FROM CODE-BASED CRYPTOGRAPHY TO PACKING BOUNDS

wcc 2024

André Chailloux, Thomas Debris-Alazard, Léo Ducas, Nicolas Resch and Jean-Pierre Tillich

June 17, 2024

Inria, École Polytechnique

Code-Based Cryptography:

Building cryptographic primitives whose security relies on

hardness of decoding a random code

But how to ensure the hardness of decoding a random code?

- Test of time,
- **Reduction:** prove that decoding is harder than another hard problem.

$\longrightarrow$ We will focus on reductions

ROADMAP

1. Decoding Random Codes: an Average Case
2. Worst-to-Average-Case Reduction: Framework
3. Smoothing Parameter
4. Packing Bounds

THE AVERAGE DECODING PROBLEM

LINEAR CODES AND HAMMING WEIGHT

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,

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

The vector/matrix multiplication $\mathbf{sA}$ is the collection of inner-products

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

Hamming Weight:

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

BERNOULLI RANDOM VARIABLE

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

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

AVERAGE DECODING PROBLEM

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

- **Input:** $(\mathbf{A}, \mathbf{s}\mathbf{A}+\mathbf{t})$ 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 A-DP in time T and probability ε means

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

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

YOU SAID AVERAGE CASE?

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

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

Law of total probability:

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

$\longrightarrow \varepsilon$: **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:

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

TEST OF TIME, WHAT ELSE?

To ensure hardness of decoding a random code (average hardness):

1. Test of time,
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

OUR GOAL

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 A-DP with probability ε

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

Key-idea:

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

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

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

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 $\mathbf{t} = (\langle \mathbf{r}, \mathbf{e}_i \rangle)_i$

THE ISSUE

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

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

Packing Instances Together:

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

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

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

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

$$\Delta(h(\mathsf{X}), h(\mathsf{Y})) \leq \Delta(\mathsf{X}, \mathsf{Y})$$

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

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

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

True average decoding instance

1. We want the following to be small:

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

2. We feed $(\mathbf{e}_i\mathbf{G}^\top, \langle \mathbf{x}, \mathbf{e}_i\mathbf{G}^\top\rangle + \langle \mathbf{r}, \mathbf{e}_i\rangle)$ to the decoding-solver $\mathcal{A}$ with succ prob. ε

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

Simplification:

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

A GEOMETRICAL INTERPRETATION: PRIMAL REPRESENTATION

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* $\mathcal{C}$

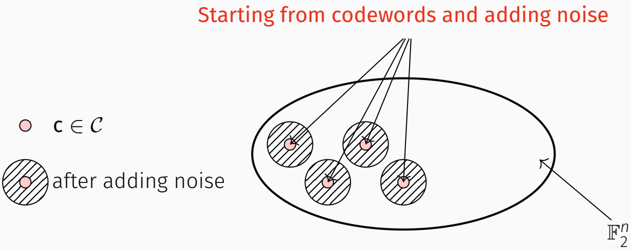
Primal representation:

$\mathbf{e}\mathbf{G}^\top$ 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 $\mathbf{e} \longleftarrow \mathcal{D}$

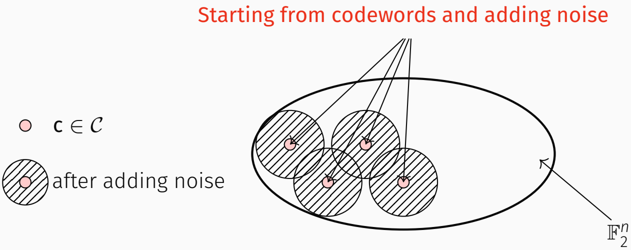
POTATOES

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



POTATOES

$\mathbf{c} + \mathbf{e}$ 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!

COMBINATORICS POINT OF VIEW: GILBERT-VARSHAMOV RADIUS

$\mathbf{c} + \mathbf{e}$ 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: $\sharp\mathcal{C} \cdot \binom{n}{t} \geq 2^n$

COMBINATORICS POINT OF VIEW: GILBERT-VARSHAMOV RADIUS

$\mathbf{c} + \mathbf{e}$ 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?

