ACCQ204 Nov. 9th 2018

Cours 1

Enseignant: Aslan Tchamkerten *Crédit: Toni Franceschelli*

1 Un peu d'histoire...

La Théorie du codage date des années '50. Claude Elwood Shannon (1916-2001) et Richard Hamming (1915-1998) en sont les pionniers.

Le premier, considéré comme le père de l'ère digitale, s'est intéressé principalement aux *limites fondamentales* de communication en terme de:

- stockage de données: limite ultime de compression.
- transmission de données: limite ultime de vitesse de transmission fiable de données.

De façon complémentaire, Hamming s'est intéressé aux *algorithmes* permettant au mieux de corriger et détecter des erreurs. Le papier de Shannon *A mathematical theory of communication* (1948) et celui de Hamming *Error detecting and error correcting codes* (1950) établirent les domaine de la théorie de l'information et le domaine du codage, respectivement. A noter que Hamming considère un modéle de communication quelque peu différent de celui de Shannon.

Problème de Hamming, exemple:

- On veut stocker des bits sur un support magnétique.
- Les bits sur le support peuvent se corrompre mais très rarement (au pire 1 bit sur 63).

1.1 Une solution naïve

Une première solution naïve consiste à répéter chaque bit 3 fois. La taille du mot code est donc 3 fois plus grande que celle du message. Exemple : message $\Rightarrow$ 0100 ; mot code $\Rightarrow$ 000111000000.

Performances:

- Complexité de codage et décodage: linéaire en la taille du message
- Taux de codage = $\frac{Taille\ message}{Taille\ mot\ code} = \frac{1}{3}$

Ce codage protège d'une erreur. Pour le décodage, on utilise la règle de la majorité sur 3 bits consécutifs.

1.2 Solution 1 de Hamming

On découpe le message en blocs de 4 bits chacun.
On associe à chaque bloc m un mot code $m \cdot G$ où $m \in \{0,1\}^4$ et

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Propriété:
$\forall m_1 \neq m_2 \in \{0,1\}^4$, $m_1 \cdot G$ et $m_2 \cdot G$ diffèrent d'au moins 3 positions.

Taux: $\frac{4}{7}$

Décodage:
Soit $y \in \{0,1\}^7$ contenant au plus 1 erreur.
$y \cdot H$ donne l'index du bit corrompu de y avec

$$H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

1.3 Solution 2 de Hamming

$\exists$ deux matrices $G \in \mathcal{M}_{57,63}$ et $H \in \mathcal{M}_{63,6}$ possédant les propriétés suivantes:

- $\forall m_1 \neq m_2 \in \{0,1\}^{57}$, $m_1 \cdot G$ et $m_2 \cdot G$ diffèrent d'au moins 3 positions;

- si y est un mot "corrompu" d'au plus 1 erreur alors $y \cdot H$ donne l'index du bit corrompu!

Taux: $\frac{57}{63} > \frac{4}{7}$. Aucun schéma qui corrige une erreur ne peut atteindre un taux supérieur à $\frac{57}{63}$, comme on le verra plus bas.

2 Notions de Hamming

2.1 Distance de Hamming

Soit Σ un alphabet de cardinalité $q < \infty$.
Soit Σ^n l'ensemble des mots de n lettres sur Σ.

On appelle distance de Hamming $\Delta(x, y)$, avec $x, y \in \Sigma$, le nombre de coordonnées où x et y diffèrent.
On note $\delta(x, y)$ la distance normalisée de Hamming: $\delta(x, y) \stackrel{\text{def}}{=} \frac{\Delta(x,y)}{n}$.

Fait: La distance de Hamming est une métrique.

1. $\Delta(x, y) \geq 0, \forall x, y \in \Sigma^n$, avec égalité ssi $x = y$
2. $\Delta(x, y) = \Delta(y, x)$
3. $\Delta(x, y) + \Delta(y, z) \geq \Delta(x, z)$

2.1.1 Codes

Soit $\mathcal{C} \subseteq \Sigma^n$.

