Mémoire

Auteur : Talmas, Emeline
Promoteur(s) : Leroy, Julien
Faculté : Faculté des Sciences
Diplôme : Master en sciences mathématiques, à finalité didactique
Année académique : 2020-2021
URI/URL : http://hdl.handle.net/2268.2/11103

Université de Liège
Faculté des Sciences
Département de Mathématique

Introduction à la théorie des codes

Mémoire présenté en vue de l'obtention du grade de Master en sciences mathématiques à finalité didactique

Promoteur : Julien Leroy

Emeline Talmas

Année académique 2020–2021

Remerciements

Le succès est la somme de petits efforts, répétés jour après jour.

— Leo Robert Collier

Je tiens tout d'abord à remercier chaleureusement mon promoteur Julien Leroy pour m'avoir proposé ce sujet de mémoire et pour son encadrement tout au long de l'élaboration de ce travail.

J'adresse un grand merci à ma maman et à ma sœur qui ont relu courageusement ce mémoire qui leur était totalement incompréhensible.

Je remercie également ma famille et mes amis, peut-être plus particulièrement ceux qui ont partagé avec moi ces études, pour leur soutien durant ces dernières années.

Finalement je remercie « mon » Julien d'avoir supporté mes moments de doute, de stress et d'avoir toujours cru en moi. Son soutien a sans doute été le plus précieux.

Table des matières

Introduction

On doit la naissance de la théorie des codes à la théorie de l'information de Claude Shannon, développée dans les années 1950. Cette dernière, dont le sujet premier est de quantifier le contenu en information d'un ensemble de données, s'intéresse notamment à la quantité d'information transmissible. C'est dans ce contexte qu'apparaissent les notions de code et de codage. Un *codage* est une application injective d'un alphabet A dans l'ensemble des mots non vides sur un alphabet B. Un codage s'étend en un morphisme de A^* dans B^* ; si ce morphisme est également injectif, le codage est dit *à déchiffrage unique*. Dans ce cas, nous appellerons *code* l'image du codage.

La théorie des codes émerge donc de la théorie de l'information, mais peut être étudiée indépendamment de celle-ci, pour ses propriétés mathématiques intrinsèques. C'est d'ailleurs l'approche que nous adopterons dans ce mémoire. Nous définirons donc la notion de code d'une manière légèrement différente, bien qu'équivalente comme nous le verrons ultérieurement : un *code* est un ensemble de mots tel que tout mot admet au plus une décomposition en facteurs de cet ensemble. De manière générale, nous pouvons voir la théorie des codes comme une étude des factorisations de mots en facteurs issus d'un ensemble donné. De ce point de vue, et au vu des problèmes auxquels elle s'intéresse, la théorie des codes se place au sein de l'informatique théorique.

La définition d'un code telle qu'elle est considérée dans ce mémoire laisse penser qu'il s'agit d'un concept purement combinatoire. Cependant, comme nous le verrons, la notion de code est essentiellement équivalente à celle de morphisme injectif d'un monoïde libre dans un autre. L'étude des codes pourrait donc être finalement considérée comme un problème d'algèbre. Ce n'est pas dans ce sens que nous traiterons ce problème puisque nous faisons le choix d'utiliser principalement des outils de combinatoire des mots, de théorie des automates et de théorie des langages formels.

Ce travail s'inspire principalement de l'ouvrage [3] *Codes and Automata* de Jean Berstel, Dominique Perrin et Christophe Reutenauer, et s'articule de la façon suivante.

Le Chapitre 1 sera consacré à une introduction générale des codes. Nous commencerons par définir la notion de code, sur laquelle repose tout ce travail. Nous en donnerons ensuite quelques équivalences. Les codes pouvant être regroupés en différentes familles, nous

introduirons notamment les codes préfixes, suffixes, bifixes ou encore les codes maximaux. Nous reviendrons sur la notion de code maximal au Chapitre 2 et l'étude des codes préfixes se fera plus en détail au Chapitre 3. De plus, nous fournirons une méthode algorithmique nous permettant de vérifier qu'un ensemble de mots donné est un code. Finalement nous nous intéresserons aux sous-monoïdes libres et aux propriétés qui les lient aux codes.

Le Chapitre 2 sera dédié notamment à la mesure des codes. Après avoir introduit les lois de probabilité sur un alphabet, nous nous pencherons plus précisément sur les lois de Bernoulli. Ces dernières seront utiles non seulement pour établir des conditions nécessaires pour qu'un ensemble de mots soit un code, mais également pour comprendre qu'un code ne peut contenir que « peu » de « petits » mots. Nous aborderons ensuite des notions de densité et introduirons entre autres les notions d'ensembles complets, de codes maximaux et de codes fins. Les notions qui seront présentées dans cette section nous permettront d'établir quelques résultats à propos des codes maximaux. Ces derniers constituent une famille privilégiée de codes qu'il est intéressant d'étudier puisque nous verrons que tout sous-ensemble d'un code est encore un code.

