1.3 Exercices TD 1

Exercice 1.1.

1) Établir que

$$\forall\, v, w \in \{0,1\}^m,\ d_H(v,w) = H(v+w) \leq H(v) + H(w).$$

2) En déduire que

$$\forall\, u, v, w \in \{0,1\}^m,\ d_H(u,v) = d_H(u+w, v+w) \text{ et } d_H(u,v) = d_H(w, u+v+w).$$

3) Vérifier que d_H est bien une distance sur $\{0,1\}^m$.

4) Représenter dans l'espace à 3 dimensions l'ensemble $L = \{0,1\}^3$. À quoi correspondent les arêtes du volume obtenu ? Où se situent les mots distants de 2 au sens de la distance de

Hamming ? Combien il y a t-il de couples différents de tels mots (attention au fait que $(u, v) \neq (v, u)$) ?

Exercice 1.2.

Cet exercice propose entre autre une démonstration du théorème 1.1. On considère donc un code $\mathcal{C}$ avec correction suivant la méthode du maximum de vraisemblance et on note $k = \lfloor(\delta(\mathcal{C}) - 1)/2\rfloor$.

On note $w \in \mathcal{C}(\{0,1\}^m)$ le mot émis et $\tilde{w}$ le mot reçu, tels que $d_H(w, \tilde{w}) \leq k$. Soient $\widehat{w_1}, \widehat{w_2} \in \mathcal{C}(\{0,1\}^m)$ vérifiant la propriété

$$\forall\, v \in \mathcal{C}(\{0,1\}^m),\ d_H(\widehat{w_1}, \tilde{w}) = d_H(\widehat{w_2}, \tilde{w}) \leq d_H(v, \tilde{w}). \tag{1.11}$$

1) Montrer que $d_H(\widehat{w_1}, \widehat{w_2}) \leq 2k$.

2) En déduire que $\widehat{w_1} = \widehat{w_2}$.

3) En déduire que la méthode du maximum de vraisemblance donnée par l'équation (1.6) du cours est bien définie. Cela suffit-il à démontrer le théorème 1.1 ?

4) Démontrer le théorème 1.1.

5) Que peut-on dire quand

$$\left\lfloor \frac{\delta(\mathcal{C}) - 1}{2} \right\rfloor \ < \ d_H(w, \tilde{w}) \ \leq \ \delta(\mathcal{C}) - 1 \ ?$$

Donner deux exemples de mots codés illustrant deux situations problématiques différentes.

Correction

1) Comme $w \in \mathcal{C}(\{0,1\}^m)$, on peut prendre dans (1.11) $v = w$, ce qui montre que

$$d_H(\widehat{w_1}, \tilde{w}) = d_H(\widehat{w_2}, \tilde{w}) \leq k.$$

L'inégalité triangulaire permet alors de conclure :

$$d_H(\widehat{w_1}, \widehat{w_2}) \leq d_H(\widehat{w_1}, \tilde{w}) + d_H(\widehat{w_2}, \tilde{w}) \leq 2k.$$

2) On a ainsi

$$d_H(\widehat{w_1}, \widehat{w_2}) \leq 2k \leq \delta(\mathcal{C}) - 1 < \delta(\mathcal{C}).$$

Comme $\widehat{w_1}, \widehat{w_2} \in \mathcal{C}(\{0,1\}^m)$, par la définition 1.8 on déduit que $\widehat{w_1} = \widehat{w_2}$.

3) La question 2) montre que le problème d'optimisation $\min_{v \in \mathcal{C}(\{0,1\}^m)} d_H(\tilde{w}, v)$ admet une unique solution, notée $\widehat{w}$. Ceci montre que la méthode du maximum de vraisemblance donnée par (1.6) dans le cours est bien définie et qu'on peut donc l'appliquer. Mais pour démontrer le théorème 1.1 il faut également établir que cette méthode permet de retrouver w, c'est-à-dire que $\widehat{w} = w$.

4) On a par hypothèse $d_H(w, \tilde{w}) \leq k$ et la question 1) a établi que $d_H(\widehat{w}, \tilde{w}) \leq k$. Or, par le même raisonnement qu'en 1)-2), on voit que w est l'unique mot du code vérifiant $d_H(w, \tilde{w}) \leq k$: on en déduit que $\widehat{w} = w$, c'est-à-dire que la procédure d'optimisation donnée par (1.6) dans le cours conduit à la solution w.