1. $\mathcal{C}$ corrige t erreurs si tout motif de au plus t erreurs peut être corrigé (par un décodage possiblement inefficace).
   Formellement:
   - $B(x, t) \stackrel{\text{def}}{=} \{y \in \Sigma^n : \Delta(x, y) \leq t\}$
   - $\mathcal{C}$ corrige t erreurs si $\forall x, y \in \mathcal{C}$ avec $x \neq y$, $B(x, t) \cap B(y, t) = \emptyset$.
2. $\mathcal{C}$ détecte $e \geq 1$ erreurs si tout motif d'au plus t erreurs peut être détecté.
   Formellement:
   $\forall x \in \mathcal{C}, B(x, e) \cap C = \{x\}$

3. On appelle distance d'un code $\Delta(\mathcal{C})$, la distance minimale qui sépare deux mots d'un code:

$$\Delta(\mathcal{C}) = \min_{x,y \in \mathcal{C}, x \neq y} \Delta(x, y)$$

Proposition 1 *Les conditions suivantes sont équivalentes:*

1. $\mathcal{C}$ *corrige* t *erreurs*
2. $\mathcal{C}$ *détecte* $2t$ *erreurs*
3. $\Delta(\mathcal{C}) \geq 2t + 1$

Preuve

- $3 \rightarrow 1$:
  $\Delta(\mathcal{C}) \geq 2t + 1 \Rightarrow$ Les boules $B(x, t)$ ne se recouvrent pas $\Rightarrow$ on associe à $y \in \Sigma^n$ le décodage "plus proche voisin"

  $$\Phi(y) = \arg\min_{x \in \mathcal{C}} \Delta(x, y)$$

  Ce décodeur corrige bien t erreurs.

- $\neg 3 \rightarrow \neg 1$:
  $\Delta(\mathcal{C}) \leq 2t \Rightarrow \exists$ 2 mots codes x_1 et $x_2 \in \mathcal{C}$ dont les boules de rayon t se recouvrent: $B(x_1, t) \cap B(x_2, t) \neq \emptyset$.
  Si y appartient à cette intersection $\rightarrow$ problème pour décoder.

- $3 \rightarrow 2$:
  $\forall x \in \mathcal{C}, B(x, 2t) \cap C = \{x\}$
  On considère le décodage:
  Si $\left|\begin{array}{l} y^n = x^n \in \mathcal{C}, \text{ on déclare } x^n. \\ y^n \in \cup_x B(x, 2t) \backslash \mathcal{C}, \text{ on déclare "erreur".} \\ y \notin \cup_x B(x, 2t), \text{ on déclare n'importe quel mot code.} \end{array}\right.$
  Ce décodeur détecte bien $2t$ erreurs.

- $\neg 3 \rightarrow \neg 2$:
  $\Delta(\mathcal{C}) \leq 2t \Rightarrow$ 2 mots codes appartiennent à une même boule $\Rightarrow$. Si y est égal à l'un de ces mots codes il n'est pas possible de savoir si y correspond à un mot code où s'il s'agit d'une version bruitée d'un mot code.

■

Proposition 2 *Soit q, n des entieres t.q. $q \geq 2$ et $n \geq 1$.*

1. $\left|B_q^n(x,t)\right| = \sum_{i=0}^{t} \binom{n}{i}(q-1)^i \stackrel{\text{def}}{=} \text{Vol}_q(n,t)$
2. *Si $\mathcal{C}$ corrige t erreurs* $\Rightarrow |\mathcal{C}| \leq \frac{q^n}{\text{Vol}_q(n,t)}$
3. *Soit*
$$H_q(p) \triangleq p\log_q(q-1) - p\cdot\log_q(p) - \overline{p}\cdot\log_q(\overline{p})$$
(ici $\overline{p} = 1-p$). Alors pour $0 \leq p \leq 1 - 1/q$
   (a) $\text{Vol}_q(n, np) \leq q^{nH_q(p)}$ *pour tout np entier*
   (b) $\text{Vol}_q(n, np) \geq q^{n(H_q(p)-o(1))}$ *pour n suffisament grand*

Observation 3 *Pour $q = 2$, $n = 63$, $t = 1$ on a* $\text{Vol}(63,1) = 64 \Rightarrow |\mathcal{C}| \leq \frac{2^{63}}{64} = 2^{57}$
$\Rightarrow$ *Taux $\frac{57}{63}$ optimal (Solution 2 Hamming).*

Preuve

1. $\binom{n}{i}(q-1)^i$ représente le nombre de séquences de longueur n qui diffèrent d'une séquence donnée sur i coordonnées exactement.
2. Si $\mathcal{C}$ corrige t erreurs alors pour tout $x, y \in \mathcal{C}$ on a $B_q^n(x,t) \cap B_q^n(y,t) = \emptyset$ d'où
$$q^n \geq |\cup_{x\in\mathcal{C}} B_q^n(x,t)| = |\mathcal{C}|\cdot\text{Vol}_q(n,t)$$
3. (a) Puisque $0 \leq p \leq 1 - 1/q$ on a
$$0 \leq \frac{p}{(q-1)\bar{p}} \leq 1.$$