Le Chapitre 3, comme annoncé précédemment, sera axé sur les codes préfixes. Il s'agit sans doute de la famille de codes la plus facile à reconnaitre et à construire. Nous commencerons par en donner quelques propriétés de base. Ensuite, nous établirons des liens entre ces codes particuliers et d'autres branches des mathématiques discrètes comme les séries formelles ou la théorie des automates. Nous nous intéresserons également à la maximalité des codes préfixes. Pour ce faire, nous adapterons les notions de densité présentées au Chapitre 2 à ce cas particulier. Finalement, nous étudierons un sous-ensemble spécifique des codes préfixes : les codes sémaphores. De manière évidente, les résultats qui seront présentés dans ce chapitre pourront être adaptés au cas des codes suffixes.

Chapitre 1

Codes

Ce premier chapitre sera dédié à l'introduction de la notion de code et de leurs propriétés principales.

Tout d'abord, nous définirons les codes, en donnerons des exemples et contre-exemples, et fournirons une caractérisation des codes et ses conséquences. Nous définirons également des familles particulières de codes : codes préfixes, suffixes, bifixes, maximaux.

Ensuite, nous présenterons un algorithme permettant de déterminer si un ensemble donné est un code.

Finalement, nous étudierons les liens entre codes et sous-monoïdes libres et démontrerons le Théorème du défaut, qui fournit une condition suffisante pour être un code.

1.1 Rappels

1.1.1 Semi-groupes, monoïdes et mots

Un *semi-groupe* est un ensemble muni d'une opération binaire associative. Un *monoïde* M est un semi-groupe qui possède un élément neutre, noté 1_M. Un *sous-monoïde* d'un monoïde M est une partie N de M qui est stable pour l'opération et qui contient l'élément neutre de M. Un *morphisme de monoïdes* entre deux monoïdes M et N est une application $\varphi : M \to N$ qui satisfait $\varphi(mm') = \varphi(m)\varphi(m')$ pour tous $m, m' \in M$ et $\varphi(1_M) = 1_N$.

Soit A un ensemble que nous appellons *alphabet*. Dans la suite, sauf mention explicite du contraire, un alphabet sera toujours supposé fini. Les éléments de A sont appelés *lettres*. Un *mot* w sur l'alphabet A est une suite finie d'éléments de A : $w = a_1 a_2 \cdots a_n$ avec $a_i \in A$. Le *mot vide* est noté ε. L'ensemble de tous les mots sur l'alphabet A est noté A^*. Muni de la concaténation, cet ensemble possède une structure de monoïde ayant ε pour neutre. Ce monoïde est appelé *monoïde libre* sur A. L'ensemble des mots non vides sur A est noté A^+. La *longueur* d'un mot $w = a_1 a_2 \cdots a_n$ avec $a_i \in A$ est le nombre n de lettres dans w. Nous la notons $|w|$. Par définition, $|\varepsilon| = 0$. Un *facteur* d'un mot $w \in A^*$ est un mot $x \in A^*$ tel qu'il existe $u, v \in A^*$ tels que $w = uxv$. Il est dit *propre* si $w \neq x$. Un mot x est un *préfixe* (resp. *suffixe*) d'un mot $w \in A^*$ s'il existe un mot $u \in A^*$ tel que $w = xu$ (resp. $w = ux$).

Un ensemble X est dit *fermé par préfixe* (resp. *fermé par suffixe*) s'il contient tous les préfixes (resp. suffixes) de ses éléments. Pour un sous-ensemble X de A^*, nous notons X^* le *sous-monoïde engendré par* X, i.e.

$$X^* = \{x_1x_2\cdots x_n \mid n \geq 0, x_i \in X\}.$$

De la même façon, nous notons X^+ le semi-groupe engendré par X, i.e.

$$X^+ = \{x_1x_2\cdots x_n \mid n \geq 1, x_i \in X\}.$$

Soient $X, Y \subset A^*$. On définit le produit de X et Y par

$$XY = \{xy \mid x \in X, y \in Y\}.$$

Le produit XY est *non ambigu* si lorsque l'on a $xy = x'y'$ avec $x, x' \in X$ et $y, y' \in Y$, on a $x = x'$ et $y = y'$.

