LECTURE 8
RANDOM LINEAR CODES AND SHANNON'S THEOREM

Information Theory

Thomas Debris-Alazard

Inria, École Polytechnique

COMMUNICATION OVER A NOISY CHANNEL

Lecture 6:

It is possible to communicate Rn bits by sending n bits through a noisy channel Q if and only if

$$R \leq C(Q) \left(\textit{capacity}\right)$$

$\longrightarrow$ The proof relies on the use of block-codes:

we encode a symbol into a block-code which adds redundancy

Issue:

Shannon's proof does not give an efficient algorithm to communicate: even encoding is non-efficient with block-codes

MOTIVATION: LINEAR CODES OFFER AN EFFICIENT ENCODING

Lecture 7:

*We introduced **linear** codes which enable at least an efficient encoding*

Issue:

Linear codes form a **sub-class** of block-codes: we don't know if they reach the capacity of noisy channels

OBJECTIVE

Linear codes reach the capacity of the $q-$ary Symmetric Channel $\Big(q\mathrm{SC}(p)\Big)$, but. . .

- The proof which is a variation of Shannon's proof from Lecture 6 does not exhibit an efficient decoding algorithmm $\Big($a priori the decoding algorithm requires computations with an exponential cost$\Big)$

COURSE OUTLINE

1. Shannon's Theorem for Linear Codes
2. About Random Codes
3. Proof of Shannon's Theorem for Linear Codes
4. Random Codes: a Powerful Tool
5. A Little Bit of Cryptography

SHANNON'S THEOREM

q-ary symmetric $\Big(q\mathrm{SC}(p)\Big)$ channels:

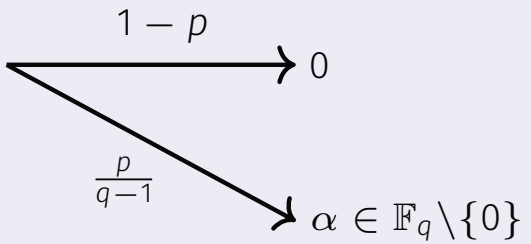
Memoryless channel $\Big(\mathbb{F}_q, \mathbb{F}_q, p(y \mid x)\Big)$ where,

$$\forall x, y \in \mathbb{F}_q, \quad p(y \mid x) = \begin{cases} 1-p & \text{if } x = y \\ \frac{p}{q-1} & \text{otherwise} \end{cases}$$

p: probability of error ; $\frac{p}{q-1}$ transition probability

When sending $\mathbf{c} \in \mathbb{F}_q^n$ through the channel

$$\mathbf{y} = \mathbf{c} + \mathbf{e} \quad \text{where the } e_i \text{ are i.i.d and } p(e_i = x) = \begin{cases} 1-p & \text{if } x = 0 \\ \frac{p}{q-1} & \text{otherwise} \end{cases}$$



Decoder:

A decoder $\mathcal{D}$ is defined as a mapping $\mathcal{D} : \mathbb{F}_q^n \mapsto \mathbb{F}_q^n \cup \{\perp\}$ where $\perp$ is a special symbol used to denote the potential case where the decoder fails to output something

Given a code $\mathcal{C}$, we can associate a decoder $\mathcal{D}$ to form the pair $(\mathcal{C}, \mathcal{D})$. The failure probability of $\mathcal{D}$ relatively to $\mathcal{C}$ is defined as

$$\mathbb{P}_{\text{fail}}(\mathcal{C}, \mathcal{D}) \stackrel{\text{def}}{=} \mathbb{P}(\mathcal{D}(\mathbf{c} + \mathbf{e}) \neq \mathbf{c})$$

where the probability is computed over: (i) the internal randomness of $\mathcal{D}$, (ii) $\mathbf{c}$ be chosen uniformly at random among $\mathcal{C}$ and (iii) $\mathbf{e}$ be the error induced by the qSC(p), *i.e.,*

$$\mathbf{e} = (e_1, \ldots, e_n) \text{ where the } e_i \text{ are i.i.d and } p(e_i = x) = \begin{cases} 1 - p & \text{if } x = 0 \\ \frac{p}{q-1} & \text{otherwise} \end{cases}$$

Maximum likelihood decoder:

Given the qSC(p) and $\mathcal{C} \subseteq \mathbb{F}_q^n$, we call the maximum likelihood decoder the map

$$\mathcal{D}_{\mathrm{ML}} : \mathbb{F}_q^n \longrightarrow \mathcal{C}$$

such that given $\mathbf{y} \in \mathbb{F}_q^n$, it outputs the codeword $\mathbf{c} \in \mathcal{C}$ maximizing the transition probabilities

$$\mathcal{D}_{\mathrm{ML}}(\mathbf{y}) \stackrel{\text{def}}{=} \underset{\mathbf{c}\in\mathcal{C}}{\arg\max}\, \mathbb{P}(\mathbf{c} \mid \mathbf{y}) = \underset{\mathbf{c}\in\mathcal{C}}{\arg\max} \prod_{i=1}^{n} \mathbb{P}(c_i \mid y_i)$$

Proposition (from Lecture 7):

In a qSC(p) with probability of transition $p/(q-1) < 1/q$, if codewords $\mathbf{c} \in \mathcal{C}$ are chosen uniformly at random among $\mathcal{C}$, then

$$\forall \mathbf{y} \in \mathbb{F}_q^n, \quad \mathcal{D}_{\mathrm{ML}}(\mathbf{y}) = \mathbf{c} \in \mathcal{C} \quad \text{such that } \mathbf{c} = \underset{\mathbf{d}\in\mathcal{C}}{\arg\min}\, d_{\mathrm{H}}(\mathbf{y}, \mathbf{d})$$

where $d_{\mathrm{H}}(\cdot,\cdot)$ is the Hamming distance,

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n, \quad d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \sharp\{i \in [1, n],\ x_i \neq y_i\}$$

Remember, an $[n, k]_q$-code is a subspace of $\mathbb{F}_q^n$ with dimension k

Asymptotic sequence of codes:

A sequence of codes $(\mathcal{C}_n)_{n\in\mathbb{N}}$ is defined as a series of $[n, k]_q$-codes where k is a function of n, *i.e.*,

$$k : n \in \mathbb{N} \longmapsto k(n) \in [\![0, n]\!]$$

$\longrightarrow$ Be careful, abuse of notation: we write k instead of $k(n)$

Rate of asymptotic sequence of codes:

