

# Théorie des codes

Silvain Rideau

8 janvier 2009

Coder, c'est-à-dire remplacer des lettres par d'autres lettres ou ensembles de lettres, est non seulement utile en cryptologie où il sert à cacher le sens des messages, mais est aussi la base de la compression de données. Ce document a pour but de formaliser cette notion de code avant de présenter quelques résultats liés entre autres à la reconnaissance des codes mais aussi de certaines de leurs propriétés comme la maximalité ou la complétude. Tous ces résultats sont extraits de [1].

## Table des matières

|   |          |
|---|----------|
| <b>1 Définitions</b>                            | <b>1</b> |
| <b>2 Algorithme de reconnaissance des codes</b> | <b>3</b> |
| <b>3 Mesure d'un code</b>                       | <b>5</b> |
| <b>4 Codes complets</b>                         | <b>9</b> |

## 1 Définitions

**Définition 1** (Code). Soit  $A$  un alphabet. Un sous-ensemble  $X$  de  $A^*$  est un code sur  $A$  si  $\forall m, n \geq 1$  et  $\forall (x_i)_{i=1\dots n}, (x'_i)_{i=1\dots m} \in X$  on a :

$$x_1x_2 \cdots x_n = x'_1x'_2 \cdots x'_m \Rightarrow n = m \text{ et } \forall i \in 1 \dots n \ x_i = x'_i$$

**Propriété 2** (Premières propriétés des codes). Soit  $X$  un code sur  $A$  :

- i.  $\epsilon \notin X$ .
- ii.  $\forall Y \subset X, Y$  est un code.
- iii. Soit  $B$  un alphabet, tout morphisme  $\phi : B^* \rightarrow A^*$  qui induit une bijection de  $B$  sur  $X$  est injectif.  
Réciproquement, s'il existe un morphisme injectif  $\phi : B^* \rightarrow A^*$  tel que  $X = \phi(B)$  alors  $X$  est un code.

Cette dernière propriété (qui pourrait aussi servir de définition aux codes) traduit la notion intuitive de code. En effet le morphisme de codage  $\phi$  permet de coder les mots de  $B^*$  dans  $A^*$  et l'injectivité permet d'assurer que le décodage est possible.

*Démonstration.* i.  $\epsilon = \epsilon\epsilon$  donc  $\epsilon$  n'a pas une unique factorisation.

- ii. Toute factorisation d'un mot  $w$  dans  $Y$  est une factorisation dans  $X$  et est donc unique.
- iii. Soit  $\phi : B^* \rightarrow A^*$  qui induit une bijection de  $B$  sur  $X$ . Soient  $u, v \in (B^*)^2$  tels que  $\phi(u) = \phi(v)$ .

Si  $u = \epsilon$ , supposons  $v \neq \epsilon$  alors  $v$  contient au moins une lettre  $b$  et par hypothèse  $\phi(b) \in X$  or  $\epsilon \notin X$  donc  $|\phi(b)| > 0$ . On en déduit que  $|\phi(v)| > 0$  ce qui est absurde car  $\phi(u) = \epsilon$

Sinon  $u = b_1 \cdots b_n$  et  $v = b'_1 \cdots b'_m$ . On a alors  $\phi(b_1) \cdots \phi(b_n) = \phi(b'_1) \cdots \phi(b'_m)$  avec  $\phi(b_i), \phi(b'_j) \in X$ . Or  $X$  est un code donc  $n = m$  et  $\forall i \phi(b_i) = \phi(b'_i)$  or  $\phi$  induit une bijection de  $B$  sur  $X$  donc  $\forall i b_i = b'_i$  i.e  $u = v$ . Donc  $\phi$  est injective.