Soient $X \subset A^*$ et $w \in A^*$. L'ensemble $w^{-1}X$ (resp. Xw^{-1}) est l'ensemble des mots qui concaténés à droite (resp. gauche) de w sont dans X, i.e.

$$w^{-1}X = \{u \in A^* \mid wu \in X\} \quad (\text{resp. } Xw^{-1} = \{u \in A^* \mid uw \in X\}).$$

Si Y est une partie de A^*, nous notons

$$Y^{-1}X = \bigcup_{w\in Y} w^{-1}X \text{ et } XY^{-1} = \bigcup_{w\in Y} Xw^{-1}.$$

Soit M un monoïde. Un *idéal à droite* (resp. *à gauche*, *bilatère*) est un sous-ensemble non vide I de M tel que $IM \subset I$ (resp. $MI \subset I$, $MIM \subset I$). Dans la suite, nous désignerons par *idéal* un idéal bilatère. Si X est inclus dans M, nous appellerons XM (resp. MX, MXM) l'idéal à droite (resp. à gauche, bilatère) engendré par X.

1.1.2 Séries formelles

Soient A un alphabet et K un semi-anneau. Une *série formelle* sur A à coefficients dans K est une application

$$S : A^* \to K.$$

On note (S, w) l'image d'un mot $w \in A^*$ par l'application S. On note $K\langle\langle A\rangle\rangle$ l'ensemble des séries formelles sur A. Soient $S, T \in K\langle\langle A\rangle\rangle$ et $k \in K$, on définit les opérations suivantes :

1. La somme :
$$S + T : A^* \to K, w \mapsto (S, w) + (T, w).$$
Ainsi, $(S + T, w) = (S, w) + (T, w)$ pour tout $w \in A^*$.
2. Le produit de Cauchy :
$$ST : A^* \to K, w \mapsto \sum_{uv=w} (S, u)(T, v).$$
Ainsi, $(ST, w) = \sum_{uv=w} (S, u)(T, v)$ pour tout $w \in A^*$.

Pour $X \subset A^*$, on écrit $\underline{X}$ pour désigner la *série caractéristique* de X définie par

$$(\underline{X}, w) = \begin{cases} 1 \text{ si } w \in X \\ 0 \text{ sinon.} \end{cases}$$

Dans le cas où $X = \{w\}$, on note $\underline{w}$ la série caractéristique de X.

Muni de ces opérations, $K\langle\langle A\rangle\rangle$ est un anneau. Le neutre pour le produit est la série $\underline{\varepsilon}$.

La série $S \in K\langle\langle A\rangle\rangle$ est dite *propre* si $(S, \varepsilon) = 0$. Lorsque S est propre on note

$$S^* = \sum_{n\geq 0} S^n \text{ et } S^+ = \sum_{n\geq 1} S^n.$$

Dans ce cas, ces sommes sont bien définies puisque, pour w fixé, on a $(S^n, w) = 0$ pour n suffisamment grand. On a alors $S^* = \underline{\varepsilon} + S^+$ et $S^*S = SS^* = S^+$.

Proposition 1.1.1. *Soient K un anneau et $S \in K\langle\langle A\rangle\rangle$. La série S est inversible dans $K\langle\langle A\rangle\rangle$ si et seulement si son terme indépendant est inversible dans K.*

Proposition 1.1.2. *Soit K un anneau et $S \in K\langle\langle A\rangle\rangle$ une série propre. La série $\underline{\varepsilon} - S$ est inversible et $S^* = (\underline{\varepsilon} - S)^{-1}$.*

Pour toute série formelle $S \in K\langle\langle A\rangle\rangle$ nous notons $S = \sum_{w\in A^*} (S, w)w$.

Proposition 1.1.3. *Soient $X, Y \in A^*$. On a*

$$(\underline{X} + \underline{Y}, w) = \begin{cases} 2 \text{ si } w \in X \cap Y \\ 1 \text{ si } w \in (X\backslash Y) \cup (Y\backslash X) \\ 0 \text{ si } w \notin X \cup Y. \end{cases}$$

En particulier, si X et Y sont disjoints, alors $\underline{X} + \underline{Y} = \underline{X \cup Y}$.

Proposition 1.1.4. *Soient $X, Y \in A^*$. On a*

$$(\underline{X}\ \underline{Y}, w) = \text{Card}\{(x, y) \in X \times Y \mid w = xy\}.$$

En particulier, on a $\underline{XY} = \underline{X}\ \underline{Y}$ si et seulement si le produit XY est non ambigu.