5) Sous ces hypothèses, la proposition 1.2 assure qu'on sait détecter tout mot reçu comportant k erreurs. Mais la méthode du maximum de vraisemblance ne s'applique pas toujours, car le problème d'optimisation n'admet pas nécessairement une unique solution (premier cas). On peut aussi rencontrer le cas où le problème d'optimisation admet une solution unique, mais qui ne correspond pas au mot émis (second cas).

Exemple pour le premier cas

On prend $m = 1$, $n = 2$ et le code $\mathcal{C}$ défini par $\mathcal{C}(0) = 00$, $\mathcal{C}(1) = 11$. On a donc $\delta(\mathcal{C}) = 2$. Soient $w = 00$ le mot émis et $\tilde{w} = 01$ le mot reçu, de sorte que $d_H(w, \tilde{w}) = k := 1$. Le mot reçu est bien détecté comme erroné, mais le problème d'optimisation admet deux solutions, puisque $d_H(00, \tilde{w}) = d_H(11, \tilde{w}) = 1$: on ne sait pas choisir entre 00 et 11.

Exemple pour le second cas

On prend $m = 1$, $n = 3$ et le code $\mathcal{C}$ défini par $\mathcal{C}(0) = 000$, $\mathcal{C}(1) = 111$. On a donc $\delta(\mathcal{C}) = 3$. Soient $w = 000$ le mot émis et $\tilde{w} = 011$ le mot reçu, de sorte que $d_H(w, \tilde{w}) = k := 2$. Le mot reçu est bien détecté comme erroné et le problème d'optimisation admet la solution unique $\widehat{w} = 111$, mais le mot décodé n'est pas le bon.

Exercice 1.3.

On considère l'application $\mathcal{C}$ de $\{0, 1\} \mapsto \{0, 1\}^2$ définie par $\mathcal{C}(0) = 00$, $\mathcal{C}(1) = 11$.

1) Pourquoi cette application est-elle un code ? Quel est l'ensemble des mots du code ?

2) Déterminer sa redondance, son rendement et sa distance minimale.

3) Que peut-on dire sur la capacité de ce code à détecter les erreurs de transmission ?

4) Que peut-on dire sur la capacité de ce code à corriger les erreurs de transmission ?

5) On modélise le bruit du canal de transmission de la manière suivante, appelée *schéma de Bernoulli* : la probabilité p que le bit d'un mot de code soit mal transmis est identique pour tout bit et tout mot et les erreurs sur les bits sont indépendantes les unes des autres. En outre, on suppose que les mots source sont équiprobables. L'expérience aléatoire consiste à appliquer le code à un mot source, à transmettre le mot de code dans le canal bruité et à observer le mot reçu.

Déterminer l'ensemble des résultats possibles de l'expérience, appelé *univers*

$$\Omega = \{e_1, e_2, \ldots, e_N\},$$

les éléments $(e_i)_{i \in [\![1,N]\!]}$ étant les évènements élémentaires. Calculer les probabilités $(P(e_i))_{i \in [\![1,N]\!]}$ de ces évènements élémentaires.

6) On note D l'évènement "un mot reçu est détecté erroné" et E l'évènement "un mot reçu

est erroné". Déterminer ces évènements ainsi que l'évènement $D \cap E$ comme parties de Ω et en déduire leur probabilité.

7) Exprimer en fonction de p la probabilité qu'un mot reçu erroné soit effectivement détecté comme étant erroné. Tracer le graphe de la fonction obtenue.

Correction

1) L'application $\mathcal{C} = \mathcal{C}_{2,1}$ ($m = 1$,$n = 2$) ainsi définie étant injective, elle est bien un code au sens de la définition 1.4. L'ensemble des mots du code est $\mathcal{C}(\{0,1\}) = \{00, 11\}$.

On peut remarquer qu'il s'agit d'un code de parité paire (cf. exemple 1.7) et aussi d'un code de répétition (cf. exercice 1.4).

2) Sa redondance est $n - m = 1$, son rendement $\rho = m/n = 1/2$ et sa distance minimale $\delta(\mathcal{C}) = 2$.

3) La proposition 1.2 assure que ce code est un 1-détecteur : il détecte toute transmission de mots erronés dont l'erreur porte sur 1 bit. Si par contre l'erreur porte sur les 2 bits, elle n'est pas détectée et le mot décodé est faux.