Réciproquement, soit  $\phi : A^* \rightarrow B^*$  morphisme injectif, supposons qu'on a  $n, m \in (N)$  et  $(x_i)_{i=1..n}, (x'_j)_{j=1..m} \in X = \phi(B)$  tels que  $x_1 x_2 \cdots x_n = x'_1 x'_2 \cdots x'_m$ . Soient  $(b_i)_{i=1..n}, (b'_j)_{j=1..m} \in B$  tels que  $\forall i x_i = \phi(b_i)$  et  $\forall j x_j = \phi(b'_j)$ . On a donc  $\phi(b_1 \cdots b_n) = \phi(b'_1 \cdots b'_m)$  or  $\phi$  est injective donc  $b_1 \cdots b_n = b'_1 \cdots b'_m$ . D'où  $n = m$  et  $\forall i b_i = b'_i$  et donc  $\forall i x_i = \phi(b_i) = \phi(b'_i) = x'_i$

□

On va maintenant introduire la notion d'ensemble préfixe (on a exactement la même chose en remplaçant préfixe par suffixe), qui donne un premier critère pour trouver des codes.

**Définition 3** (Ensembles préfixes).  $P \subset A^*$  est dit préfixe si  $\forall x, x' \in P x \leq x' \Rightarrow x = x'$  où  $\leq$  signifie être un facteur gauche (ou préfixe).

**Proposition 4.**  $\forall X \subset A^*, X \text{ préfixe} \Rightarrow X \text{ est un code.}$

*Démonstration.* Supposons que  $X$  n'est pas un code. Soit  $w$  de longueur minimale tel que  $w$  ait deux factorisations dans  $X$ . On a donc  $n, m \in (\mathbb{N})$  et  $(x_i)_{i=1\dots n}, (x'_j)_{j=1\dots m} \in X$  tels que  $w = x_1x_2 \cdots x_n = x'_1x'_2 \cdots x'_m$ . Comme  $w$  est de longueur minimale, on a  $x_1 \neq x'_1$  et donc  $x_1 < x'_1$  ou  $x'_1 < x_1$  ce qui rentre en contradiction avec  $X$  préfixe.  $\square$

*Exemple 5.* Soit  $A = \{a, b\}$  et  $X_1 = \bigsqcup_{n \geq 0} a^n b A^n$ .  $X_1$  est préfixe car  $a^n b u = a^m b v \Rightarrow m = n$  et donc  $u = v$ . C'est donc un code sur  $A$ .

**Définition 6** (Code maximal). Un code  $X$  sur  $A$  est dit maximal s'il n'est pas strictement inclus dans un autre code sur  $A$ .

## 2 Algorithmes de reconnaissance des codes

Reconnaitre si un ensemble donné est un code n'est pas toujours chose facile, mais il existe un algorithme simple qui permet de le décider. Les deux propositions qui suivent sont la preuve de correction et de terminaison de cet algorithme.

**Proposition 7.** Soit  $X \subset A^*$ . On définit  $U_1 = X^{-1}X \setminus \{\epsilon\}$  et par récurrence, pour tout  $n \geq 1$   $U_{n+1} = X^{-1}U_n \cup U_n^{-1}X$ . On a alors :

$$X \text{ est un code} \iff \forall n \geq 1 \epsilon \notin U_n$$

La démonstration nécessite le lemme suivant :

**Lemme 8.** Soit  $X \subset A^+$ .  $\forall n \geq 1 \forall k \in \{1 \dots n\}$  on a

$$\epsilon \in U_n \iff \exists u \in U_k \exists i, j \in \mathbb{N}^2 uX^i \cap X^j \neq \emptyset$$

avec  $i + j + k = n$

*Démonstration.* On prouve le lemme à  $n$  fixé par récurrence descendante sur  $k$ .

Si  $k = n$ , on a évidemment  $i = j = 0$ . Si  $\epsilon \in U_n$  on pose  $u = \epsilon$  et on a bien  $\epsilon U_n^0 \cap U_n^0 = \{\epsilon\}$ . Réciproquement si on a  $u \in U_n$  tel que  $u U_n^0 \cap U_n^0 = \{u\} \cap \{\epsilon\} \neq \emptyset$  alors  $u = \epsilon$  et donc  $\epsilon \in U_n$ .

Soit  $1 \leq k < n$ , supposons la propriété vérifiée pour  $k + 1$ . Si  $\epsilon \in U_n$  par hypothèse de récurrence,  $\exists v \in U_{k+1}, \exists i, j \in \mathbb{N}^2$  tels que  $i + j + k + 1 = n$  et

$\exists x, y \in X^i \times X^j$  tels que  $vx = y \in vX^i \cap X^j$ . Comme  $U_{k+1} = X^{-1}U_k \cup U_k^{-1}X$ , on a  $z, u \in X \times U_k$  tel que soit  $zv = u$  soit  $z = uv$ .