Proposition 1.1.5. *Soient $X \in A^*$. On a*

$$((\underline{X})^*, w) = \text{Card}\{(x_1, x_2, \ldots, x_n) \mid n \geq 0, x_i \in X, w = x_1 \cdots x_n\}.$$

1.1.3 Automates

Un *automate déterministe* est la donnée d'un quintuple

$$\mathcal{A} = \{Q, q_0, F, A, \delta\}$$

où Q est un ensemble dont les éléments sont les *états* de $\mathcal{A}$; $q_0 \in Q$ est un état privilégié appelé *état initial* ; $F \subset Q$ est l'ensemble des états finals ; A est l'alphabet de l'automate ; $\delta : Q \times A \to Q$ est la fonction de transition de $\mathcal{A}$. Nous supposerons que δ est une fonction totale. On étend la fonction δ à $Q \times A^*$ par récurrence en posant $\delta(q, \varepsilon) = q$ et pour tous $w \in A^*$ et $a \in A$:

$$\delta(q, wa) = \delta(\delta(q, w), a).$$

On définit l'automate minimal

$$\mathcal{A}_X = (Q_X, q_{0,X}, F_X, A, \delta_X)$$

d'un sous-ensemble X de A^* de la manière suivante :

- $Q_X = \{w^{-1}X \mid w \in A^*\}$,
- $q_{0,X} = \varepsilon^{-1}X = X$,
- $F_X = \{w^{-1}X \mid w \in X\}$,
- $\delta_X(q, a) = a^{-1}q$ pour tout $q \in Q_X, a \in A$.

L'automate minimal d'un sous-ensemble X de A^* accepte X.

Un automate déterministe $\mathcal{A} = \{Q, q_0, F, A, \delta\}$ est *accessible* si pour tout état $q \in Q$, il existe un mot $w \in A^*$ tel que $\delta(q_0, w) = q$. Il est *coaccessible* si pour tout état $q \in Q$, il existe un mot $w \in A^*$ tel que $\delta(q, w) \in F$. Il est *réduit* si pour tous $p, q \in Q$

$$\{w \in A^* \mid \delta(p, w) \in F\} = \{w \in A^* \mid \delta(q, w) \in F\} \Rightarrow p = q.$$

Autrement dit, l'automate est réduit si pour tout couple (p, q) d'états avec $p \neq q$ il existe un mot $w \in A^*$ tel que

$$\delta(p, w) \in F \text{ et } \delta(q, w) \notin F$$

ou

$$\delta(p, w) \notin F \text{ et } \delta(q, w) \in F.$$

Dans ce cas, on dit que le mot w *distingue* les états p et q et que ces états sont *distingués*.

L'automate minimal $\mathcal{A}_X$ d'un ensemble $X \subset A^*$ est déterministe, accessible et réduit.

1.2 Premières définitions et propriétés

1.2.1 Qu'est-ce qu'un code ?

Définition 1.2.1. Soit A un alphabet. Un sous-ensemble X du monoïde libre A^* est un *code* sur A si pour tous $m, n \geq 1$ et $x_1, \ldots, x_m, y_1, \ldots, y_n \in X$, la condition

$$x_1 x_2 \cdots x_m = y_1 y_2 \cdots y_n$$

implique $m = n$ et $x_i = y_i$ pour $i = 1, \ldots, m$.
Un élément du code sera simplement appelé *mot du code*.

Autrement dit, un ensemble X est un code si tout mot de X^* possède une unique factorisation en mots de X.
Étant donné cette définition, il est évident qu'un code ne contient pas le mot vide ε.

Remarque 1.2.2. Tout alphabet A est trivialement un code.

Exemple 1.2.3. Soit l'alphabet $A = \{a, b\}$. L'ensemble $X = \{aa, baa, ba\}$ est un code sur A.
Procédons par l'absurde pour le montrer.
Supposons qu'il existe un mot w de X^+, de longueur minimale, qui possède deux factorisations distinctes :

$$\begin{aligned} w &= x_1 x_2 \cdots x_m \\ &= y_1 y_2 \cdots y_n \end{aligned}$$

avec $m, n \geq 1$ et $x_i, y_j \in X$.
Puisque l'on a supposé que w était de longueur minimale, on a forcément $x_1 \neq y_1$, ce qui implique que soit x_1 est préfixe de y_1, soit y_1 est préfixe de x_1. Supposons que x_1 est préfixe de y_1. Vu la définition de notre ensemble X, on a $x_1 = ba$ et $y_1 = baa$. Ainsi a doit être un préfixe de x_2, impliquant alors que $x_2 = aa$. Ensuite, on voit que a doit être un préfixe de y_2 également, ce qui entraine que $y_2 = aa$:

$$\begin{aligned} w &= ba\ aa \cdots x_m \\ &= baa\ aa \cdots y_n \end{aligned}$$

Ainsi, on a que $y_1 = x_1 a$ et $y_1 y_2 = x_1 x_2 a$.
Si l'on suppose avoir $y_1 y_2 \cdots y_p = x_1 x_2 \cdots x_p a$ (avec $p \leq \min\{m, n\}$) , on aura forcément $y_{p+1} = aa$ puis $x_{p+1} = aa$. On obtient alors

$$y_1 y_2 \cdots y_{p+1} = x_1 x_2 \cdots x_{p+1} a.$$

Finalement, il vient

$$w = y_1 y_2 \cdots y_n = x_1 x_2 \cdots x_n a,$$

or $a \notin X$, ce qui contredit le fait que w possède deux factorisations distinctes en mots de X.