Given a sequence of codes $(\mathcal{C}_n)_{n\in\mathbb{N}}$, its rate is defined as $\Big($if the following limit exists$\Big)$,

$$R \stackrel{\text{def}}{=} \lim_{n\to+\infty} \frac{k}{n} \quad \left(= \lim_{n\to+\infty} \frac{k(n)}{n} \right)$$

q-ary entropy

The q-ary entropy is defined as,

$$h_q : x \in [0, 1] \longmapsto -x \log_q \frac{x}{q-1} - (1-x) \log_q (1-x)$$

Exercise:

Let $\mathsf{X} \in [\![0, q-1]\!]$ where,

$$\mathbb{P}(\mathsf{X} = x) = \begin{cases} 1-p \text{ if } x = 0 \\ \frac{p}{q-1} \quad \text{if } x \neq 0 \end{cases}$$

Show that the q-ary entropy is the entropy of X up to $\log_2 q$ factor, *i.e.*,

$$h_q(p) = \frac{1}{\log_2 q} \cdot H(\mathsf{X})$$

SHANNON'S THEOREM STATEMENT

Shannon's Theorem for Linear Codes:

For all $0 \leq p < \frac{q-1}{q}$ and $\varepsilon > 0$ it holds that,

1. It exists $\delta > 0$ such that for any n large enough it exists $(\mathcal{C}, \mathcal{D})$ where $\mathcal{C}$ is a linear code of length n and dimension $n\,(1 - h_q(p) - \varepsilon)$ and
$$\mathbb{P}_{\text{fail}}\,(\mathcal{C}, \mathcal{D}) \leq \tfrac{1}{q^{\delta n}}$$
2. For all $\delta > 0$ and n large enough, and all pair $(\mathcal{C}, \mathcal{D})$, where $\mathcal{C}$ is a linear code with length n and dimension $n\,(1 - h_q(p) + \varepsilon)$, we have
$$\mathbb{P}_{\text{fail}}\,(\mathcal{C}, \mathcal{D}) \geq 1 - \delta$$

$\longrightarrow$ In particular: the proof of Shannon's theorem for linear codes uses the maximum likelihood decoder, not the jointly typical decoder!

An important remark:

Linear codes enable to achieve reliable communication over $q\text{SC}(p)$ but under the necessary and sufficient condition to have a transmission rate
$$R = k/n < 1 - h_q(p)$$
But $\Big(1 - h_q(p)\Big)$ is exactly the capacity, as defined in Lecture 6, of the $q\text{SC}(p)$ channel

1. As with block codes we will used the average code trick but by picking them as uniform $[n, k]_q$-codes $\Big($not a coincidence, capacity of $q\mathrm{SC}(p) = \max_{\mathsf{X}} I(\mathsf{X}, \mathsf{Y})$ achieved for a uniform $\mathsf{X}\Big)$
2. Our decoder choice will be the maximum likelihood decoder
3. We will show that errors of the $q\mathrm{SC}(p)$ concentrate over Hamming ball of radius np
4. Then for "random codes" balls centered at codewords and with radius np do not typically intersect at the condition
$$\sharp\mathcal{C} \cdot \mathrm{Vol}\left(\mathcal{B}(np)\right) = q^k \cdot \mathrm{Vol}\left(\mathcal{B}(np)\right) \leq \sharp\mathbb{F}_q^n = q^n \iff k < n(1 - h_q(p))$$
5. As above balls do not intersect under the good condition, the maximum likelihood decoder works because it outputs the closest codeword $\Big($for the Hamming distance$\Big)$

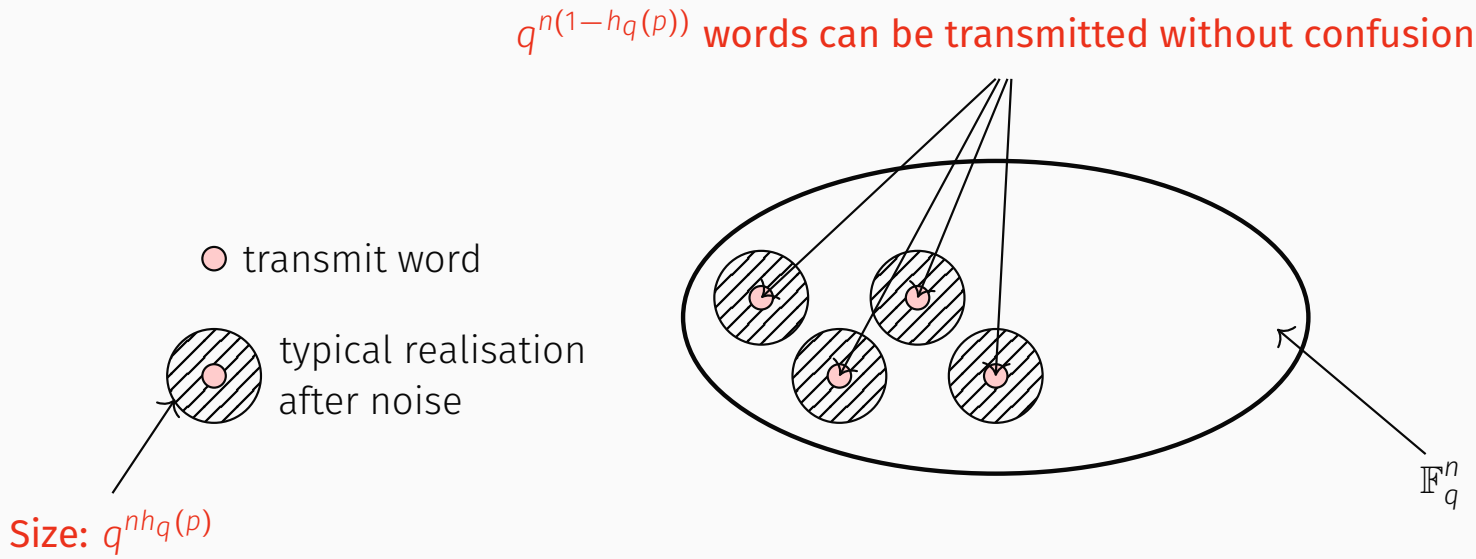
The average trick is the crucial idea of the proof

$\longrightarrow$ If a quantity is $< \varepsilon$ in average over $\{A\}$, then it exists an A_0 s.t the quantity is $< \varepsilon$ for this A_0

SOME PICTURE: AVOID CONFUSION

Errors $\mathbf{e}$ in the $q\mathrm{SC}(p)$ are such that $|\mathbf{e}| \approx np$

$\longrightarrow$ There are $\approx q^{nh_q(p)}$ vectors of Hamming weight $\approx np$



In Shannon's theorem we use the maximum likelihood-decoder: from $\mathbf{y} \in \mathbb{F}_q^n$ it outputs $\mathbf{c} \in \mathcal{C}$, the closest codeword for the Hamming distance!