Dans le premier cas, on a  $ux = zvx = zy$  comme  $z, y \in X \times X^j$ ,  $zy \in X^{j+1}$  et  $uX^i \cap X^{j+1} \neq \emptyset$ . Dans le deuxième cas  $uy = uvx = zx \in X^{i+1}$  donc  $uX^j \cap X^{i+1} \neq \emptyset$ , avec à chaque fois  $u \in U_k$ .

Réciproquement, soient  $w \in uX^i \cap X^j$  où  $i + j + k = n$ . Si  $j = 0$ , alors l'intersection est vide à moins d'avoir  $u = \epsilon$  et  $i = 0$  car  $\epsilon \notin X$ , on a alors  $k = n$  mais on a supposé  $k < n$  donc  $j \geq 1$ . On a donc  $v, x, v' \in X^i \times X \times X^{j-1}$  tels que  $uv = xv'$ . On distingue ensuite 2 cas suivant les longueurs comparées de  $u$  et  $x$ .

Si  $|u| \leq |x|$  alors  $\exists u' \in A^*$   $uu' = x$  et alors  $u' \in U_k^{-1}X \subset U_{k+1}$ . De plus  $v = u'v'$  donc  $u'X^{j-1} \cap X^i \neq \emptyset$  et par hypothèse de récurrence  $\epsilon \in U_n$ .

Sinon  $\exists x' \in A^+$   $u = xx'$  avec  $x' \in X^{-1}U_k \subset U_{k+1}$  et  $x'v = v' \in x'X^i \cap X^{j-1}$ . D'après l'hypothèse de récurrence  $\epsilon \in U_n$ .  $\square$

*Démonstration.* On peut maintenant démontrer la proposition 7.

Supposons que  $X$  ne soit pas un code. Soit  $w$  de longueur minimale tel que  $w$  ait deux factorisations dans  $X$ . On a donc  $n, m \in (N)$  et  $(x_i), (x'_j) \in X$  tels que  $w = x_1x_2 \cdots x_n = x'_1x'_2 \cdots x'_m$  avec  $x_1$ . On peut supposer sans perdre de généralité que  $|x_1| < |x'_1|$  car  $w$  est de longueur minimale. On a alors  $\exists u \in A^+$   $x_1u = x'_1$  avec  $u \in X^{-1}X \setminus \{\epsilon\} = U_1$  et  $uX^{m-1} \cap X^{n-1}$  et donc  $\epsilon \in U_{n+m-1}$  d'après le lemme 8

Réciproquement, si  $\epsilon \in U_n$  pour un certain  $n$ , on applique le lemme avec  $k = 1$ .  $\exists u \in U_1$ ,  $i, j \in \mathbb{N}^2$  et  $v, w \in X^i \times X^j$  tels que  $uX^i \cap X^j \neq \emptyset$ . De plus comme  $U_1 = X^{-1}X \setminus \{\epsilon\}$   $\exists x, y \in X^2$  tels que  $xu = y$  avec  $x \neq y$  car  $u \neq \epsilon$ . d'où  $yX^i \cap xX^j = xuX^i \cap xX^j \neq \emptyset$  ce qui fait que  $X$  ne peut être un code.  $\square$

La construction des  $U_n$  donne donc un algorithme pour déterminer si un ensemble est un code. La propriété suivante conclut quant à la terminaison de l'algorithme dans le cas où  $X$  est rationnel.