$$\begin{aligned}
\mathrm{Vol}_q(n, np) \cdot q^{-nH_q(p)} &= \sum_{i=0}^{np} \binom{n}{i} (q-1)^i (q-1)^{-np} p^{np} \bar{p}^{n\bar{p}} \\
&= \sum_{i=0}^{np} \binom{n}{i} (q-1)^i \bar{p}^n \left[\frac{p}{(q-1)\bar{p}}\right]^{np} \\
&\leq \sum_{i=0}^{np} \binom{n}{i} (q-1)^i \bar{p}^n \left[\frac{p}{(q-1)\bar{p}}\right]^{i} \\
&= \sum_{i=0}^{np} \binom{n}{i} \bar{p}^{n-i} p^i \\
&\leq \sum_{i=0}^{n} \binom{n}{i} \bar{p}^{n-i} p^i \\
&= 1.
\end{aligned}$$

où la dernière inégalité vient de l'identité

$$\sum_{i=0}^{n} \binom{n}{i} \cdot p^i \cdot \overline{p}^{n-i} = 1.$$

(b) En utilisant une version grossière de la formule de Stirling

$$k! = k^k \cdot e^{-k} \mathrm{poly}(k)$$

où $\mathrm{poly}(k)$ est un terme polynomiale en k (i.e., $k^\alpha \leq \mathrm{poly}(k) \leq k^\beta$ pour certains $0 < \alpha \leq \beta$ et k suffisamment grand) on a

$$\begin{aligned}
\sum_{i=0}^{np} \binom{n}{i} (q-1)^i &\geq \binom{n}{np} (q-1)^{np} \\
&= \left(\frac{1}{p}\right)^{pn} \left(\frac{1}{\bar{p}}\right)^{\bar{p}n} (q-1)^{np} \mathrm{poly}(n) \\
&= 2^{nH_q(p)} \mathrm{poly}(n) \\
&\geq 2^{n(H_q(p) - o(1))}
\end{aligned}$$

pour n suffisamment grand.

■

3 Bornes fondamentales sur les codes

Un code $\mathcal{C} \subseteq \Sigma^n$ sur un alphabet Σ est noté $(n, k, d)_q$ où

- $q = |\Sigma|$ est la taille de l'alphabet
- $|\mathcal{C}| \geq q^k$
- $\Delta(\mathcal{C}) \geq d$.

Parfois on utilise la notation (n, M, d) avec $M = q^k$.

Le but ici est la charactérisation de la région de faisabilité de $(n, k, d)_q$. Bien que ce problème reste partiellement ouvert, on établira des conditions nécessaires et des conditions suffisantes pour l'existence de codes pour des paramétres donnés. En particulier, on s'intéressera aux paires (R, δ) asymptotiquement atteignable, i.e., pour lesquels il existe des suites de codes $\{(n, k(n), d(n))\}_{n\geq}$ où $R \triangleq \liminf_{n\to\infty} \frac{k(n)}{n}$ et $\delta \triangleq \liminf_{n\to\infty} \frac{d(n)}{n}$.[1]

3.1 Bornes Supérieures

Theorem 4 (Singleton) *Pour tout $q \geq 2$ on a $k+d \leq n+1$ d'où $R+\delta \leq 1$.*

Preuve
Soit $(n, k, d)_q$ un code. On définit la projection sur les $k-1$ premières composantes

$$\pi : \Sigma^n \to \Sigma^{k-1} \quad \pi(x^n) \triangleq x_1, x_2 ... x_{k-1}.$$

Puisque $|\mathcal{C}| \geq q^k$ on a $|\mathcal{C}| > q^{k-1}$ et par le principe des niches de pigeons il existe x^n et y^n tels que $\pi(x^n) = \pi(y^n)$ et donc tels que $\Delta(x^n, y^n) \leq n - (k-1) = n - k + 1$. Par suite

$$d \leq \Delta(\mathcal{C}) \leq \Delta(x^n, y^n) \leq n - k + 1.$$

■

Theorem 5 (Hamming) *Pour tout entier $q \geq 2$ et $R, \delta \in [0, 1]$ on a*

$$R + H_q(\delta/2) \leq 1.$$

[1] Il est clair que R est une fonction non-croissante de δ.