As soon as the code size is $> q^{n(1-h_q(p))}$ an exponential number of balls intersect:

it is impossible to recover the sent codeword without ambiguity

ABOUT RANDOM CODES

Random code(s):

It is defined as

- $\mathcal{C} = \left\{ \mathbf{m}\mathbf{G}_u : \ \mathbf{m} \in \mathbb{F}_q^k \right\}$ where $\mathbf{G}_u \leftarrow \text{Unif}\left(\mathbb{F}_q^{k \times n}\right)$

or,

- $\mathcal{C} = \left\{ \mathbf{c} \in \mathbb{F}_q^n : \ \mathbf{H}_u \mathbf{c}^\intercal = \mathbf{0} \right\}$ where $\mathbf{H}_u \leftarrow \text{Unif}\left(\mathbb{F}_q^{(n-k) \times n}\right)$

where $\text{Unif}\left(\mathbb{F}_q^{m \times n}\right)$ means the uniform distribution over $\mathbb{F}_q^{m \times n}$, *i.e.,* coefficients the matrix are i.i.d. and uniform over $\mathbb{F}_q$

Exercise:

Are these models equivalent? Do they define a uniform $[n, k]_q$-code?

Random code(s):

- $\mathcal{C} = \left\{ \mathbf{m}\mathbf{G}_u : \ \mathbf{m} \in \mathbb{F}_q^k \right\}$ where $\mathbf{G}_u \leftarrow \text{Unif}\left(\mathbb{F}_q^{k \times n}\right)$

$$\longrightarrow \dim \mathcal{C} \leq k \text{ as } \text{rank}(\mathbf{G}_u) \leq k$$

- $\mathcal{C} = \left\{ \mathbf{c} \in \mathbb{F}_q^n : \ \mathbf{H}_u \mathbf{c}^\intercal = \mathbf{0} \right\}$ where $\mathbf{H}_u \leftarrow \text{Unif}\left(\mathbb{F}_q^{(n-k) \times n}\right)$

$$\longrightarrow \dim \mathcal{C} \geq k \text{ as } \text{rank}(\mathbf{H}_u) \leq n - k$$

Both models do not seem to be equivalent... (Spoiler: they "are"!)

and they don't define uniform $[n, k]_q$-codes...

Random code(s):

It is defined as

- $\mathcal{C} = \left\{ \mathbf{m}\mathbf{G}_k : \ \mathbf{m} \in \mathbb{F}_q^k \right\}$ where $\mathbf{G}_k \leftarrow \text{Unif}\left\{ \mathbf{M} \in \mathbb{F}_q^{k \times n} \text{ with rank } k \right\}$

or,

- $\mathcal{C} = \left\{ \mathbf{c} \in \mathbb{F}_q^n : \ \mathbf{H}_k \mathbf{c}^\intercal = \mathbf{0} \right\}$ where $\mathbf{H}_k \leftarrow \text{Unif}\left\{ \mathbf{M} \in \mathbb{F}_q^{(n-k) \times n} \text{ with rank } n-k \right\}$

$\longrightarrow$ Both models are equivalent: they pick a code uniformly at random among $[n, k]_q$-codes

Issue:

These models imply in many situations unnecessarily complex computations contrary to the cases where we don't fix the rank of the matrix. . .

But could it be that all these models are "equivalent"?

AN IMPORTANT TOOL: STATISTICAL DISTANCE

Statistical distance:

Let X and Y be random variables,

$$\Delta(\mathsf{X}, \mathsf{Y}) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{a \in \mathcal{E}} \left| \mathbb{P}(\mathsf{X} = a) - \mathbb{P}(\mathsf{Y} = a) \right|$$

A crucial property: data processing inequality

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

Consequence: $\forall \mathcal{A}$ algorithm

$$\left| \mathbb{P}_{\mathsf{X}}\Big(\mathcal{A}(\mathsf{X}) = \text{“success”}\Big) - \mathbb{P}_{\mathsf{Y}}\Big(\mathcal{A}(\mathsf{Y}) = \text{“success”}\Big) \right| \leq \Delta(\mathsf{X}, \mathsf{Y})$$

Typical situation:

We want to analyse the success probability of $\mathcal{A}(\mathsf{m})$ when its input m is picked according to X. But it implies ugly computations. . . However, we know how to analyse this success probability when m picked according to Y. If $\Delta(\mathsf{X}, \mathsf{Y})$ is small we are done! We perform computations with Y, what we loose in our estimation of the success probability is $\pm\Delta(\mathsf{X}, \mathsf{Y})$

G_u or H_u-models $\iff$ draw uniformly an $[n,k]$-code:

$\mathsf{G}_k \in \mathbb{F}_q^{k\times n}$ $\left(\mathsf{H}_k \in \mathbb{F}_q^{(n-k)\times n}\right)$ be uniform of rank k $\left(\textit{resp. } n-k\right)$:

$$\Delta\left(\mathsf{G}_u, \mathsf{G}_k\right) = O\left(q^{-(n-k)}\right) \quad \left(\textit{resp. } \Delta\left(\mathsf{H}_u, \mathsf{H}_k\right) = O\left(q^{-k}\right)\right)$$

Computation are the same in G_u and H_u-models:

Let $\mathcal{E}$ be a set of codes $\left(\text{defined as an event}\right)$. We have,

1. $$\left|\mathbb{P}_{\mathsf{G}_k}(\mathcal{E}) - \mathbb{P}_{\mathsf{H}_k}(\mathcal{E})\right| = 0$$
2. $$\left|\mathbb{P}_{\mathsf{G}_u}(\mathcal{E}) - \mathbb{P}_{\mathsf{H}_u}(\mathcal{E})\right| = O\left(q^{-\min(k,n-k)}\right)$$

Proof:

$$\left|\mathbb{P}_{\mathbf{G}_u}(\mathcal{E}) - \mathbb{P}_{\mathbf{H}_u}(\mathcal{E})\right| \leq \left|\mathbb{P}_{\mathbf{G}_u}(\mathcal{E}) - \mathbb{P}_{\mathbf{G}_k}(\mathcal{E})\right| + \left|\mathbb{P}_{\mathbf{H}_k}(\mathcal{E}) - \mathbb{P}_{\mathbf{H}_u}(\mathcal{E}) + \left|\mathbb{P}_{\mathbf{G}_k}(\mathcal{E}) - \mathbb{P}_{\mathbf{H}_k}(\mathcal{E})\right|\right.$$