THE REDUCTION IN A NUTSHELL

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 $(\mathbf{G}, \mathbf{x}\mathbf{G} + \mathbf{r})$

THE REDUCTION IN A NUTSHELL

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 $(\mathbf{G}, \mathbf{xG} + \mathbf{r})$

THE REDUCTION IN A NUTSHELL

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 $(\mathbf{G}, \mathbf{x}\mathbf{G} + \mathbf{r})$

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

THE NOISE: OUR BEST FRIEND TO UNIFORMIZE

Our aim:

To find $\mathbf{e} \longleftarrow \mathcal{D}$ such that $\mathbf{c} + \mathbf{e}$ is close (stat. 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 $\mathbf{e}$!

OUR GOAL

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 $t \geq t_{\mathsf{GV}}$

SOME NOTATION

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 s.t,

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

Our Dream:

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

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

A Good Idea: Cauchy-Schwarz

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

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

A natural approach: Parseval's identity

FOURIER TRANSFORM IN THE HAMMING CUBE

- 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 $f : \mathbb{F}_2 \to \mathbb{C}$,

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

- Convolution:

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

Parseval Identity: Fourier Transform Isometry for L_2

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

$\longrightarrow$ We need to compute $\widehat{1_{\mathcal{C}}}$

Dual Code:

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

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

Fourier Transform of the Code Indicator:

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

SMOOTHING PARAMETER: AN UPPER-BOUND

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

$$= \sqrt{2^n}\,\left\|\frac{\sqrt{2^n}}{\sqrt{2^n}\cdot\sharp\mathcal{C}}\cdot\sharp\mathcal{C}\cdot 1_{\mathcal{C}^\perp}\cdot\widehat{f_t} - \frac{1}{\sqrt{2^n}}\delta_{\mathbf{0}}\right\|_2$$

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

Upper-Bound:

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

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

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

where,

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

AN OPTIMAL UPPER-BOUND: THE RANDOM CASE

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

AN OPTIMAL UPPER-BOUND: THE RANDOM CASE

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!

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

Bernoulli: our dream comes false

Choosing $f_t(\mathbf{x}) = p^{|\mathbf{x}|}(1-p)^{n-|\mathbf{x}|}$ concentrating over words of Hamming weight $t = pn$ 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 $\mathbb{E}_{\mathcal{C}^{\perp}}\left(\Delta\left(\frac{1_{\mathcal{C}}}{\sharp\mathcal{C}}\star f, \text{unif}\right)\right)$ negligible while

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

UNIFORM DISTRIBUTION OVER A SPHERE

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

UNIFORM DISTRIBUTION OVER A SPHERE

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: $\mathbb{E}_{\mathcal{C}^\perp}\left(\sharp\left\{\mathbf{c}^\perp \in \mathcal{C}^\perp : |\mathbf{c}^\perp| = a\right\}\right) = \frac{\binom{n}{t}}{\sharp\mathcal{C}}$

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

We will use

Linear Programming Bounds:

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

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

PACKING BOUNDS

What we need: to bound $N_a(\mathcal{C})$ when the minimum distance of $\mathcal{C}$ is fixed

Simplification: Packing Bound

We will instead bound $\sharp\mathcal{C}$ as function of its minimum distance

$$\max\left\{\sharp\mathcal{C} : \ \mathcal{C} \subseteq \mathbb{F}_2^n \text{ and minimum distance } d\right\}$$

The most fruitful approach to get the best (asymptotic) packing bounds:

theory of association schemes

Metric Space and Adjacency Matrix:

(X, τ, n) be a finite metric space with $\tau : X \times X \to \{0, \dots, n\}$.

It associated adjacency matrices $\mathbf{D}_i \in \mathbb{C}^{|X| \times |X|}$ are,

$$\forall x, y \in X, \quad \mathbf{D}_i(x,y) \stackrel{\text{def}}{=} \begin{cases} 1 \text{ if } \tau(x,y) = i \\ 0 \text{ otherwise} \end{cases}$$

- Typical cases: $X = \mathbb{F}_2^n$ **Hamming scheme** or $X = \mathcal{S}_t$ (Hamming sphere with radius t) **Johnson scheme** for the Hamming distance

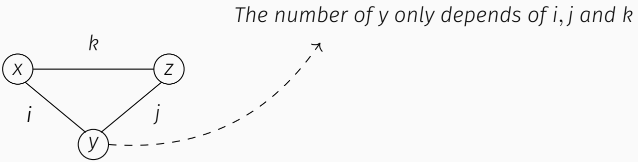
ASSOCIATION SCHEMES: A TRIANGLE CONDITION

Association Scheme:

(X, τ, n) is a (metric) association scheme if there exists an integer $p_{i,j}^k$ s.t,

$$\forall x, z \in X \text{ s.t } \tau(x,z) = k, \quad p_{i,j}^k = \sharp\{y \in X : \tau(x,y) = i \text{ and } \tau(y,z) = j\}$$

and $p_{1,k}^{k+1} \neq 0$.