Preuve Soit $d = \lfloor \delta n \rfloor$. Pour tout $x, y \in \mathcal{C}$ on a

$$B_q^n(x, \lfloor (d-1)/2 \rfloor) \cap B_q^n(y, \lfloor (d-1)/2 \rfloor) = \emptyset$$

d'où $|\mathcal{C}| \cdot \mathrm{Vol}_q(n, \lfloor (d-1)/2 \rfloor]) \leq q^n$ et donc

$$q^k \cdot 2^{n(H_q(\frac{\delta}{2}) - o(1))} \leq q^n.$$

En fixant le rapport $R = k/n$ et en prenant la limite $n \to \infty$ le résultat suit. ■

Theorem 6 (Plotkin) *Pour tout* $R, \delta \in [0, 1]$

$$R \leq \begin{cases} 1 - \frac{\delta}{\theta} & \delta \leq \theta \\ 0 & \delta > \theta \end{cases}$$

avec $\theta \triangleq 1 - \frac{1}{q}$.

Preuve
Cas $\delta > \theta$: Soit $\Delta(\mathcal{C}) \geq d$ et $|\mathcal{C}| = M$. On considère la quantité auxiliaire

$$S \triangleq \sum_{x,y \in \mathcal{C}} \Delta(x, y) \geq d \cdot M \cdot (M-1). \tag{1}$$

On remarque que S est une somme de contributions de colonnes de la matrice $q^k \times n$ correspondant aux q^k mots codes écrit en ligne:

$$\begin{pmatrix} x_1(1) & x_2(1) & . & . & . & x_n(1) \\ x_1(2) & . & . & . & . & x_n(2) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_1(M) & . & . & . & . & x_n(M) \end{pmatrix}$$

On peut donc écrire

$$S = S_1 + S_2 + ... + S_l + ... + S_n$$

avec S_l la contribution de la l-ième colonne.

Calcul de S_l: Soit ${n_i}^l$ le nombre de fois où l'élément i apparait dans la colonne l. Afin de calculer S_l on somme la contribution de chaque élément

$i \in \Sigma$ de la colonne l. Cette contribution est le produit entre le nombre d'apparitions de i dans la colonne l et le nombre d'apparitions d'éléments différents de i dans la colonne l. Il suit que

$$S_l = \sum_{i=1}^{q} {n_i}^l \cdot (M - {n_i}^l) = \sum_{i=1}^{q} {n_i}^l \cdot M - \sum_{i=1}^{q} {n_i}^2 = M^2 - \sum_{i=1}^{q} {n_i}^2.$$

D'après l'inégalité de Cauchy-Schwarz on a $\left|\sum_{i=1}^{q} n_i m_i\right|^2 \leq \sum_{i=1}^{q} |n_i|^2 \cdot \sum_{i=1}^{q} |m_i|^2$ et donc pour $m_i = 1$ l'inégalité s'écrit $\left|\sum_{i=1}^{q} n_i\right|^2 \leq \sum_{i=1}^{q} |n_i|^2 \cdot n$. Cette inégalité donne $S_l \leq M^2 - \frac{M^2}{q}$ et on conclut

$$S = \sum_{l=1}^{n} S_l \leq n(M^2 - \frac{M^2}{q}) = nM^2(1 - \frac{1}{q}). \tag{2}$$

De (1) et (2) on déduit

$$nM^2(1 - \frac{1}{q}) \geq dM(M-1)$$

ce qui implique avec $M \geq q^k$ que

$$q^k \leq \frac{d}{d - \theta n} = \frac{qd}{qd - (q-1)n}. \tag{3}$$

Donc pour $\theta < \delta$ on a $R = 0$.
Cas $\delta \leq \theta$: On prend n' tel que $\theta n' \approx d$, plus précisément on définit $n' = \lfloor \frac{d}{\theta} - \frac{1}{q-1} \rfloor$.

On groupe les mots code qui sont les même sur les $n - n'$ premières positions et on définit les sous-codes

$$\mathcal{C}_x \triangleq \{(c_{n-n'+1}, \ldots, c_n) : (c_1, c_2, \ldots, c_n) \in \mathcal{C}\ , (c_1, c_2, \ldots, c_{n-n'}) = x\}$$

En appliquant (3) au code $\mathcal{C}_x$ en remplaçant q^k par $|\mathcal{C}_x|$ et n par n' on obtient[2]

$$|\mathcal{C}_x| \leq \frac{qd}{qd - (q-1)n'} \leq qd$$

[2]On peut appliquer (3) car clairement $\Delta(\mathcal{C}_x) \geq d$ et notre choix de n' satisfait $\theta n' < d$.

où la deuxième inégalite suit de la définition de n' qui guarantit que $qd - (q-1)n' \geq 1$. On déduit que

$$|\mathcal{C}| = \sum_{x \in q^{n-n'}} |\mathcal{C}_x| \leq qd \cdot q^{n-n'} = q^{n - \frac{d}{\theta} + O(\log_q d)}$$

d'où $R \leq 1 - \frac{\delta}{\theta}$. ■

3.2 Bornes Inférieures

Theorem 7 (Gilbert-Varshamov) *$R \geq 1 - H_q(\delta)$ pour $0 \leq \delta \leq 1 - 1/q$.*