- $\left|\mathbb{P}_{\mathbf{G}_u}(\mathcal{E}) - \mathbb{P}_{\mathbf{H}_u}(\mathcal{E})\right|$ and $\left|\mathbb{P}_{\mathbf{H}_k}(\mathcal{E}) - \mathbb{P}_{\mathbf{H}_u}(\mathcal{E})\right|$ are $O(q^{-\min(k,n-k)})$ because of the statistical distance
- $\mathbb{P}_{\mathbf{G}_k}(\mathcal{E}) = \mathbb{P}_{\mathbf{H}_k}(\mathcal{E})$ because codes defined by $\mathbf{G}_k$ and $\mathbf{H}_k$ have the same distribution: uniform over $[n, k]_q$-codes

Shannon's theorem: proof

Hamming ball:

The Hamming ball of center $\mathbf{x}$ and radius $r \in [\![0, n]\!]$ is defined as,

$$\mathcal{B}_{\mathrm{H}}(\mathbf{x}, r) = \left\{ \mathbf{y} \in \mathbb{F}_q^n,\ d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) \leq r \right\}$$

$\longrightarrow$ The volume of $\mathcal{B}_{\mathrm{H}}(\mathbf{x}, r)$ is given by $\sum_{j=0}^{r} \binom{n}{j} (q-1)^j$, it is independent from $\mathbf{x}$

Lemma (see Exercise Session 5)**:**

for all $0 \leq r \leq \frac{q-1}{q} n$,

$$\tfrac{1}{n+1} q^{nh_q(r/n)} \leq \binom{n}{r} (q-1)^r \leq q^{nh_q(r/n)}, \quad \text{then} \quad \sharp \mathcal{B}_{\mathrm{H}}(\mathbf{x}, r) = q^{n\left(h_q(r/n)+o(1)\right)}$$

Intuitive proof: a q-ary Bernoulli with parameter concentrates exactly over the shell of $\mathcal{B}_{\mathrm{H}}(\mathbf{0}, r)$ (Chernoff bound) but we know that the number of typical sequence of this distribution is $q^{nh_q(r/n)}$. To conclude, the volume of a ball is $\approx$ given by its shell (or $x \mapsto h_q(x)$ is increasing over $[0, (q-1)/q]$)

From Lecture 7:

Proposition:

In a qSC(p) with probability of transition $p/(q-1) < 1/q$, if codewords $\mathbf{c} \in \mathcal{C}$ are chosen uniformly at random among $\mathcal{C}$, then

$$\forall \mathbf{y} \in \mathbb{F}_q^n, \quad \mathcal{D}_{\mathrm{ML}}(\mathbf{y}) = \underset{\mathbf{c}_0 \in \mathcal{C}}{\arg\max}\, \mathbb{P}(\mathbf{c}_0 \mid \mathbf{y}) = \mathbf{c} \in \mathcal{C} \quad \text{such that } \mathbf{c} = \underset{\mathbf{d} \in \mathcal{C}}{\arg\min}\, d_{\mathrm{H}}(\mathbf{y}, \mathbf{d})$$

where $d_{\mathrm{H}}(\cdot, \cdot)$ is the Hamming distance,

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n, \quad d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) \overset{\text{def}}{=} \sharp \{i \in [1, n],\ x_i \neq y_i\}$$

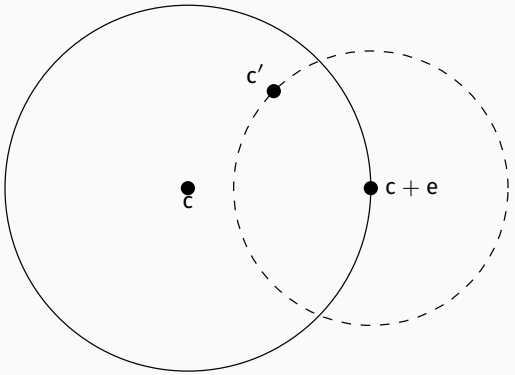
Lemma: maximum likelihood decoder and the linearity

Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a linear code, and $\mathbf{e}$ be the error induced by qSC(p),

$$\mathbb{P}_{\text{fail}}\left(\mathcal{C}, \mathcal{D}_{\mathrm{ML}}\right) = \mathbb{P}\left(\mathcal{D}_{\mathrm{ML}}(\mathbf{c} + \mathbf{e}) \neq \mathbf{c}\right) = \mathbb{P}_{\mathbf{e}}\left(\mathcal{D}_{\mathrm{ML}}(\mathbf{e}) \neq 0\right)$$

or equivalently,

$$\mathbb{P}_{\text{fail}}\left(\mathcal{C}, \mathcal{D}_{\mathrm{ML}}\right) = \mathbb{P}_{\mathbf{e}}\Big(\exists \mathbf{c} \in \mathcal{B}_{\mathrm{H}}\left(\mathbf{e}, |\mathbf{e}|\right) \cap \mathcal{C} \backslash \{0\}\Big)$$



Proof:

Given $\mathbf{c} \in \mathcal{C}$ be chose uniformly at random and $\mathbf{e}$ be the error induced by $q\mathrm{SC}(p)$. The maximum likelihood decoder $\mathcal{D}_{\mathrm{ML}}$ will fail given $\mathbf{y} \stackrel{\mathrm{def}}{=} \mathbf{c} + \mathbf{e}$ if there is another codeword $\mathbf{c}' \in \mathcal{C}$ such that $|\mathbf{y} - \mathbf{c}'| \leq |\mathbf{y} - \mathbf{c}|$

$$\begin{aligned}
\mathbb{P}_{\mathrm{fail}}\left(\mathcal{C}, \mathcal{D}_{\mathrm{ML}}\right) &= \mathbb{P}\Big(\exists \mathbf{c}' \in \mathcal{B}_{\mathrm{H}}\left(\mathbf{c} + \mathbf{e}, |\mathbf{e}|\right) \cap \mathcal{C} \backslash \{\mathbf{c}\}\Big) \\
&= \frac{1}{q^k} \sum_{\mathbf{c}_0 \in \mathcal{C}} \mathbb{P}\Big(\exists \mathbf{c}' \in \mathcal{B}_{\mathrm{H}}\left(\mathbf{c}_0 + \mathbf{e}, |\mathbf{e}|\right) \cap \mathcal{C} \backslash \{\mathbf{c}_0\}\Big) \quad \Big(\text{law of total probability}\Big) \\
&= \frac{1}{q^k} \sum_{\mathbf{c}_0 \in \mathcal{C}} \mathbb{P}\Big(\exists \mathbf{c}' \in \mathcal{B}_{\mathrm{H}}\left(\mathbf{e}, |\mathbf{e}|\right) \cap \mathcal{C} \backslash \{\mathbf{0}\}\Big)
\end{aligned}$$

where in the last equality we used that fact that $\mathcal{C}$ is linear by applying a translation $-\mathbf{c}_0$. Then,

$$\mathbb{P}_{\mathrm{fail}}\left(\mathcal{C}, \mathcal{D}_{\mathrm{ML}}\right) = \mathbb{P}_{\mathbf{e}}\Big(\exists \mathbf{u} \in \mathcal{B}_{\mathrm{H}}\left(\mathbf{e}, |\mathbf{e}|\right) \cap \mathcal{C} \backslash \{\mathbf{0}\}\Big) = \mathbb{P}_{\mathbf{e}}\left(\mathcal{D}_{\mathrm{ML}}(\mathbf{e}) \neq \mathbf{0}\right)$$