**Proposition 9.** *Si  $X$  est rationnel l'ensemble de ses  $U_n$  est fini.*

*Démonstration.* On rappelle qu'un langage est rationnel est équivalent au fait que le nombre de ses quotients à gauche est fini. On montre par récurrence sur  $n$  que les  $U_n$  sont des unions finies de quotients à gauche de  $X$  auxquelles on peut avoir retiré  $\epsilon$ . Pour  $n = 1$ ,  $U_1 = X^{-1}X \setminus \{\epsilon\} = (\bigcup_{x \in X} x^{-1}X) \setminus \{\epsilon\}$ . Supposons que c'est vrai au rang  $n - 1$ , alors comme le quotients à gauche

d'un quotient à gauche de  $X$  est un quotient à gauche de  $X X^{-1}U_n$  est bien une union de quotients à gauche de  $X$  (l'absence ou la présence de  $\epsilon$  ne change pas grand chose juste de nombreuses disjonctions de cas si on veut rentrer dans les détails). De plus  $(U_n)^{-1}X$  est bien sur une union de quotient à gauche de  $X$

Comme le nombre de quotients à gauche de  $X$  est fini leurs unions sont aussi en nombre fini (retirer  $\epsilon$  ne fait que doubler ce nombre au pire).  $\square$

Cette proposition indique que l'algorithme s'arrête quand on retombe sur un  $U_n$  déjà construit et que cela arrive forcément si  $X$  est rationnel (ce qui est relativement raisonnable comme hypothèse).

*Exemple 10.* Soit  $A = \{a, b\}$ , considérons  $X_2 = \{aa, ba, bb, baa, bba\}$ .  $X$  n'est pas préfixe et considérer des factorisations dans  $X$  n'est pas vraiment envisageable. Les  $U_n$  permettent pourtant de conclure très vite. En effet  $U_1 = \{a\} = U_2$ .

### 3 Mesure d'un code

Cette partie introduit une mesure sur les parties de  $A^*$ , et certaines conséquences, quand ces ensembles sont des codes, quant à la mesure qu'il peuvent avoir.

**Définition 11** (Distribution de Bernoulli). Soit  $A$  un alphabet, une distribution de Bernoulli sur  $A^*$  est un morphisme  $\pi : A^* \rightarrow \mathbb{R}_+$  (où  $\mathbb{R}_+$  est considéré comme un monoïde multiplicatif) tel que :

$$\sum_{a \in A} \pi(a) = 1$$

Une distribution est dite positive si :  $\forall a \in A \pi(a) > 0$

**Propriété 12.** Pour tout  $n \geq 1$  :

$$\sum_{u \in A^n} \pi(u) = 1$$

*Démonstration.* Par récurrence sur  $n$ .

Pour  $u \in A^n$ , on a  $\sum_{a \in A} \pi(ua) = \pi(u) \sum_{a \in A} \pi(a) = \pi(u)$ , et donc

$$\sum_{u \in A^{n+1}} \pi(u) = \sum_{v \in A^n} \sum_{a \in A} \pi(va) = \sum_{v \in A^n} \pi(v) = 1$$

□

**Définition 13** (Mesure d'une partie). On étend  $\pi$  à  $\mathcal{P}(A^*)$  (les parties de  $A^*$ ) en posant pour tout  $L \subset A^*$  :

$$\pi(L) = \sum_{l \in L} \pi(l)$$

**Propriété 14.**  $\pi : \mathcal{P}(A^*) \rightarrow \mathbb{R}_+$  a les propriétés immédiates suivantes. Les trois premières propriétés en font une mesure sur  $\mathcal{P}(A^*)$  :

- i.  $\forall L \subset A^* \pi(L) \geq 0$
- ii.  $\pi(\emptyset) = 0$
- iii. Pour toute famille  $(L_i)_{i \in I}$  de sous-ensembles de  $A^*$  deux à deux disjoints :

$$\pi\left(\bigsqcup_{i \in I} L_i\right) = \sum_{i \in I} \pi(L_i)$$

iv.  $\pi(A^*) = \pi\left(\bigsqcup_{n \in \mathbb{N}} A^n\right) = \sum_{n \in \mathbb{N}} \pi(A^n) = \infty$

- v. Si les  $(L_i)_{i \in I}$  ne sont pas deux à deux disjoints :

$$\pi\left(\bigcup_{i \in I} L_i\right) \leq \sum_{i \in I} \pi(L_i)$$

- vi. Soit  $L \subset A^*$ , on pose pour  $n \in \mathbb{N}$   $s_n = \pi(\{w \in L \mid |w| \geq n\})$ . On a alors comme  $\pi$  est une mesure :

$$\pi(L) = \sup_{n \geq 0} s_n$$

- vii. Soient  $L$  et  $M$  des langages sur  $A$ , comme  $LM = \bigcup_{l \in L} \bigcup_{m \in M} lm$  :

$$\pi(LM) \leq \sum_{l \in L} \sum_{m \in M} \pi(lm) = \sum_{l \in L} \pi(l) \sum_{m \in M} \pi(m) = \pi(L)\pi(M)$$