Preuve
Fixer $0 \leq \delta \leq 1 - 1/q$ et considérer $d = \delta n$ (n suffisament grand).
On considère la construction suivante:

- Initialisation : $\mathcal{C} \leftarrow \emptyset$, $S = \Sigma^n$ = espace entier
- while $S \neq \emptyset$ do
  - choisir $x \in S$
  - $\mathcal{C} \leftarrow \mathcal{C} \cup \{x\}$
  - $S \leftarrow S \backslash B_q^n(x, d-1)$
- end while
- output $\mathcal{C}$

Propriétés:

- $\Delta(\mathcal{C}) \geq d$
- $|\mathcal{C}| \geq \frac{q^n}{\mathrm{Vol}_q(n, d-1)}$

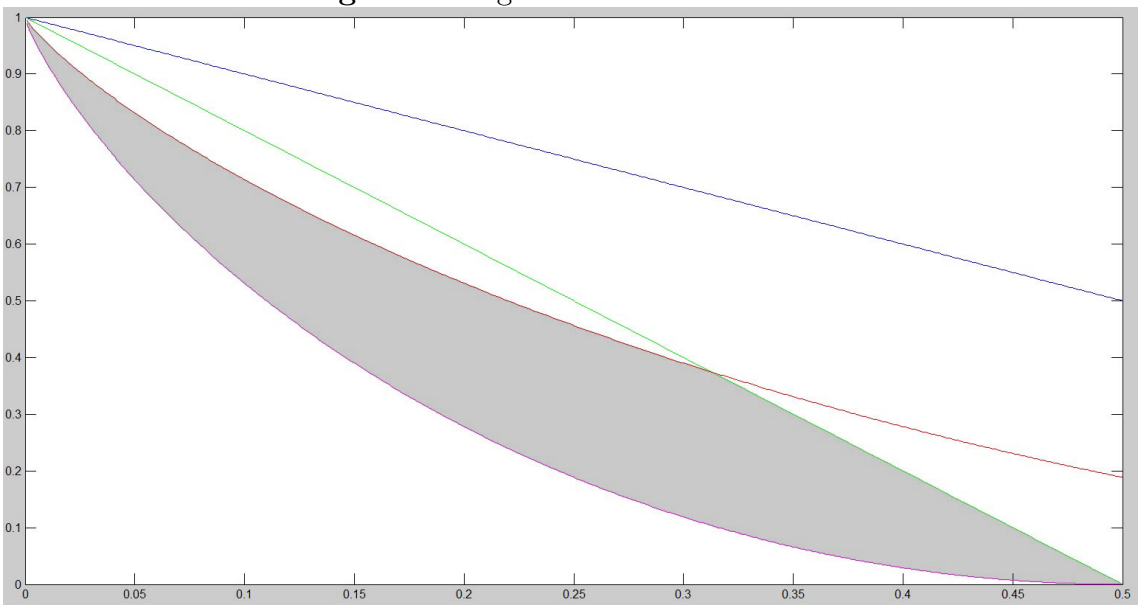
Puisque $\mathrm{Vol}_q(n, d-1) \doteq q^{nH_q(\delta)}$ on a $R \geq 1 - H_q(\delta)$. ■

4 Comparaison pour $q = 2$

- Singleton: $R + \delta \leq 1$ (bleu)
- Hamming: R + $H(\delta/2) \leq 1$ (rouge)
- Plotkin: R $\leq \max\{1 - 2\delta, 0\}$ (vert)
- Gilbert-Varshamov: $R \geq 1 - H(\delta)$ (courbe jaune)

La frontière entre (R, δ) atteignables et non-atteignables est inclue dans la région grisée. Notons qu'il existe de meilleurs bornes, par exemple la borne supérieure d'Elias-Bassalygo est meilleure que la borne Plotkin-Hamming.

Figure 1: Région de faisabilité



Remark

- Modèle pire-des-scénarios de Hamming: on peut corriger une fraction $\frac{\delta}{2}$ d'erreurs à taux $R \geq 1 - H(\delta)$ (Gilbert-Varshamov). Par Hamming, tout code corrigeant $\leq \frac{\delta}{2}$ erreurs a un taux $R \leq 1 - H(\frac{\delta}{2})$.
- Modèle probabiliste de Shannon: on peut corriger $\frac{\delta}{2} \pm \varepsilon$ erreurs avec probabilité $1 - \varepsilon$ et taux maximal $1 - H(\frac{\delta}{2})$.