Exemple 1.2.4. Soit l'alphabet $A = \{a, b\}$. L'ensemble $X = \{a, ab, ba\}$ n'est, quant à lui, pas un code puisque le mot $w = aba$ possède deux factorisations distinctes en mots de X :

$$w = ab \cdot a = a \cdot ba.$$

Proposition 1.2.5. *Tout sous-ensemble d'un code est encore un code.*

Définition 1.2.6. Si p est un naturel non nul, l'ensemble $X = A^p$ est appelé *le code uniforme des mots de longueur p.*

Il s'agit en effet d'un code : soit $w \in X^*$ tel que

$$w = x_1 x_2 \cdots x_m = y_1 y_2 \cdots y_n,$$

avec $m, n \geq 1$ et $x_i, y_j \in X$. Puisque les mots de X sont de longueur constante p, la seule manière de factoriser w en mots de X est de considérer des blocs de longueur p, ce qui implique donc que $m = n$ et $x_i = y_i$ pour $i = 1, \ldots, m$.

Le résultat suivant nous permet de reformuler la problématique de l'étude des codes en un problème purement algébrique : les codes correspondent en effet exactement aux morphismes injectifs d'un monoïde libre dans un autre.

Théorème 1.2.7. *Si un sous-ensemble X de A^* est un code alors toute bijection d'un alphabet B (éventuellement infini) dans X s'étend en un morphisme injectif de B^* dans A^*.*

Inversement, s'il existe un morphisme injectif $\beta : B^* \to A^*$ tel que $X = \beta(B)$ alors X est un code.

Démonstration. Soit $f : B \to X$ une bijection. On définit l'application $\beta : B^* \to A^*$ de la façon suivante :

$$\beta(\varepsilon) = \varepsilon \text{ et } \beta(b) = f(b_1)f(b_2)\cdots f(b_n)$$

pour tout $b \in B^*$ tel que $b = b_1 b_2 \cdots b_n$, avec $b_i \in B$ pour $i = 1, \ldots, n$.
Montrons que β est un morphisme injectif étendant f :

- β est un morphisme :
  En effet, vu la définition de β on a $\beta(\varepsilon) = \varepsilon$.
  De plus, soient $b, b' \in B^*$ tels que $b = b_1 b_2 \cdots b_n$ et $b' = b'_1 b'_2 \cdots b'_m$. On a alors :

$$\begin{aligned} \beta(b \cdot b') &= \beta(b_1 b_2 \cdots b_n b'_1 b'_2 \cdots b'_m) \\ &= f(b_1)f(b_2)\cdots f(b_n)f(b'_1)f(b'_2)\cdots f(b'_m) \\ &= \beta(b)\beta(b'). \end{aligned}$$

- β étend f :
  Si $b \in B$, alors on a simplement $\beta(b) = f(b)$.

- β est injectif :
  Soient $u, v \in B^*$ des mots tels que $\beta(u) = \beta(v)$. Notons

$$u = b_1 b_2 \cdots b_m \text{ et } v = b'_1 b'_2 \cdots b'_n,$$

  avec $m, n \geq 1$ et $b_i, b'_j \in B$.
  Par définition de β, on obtient que

$$\beta(u) = f(b_1)f(b_2)\cdots f(b_m) = f(b'_1)f(b'_2)\cdots f(b'_n) = \beta(v).$$

avec $f(b_i), f(b'_j) \in X$.
L'ensemble X étant un code, il vient $m = n$ et $f(b_i) = f(b'_i)$ pour $i = 1, \dots m$.
On a alors que $b_i = b'_i$ pour $i = 1, \dots, m$, puisque $f : B \to X$ est en particulier injectif. Finalement on obtient bien que $u = v$.