In the proof of Shannon's theorem will use as model of random codes:

$\mathbf{G}_k$ *be a uniform matrix over* $\mathbb{F}_q^{k\times n}$ ***with rank*** k

Lemma:

For all $\left(\text{fixed}\right)$ $\mathbf{m} \in \mathbb{F}_q^k \backslash \{\mathbf{0}\}$, the random variable $\mathbf{m}\mathbf{G}_k$ is uniform over $\mathbb{F}_q^n \backslash \{\mathbf{0}\}$

Proof:

Let $\mathbf{y}, \mathbf{y}' \in \mathbb{F}_q^n \backslash \{\mathbf{0}\}$: we will prove that $\mathbb{P}\left(\mathbf{m}\mathbf{G}_k = \mathbf{y}\right) = \mathbf{Pr}\left(\mathbf{m}\mathbf{G}_k = \mathbf{y}'\right)$

First, it exists an invertible $n \times n$ matrix $\mathbf{M}$ such that $\mathbf{y}\mathbf{M} = \mathbf{y}'$. Therefore,

$$\mathbb{P}\left(\mathbf{x}\mathbf{G}_k = \mathbf{y}\right) = \mathbb{P}\left(\mathbf{x}\mathbf{G}_k\mathbf{M} = \mathbf{y}'\right)$$

Since $\mathbf{A} \mapsto \mathbf{A}\mathbf{M}$ is a bijection from the set of full rank $k \times n$ matrices onto itself, $\mathbf{G}_k\mathbf{M}$ is a uniformly random full rank $k \times n$ matrix and,

$$\mathbb{P}\left(\mathbf{x}\mathbf{G}_k\mathbf{M} = \mathbf{y}'\right) = \mathbb{P}\left(\mathbf{x}\mathbf{G}_k = \mathbf{y}'\right)$$

Combining the previous equality concludes the proof

Proof of Shannon's theorem:

Let $k = \lfloor (1 - h_q(p) - \varepsilon)n \rfloor$. Let $\mathcal{C}$ be an $[n, k]_q$-code and $\mathbf{G}$ be a generator matrix for $\mathcal{C}$. We have proved,

$$\mathbb{P}_{\text{fail}}(\mathcal{C}, \mathcal{D}_{\text{ML}}) = \mathop{\mathbb{P}}_{\mathbf{e} \leftarrow q\text{SC}(p)} \Big(\exists \mathbf{m} \in \mathbb{F}_q^k \backslash \{0\},\ \mathbf{mG} \in \mathcal{B}_{\text{H}}(\mathbf{e}, |\mathbf{e}|) \Big)$$

Let $\mathcal{E}(\mathbf{G}, \mathbf{e})$ be the event: $\mathcal{E}(\mathbf{G}, \mathbf{e}) \stackrel{\text{def}}{=} \Big\{ \exists \mathbf{m} \in (\mathbb{F}_q^k \backslash \{0\}),\ \mathbf{mG} \in \mathcal{B}_{\text{H}}(\mathbf{e}, |\mathbf{e}|) \Big\}$. Let $\gamma > 0$, then,

$$\begin{aligned}\mathbb{P}_{\text{fail}}(\mathcal{C}, \mathcal{D}_{\text{ML}}) = \mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}, \mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)\mathbb{P}_{\mathbf{e}}\Big(|\mathbf{e}| \leq (p+\gamma)n\Big) \\ + \mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}, \mathbf{e}) \mid |\mathbf{e}| > (p+\gamma)n\Big)\mathbb{P}_{\mathbf{e}}\Big(|\mathbf{e}| > (p+\gamma)n\Big)\end{aligned}$$

Since probabilities are always less than or equal to 1, we get

$$\mathbb{P}_{\text{fail}}(\mathcal{C}, \mathcal{D}_{\text{ML}}) \leq \mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}, \mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big) + \mathbb{P}_{\mathbf{e}}\Big(|\mathbf{e}| > (p+\gamma)n\Big)$$

Thanks to Chernoff bound,

$$\mathbb{P}_{\mathbf{e}}\Big(|\mathbf{e}| > (p+\gamma)n\Big) \leq 2\mathrm{e}^{-2pn\gamma^2}$$

Conclusion:

To conclude it remains to upper-bound: $\mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}, \mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)$

$\longrightarrow$ Use the average trick over $\mathbf{G}$

It remains to get an upper bound on $\mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G},\mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)$ but it is hopeless to get

a uniform and sharp upper bound for all $\mathbf{G}$

Proof of Shannon's theorem

Let $\mathbf{G}_k$ be uniformly picked over $\mathcal{G}_{n,k}$ the set of matrices $\mathbb{F}_q^{k\times n}$ with rank k

$$
\begin{aligned}
M &\stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{G}_k}\Big(\mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}_k,\mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)\Big)\\
&= \frac{1}{|\mathcal{G}_{k,n}|}\sum_{\mathbf{G}\in\mathcal{G}_{k,n}} \mathbb{P}_{\mathbf{e}}\Big(\exists \mathbf{m}\in\mathbb{F}_q^k\setminus\{0\},\ \mathbf{mG}\in\mathcal{B}_{\mathrm{H}}(\mathbf{e},|\mathbf{e}|) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)\\
&\leq \frac{1}{|\mathcal{G}_{k,n}|}\sum_{\mathbf{G}\in\mathcal{G}_{k,n}} \mathbb{P}_{\mathbf{e}}\Big(\exists \mathbf{m}\in\mathbb{F}_q^k\setminus\{0\},\ \mathbf{mG}\in\mathcal{B}_{\mathrm{H}}(\mathbf{e},(p+\gamma)n) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)
\end{aligned}
$$

By the union bound, we get

$$
M \leq \tfrac{1}{|\mathcal{G}_{k,n}|}\textstyle\sum_{\mathbf{G}\in\mathcal{G}_{k,n}}\sum_{\mathbf{m}\in\mathbb{F}_q^k\setminus\{0\}} \mathbb{P}_{\mathbf{e}}\Big(\mathbf{mG}\in\mathcal{B}_{\mathrm{H}}(\mathbf{e},(p+\gamma)n) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)
$$