- viii. On en déduit, pour tout  $X \subset A^*$  :

$$\pi(X^*) \leq \sum_{n \geq 0} \pi(X^n) \leq \sum_{n \geq 0} \pi(X)^n$$

Certaines de ces inégalités deviennent des égalités dans le cas des codes (et seulement dans le cas des codes comme l'indique la réciproque; elle est donnée ici juste pour la symétrie, elle ne sera pas utile plus loin).

**Proposition 15.** *Soit  $X$  un code alors :*

$$\forall n \geq 1 \quad \pi(X^n) = \pi(X)^n$$

$$\pi(X^*) = \sum_{n \geq 0} \pi(X)^n$$

En particulier  $\pi(X^*) < \infty \iff \pi(X) < 1$

Réciproquement, si  $\pi$  est positive, si  $\pi(X) < \infty$  et  $\forall n \geq 1 \quad \pi(X^n) = \pi(X)^n$  alors  $X$  est un code.

*Démonstration.* Supposons que  $X$  est un code.

On notera  $X^{(n)} = X \times X \times \dots \times X$  le produit cartésien de  $X$  par  $X$ ,  $n$  fois. Comme  $X$  est un code, la fonction  $\psi$  :

$$\begin{aligned} X^{(n)} &\rightarrow X^n \\ \underline{x} = (x_1, \dots, x_n) &\mapsto x_1 \cdots x_n \end{aligned}$$

est bijective (son image est  $X^n$  par définition et son injectivité découle directement de la définition d'un code).

On en déduit que :

$$\begin{aligned} \pi(X)^n &= \left( \sum_{x \in X} \pi(x) \right)^n = \sum_{(x_1, \dots, x_n) \in X^{(n)}} \pi(x_1) \cdots \pi(x_n) \\ &= \sum_{\underline{x} \in X^{(n)}} \pi(\psi(\underline{x})) \\ &= \sum_{x \in X^n} \pi(x) = \pi(X^n) \end{aligned}$$

Le changement de numérotation dans la somme découle de la bijectivité de  $\psi$ .

De plus les ensembles  $X^n$  sont disjoints car  $X$  est un code donc par la propriété 14.iii et l'égalité précédente,

$$\pi(X^*) = \pi\left(\bigsqcup_{n \geq 0} X^n\right) = \sum_{n \geq 0} \pi(X^n) = \sum_{n \geq 0} \pi(X)^n$$

L'équivalence  $\pi(X^*) < \infty \iff \pi(X) < 1$  est une conséquence directe du fait que  $\pi(X^*)$  est une somme géométrique de raison  $\pi(X)$ .

Supposons maintenant que  $X$  n'est pas un code mais qu'on a bien les hypothèses de la réciproque. Il existe donc un mot  $u \in X^+$  qui a deux factorisations dans  $X$  :  $u = x_1 \cdots x_n = x'_1 \cdots x'_m$ . Le mot  $uu = x'_1 \cdots x'_m x_1 \cdots x_n = x_1 \cdots x_n x_1 \cdots x_n$  a alors 2 factorisations en  $m + n$  mots de  $X$ . D'où

$$\pi(X)^k = \sum_{\substack{\underline{x} \in X^{(n+m)} \\ \psi(\underline{x}) \neq uu}} \pi(\psi(\underline{x})) + 2\psi(uu) \geq \pi(X^{m+n}) + \pi(uu)$$

Comme  $\pi(X) < \infty$  et  $\pi(X^{m+n}) = \pi(X)^{m+n}$  on a  $\pi(uu) \leq 0$  ce qui contredit la positivité de  $\pi$ .  $\square$

**Proposition 16.** *Soit  $X$  un code sur  $A$ . Pour toute distribution de Bernoulli  $\pi$  sur  $A^*$ , on a  $\pi(X) \leq 1$*