Inversement, soit $\beta : B^* \to A^*$ un morphisme injectif. Si on a

$$x_1 x_2 \cdots x_m = y_1 y_2 \cdots y_n$$

pour $m, n \geq 1$ et $x_i, y_j \in X = \beta(B)$, alors il existe des $b_i, b'_j \in B$ tels que $x_i = \beta(b_i)$ et $y_j = \beta(b'_j)$. Il vient alors

$$x_1 x_2 \cdots x_m = \beta(b_1 b_2 \cdots b_m)$$

et

$$y_1 y_2 \cdots y_n = \beta(b'_1 b'_2 \cdots b'_n).$$

Ainsi, par injectivité de β, on a $b_1 b_2 \cdots b_m = b'_1 b'_2 \cdots b'_n$. Puisque les b_i, b'_j sont des lettres, il vient que $m = n$ et $b_i = b'_i$ pour $i = 1, \dots, m$.
Finalement on obtient $x_i = y_i$ pour $i = 1, \dots, m$, ce qui conclut que X est bien un code.
□

Définition 1.2.8. Soient A, B deux alphabets, B éventuellement infini, et $X \subset A^*$. Un morphisme $\beta : B^* \to A^*$ qui est injectif et tel que $X = \beta(B)$ est appelé un *codage* pour X.

Le théorème et la définition qui précèdent nous permettent de faire le lien entre nos codes et ceux utilisés dans le cadre de la théorie de l'information. Ceci montre en effet que les codes tels que nous les étudions sont précisément les images des codages à déchiffrage unique.

Exemple 1.2.9. Grâce au résultat précédent, on obtient une autre façon de montrer que l'ensemble X de l'Exemple 1.2.4. n'est pas un code.
En effet, considérons le morphisme $\varphi : B^* = \{1, 2, 3\}^* \to A^* = \{a, b\}^*$ tel que

$$\varphi(1) = a, \varphi(2) = ab, \varphi(3) = ba.$$

Il n'est pas injectif puisque $\varphi(13) = aba = \varphi(21)$. Ainsi, $\varphi(B) = X = \{a, ab, ba\}$ n'est pas un code.

Le corollaire suivant nous montre que l'image (resp. l'image inverse) d'un code par un morphisme injectif est encore un code.

Corollaire 1.2.10. *Soit $\alpha : A^* \to C^*$ un morphisme injectif. Si X est un code sur A alors $\alpha(X)$ est un code sur C et si Y est un code sur C alors $\alpha^{-1}(Y)$ est un code sur A.*

Démonstration. Soit $\beta : B^* \to A^*$ un codage pour X.
On a que

$$\alpha \circ \beta : B^* \to C^*$$

est un morphisme injectif.
En appliquant le Théorème 1.2.7, on obtient que $\alpha(\beta(B)) = \alpha(X)$ est un code.

Soit $X = \alpha^{-1}(Y)$. Soient $m, n \geq 1$, $x_1, \ldots, x_m, x'_1, \ldots x'_n \in X$ tels que

$$x_1 x_2 \cdots x_m = x'_1 x'_2 \cdots x'_n.$$

Il vient alors que

$$\begin{aligned} &\alpha(x_1 x_2 \cdots x_m) = \alpha(x'_1 x'_2 \cdots x'_n) \\ \Leftrightarrow\ &\alpha(x_1)\alpha(x_2)\cdots\alpha(x_m) = \alpha(x'_1)\alpha(x'_2)\cdots\alpha(x'_n) \end{aligned}$$

avec $\alpha(x_i), \alpha(x'_j) \in Y$.
Puisque Y est un code, on a $m = n$ et $\alpha(x_i) = \alpha(x'_i)$. Par injectivité, on en tire que $x_i = x'_i$ pour $i = 1, \ldots, m$. Ainsi $X = \alpha^{-1}(Y)$ est bien un code.

□

Exemple 1.2.11. L'ensemble $X = \{00, 01, 02, 1, 2\}$ est un code [1] sur l'alphabet $B = \{0, 1, 2\}$.
Soit $\alpha : B^* \to A^*$, avec $A = \{a, b\}$ un morphisme tel que

$$\alpha(0) = a, \alpha(1) = ab, \alpha(2) = bb.$$

Puisque l'ensemble $X' = \{a, ab, bb\}$ est un code [2], nous savons, grâce au Théorème 1.2.7, que α est injectif.
Ainsi l'ensemble $\alpha(X) = \{aa, aab, abb, ab, bb\}$ est un code.

Corollaire 1.2.12. *Si $X \subset A^*$ est un code, alors X^n est un code pour tout naturel non nul n.*

Démonstration. Soit $\beta : B^* \to A^*$ un codage pour X. On a [3] $X^n = (\beta(B))^n = (\beta(B^n))$.

Or on sait que B^n est un code uniforme. Ainsi, par le Corollaire 1.2.10, on sait que X^n est un code puisqu'il s'agit de l'image d'un code par un morphisme injectif.

□

Remarque 1.2.13. Le corollaire précédent nous montre que toute puissance d'un code est encore un code. Cependant le produit de deux codes distincts n'est pas nécessairement un code.