Crucial Consequences:

- $p_{i,j}^k = p_{j,i}^k$ because τ symmetric as distance
- $\mathbf{D}_i \cdot \mathbf{D}_j = \sum_{k=0}^{n} p_{i,j}^k \cdot \mathbf{D}_k$

$\longrightarrow$ Vect $(\mathbf{D}_i : 0 \leq i \leq n)$ forms a commutative algebra $\subseteq \mathbb{C}^{|X| \times |X|}$

Vect $(\mathbf{D}_i : \; 0 \leq i \leq n)$ forms a **commutative** algebra $\subseteq \mathbb{C}^{|X|\times|X|}$ and the $\mathbf{D}_i$ are symmetric

$\longrightarrow$ The $\mathbf{D}_i$ share common orthogonal eigenspaces!

The matrices E_j:

There exists **orthogonal projectors** $\mathbf{E}_j \in \mathbb{C}^{|X|\times|X|}$ such that,

$$\forall i \in \{0, \ldots, n\}, \quad \mathbf{D}_i = \sum_{j=0}^{n} p_i(j)\mathbf{E}_j$$

$\longrightarrow$ Matrices $\mathbf{D}_i$ and $\mathbf{E}_j$ generate the same space!

q-numbers:

$$\forall j \in \{0, \ldots, n\}, \quad \mathbf{E}_j = \frac{1}{|X|} \sum_{i=0}^{n} q_j(i)\mathbf{D}_i$$

$$\mathsf{D}_i = \sum_{j=0}^{n} p_i(j)\mathsf{E}_j \quad \text{and} \qquad \mathsf{E}_j = \frac{1}{|X|}\sum_{i=0}^{n} q_j(i)\mathsf{D}_i$$

Fourier Transform and Its Inverse: given $f : \{0,\dots,n\} \to \mathbb{C}$

$$\widehat{f}(x) \stackrel{\text{def}}{=} \sum_{y=0}^{n} f(y)p_y(x) \quad \text{and} \quad \widetilde{f}(x) \stackrel{\text{def}}{=} \sum_{y=0}^{n} f(y)q_y(x)$$

$$\mathsf{D}^f \stackrel{\text{def}}{=} \sum_{x=0}^{n} f(x)\mathsf{D}_x \quad ; \quad \mathsf{E}^g \stackrel{\text{def}}{=} \sum_{x=0}^{n} g(x)\mathsf{E}_x$$

From the decomposition of the D_i and E_j in each basis

$$\mathsf{D}^f = \mathsf{E}^{\widehat{f}} \quad \text{and} \quad \mathsf{E}^f = \mathsf{D}^{\widetilde{f}}$$

CODE, DIRAC NOTATION AND WEIGHT DISTRIBUTION

Our aim is to upper-bound the size of a code with minimum distance d

- **Code**: given (X, τ, n) an association scheme, **a code** $\mathcal{C}$ is a subset of X.

$$d \stackrel{\text{def}}{=} \min\left(\tau(c,c') : \; c,c' \in \mathcal{C} \text{ and } c \neq c'\right)$$

Dirac/Bra-ket Notation

$\mathrm{X} = \{x_1, \ldots, x_N\}$. For any x_i, the vector $|x_i\rangle$ is zero except at the ith entry where it is 1.

$$|v\rangle = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} = \sum_{i=1}^{N} v_i \, |x_i\rangle \quad \text{and} \quad \langle v| = \begin{pmatrix} \overline{v_1} & \ldots & \overline{v_N} \end{pmatrix}$$

Given a code $\mathcal{C} \subseteq \mathrm{X}$,

$$|\psi_{\mathcal{C}}\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{\sharp \mathcal{C}}} \sum_{c \in \mathcal{C}} |c\rangle$$

- **Weight Distribution:** $a(t) = \frac{1}{\sharp \mathcal{C}} \cdot \sharp\left((c,c') \in \mathcal{C}^2 : \; \tau(c,c') = t\right) = \langle \psi_{\mathcal{C}} | \, \mathrm{D}_t \, | \psi_{\mathcal{C}} \rangle$

$$\sharp \mathcal{C} = \sum_{t=0}^{n} a(t), \quad a(0) = 1 \quad \text{and} \quad a(t) = 0 \text{ if } t \in \{1, \ldots, d-1\}$$

