbf{r}, \mathbf{e}_i \rangle \quad \text{where } \mathbf{e}_i \longleftarrow \mathcal{D}$$

implies an algorithm solving the fixed decoding problem $(\mathbf{G}, \mathbf{xG} + \mathbf{r})$

The average decoding problem with noise

$$e_i = \langle \mathbf{r}, \mathbf{e}_i \rangle \quad \text{where } \mathbf{e}_i \longleftarrow \mathcal{D}$$

is harder than solving the fixed decoding problem $(\mathbf{G}, \mathbf{x}\mathbf{G} + \mathbf{r})$

The average decoding problem with noise

$$e_i = \langle \mathbf{r}, \mathbf{e}_i \rangle \quad \text{where } \mathbf{e}_i \longleftarrow \mathcal{D}$$

is harder than solving the fixed decoding problem $(\mathbf{G}, \mathbf{xG} + \mathbf{r})$

Ideal Situation:

The reduction works with $\mathbb{P}\left(\langle \mathbf{r}, \mathbf{e}_i \rangle = 1\right)$ is small

Because in cryptography we use the assumption that average decoding is hard for a noise e with $\mathbb{P}(e = 1)$ small

$\longrightarrow$ To ensure $\mathbb{P}\left(\langle \mathbf{r}, \mathbf{e}_i \rangle = 1\right)$ is small we need to choose $\mathbf{e}_i$ concentrating over words of small Hamming weight

ABOUT THE NOISE DISTRIBUTION

Our Aim:

To find $\mathbf{e} \longleftarrow \mathcal{D}$ such that $\mathbf{c} + \mathbf{e}$ is close (statistical distance) to uniform when $\mathbf{c} \xleftarrow{unif} \mathcal{C}$

A First Approach:

Choose each bit of $\mathbf{e}$ with probability 1/2, then $\mathbf{c} + \mathbf{e}$ is uniform

But, doing this is useless: $\langle \mathbf{r}, \mathbf{e} \rangle$ will be a uniform noise...

Therefore, impossible to solve $(\mathbf{e}\mathbf{G}^{\top}, \langle \mathbf{x}, \mathbf{e}\mathbf{G}^{\top} \rangle + \underbrace{\langle \mathbf{r}, \mathbf{e} \rangle}_{\text{noise}})$

$\longrightarrow$ We need to carefully choose $\mathbf{e}$!

Given a Linear Code $\mathcal{C} \subseteq \mathbb{F}_2^n$: we want

$\mathbf{c} + \mathbf{e}$ to be uniform where $\mathbf{c} \xleftarrow{unif} \mathcal{C}$ and $\mathbf{e} \leftarrow \mathcal{D}$ (free choice in the reduction)

$\mathcal{S}_t$ be the Hamming-sphere with radius t

If $\mathcal{D}$ concentrates over $\mathcal{S}_t$,

$$\sharp\mathcal{C} \cdot \binom{n}{t} \geq 2^n \iff t \geq t_{\mathsf{GV}}$$

A Lower-Bound on the Amount of Noise:

If noise concentrates on sphere with radius t: necessarily $t \geq t_{\mathsf{GV}}$

Notation:

- unif: uniform distribution of $\mathbb{F}_2^n$
- $1_{\mathcal{C}}$: indicator function of $\mathcal{C}$
- Convolution, $f \star g(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{\mathbf{y} \in \mathbb{F}_2^n} f(\mathbf{y}) g(\mathbf{x}-\mathbf{y})$

If $\mathbf{X} \leftarrow f$ and $\mathbf{Y} \leftarrow g$ are independent, then $\mathbf{X}+\mathbf{Y} \leftarrow f \star g$

Smoothing Parameter:

If f_t concentrates over words of weight t. Smoothing parameter is the smallest t such that,

$$\Delta\left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t, \text{unif}\right) = \frac{1}{2} \sum_{\mathbf{x} \in \mathbb{F}_2^n} \left| \frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t(\mathbf{x}) - \text{unif}(\mathbf{x}) \right| \quad \text{is negligible}$$

Our Dream:

$\Delta\left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t, \text{unif}\right)$ is negligible as soon as $t = t_{\text{GV}}(1+o(1))$, We want: $\frac{1_{\mathcal{C}}}{\sharp \mathcal{C}} \star f_t$ close to uniform

So, $\mathsf{x} \mapsto \left| \frac{1_{\mathcal{C}}}{\sharp \mathcal{C}} \star f_t(\mathsf{x}) - \mathrm{unif}(\mathsf{x}) \right|$ will be roughly constant!