En effet, considérons les codes suivants :

$$X = \{a, ba\} \text{ et } Y = \{a, ab\}$$

1. Nous vérifierons aisément que l'ensemble X est bel et bien un code lorsque nous aurons défini les ensembles préfixes.
2. De la même façon, nous le vérifierons plus facilement après avoir défini les ensembles suffixes.
3. L'égalité $(\beta(B))^n = \beta(B^n)$ est assez directe en procédant par double inclusion.

sur l'alphabet $A = \{a, b\}$. Posons $Z = XY$, il vient

$$Z = \{aa, aab, baa, baab\}.$$

Considérons le mot $w = aabaab \in Z^*$. Il possède deux factorisations distinctes en mots de Z :

$$\begin{aligned} w &= aab \cdot aab \\ &= aa \cdot baab \end{aligned}$$

1.2.2 Ensembles préfixes, suffixes, bifixes

Soient $x, y \in A^*$. Nous notons $x \preceq y$ (resp. $x \prec y$) le fait que x est un préfixe (resp. préfixe propre) de y. La relation $\preceq$ est un ordre appelé *ordre préfixe*.

Définition 1.2.14. Un sous-ensemble X de A^* est dit *préfixe* si aucun élément de X n'est préfixe propre d'un autre élément de X.

Autrement dit, X est préfixe si pour tous $x, y \in X$, la relation $x \preceq y$ entraine $x = y$.

Remarque 1.2.15. Si un ensemble préfixe X contient le mot vide ε, alors on a $X = \{\varepsilon\}$.

Définition 1.2.16. Un sous-ensemble X de A^* est dit *suffixe* si aucun élément de X n'est suffixe propre d'un autre élément de X.

Définition 1.2.17. Un ensemble est dit *bifixe* s'il est à la fois préfixe et suffixe.

Proposition 1.2.18. *Tout ensemble préfixe (suffixe, bifixe) $X \neq \{\varepsilon\}$ est un code.*

Démonstration. Étant donné que $X \neq \{\varepsilon\}$, on sait déjà que $\varepsilon \notin X$.
Procédons alors par l'absurde. Supposons que l'ensemble X n'est pas un code.
Il existe donc un mot $w \in X^*$, que l'on suppose être de longueur minimale, qui possède deux factorisations distinctes en mots de X :

$$\begin{aligned} w &= x_1 x_2 \cdots x_m \\ &= y_1 y_2 \cdots y_n, \end{aligned}$$

avec $m, n \geq 1$ et $x_i, y_j \in X$.
On a donc $x_1 \neq y_1$ et par conséquent, on a soit $x_1 \prec y_1$, soit $y_1 \prec x_1$, ce qui contredit le fait que X est un ensemble préfixe.

□

Définition 1.2.19. Un *code préfixe (suffixe, bifixe)* est un ensemble préfixe (suffixe, bifixe) qui est un code, c'est-à-dire distinct de $\{\varepsilon\}$.

Nous vérifions maintenant que les ensembles de l'Exemple 1.2.11 sont effectivement des codes puisque

— l'ensemble $X = \{00, 01, 02, 1, 2\}$ est préfixe,

— l'ensemble $X' = \{a, ab, bb\}$ est suffixe.

Exemple 1.2.20. Les codes uniformes sont des codes bifixes.

1.2.3 Codes maximaux

Définition 1.2.21. Un code X est *maximal* sur un alphabet A si X n'est pas strictement inclus dans un autre code sur A, i.e. si $X \subset X'$ avec X' un code sur A implique $X = X'$.

Remarque 1.2.22. Le caractère maximal d'un code dépend évidemment de l'alphabet sur lequel il est défini. En effet, si $X \subset A^*$ et $A \subset B$, on a $X \subset B^*$. Ce n'est pas parce que X est maximal sur A qu'il l'est sur B.

Proposition 1.2.23. *Les codes uniformes sont maximaux.*

Démonstration. Soit un alphabet A. Considérons le code uniforme A^n pour un $n > 0$.
Procédons par l'absurde et supposons que A^n n'est pas maximal.
Il existe donc un mot $u \in A^+ \backslash A^n$ tel que $Y = A^n \cup \{u\}$ est un code. Considérons le mot $w = u^n$. On a $w \in Y^*$ mais également $w \in (A^n)^*$ puisque sa longueur est un multiple de n.
On a alors

$$w = u^n = x_1 x_2 \cdots x_{|u|}$$

avec $x_1, x_2, \ldots, x_{|u|} \in A^n \subset Y$.
Or $u \notin A^n$ donc w possède deux factorisations distinctes en mots de Y, ce qui contredit le fait que Y est un code.