4) Ce code ne peut corriger aucune erreur de transmission, car il n'est pas possible de savoir quel bit est erroné.

5) Les évènements élémentaires sont constitués de toutes les possibilités de mot reçu associé à un mot source. On a ainsi, un couple signifiant (mot source, mot reçu),

$$\Omega = \{0,1\} \times \{0,1\}^2 = \{(0,00), (0,01), (0,10), (0,11), (1,00), (1,01), (1,10), (1,11)\}$$

et

$$N = |\Omega| = 2 \times 2^2 = 8.$$

Attention au fait que ces évènements élémentaires ne sont pas équiprobables, puisque le nombre de bits mal reçu est variable d'un couple à l'autre. En considérant que la probabilité d'apparition du mot source 0 est la même que celle du mot source 1, soit $1/2$, et en posant $q = 1 - p$ la probabilité qu'un bit du mot de code soit bien transmis, il vient

Évènement élémentaire	Probabilité
$e_1 = (0,00)$	$q^2/2$
$e_2 = (0,01)$	$pq/2$
$e_3 = (0,10)$	$pq/2$
$e_4 = (0,11)$	$p^2/2$
$e_5 = (1,00)$	$p^2/2$
$e_6 = (1,01)$	$pq/2$
$e_7 = (1,10)$	$pq/2$
$e_8 = (1,11)$	$q^2/2$

On vérifie que l'on a bien pour tout $p \in [0,1]$,

$$\forall\, i \in [\![1, N]\!],\ 0 \leq P(e_i) \leq 1 \text{ et } \sum_{i=1}^{N} e_i = 1.$$

6) On a

$$E = \{(0,01),(0,10),(0,11),(1,00),(1,01),(1,10)\} = \Omega \setminus \{(0,00),(1,11)\}$$

et

$$D = \{(0,01),(0,10),(1,01),(1,10)\} = D \cap E.$$

Les évènements élémentaires étant disjoints, on en déduit que

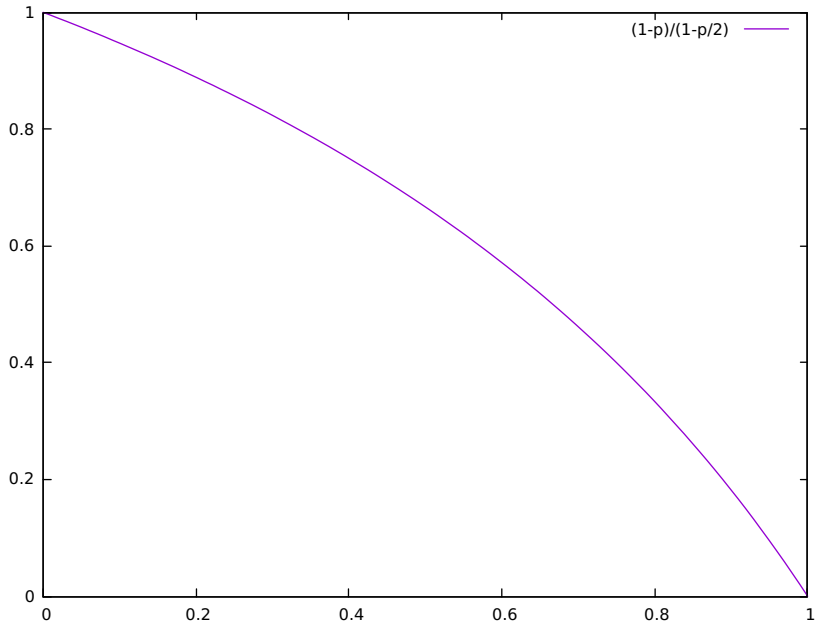
$$P(E) = 1 - P(\{(0,00),(1,11)\}) = 1 - \left(\frac{q^2}{2} + \frac{q^2}{2}\right) = 1 - q^2;$$
$$P(D) = P(D \cap E) = pq/2 + pq/2 + pq/2 + pq/2 = 2pq.$$

7) Il s'agit d'exprimer la probabilité qu'un mot reçu soit détecté erroné alors qu'il est erroné, soit la probabilité conditionnelle de l'évènement D sachant que l'évènement E s'est réalisé :

$$p_{de}(p) := P(D|E) = \frac{P(D \cap E)}{P(E)} = \frac{2pq}{1-q^2} = \frac{2p(1-p)}{1-(1-p)^2} = \frac{2p(1-p)}{2p-p^2} = \frac{1-p}{1-p/2}.$$

Le graphe de cette fonction est tracé à la figure 1.2.