Therefore,

$$
\begin{aligned}
M &\leq \frac{1}{|\mathcal{G}_{k,n}|}\sum_{\mathbf{G}}\sum_{\mathbf{m}\in\mathbb{F}_q^k\setminus\{0\}}\sum_{\mathbf{x}:|\mathbf{x}|\leq(p+\gamma)n} \mathbb{P}_{\mathbf{e}}\Big(\mathbf{mG}\in\mathcal{B}_{\mathrm{H}}(\mathbf{x},(p+\gamma)n)\Big)\mathbb{P}_{\mathbf{e}}\Big(\mathbf{e}=\mathbf{x} \mid |\mathbf{e}| \leq (p+\gamma)n\Big)\\
&= \sum_{\mathbf{m}\in\mathbb{F}_q^k\setminus\{0\}}\sum_{\mathbf{x}:|\mathbf{x}|\leq(p+\gamma)n} \mathbb{P}_{\mathbf{G}_k}\Big(\mathbf{mG}_k\in\mathcal{B}_{\mathrm{H}}(\mathbf{x},(p+\gamma)n)\Big)\mathbb{P}_{\mathbf{e}}\Big(\mathbf{e}=\mathbf{x} \mid |\mathbf{e}| \leq (p+\gamma)n\Big)
\end{aligned}
$$

Proof of Shannon's Theorem

We have proved: if $\mathbf{m} \in \mathbb{F}_q^k \backslash \{0\}$ then $\mathbf{m}\mathbf{G}_k$ is uniform over $\mathbb{F}_q^n \backslash \{\mathbf{0}\}$. Therefore,

$$\mathbb{P}_{\mathbf{G}_k}\Big(\mathbf{m}\mathbf{G}_k \in \mathcal{B}_{\mathrm{H}}(\mathbf{x}, (p+\gamma)n)\Big) = \frac{\sharp\mathcal{B}_{\mathrm{H}}((p+\gamma)n, n) - 1}{q^n - 1} \leq q^{n(h_q(p+\gamma)-1)}$$

where the last inequality we used our inequality on the size of Hamming ball. Therefore,

$$\begin{aligned} M &\leq \sum_{\mathbf{m}\in\mathbb{F}_q^k\backslash\{0\}} \sum_{\mathbf{x}:|\mathbf{x}|\leq(p+\gamma)n} q^{n(h_q(p+\gamma)-1)}\ \mathbb{P}_{\mathbf{e}}\Big(\mathbf{e} = \mathbf{x} \mid |\mathbf{e}| \leq (p+\gamma)n\Big) \\ &= q^{n(h_q(p+\gamma)-1)} \sum_{\mathbf{m}\in\mathbb{F}_q^k\backslash\{0\}} \mathbb{P}_{\mathbf{e}}\Big(|\mathbf{e}| \leq (p+\gamma)n\Big) \\ &\leq q^{n(h_q(p+\gamma)-1)}\ q^k \end{aligned}$$

Finally, since $k \leq n(1 - h_q(p) - \varepsilon)$ and h_q is a decreasing continuous function over $\left[0, \frac{p}{q-1}\right]$, we obtain for γ small enough

$$M = \mathbb{E}_{\mathbf{G}_k}\left(\mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}_k, \mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big)\right) \leq q^{n(h_q(p+\gamma)-h_q(p)-\varepsilon)} \leq q^{-n\delta}$$

where $\delta > 0$ is some constant. Therefore, it exists at least one matrix $\mathbf{G}_0$ of rank k in $\mathbb{F}_q^{k\times n}$ s.t

$$\mathbb{P}_{\mathbf{e}}\Big(\mathcal{E}(\mathbf{G}_0, \mathbf{e}) \mid |\mathbf{e}| \leq (p+\gamma)n\Big) \leq q^{-n\delta}$$

Proof of Shannon's theorem (iv)

Proof of Shannon's theorem:

Putting everything together we obtain, where $\mathcal{C}_0$ is the code with generator matrix $\mathbf{G}_0$,

$$\mathbb{P}_{\text{fail}}(\mathcal{C}_0, \mathcal{D}_{\text{ML}}) \leq q^{-n\delta} + 2\mathrm{e}^{-pn\gamma^2} \leq q^{-\delta' n}$$

for some constant $\delta' > 0$ showing 1. in Shannon's theorem

It turns out that we have proven $\Big($where $\mathcal{C}$ is a uniform $[n,k]_q$-code$\Big)$

$$\mathbb{E}_{\mathcal{C}}\left(\mathbb{P}_{\text{fail}}\left(\mathcal{C},\mathcal{D}_{\text{ML}}\right)\right) \leq q^{-\delta' n}$$

$\longrightarrow$ Therefore, by Markov's inequality: $\mathbb{P}_{\mathcal{C}}\left(\mathbb{P}_{\text{fail}}\left(\mathcal{C},\mathcal{D}_{\text{ML}}\right) \geq q^{-\delta' n/2}\right) \leq q^{-\delta' n/2}$

Conclusion:

For almost all $[n,k]_q$-codes $\mathcal{C}$ $\Big($a proportion $1-\frac{1}{q^{\delta' n/2}}\Big)$, the maximum-likelihood decoder fails with probability exponentially close to 0, *i.e.*, $\mathbb{P}_{\text{fail}}\left(\mathcal{C},\mathcal{D}_{\text{ML}}\right) \leq q^{-\delta' n/2}$

Shannon's theorem for linear codes: negative part

For all $0 \leq p < \frac{q-1}{q}$ and $\varepsilon > 0$ it holds that,

2. For all $\delta > 0$ and n large enough, and all pair $(\mathcal{C}, \mathcal{D})$, where $\mathcal{C}$ is a linear code with length n and dimension $n\left(1 - h_q(p) + \varepsilon\right)$, we have

$$\mathbb{P}_{\text{fail}}\left(\mathcal{C}, \mathcal{D}\right) \geq 1 - \delta$$

- The capacity of the qSC(p) $\Big($according to Lecture 6$\Big)$ is given by $\Big(1 - h_q(p)\Big)$. Therefore to prove this part of the theorem we can use the negative part of Shannon's theorem from Lecture 6

RANDOM CODES: A POWERFUL TOOL

Fundamental trick:

Suppose that $\mathcal{C}$ is a uniform $[n,k]_q$-code, X be some source of randomness and

$$\mathbb{P}_{\mathcal{C},\mathsf{X}}\Big(\mathcal{C} \text{ verifies some good property } \mathcal{P}\Big) \geq C \quad \text{or} \quad \mathbb{P}_{\mathcal{C},\mathsf{X}}\Big(\mathcal{C} \text{ verifies some bad property } \mathcal{Q}\Big) \leq \varepsilon$$