*Démonstration.* On commence par prouver la proposition dans le cas où  $k = \sup_{x \in X} |x| < \infty$ . On a alors  $\forall n \geq 1 \sup_{x \in X^n} |x| = nk$  ce qui se traduit par  $X^n \subset \bigsqcup_{m=0}^{nk} A^m$ . D'où  $\pi(X^n) \leq \sum_{m=0}^{nk} \pi(A^m) = nk$  car  $\pi(A^m) = 1$

Supposons maintenant, en raisonnant pas l'absurde, que  $\exists \epsilon > 0 \pi(X) = 1 + \epsilon$ . On a alors d'après la proposition 15  $\forall n (1 + \epsilon)^n = (\pi(X))^n \leq kn$  ce qui est impossible car  $kn = o((1 + \epsilon)^n)$  quand  $n \rightarrow \infty$ .

D'où  $\pi(X) \leq 1$ .

Si  $X$  est un code quelconque, on pose  $X_n = \{x \in X \mid |x| \leq n\}$ .  $X_n$  est un code (propriété 2.ii) donc  $\pi(X_n) \leq 1$  et par la propriété 14.vi :

$$\pi(X) = \sup_{n \geq 1} \pi(X_n) \leq 1$$

.  $\square$

**Proposition 17.** *Soit  $X$  un code sur  $A$ . S'il existe une distribution de Bernoulli  $\pi$  positive sur  $A^*$  telle que  $\pi(X) = 1$  alors  $X$  est maximal.*

*Démonstration.* Supposons que  $X$  n'est pas maximal. Il existe alors  $y \in X$  tel que  $Y = X \sqcup \{y\}$  est un code. D'après la proposition 16, on a  $\pi(Y) \leq 1$ . De plus  $\pi(Y) = \pi(X) + \pi(y) = 1 + \pi(y)$ . Donc  $\pi(y) = 0$  ce qui contredit  $\pi$  positive.  $\square$

*Exemple 18.* On va montrer que  $X_1$  défini à l'exemple 5 est maximal. On pose  $\pi(a) = p < 1$  et donc  $\pi(b) = 1 - p$ .

$$\pi(X_1) = \sum_{n \geq 0} \pi(a^n b A^n) = \sum_{n \geq 0} p^n (1 - p) \pi(A^n) = (1 - p) \sum_{n \geq 0} p^n = 1$$

Donc  $\pi(X_1) = 1$  pour toute distribution de Bernoulli positive, on a donc une hypothèse plus forte que nécessaire dans la proposition 17. On en déduit que  $X_1$  est un code maximal.

## 4 Codes complets

On en arrive enfin à la notion de complétude, dernière des 3 notions que le théorème 28 met en rapport.

**Définition 19** (Eléments complétables). Soient  $M$  un monoïde et  $P$  un sous-ensemble de  $M$ . Un élément  $m$  de  $M$  est dit complétable dans  $P$  si

$$\exists u, v \in M^2 \text{ } umv \in P$$

On notera  $F(P) = M^{-1}PM^{-1}$  l'ensemble des mots complétables dans  $P$ .

**Définition 20** (Ensembles denses et complets et maigres).  $P \subset M$  est dense dans  $M$  si  $M = F(P)$ .

Si  $P$  n'est pas dense, on dit que  $P$  est maigre.

$P \subset M$  est complet dans  $M$  si le monoïde généré par  $P$  est dense.

Pour les codes cela se traduit par : le code  $X$  sur  $A$  est complet si  $X^*$  est dense dans  $A^*$ .

On peut dans le cas du monoïde  $A^*$  voir l'ensemble  $F(P)$  comme l'ensemble des facteurs des mots de  $P$ .

*Exemple 21.* Le code  $X_1$  introduit à l'exemple 5 est un code dense. En effet, soit  $w \in A^*$ , le mot  $a^{|w|}bw \in X_1$ . Il est donc a fortiori complet.

La définition et les deux lemmes qui suivent servent à démontrer la proposition 25 qui met en relation complétude et maximalité d'un code.