DUAL DISTRIBUTION AND MACWILLIAMS IDENTITY

$$a(t) = \langle\psi_{\mathcal{C}}|\, \mathbf{D}_t \,|\psi_{\mathcal{C}}\rangle \text{ where } |\psi_{\mathcal{C}}\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{\mathcal{C}}} \sum_{c\in\mathcal{C}} |c\rangle$$

Dual Code Distribution:

$$a'(i) \stackrel{\text{def}}{=} \langle\psi_{\mathcal{C}}|\, \mathbf{E}_i \,|\psi_{\mathcal{C}}\rangle = \langle\psi_{\mathcal{C}}|\, \frac{1}{|\mathrm{X}|} \sum_{t=0}^{n} q_i(t)\mathbf{D}_t \,|\psi_{\mathcal{C}}\rangle = \sum_{t=0}^{n} q_i(t) a'(t)$$

If $\mathcal{C}$ is linear, then

$$a'(i) = \sharp\left((c^{\perp}, d^{\perp}) \in \mathcal{C}^{\perp} :\ \tau(c^{\perp}, d^{\perp}) = i\right)$$

Otherwise, if $\mathcal{C}$ is not linear maybe even not integers... But in any case:

MacWilliams identity:

$$\forall i \in \{0, \dots, n\}, \quad a'(i) \geq 0$$

Proof: the $\mathbf{E}_i$ are projectors, *i.e.*, $\mathbf{E}_i = \sum_t \left|v_t^{(i)}\right\rangle\!\left\langle v_t^{(i)}\right|$ and,

$$a'(i) = \langle\psi_{\mathcal{C}}| \sum_t \left|v_i^{(t)}\right\rangle\!\left\langle v_i^{(t)}\right| |\psi_{\mathcal{C}}\rangle = \sum_t \left\langle \psi_{\mathcal{C}} \middle| v_i^{(t)} \right\rangle \left\langle v_i^{(t)} \middle| \psi_{\mathcal{C}} \right\rangle = \sum_t \left| \left\langle \psi_{\mathcal{C}} \middle| v_i^{(t)} \right\rangle \right|^2 \geq 0$$

Delsarte's Linear Program:

$$\mathrm{DLP}(n,d) \overset{\mathrm{def}}{=} \sum_{t \in [\![0,n]\!]} u(t)$$

$$u(0) = 1$$

$$u(t) = 0 \text{ for } t \in \{1, \ldots, d-1\}$$

$$u(t) \geq 0 \text{ for } t \in \{d, \ldots, n\}$$

$$\sum_{t \in [\![0,n]\!]} u(t) q_i(t) \geq 0 \text{ for } i \in \{0, \ldots, n\}.$$

Given a code $\mathcal{C}$ with minimum distance d, its weight distribution $a(t)$ is a solution of Delsarte's linear program,

$$\sharp\mathcal{C} \leq A(n,d) \leq \mathrm{DLP}(n,d)$$

$\longrightarrow$ MacWilliams identity shows that the weight distribution verifies the last condition of Delsarte's Linear Program

Dual Delsarte Linear Program:

Let $d \in \{0, \ldots, n\}$ and $f : \{0, \ldots, n\} \longrightarrow \mathbb{R}$ be a function such that,

$$\widehat{f} \geq 0 \quad , \quad \widehat{f}(0) > 0 \quad , \quad \forall x \geq d,\ f(x) \leq 0.$$

Then,

$$\max\left\{\sharp\mathcal{C} :\ \mathcal{C} \subseteq \mathbb{F}_2^n \text{ and minimum distance } d\right\} \leq \mathrm{DLP}(n, d) \leq |X| \cdot \frac{f(0)}{\widehat{f}(0)}.$$

Obtained bounds via the choice of a function f with $X = \mathbb{F}_2^n$,

- Plotkin,
- Hamming,
- Elias-Bassalygo,
- MRRW1&2 (McEliece, Rodemich, Rumsay, Welch) best bounds from '77.

CONCLUSION

- Framework for a worst-to-average-case reduction in coding theory: smoothing parameter

  $\longrightarrow$ It reduces to upper-bound the weight distribution of a fixed code

- To derive upper-bounds for the weight distribution of a fixed code: use Delsarte's Linear Program approach as for packing bounds