Then,

1. It exists at least one $[n,k]_q$-code $\mathcal{C}_0$ verifying the good property with probability at least C,

$$\mathbb{P}_{\mathsf{X}}\Big(\mathcal{C}_0 \text{ verifies some good property } \mathcal{P}\Big) \geq C$$

2. The proportion of codes for which the probability of the bad property is $\geq \sqrt{\varepsilon}$ is $\leq \sqrt{\varepsilon}$,

$$\frac{\sharp\Big\{\mathcal{C}_0 \colon \mathbb{P}_{\mathsf{X}}\Big(\mathcal{C}_0 \text{ an } [n,k]_q\text{-code verifying some bad property } \mathcal{Q}\Big) \geq \sqrt{\varepsilon}\Big\}}{\sharp\Big\{\mathcal{C}_0 \text{ be an } [n,k]_q\text{-code}\Big\}} \leq \sqrt{\varepsilon}$$

Proof $\Big($nothing else than Markov's inequality$\Big)$:

As $\mathcal{C}$ is uniform over $[n,k]_q$-codes we have by the law of total number,

$$\begin{aligned}\mathbb{P}_{\mathcal{C},\mathsf{x}}\Big(\mathcal{C} \text{ verifies some property}\Big) &= \frac{1}{M}\sum_{\substack{\mathcal{C}_0 \text{ be an} \\ [n,k]_q\text{-code}}} \mathbb{P}_{\mathsf{x}}\Big(\mathcal{C}_0 \text{ verifies some property}\Big) \\ &= \mathbb{E}_{\mathcal{C}}\Big(\mathbb{P}_{\mathsf{x}}\Big(\mathcal{C} \text{ verifies some property}\Big)\Big)\end{aligned}$$

where M is the number of $[n,k]_q$-code. The above sum is enough to conclude

We have proved that we can decode $\Big($*but without an efficient decoding algorithm*$\Big)$ *a random code*

up to the channel capacity of $qSC(p)$

$\longrightarrow$ We know more things about random codes

Proposition: weight distribution of random codes

Let $\mathcal{C}$ be a random $[n, k]_q$-codes where we use the model that $\mathcal{C}$ admits as parity-check matrix a uniform $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$. We have,

$$\forall \ell > 0,\ \mathbb{E}_{\mathcal{C}}\left(\sharp\left\{\mathbf{c} \in \mathcal{C} :\ |\mathbf{c}| = \ell\right\}\right) = \frac{\binom{n}{\ell}(q-1)^{\ell}}{q^{n-k}}$$

We have proved that we can decode $\Big($*but without an efficient decoding algorithm*$\Big)$ *a random code up to the channel capacity of qSC*(p)

$\longrightarrow$ We know more things about random codes

Proposition: minimum distance of random codes

Let $\varepsilon > 0$ and $(\mathcal{C}_n)_{n\in\mathbb{N}}$ be an asymptotic sequence of random $[n,k]_q$-codes $\Big($whichever of the four models we previously discussed$\Big)$ where $\frac{k}{n} \xrightarrow[n\to+\infty]{} R$. Then,

$$\mathbb{P}_{\mathcal{C}}\left((1-\varepsilon)h_q^{-1}(1-R) \leq \frac{d_{\min}(\mathcal{C}_n)}{n} \leq (1+\varepsilon)h_q^{-1}(1-R)\right) \geq 1-q^{-\alpha n}$$

for some constant $\alpha > 0$

Consequence:

Almost all codes $[n,k]_q$-codes $\Big($via Markov's inequality$\Big)$ with rate $R \overset{\text{def}}{=} k/n$ have a minimum distance given by

$$t_{\text{GV}}(R) \overset{\text{def}}{=} h_q^{-1}(1-R)\cdot n \quad \text{known as Gilbert-Varshamov radius}$$

Proof:

See Exercise Session

We know that,

- we can decode with success probability > 0 a random code with rate R in the $q\mathrm{SC}(p)$ as soon as $R \leq 1 - h_q(p)$. But in the $q\mathrm{SC}(p)$ there are $\approx np$ errors. Notice that

$$R \leq 1 - h_q(p) \iff p \leq n h_q^{-1}(1 - R)$$

Conclusion:

We can decode a random code of rate R up to the Hamming distance $h_q^{-1}(1 - R) \cdot n$

- random codes with rate R have minimum distance $\approx h_q^{-1}(1 - R)\cdot$

Surprising conclusion:

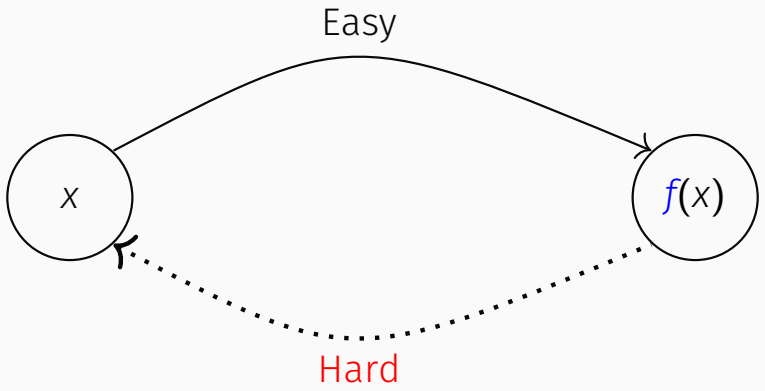
We can $\Big($theoretically, *i.e.*, non-efficiently$\Big)$ decode a random code $\mathcal{C}_{\mathrm{rand}}$ with probability > 0 as soon as the number of errors is $\leq d_{\min}(\mathcal{C}_{\mathrm{rand}})$ and not $d_{\min}(\mathcal{C}_{\mathrm{rand}})/2$

A LITTLE BIT OF CRYPTOGRAPHY

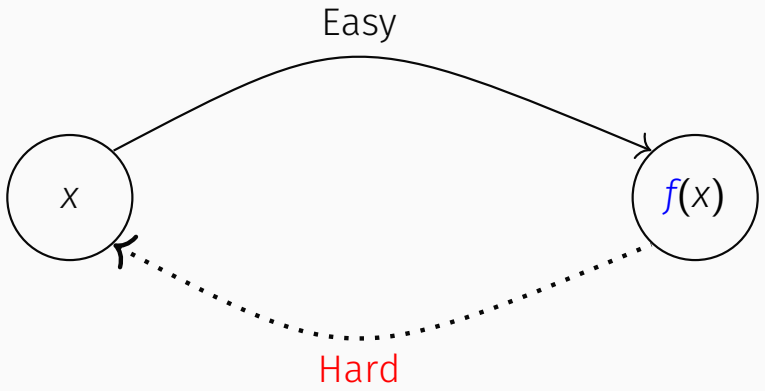
Cryptographer's goal

Find hard problems. . .