FIGURE 1.2 – Probabilité qu'un mot reçu erroné soit effectivement détecté comme étant erroné en fonction de la probabilité p qu'un bit soit mal transmis, pour le code considéré à l'exercice 1.3.



Exercice 1.4.

Soit $t \in \mathbb{N}$, $t \geq 2$. Un *code de répétition* $\mathcal{C} = \mathcal{C}_{tm,m}$ consiste à répéter t fois le mot à transmettre. Ainsi, si $u \in \{0,1\}^m$ est le mot source, le mot transmis est $w = u^t$. On a donc $n = tm$.

1) Quel est le rendement d'un code de répétition ?

2) Montrer que la distance minimale $\delta(\mathcal{C})$ vérifie $\delta(\mathcal{C}) \geq t$.

3) Déterminer la valeur $\delta(\mathcal{C})$.

4) Donner t pour que le code de répétition soit un 2-détecteur.

5) Comment le décodeur d'un tel code peut-il détecter au plus 2 erreurs ?

6) Donner t pour que le code de répétition soit un 2-correcteur.

7) Comment le décodeur d'un tel code peut-il corriger au plus 2 erreurs ?

8) Le canal de transmission est modélisé suivant le *schéma de Bernoulli* : la probabilité p que le bit d'un mot de code soit mal transmis est identique pour tout bit et tout mot et les erreurs sur les bits sont indépendantes les unes des autres. On suppose maintenant que le mot source est réduit à un bit : $m = 1$. En utilisant la variable aléatoire X du nombre de bits erronés dans un mot reçu, déterminer en fonction de t et de p la probabilité p_{de} qu'un mot reçu erroné soit effectivement détecté erroné.

9) Application numérique : calculer la probabilité p_{de} dans le cas où $t = 3$ et $p = 0.1$.

10) Pour ces mêmes valeurs numériques, quelle est la probabilité p_{me} qu'un mot reçu erroné soit mal décodé ?

Exercice 1.5.

1) Quelle est la capacité de correction maximale que l'on puisse obtenir avec un code $\mathcal{C}_{5,2}$?

2) Construire un code possédant une telle capacité de correction.

3) Quelle est la plus petite longueur n possible pour qu'un code $\mathcal{C}_{n,4}$ possède une capacité de correction d'une erreur ? Que peut-on alors dire d'un tel code ?

4) On cherche maintenant à encoder 4 mots binaires avec une capacité de correction de 2 erreurs. Quelle est la distance minimale d'un tel code ?

5) Quelle est la longueur minimale d'un tel code ?

Correction

1) La contrainte dans le cours (1.8) de la borne d'empilement des sphères affirme que si k est la capacité de correction d'un code $\mathcal{C}_{5,2}$, alors

$$1 + \binom{5}{1} + \binom{5}{2} + \cdots + \binom{5}{k} \leq 2^{5-2} = 2^3 = 8.$$

Comme $1 + \binom{5}{1} = 6$ et $1 + \binom{5}{1} + \binom{5}{2} = 16$, on voit que la plus grande valeur possible est $k = 1$.

2) La distance maximale d'un tel code est $\delta(\mathcal{C}) = 3$, puisqu'alors $1 = \lfloor(\delta(\mathcal{C}) - 1)/2\rfloor$. Il s'agit donc de trouver des mots du code qui diffèrent les uns des autres de 3 bits. L'exemple suivant vérifie cette contrainte :

$u \in \{0,1\}^2$	$\mathcal{C}(u) \in \{0,1\}^5$
00	00000
01	01011
10	10101
11	11110

3) Dans le cas $m = 4$ et $k = 1$, la contrainte dans le cours (1.8) de la borne d'empilement des sphères s'écrit

$$1 + \binom{n}{1} = 1 + n \leq 2^{n-4}$$

et le plus petit entier $n > m = 4$ qui la vérifie est $n = 7$. Dans ce cas, l'inégalité de la contrainte est une égalité : un tel code est parfait (cf. définition 1.12).