Any idea to upper-bound tightly $\sum_{\mathsf{x} \in \mathbb{F}_2^n} \left| \frac{1_{\mathcal{C}}}{\sharp \mathcal{C}} \star f_t(\mathsf{x}) - \mathrm{unif}(\mathsf{x}) \right|$?

We want: $\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t$ close to uniform

So, $\mathbf{x} \mapsto \left| \frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t(\mathbf{x}) - \mathrm{unif}(\mathbf{x}) \right|$ will be roughly constant!

Any idea to upper-bound tightly $\sum_{\mathbf{x}\in\mathbb{F}_2^n} \left| \frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t(\mathbf{x}) - \mathrm{unif}(\mathbf{x}) \right|$?

A Good Idea: Cauchy-Schwarz

$$\sum_{\mathbf{x}\in\mathbb{F}_2^n} \left| \frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t(\mathbf{x}) - \mathrm{unif}(\mathbf{x}) \right| \leq \sqrt{2^n} \sqrt{\sum_{\mathbf{x}\in\mathbb{F}_2^n} \left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t(\mathbf{x}) - \mathrm{unif}(\mathbf{x}) \right)^2}$$

$\longrightarrow$ The upper-bound: L_2-distance!

A natural approach: Parseval's identity via Fourier Theory

FOURIER TRANSFORM IN THE HAMMING CUBE

Fourier Transform (informal):

It decomposes a function in the Fourier basis

But how is defined the Fourier basis?

Fourier Transform (informal):

It decomposes a function in the Fourier basis

But how is defined the Fourier basis?

$\longrightarrow$ Basis that diagonalizes (per-block in non-abelian case) translation operators!

Hamming Cube Case:

Given the translation operator $R(\mathsf{t})$ for functions $f : \mathbb{F}_2^n \longrightarrow \mathbb{C}$,

$$R(\mathsf{t}) : f \longmapsto \left(g : \mathsf{x} \in \mathbb{F}_2^n \longmapsto g(\mathsf{x} + \mathsf{t})\right)$$

It is diagonal in the character basis $\left(\chi_{\mathsf{y}} : \mathsf{x} \longmapsto (-1)^{\langle \mathsf{x}, \mathsf{y} \rangle}\right)$,

$$R(\mathsf{t})\,(\chi_{\mathsf{y}}) = (-1)^{\langle \mathsf{y}, \mathsf{t} \rangle} \cdot \chi_{\mathsf{y}}$$

- Scalar product and associated norms:

$$\langle f, g \rangle \stackrel{\text{def}}{=} \frac{1}{2^n} \sum_{\mathbf{y} \in \mathbb{F}_2^n} f(\mathbf{y}) g(\mathbf{y}) \quad \text{and} \quad ||f||_2 \stackrel{\text{def}}{=} \sqrt{\langle f, f \rangle}$$

- An orthonormal basis, characters:

$$\chi_{\mathbf{x}}(\mathbf{y}) \stackrel{\text{def}}{=} (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

Fourier Transform:

Given $f : \mathbb{F}_2 \rightarrow \mathbb{C}$,

$$\widehat{f}(\mathbf{x}) = \frac{1}{\sqrt{2^n}} \sum_{\mathbf{y} \in \mathbb{F}_2^n} f(\mathbf{y}) \chi_{\mathbf{x}}(\mathbf{y}) = \sqrt{2^n} \, \langle f, \chi_{\mathbf{x}} \rangle$$

- Convolution:

$$\widehat{f \star g} = \sqrt{2^n} \, \widehat{f} \cdot \widehat{g}$$

Parseval Identity: Fourier Transform Isometry for L_2

$$\|f - g\|_2 = \|\widehat{f} - \widehat{g}\|_2$$

Proof.