Hard problem: given some function f such that



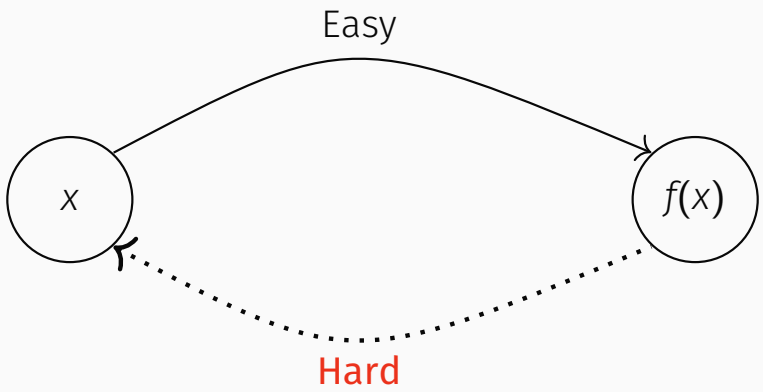
Hard problem: given some function f such that



- What is Mr. Rodemich number?
- Who is owner of +33 631053595?

$\longrightarrow$ Can we really do cryptography with one-way functions?

Cryptography asks for particular one-way functions: trapdoor one-way functions



Easy with trapdoor

Alice wants to receive secret data

Public-Key Encryption:

- Alice reveals f to the world but Alice owes the trapdoor
- Bob wants to share secretly with Alice some number 711327: he computes $f(711327)$
- Alice receives $f(711327)$ and she recovers 711327 while it is hard for everyone else

To build a public-key encryption scheme we need:

- first, a one-way function f
- then to find a trapdoor for this function

Public-Key: description of f ; Secret-Key: its associated trapdoor

WHAT ABOUT DECODING A RANDOM CODE?

Shannon's proof relied on the following decoding problem:

$$\text{Given } \mathbf{y} \stackrel{\text{def}}{=} \mathbf{c} + \mathbf{e} \text{ where } \begin{cases} \mathbf{c} \in \mathcal{C} \text{ with } \mathcal{C} \text{ being a random code} \\ \mathbf{e} \text{ is such that } |\mathbf{e}| \approx t \end{cases},$$

we have to recover the closest codeword from $\mathbf{y}$

It seems to be a hard problem, Shannon did not give any efficient algorithm to solve this problem!

Shannon only proved that the closest codeword from $\mathbf{y}$ is $\mathbf{c}$ if and only if $\sharp\mathcal{C}$ is small enough

$$\Big(\text{capacity condition}\Big)$$

Given $\mathbf{G} \in \mathbb{F}_q^{k \times n}$,

$$f_{\mathbf{G}} : (\mathbf{m}, \mathbf{e}) \in \mathbb{F}_q^k \times \left\{ \mathbf{x} \in \mathbb{F}_q^n : \ |\mathbf{x}| = t \right\} \longmapsto \mathbf{mG} + \mathbf{e}$$

- The knowledge of $\mathbf{G}$ which defines an $[n, k]_q$-code $\mathcal{C}$ and t is enough to compute $f_{\mathbf{G}}$
- It is easy to compute $f_{\mathbf{G}}(\mathbf{m}, \mathbf{e})$ $\Big($only linear algebra$\Big)$
- If $\mathbf{G}$ has been chosen uniformly,inverting $f_{\mathbf{G}}$ amount to decode a random code at distance t which is believed to be hard

But we need a trapdoor to build public-key encryptions!

McEliece in '78 had the idea of introducing the following trapdoor:

The underlying code in $f_{\mathbf{G}}$ is chosen as a code that we know how to decode and the quantities which enable to decode form the secret key

$\longrightarrow$ McEliece proposed in '78 to choose $\mathbf{G}$ as the generator matrix of a *Goppa code*

*In Lecture 7 we have studied codes that we know how to decode: **Reed-Solomon** codes*

Berlekamp-Welsh Algorithm:

We can decode $\mathrm{RS}_k(\mathbf{x}, \mathbf{z})$ at any distance $< \frac{n-k}{2}$

- Public Key: a representation of $\mathrm{RS}_k(\mathbf{x}, \mathbf{z})$ which admits as generator matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{pmatrix}$$

- Secret Key:

What is the secret key? Can we give the above matrix as a public key?

*In Lecture 7 we have studied codes that we know how to decode: **Reed-Solomon** codes*

Berlekamp-Welsh Algorithm:

We can decode $\mathrm{RS}_k(\mathbf{x}, \mathbf{z})$ at any distance $< \frac{n-k}{2}$

- Public Key: a representation of $\mathrm{RS}_k(\mathbf{x}, \mathbf{z})$ which admits as generator matrix

$$\mathbf{G}_{\mathrm{pk}} = \mathbf{S} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{pmatrix}$$ where $\mathbf{S}$ non-singular picks random basis

- Secret Key: $T = \mathbf{x}$

- To encrypt a message $\mathbf{m}$ with the public-key $\mathbf{G}_{\mathrm{pk}}$: compute $\mathbf{mG} + \mathbf{e}$ where $|\mathbf{e}| < \frac{n-k}{2}$
- To decrypt $\mathbf{mG}_{\mathrm{pk}} + \mathbf{e}$ with the knowledge of the secret-key T: use Berlekamp-Welsh algorithm

But in '92, Sidelnikov and Shestakov have shown how to break this instantiation of McEliece's encryption

$\longrightarrow$ From the knowledge of $\mathsf{G}_{\mathsf{pk}} = \mathsf{S} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{pmatrix}$ we can easily recover the secret $\mathbf{x}$

The above attack shows that the security of McEliece's encryption scheme does not strictly relies on the hardness of decoding a random code

But under which assumptions McEliece's scheme is secure?

The above attack shows that the security of McEliece's encryption scheme does not strictly relies on the hardness of decoding a random code

But under which assumptions McEliece's scheme is secure?

McEliece security assumption:

We can prove that McEliece's encryption is secure under the following assumptions:

1. It is hard to decode a random code
2. It is hard to distinguish the public-key $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ and a uniform matrix with the same size

Can we build an encryption scheme whose security is only based on the hardness of decoding a random code?

Yes, a public-key encryption scheme whose security is only based on the hardness of decoding a random code is known since '03:

Alekhnovich's cryptosystem

- Lecture notes by Gilles Zémor about Alekhnovich's cryptosystem:

  https://www.math.u-bordeaux.fr/~gzemor/alekhnovich.pdf

EXERCISE SESSION