**Définition 22** (Mots sans bords). Un mot  $w \in A^*$  est sans bords si aucun facteur gauche propre de  $w$  est aussi un facteur droit propre. En d'autres termes :

$$\forall u \in A^* \quad w \in (uA^* \cap A^*u) \Rightarrow (u = \epsilon \vee u = w)$$

**Lemme 23.** Soient  $X \subset A^+$  un code,  $y \in A^*$  un mot sans bords tel que  $y \notin F(X^*)$ . Alors l'ensemble  $Y = X \sqcup \{y\}$  est un code.

*Démonstration.* On suppose que  $Y$  n'est pas un code sur  $A$ . Soit  $w$  le plus petit mot qui a 2 factorisations dans  $Y$  on a alors  $n, m \in \mathbb{N}$  et  $(y_i)_{i=1\dots n}, (y'_j)_{j=1\dots m} \in Y$  tels que  $w = y_1y_2 \cdots y_n = y'_1y'_2 \cdots y'_m$ .

Si  $y \notin \{y_i\}_{i=1\dots n} \cup \{y'_j\}_{j=1\dots m}$  alors c'est une factorisation dans  $X$  qui est un code donc elles sont identiques ce qui est impossible.

Si  $y \notin \{y_i\}_{i=1\dots n}$  mais  $y \in \{y'_j\}_{j=1\dots m}$  (ou vice versa) alors  $y \in F(X)$  ce qui est impossible.

Enfin si  $y \in \{y_i\}_{i=1\dots n} \cap \{y'_j\}_{j=1\dots m}$  soient  $i_0$  et  $j_0$  les plus petits indices tels que  $y$  apparaît dans  $(y_i)$  et  $(y'_j)$ . On ne peut pas avoir  $y_1 \cdots y_{i_0-1} = y'_1 \cdots y'_{j_0-1}$  sinon cela contredirait la minimalité de  $w$ . On peut supposer  $y_1 \cdots y_{i_0-1} < y'_1 \cdots y'_{j_0-1}$ . Si  $y_1 \cdots y_{i_0-1}y < y'_1 \cdots y'_{j_0-1}y$  alors  $y \in F(X)$  ce qui contredit les hypothèses. Sinon les deux occurrences de  $y$  ont des lettres en commun (mais pas toutes) ce qui contredit le fait que  $y$  est sans bords.  $\square$

**Lemme 24** (Complétion en un mot sans bords). Soit  $A$  un alphabet contenant au moins 2 lettres.  $\forall u \in A^+ \exists v \in A^*$  tel que  $uv$  est sans bords.

*Démonstration.* Soient  $a$  la première lettre de  $u$  et  $b \in A \setminus \{a\}$ . Montrons que  $w = uab^{|u|}$  est sans bords. Soit  $t$  un facteur gauche non vide de  $w$ . Il commence par un  $a$  donc  $t$  ne peut pas être un facteur droit à moins que  $|t| > |u|$  vu que les  $|u|$  dernières lettres de  $w$  sont des  $b$ . Alors  $\exists v \in A^*$  tel que  $t = vab^{|u|} = uab^{|v|}$  on donc forcément  $|v| = |u|$  et donc  $t = w$ .  $\square$

**Proposition 25.** Tout code maximal est complet

*Démonstration.* Si  $|A| = 1$  alors on vérifie facilement que les seuls codes sont les  $a^n$  pour  $n \in \mathbb{N}$  qui sont complets et  $\emptyset$  qui n'est pas maximal.

Sinon soient  $X \subset A^+$  un code qui n'est pas complet et  $u \notin F(X^*)$ . D'après le lemme 24 (on a bien  $|A| \geq 2$ )  $\exists v \in A^*$  tel que  $uv$  soit sans bords. Comme  $u$  est un facteur de  $uv$  on a toujours  $uv \notin F(X^*)$ . On déduit du lemme 23 que  $X \sqcup \{y\}$  est un code. Donc  $X$  n'est pas maximal.  $\square$

Avant d'arriver à la proposition 27 qui relie complétude et mesure d'un code, on démontre le lemme technique suivant :