Given any function $h : \mathbb{F}_2^n \longrightarrow \mathbb{C}$, as $(\chi_{\mathrm{x}})_{\mathrm{x} \in \mathbb{F}_2^n}$ is an orthonormal basis,

$$h = \sum_{\mathrm{x} \in \mathbb{F}_2^n} \langle h, \chi_{\mathrm{x}} \rangle \cdot \chi_{\mathrm{x}} \quad \text{and} \quad \|h\|_2^2 = \sum_{\mathrm{x} \in \mathbb{F}_2^n} \left| \langle h, \chi_{\mathrm{x}} \rangle \right|^2 = \frac{1}{2^n} \sum_{\mathrm{x} \in \mathbb{F}_2^n} \left| \widehat{h}(\mathrm{x}) \right|^2 = \|\widehat{h}\|_2^2$$

□

$\longrightarrow$ For our purpose: we need to compute $\widehat{1_{\mathcal{C}}}$

Dual Code:

Given $\mathcal{C} \subseteq \mathbb{F}_2^n$,

$$\mathcal{C}^{\perp} \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{F}_2^n : \forall \mathbf{y} \in \mathbb{F}_2^n, \sum_{i=1}^{n} x_i y_i = 0 \right\} = \left\{ \mathbf{x} \in \mathbb{F}_2^n : \forall \mathbf{y} \in \mathcal{C}, \chi_{\mathbf{x}}(\mathbf{y}) = 1 \right\}$$

Fourier Transform of the Code Indicator:

$$\widehat{1_{\mathcal{C}}} = \frac{\sharp \mathcal{C}}{\sqrt{2^n}} \, 1_{\mathcal{C}^{\perp}}$$

$\longrightarrow$ This result is known as “Poisson summation” formula!

$\longrightarrow$ We also need to compute $\widehat{\text{unif}}$ where $\text{unif}(\mathbf{x}) = \frac{1}{2^n}$ for any $\mathbf{x} \in \mathbb{F}_2^n$

Fourier Transform of the Uniform Function:

$$\widehat{\text{unif}} = \frac{1}{\sqrt{2^n}} \cdot \delta_0 \quad \text{where} \quad \delta_0(\mathbf{x}) = 0 \text{ if } \mathbf{x} \neq \mathbf{0} \text{ and } 1 \text{ otherwise } \Big(\text{Kronecker delta}\Big)$$

Proof.

$$\sqrt{2^n} \cdot \widehat{\text{unif}}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathbb{F}_2^n} \text{unif}(\mathbf{y}) \chi_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{y} \in \mathbb{F}_2^n} \frac{(-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}}{2^n}$$

But,

$$\sum_{\mathbf{y} \in \mathbb{F}_2^n} (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle} = 0 \text{ when } \mathbf{x} \neq \mathbf{0}.$$

Indeed, when $\mathbf{x} \neq \mathbf{0}$, it exists $\mathbf{z} \neq \mathbf{0}$ such that $\langle \mathbf{x}, \mathbf{z} \rangle \neq 0 \bmod 2$ and

$$\sum_{\mathbf{y} \in \mathbb{F}_2^n} (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle} = \sum_{\mathbf{y} \in \mathbb{F}_2^n} (-1)^{\langle \mathbf{x}, (\mathbf{y}+\mathbf{z}) \rangle} = (-1)^{\langle \mathbf{x}, \mathbf{z} \rangle} \sum_{\mathbf{y} \in \mathbb{F}_2^n} (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

As $(-1)^{\langle \mathbf{x}, \mathbf{z} \rangle} \neq 1$, the above equality is only possible if $\sum_{\mathbf{y} \in \mathbb{F}_2^n} (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle} = 0$. □

$$\Delta\left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t, \text{unif}\right) \leq \sqrt{2^n}\ \left\|\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t - \text{unif}\right\|_2 = \sqrt{2^n}\ \left\|\frac{\sqrt{2^n}}{\sharp\mathcal{C}}\ \widehat{1_{\mathcal{C}}} \cdot \widehat{f_t} - \widehat{\text{unif}}\right\|_2$$

$$= \sqrt{2^n}\ \left\|\frac{\sqrt{2^n}}{\sqrt{2^n}\cdot\sharp\mathcal{C}} \cdot \sharp\mathcal{C} \cdot 1_{\mathcal{C}^\perp} \cdot \widehat{f_t} - \frac{1}{\sqrt{2^n}}\delta_0\right\|_2$$

$$= \sqrt{2^n}\ \sqrt{\sum_{\mathbf{c}^\perp \in \mathcal{C}^\perp \setminus \{\mathbf{0}\}} |\widehat{f_t}(\mathbf{x})|^2}$$

