Solutions des Exercices du cours de Théorie de l'Information et Codage
cours 2 du 22 février 2011.

1. Soit une source discrète sans mémoire d'entropie $H(U)$. On considère un encodage des suites de k lettres de la source dans des mots code de longueur n dans un alphabet de taille D. La fonction d'encodage doit être injective et P_E est la probabilité de l'événement: une suite de la source ne correspond à aucun mot code (la fonction d'encodage n'étant pas surjective). On a vu au premier cours que pour tout $\delta > 0$, si $n/k \geq \frac{H(U)+\delta}{\log D}$, alors on peut trouver une fonction d'encodage avec $P_E \to 0$ quand $k \to \infty$. Montrer que si $n/k \leq \frac{H(U)-\delta}{\log D}$ alors quelque soit l'encodage, on a $P_E \to 1$ quand $k \to \infty$.

   - Le nombre de mots code est au plus $D^n \leq 2^{k(H(U)-\delta)}$. Comme la probabilité de chaque $u^{(k)} \in B(k, \delta/2)$ (l'ensemble $B(k, \delta)$ a été défini en cours) est majorée par $2^{-k(H(U)-\delta/2)}$, la probabilité totale de toutes les suites dans $B(k, \delta)$ pour lesquelles un mot code est disponible est au plus $2^{-k(H(U)-\delta/2)}2^{k(H(U)-\delta)} = 2^{-k\delta/2}$. La probabilité totale des suites hors de $B(k, \delta/2)$ est $1 - P(B(k, \delta/2))$, donc la probabilité de l'ensemble des suites (dans $B(k, \delta/2)$ ou en dehors) pour lesquelles un mot code est disponible est majorée par:

$$1 - P_E \leq 1 - P(B(k, \delta/2)) + 2^{-k\delta/2}.$$

2. **Un test pour les codes non-ambigus.** Le but de cet exercice est de donner un algorithme qui permet de vérifier si un code est non-ambigu. Voici un exemple d'un code binaire ambigu:

$$C = \{1, 011, 01110, 1110, 10011\}. \tag{1}$$

   Le mot $w = 011101110011$ a deux factorisations:

$$w = (01110)(1110)(011) = (011)(1)(011)(10011).$$

   L'alphabet D-aire est noté $\mathcal{D}$. L'ensemble des mots sur $\mathcal{D}$ est noté $\mathcal{D}^*$. Pour $x, y \in \mathcal{D}^*$, on définit:

$$x^{-1}y = \{z \in \mathcal{D}^*;\ xz = y\} \text{ et, } xy^{-1} = \{z \in \mathcal{D}^*;\ x = zy\}.$$

   Pour des ensembles X, Y de $\mathcal{D}^*$, on étend ces définitions comme suit:

$$X^{-1}Y = \cup_{x\in X} \cup_{y\in Y} x^{-1}y \text{ et, } XY^{-1} = \cup_{x\in X} \cup_{y\in Y} xy^{-1}.$$

   Les puissances de X sont définies par $X^0 = \{e\}$ où e est le mot vide, $X^1 = X$ et $X^{n+1} = XX^n = \{xy,\ x \in X, y \in X^n\}$, pour $n \geq 1$.
   On voit un code D-aire C comme un sous-ensemble de $\mathcal{D}^+ = \mathcal{D}^* - e$. On définit alors

$$\begin{aligned} U_1 &= C^{-1}C - e, \\ U_{n+1} &= C^{-1}U_n \cup U_n^{-1}C, \text{ pour } n \geq 1. \end{aligned}$$

   Nous allons montrer le théorème suivant:

Théorème 0.1 *Le code $C \subset \mathcal{D}^+$ est un code non-ambigu si et seulement si aucun des ensembles U_n définis ci-dessus ne contient le mot vide.*

1) Ecrire U_1, U_2, U_3 pour le code binaire donné par (1). Que vaut U_1 pour un code instantané? Que valent les U_n pour l'exemple de code binaire vu en cours: $\{10, 00, 11, 110\}$?