**Lemme 26.** Soient  $X \subset A^*$  un ensemble maigre et complet. Soit  $w$  qui n'est pas complétable dans  $X$ . Alors, en notant  $D$  (resp.  $G$ ) les facteurs gauches (resp. droits) de  $w$ , on a :

$$A^* = \bigcup_{\substack{d \in D \\ g \in G}} d^{-1}X^*g^{-1} = D^{-1}X^*G^{-1}$$

*Démonstration.* Soit  $z \in A^*$ . Comme  $X^*$  est dense, le mot  $wzw$  est complétable dans  $X^*$  et donc  $\exists u, v \in (A^*)^2$   $uwzvw \in X^*$ . Par hypothèse  $w$  ne peut pas être facteur d'un mot de  $X^*$  si on coupe  $uwzvw$  en mots de  $X$  on a donc forcément une coupure dans chacun des  $w$ . On a donc deux factorisations de  $w$ ,  $w = gd = g'd'$  telles que  $ug, dzg', d'v \in (X^*)^3$ .

On a donc  $z \in d^{-1}X^*g^{-1}$  avec  $d, g \in D \times G$  □

**Proposition 27.** Soient  $X$  un sous-ensemble maigre et complet de  $A^*$  et  $\pi$  une distribution de Bernoulli positive sur  $A^*$ , on a :

$$\pi(X) \geq 1$$

*Démonstration.* On a déjà montré  $\pi(A^*) = \infty$  (Propriété 14.iv). D'après le résultat de la proposition 26,  $\exists (d, g) \in (A^*)^2$  tels que  $\pi(d^{-1}X^*g^{-1}) = \infty$ . En effet, soit  $w$  qui n'est pas complétable dans  $X$  on a dans le cas contraire

$$\infty = \pi(A^*) = \pi\left(\bigcup_{d \in D, g \in G} d^{-1}X^*g^{-1}\right) \leq \sum_{d \in D, g \in G} \pi(d^{-1}X^*g^{-1}) < \infty$$

car les suffixes et préfixes d'un mots sont en nombre fini, mais c'est absurde.

De plus  $d(d^{-1}X^*g^{-1})g \subset X^*$  ce qui implique que  $\pi(d)\pi(d^{-1}X^*g^{-1})\pi(g) \leq \pi(X^*)$ . Comme  $\pi$  est positive  $\pi(d)\pi(g) \neq 0$  on a  $\pi(X^*) = \infty$ . La proposition 15 permet de conclure que  $\pi(X) \geq 1$  □

En rassemblant tout ce qui a été démontré auparavant, on obtient le théorème suivant qui relie 3 notions qui semblent pourtant très peu liées à première vue : la mesure, la maximalité et la complétude d'un code maigre.

**Théorème 28.** Soit  $X$  un code maigre sur un alphabet  $A$ . Les propositions suivantes sont équivalentes :

- (i)  $X$  est un code maximal.
- (ii)  $\exists \pi$  distribution de Bernoulli positive sur  $A^*$  telle que  $\pi(X) = 1$ .

(iii)  $\forall \pi$  distribution de Bernoulli positive sur  $A^*$ ,  $\pi(X) = 1$ .

(iv)  $X$  est un code complet

*Démonstration.* (i)  $\Rightarrow$  (iv) est la proposition 25.

(iv)  $\Rightarrow$  (iii) est une conséquence des propositions 16 et 27.

(iii)  $\Rightarrow$  (ii) est trivial.

(ii)  $\Rightarrow$  (i) est la proposition 17

□

*Exemple 29.* Le code  $X_1$  introduit à l'exemple 5 vérifie ces 4 propriétés sans être maigre.

*Exemple 30.*  $X_2$  (voir exemple 10) quant à lui est maigre (car fini). Posons  $\pi(a) = p$  où  $0 < p < 1$

$$\pi(X_2) = p^2 + (1-p)p + (1-p)^2 + (1-p)p^2 + (1-p)^2p = 1$$

$X_2$  est donc complet et maximal, ce qui a priori n'était pas évident.

Le théorème 28 donne un moyen très simple de savoir si un code maigre est maximal ou complet, propriétés assez fortes, il suffit de tester sur n'importe quelle distribution de Bernoulli positive et de vérifier qu'on obtient 1.

## Références

- [1] J. Berstel and D. Perrin. *Theory of Codes*. Academic Press, 1984.