4) Dans le cas $m = 2$ et $k = 2$, la distance minimale est $\delta(\mathcal{C}) = 5$, puisqu'alors $2 = \lfloor(\delta(\mathcal{C}) - 1)/2\rfloor$.

5) Toujours avec $m = 2$ et $k = 2$, la contrainte (1.8) dans le cours de la borne d'empilement des sphères s'écrit

$$1 + \binom{n}{1} + \binom{n}{2} = 1 + n + \frac{n(n-1)}{2} \leq 2^{n-2}$$

et le plus petit entier $n > m = 2$ qui la vérifie est $n = 7$. En effet, pour $n = 6$ on a

$$1 + n + \frac{n(n-1)}{2} = 7 + \frac{6 \times 5}{2} = 7 + 15 = 22 > 2^4 = 16$$

et pour $n = 7$,

$$1 + n + \frac{n(n-1)}{2} = 8 + \frac{7 \times 6}{2} = 8 + 21 = 29 \leq 2^5 = 32.$$

2.6 Exercices TD 2

Exercice 2.1.

Pour les différentes applications ci-après définies, dire s'il s'agit d'un code et s'il est linéaire (auquel cas on donnera la matrice génératrice), calculer sa distance minimale, dire s'il possède une capacité de correction et mentionner les codes appartenant à une classe connue.

$u \in \{0,1\}^3$	$\mathcal{C}_1(u)$	$\mathcal{C}_2(u)$	$\mathcal{C}_3(u)$	$\mathcal{C}_4(u)$
000	0000000	0000000	0000000	0000000
001	0010110	0010111	0010011	1010101
010	0101000	0100111	0100101	1001011
011	0111110	0110110	0110110	0011110
100	1000101	1001011	1001001	1110010
101	1010011	1011010	1011010	0100111
110	1101101	1101100	1101100	0111001
111	1111011	1111111	1111111	1101100

Exercice 2.2.

Cet exercice propose une démonstration du théorème 2.5. On considère donc $\mathcal{C} = \mathcal{C}_{n,m}$ un code linéaire et Y une matrice de contrôle de $\mathcal{C}$ représentant une fonction syndrome S.

1) Montrer que s'il existe un mot de code de poids p, alors il existe p lignes de Y linéairement dépendantes.

2) Réciproquement, montrer que s'il existe p lignes de Y linéairement dépendantes, alors il existe un mot de code de poids $p' \in [\![1, p]\!]$.

3) En déduire la preuve du théorème 2.5.

Exercice 2.3.

On considère le code $\mathcal{C}$ de matrice génératrice

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

1) Déterminer une matrice de contrôle de ce code.

2) Calculer sa capacité de correction k.

3) Sous l'hypothèse d'au plus k erreurs, décoder par syndrome les mots 101111 et 111111.

Exercice 2.4.

On considère le code $\mathcal{C}$ de matrice génératrice

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

1) Quelle est la capacité de correction k de ce code ?

2) Déterminer une matrice de contrôle Y' d'un code systématique $\mathcal{C}'$ équivalent à $\mathcal{C}$, obtenu par une permutation adéquate des colonnes de G.

3) Sous l'hypothèse d'au plus k erreurs, en déduire le décodage par $\mathcal{C}$ suivant la méthode de correction par syndrome des mots 1111100, 1100111, 0100000, 1010101 et 1110101.

Exercice 2.5.

On considère le code linéaire systématique $\mathcal{C}$ dont une matrice de contrôle est

$$Y = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Pour modéliser les erreurs de transmission, on se place dans un *schéma de Bernoulli* : la probabilité p que le bit d'un mot de code soit mal transmis est identique pour tout bit et tout mot et les erreurs sur les bits sont indépendantes les unes des autres.

1) Donner la matrice génératrice de ce code, en vérifiant que Y la définit de manière unique.

2) De quel code s'agit-il ?

3) Déterminer l'ensemble des mots du code.

4) Quelles sont les capacités en détection et en correction de ce code ?

5) Pour un mot transmis w on reçoit le mot $\tilde{w}$, lequel n'est pas détecté erroné. Quelles sont les erreurs $w - \tilde{w}$ possibles ?

6) Quelles sont les probabilités de ces différentes erreurs ?

7) En déduire la probabilité qu'un mot soit détecté erroné.

8) Quelle est la probabilité qu'un mot erroné soit détecté erroné ?