2) Montrer par induction sur k que: pour tout $n \geq 1$ et $k \in \{1, \ldots, n\}$, on a $e \in U_n$ ssi il existe un mot $u \in U_k$ et des entiers $i, j \geq 0$ tels que:

$$uC^i \cap C^j \neq \emptyset \text{ et } i + j + k = n. \tag{2}$$

3) En déduire le Théorème 0.1.

1) Pour le code donné par (1), on a:

$$\begin{aligned} U_1 &= \{10, 110, 0011\}, \quad C^{-1}U_1 = \{0, 10\}, \quad U_1^{-1}C = \{011\}; \\ U_2 &= \{0, 10, 011\}, \quad C^{-1}U_2 = \{0, e\}, \quad U_2^{-1}C = \{11, 110, 011, e, 10\}. \end{aligned}$$

donc $e \in U_3$ et C est ambigu.

Pour un code instantané, on a $U_1 = \emptyset$.

Pour l'exemple vu en cours, on a $U_1 = U_n = \{0\}$.

2) On fait une induction décroissante sur k. Pour $k = n$: si $e \in U_n$, il suffit de prendre $u = e$, $i = j = 0$. Inversement si (2) est vérifiée pour $k = n$, on a $i = j = 0$ et donc $u = e$.

Soit $n > k \geq 1$, on suppose que l'equivalence est vraie pour $n, n-1, \ldots, k+1$. Si $e \in U_n$ alos par induction, il existe $v \in U_{k+1}$ tel que $vx = y$ avec $x \in C^i$ et $y \in C^j$ et $i + j + k + 1 = n$. Par définition de U_{k+1}, on a

- soit $zv = u$ avec $z \in C$ et $u \in U_k$. Dans ce cas, $ux = zvx = zy$ avec $x \in C^i$ et $zy \in C^{j+1}$ donc $uC^i \cap C^{j+1} \neq \emptyset$.
- soit $z = uv$ avec $z \in C$ et $u \in U_k$. Dans ce cas, $zx = uvx = uy$ avec $zx \in C^{i+1}$ et $y \in C^j$ donc $C^{i+1} \cap uC^j \neq \emptyset$.

Dans les deux cas (2) est satisfaite.

Inversement, supposons qu'il existe $u \in U_k$ et $i, j \geq 0$ avec

$$uC^i \cap C^j \neq \emptyset,\ i + j + k = n.$$

On peut donc écrire: $ux_1x_2 \ldots x_i = y_1y_2 \ldots y_j$. Si $j = 0$ alors $i = 0$ et $k = n$. Pour $j \geq 1$, on distingue à nouveau deux cas selon les longueurs respectives de u et y_1:

- si $u = y_1v$ pour un $v \in \mathcal{D}^+$, alors $v \in C^{-1}U_k \subset U_{k+1}$ et de plus $vx_1x_2 \ldots x_i = y_2 \ldots y_j$. Donc $vC^i \cap C^{j-1} \neq \emptyset$ et par l'hypothèse d'induction $e \in U_n$.
- si $y_1 = uv$ pour un $v \in \mathcal{D}^+$, alors $v \in U_k^{-1}C \subset U_{k+1}$ et $x_1x_2 \ldots x_i = vy_2 \ldots y_j$. Donc $C^i \cap uC^{j-1} \neq \emptyset$ et $e \in U_n$.

3) Si C est ambigu, alors il existe une relation:

$$x_1x_2 \ldots x_p = y_1y_2 \ldots y_q, \quad x_1 \neq y_1.$$

On peut supposer sans perte de généralité que $x_1 = y_1 u$ pour un $u \in \mathcal{D}^+$. On a alors $u \in U_1$ et $uC^{p-1} \cap C^{q-1} \neq \emptyset$, d'où $e \in U_{p+q-1}$.
Inversement si $e \in U_n$. Prenons $k = 1$ dans la question précédente: il existe $u \in U_1$ et des entiers $i, j \geq 0$ tels que $uC^i \cap C^j \neq \emptyset$. Comme $u \in U_1$, on a $xu = y$ pour $x, y \in C$ et $x \neq y$ car $u \neq e$. Il découle de $xuC^i \cap xC^j \neq \emptyset$ que $yC^i \cap xC^j \neq \emptyset$ montrant que C est ambigu.