Upper-Bound:

$$\Delta\left(\tfrac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t, \text{unif}\right) \leq \sqrt{2^n}\ \sqrt{\sum_{\mathbf{c}^\perp \in \mathcal{C}^\perp \setminus \{\mathbf{0}\}} |\widehat{f_t}(\mathbf{x})|^2}$$

If $f_t(x)$ depends only on $|\mathbf{x}|$ (radial),

$$\Delta\left(\tfrac{1_{\mathcal{C}}}{\sharp\mathcal{C}} \star f_t, \text{unif}\right) \leq \sqrt{2^n}\ \sqrt{\sum_{a>0} N_a(\mathcal{C}^\perp)\ |\widehat{f_t}(a)|^2}$$

where,

$$N_a(\mathcal{C}^\perp) \stackrel{\text{def}}{=} \sharp\left\{\mathbf{c}^\perp \in \mathcal{C}^\perp\ :\ |\mathbf{c}^\perp| = a\right\}$$

We need to upper-bound $N_a\left(\mathcal{C}^{\perp}\right)$*, but how?*

We need to upper-bound $N_a\left(\mathcal{C}^\perp\right)$, *but how?*

$\longrightarrow$ To understand first if our approach is meaningful, use random codes of fixed size!

$$\mathbb{E}_{\mathcal{C}^\perp}\left(\Delta\left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}}\star f_t, \mathrm{unif}\right)\right) \leq \mathbb{E}_{\mathcal{C}^\perp}\left(\sqrt{2^n}\sqrt{\sum_{a>0} N_a(\mathcal{C}^\perp)\ |\widehat{f_t}(a)|^2}\right)$$
$$\leq \sqrt{2^n}\sqrt{\sum_{a>0}\mathbb{E}_{\mathcal{C}^\perp}\left(N_a(\mathcal{C}^\perp)\ |\widehat{f_t}(a)|^2\right)} \quad \text{(Jensen's Inequality)}$$
$$= \sqrt{2^n}\sqrt{\sum_{a>0}\frac{\binom{n}{a}}{\sharp\mathcal{C}}\ |\widehat{f}(t)|^2}$$

Bernoulli: our dream comes false

Choosing $f(\mathbf{x}) = p^{|\mathbf{x}|}(1-p)^{n-|\mathbf{x}|}$ concentrating over words of Hamming weight pn with random codes $\mathcal{C}$ of dimension k leads to:

$$np \geq \frac{n}{2}\left(1-\sqrt{2^{k/n}-1}\right)$$

To ensure $\mathbb{E}_{\mathcal{C}^\perp}\left(\Delta\left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}}\star f, \mathrm{unif}\right)\right)$ negligible while

$$\frac{n}{2}\left(1-\sqrt{2^{k/n}-1}\right) \gg t_{\mathrm{GV}}$$

Using Bernoulli seems to be non-optimal. Which other distribution concentrating over $\mathcal{S}_{pn}$ could be chosen?

Using Bernoulli seems to be non-optimal. Which other distribution concentrating over $\mathcal{S}_{pn}$ could be chosen?

$\longrightarrow 1_{\mathcal{S}_t} / \binom{n}{t}$ be the uniform distribution over $\mathcal{S}_t$

CONCLUSION

Using $f = \frac{1_{\mathcal{S}_t}}{\binom{n}{t}}$,

$$\mathbb{E}_{\mathcal{C}^\perp}\left(\Delta\left(\frac{2^n}{\sharp\mathcal{C}}1_{\mathcal{C}} \star f, \text{unif}\right)\right) \leq \sqrt{\frac{2^n}{\sharp\mathcal{C}\cdot\binom{n}{t}}}$$

$\longrightarrow$ Our dream comes true: $t \geq t_{\text{GV}}$ to ensure a negligible statistical distance

But our bound only holds on average, not for a fixed code $\mathcal{C}$ . . .

To get our upper-bound we used: $\mathbb{E}_{\mathcal{C}^\perp}\left(\sharp\left\{\mathbf{c}^\perp \in \mathcal{C}^\perp : \ |\mathbf{c}^\perp| = a\right\}\right) = \frac{\binom{n}{a}}{\sharp\mathcal{C}}$

$\longrightarrow$ What happens for a fixed code, as aimed in the reduction?

We use

Linear Programming Bounds from Delsarte's Theory $\big($Association Schemes, $\ldots\big)$:

$$N_a\left(\mathcal{C}^\perp\right) \leq F(d, a)$$

where d minimum distance of $\mathcal{C}^\perp$