□

Proposition 1.2.24. *Tout code X sur A est contenu dans un code maximal sur A.*

Démonstration. Soit $\mathcal{F}$ l'ensemble des codes sur A contenant X, ordonné par l'inclusion. Pour montrer que $\mathcal{F}$ contient un élément maximal, nous allons utiliser le Lemme de Zorn, i.e. nous allons montrer que toute chaine $\mathcal{C}$ de $\mathcal{F}$ admet un majorant. Notons que $\mathcal{F}$ est non vide puisqu'il contient X.
Considérons une chaine $\mathcal{C}$ de codes contenant X.
Posons $\mathcal{Y} = \cup_{Y \in \mathcal{C}} Y$. Montrons alors que $\mathcal{Y} \in \mathcal{F}$ et que $\mathcal{Y}$ majore $\mathcal{C}$.

- Il est évident, vu sa définition, que $\mathcal{Y}$ majore $\mathcal{C}$.
- On a $\mathcal{Y} \in \mathcal{F}$:
  En effet, par définition nous savons que $X \subset \mathcal{Y}$ puisque chaque $Y \in \mathcal{C}$ contient X. Il nous reste donc à montrer que $\mathcal{Y}$ est bien un code.
  Soient $m, n \geq 1$ et $x_1, \ldots, x_m, y_1, \ldots y_n \in \mathcal{Y}$ tels que :

  $$x_1 \cdots x_m = y_1 \cdots y_n.$$

  Il existe donc $Y_{1,1}, \ldots, Y_{1,m}, Y_{2,1}, \ldots, Y_{2,n} \in \mathcal{C}$ tels que $x_1 \in Y_{1,1}, \ldots, x_m \in Y_{1,m}, y_1 \in Y_{2,1}, \ldots y_n \in Y_{2,n}$. Nous avons alors $n + m$ éléments de $\mathcal{C}$, il y a donc un de ces éléments qui contient tous les autres puisque $\mathcal{C}$ est une chaine. Notons le Y'. Ainsi on a $x_i, y_j \in Y'$ qui est un code par définition. Il vient alors que $n = m$ et $x_i = y_i$ pour $i = 1, \ldots, m$.

□

Remarque 1.2.25. Cette proposition, comme nous le verrons dans la Section 2.2, n'est plus vraie si nous nous limitons à des codes finis, i.e. tout code fini n'est pas nécessairement inclus dans un code maximal fini.

1.3 Un algorithme pour les codes

Il n'est pas toujours facile de vérifier qu'un ensemble donné est un code. Dans cette section, nous allons décrire un algorithme qui nous permettra de déterminer si un ensemble de mots satisfait à la définition d'un code.

1.3.1 Idée de l'algorithme

Pour se familiariser avec l'algorithme, partons d'un exemple.
Soit $X = \{b, abb, abbba, bbba, baabb\}$.
Si nous souhaitions « naïvement » trouver un mot $w \in X^*$ possédant deux décompositions en mots de X de manière systématique, nous pourrions procéder de la manière suivante :

- Première étape :
  Nous commençons par chercher tous les mots de X qui sont préfixes d'autres mots de X. Vu notre exemple, nous obtenons les égalités suivantes :

$$(bbba) = (b)\underline{bba} \tag{1.1}$$
$$(baabb) = (b)\underline{aabb} \tag{1.2}$$
$$(abbba) = (abb)\underline{ba} \tag{1.3}$$

  Nous appellons les suffixes soulignés les *restes*.

- Deuxième étape :
  Nous essayons ensuite de compléter les égalités ci-dessus à partir de ces restes. Pour ce faire, nous avons deux possibilités :
  1. Soit nous cherchons un mot du code qui **a** le reste comme préfixe.
  2. Soit nous cherchons un mot du code qui **est** préfixe du reste.

  Le reste bba de l'égalité 1.1 a b comme préfixe mais n'est préfixe d'aucun mot de X, ainsi l'égalité devient

$$(bbba) = (b)(b)\underline{ba}. \tag{1.4}$$

  Le reste $aabb$ de l'égalité 1.2 n'est préfixe d'aucun mot de X et n'a aucun préfixe dans X. Cette égalité ne pouvant être complétée, elle ne peut déboucher sur une double factorisation.
  Le reste ba de l'égalité 1.3 a b comme préfixe et est préfixe de $baabb$. Nous obtenons donc ces deux nouvelles égalités :

$$(abbba) = (abb)(b)\underline{a} \tag{1.5}$$
$$(abbba)\underline{abb} = (abb)(baabb) \tag{1.6}$$

  À l'issue de cette étape nous obtenons donc de nouveaux restes.

- Troisième étape :
  La démarche est la même que celle appliquée précédemment, seuls les restes ont changé. Nous obtenons alors les égalités suivantes :

$$(bbba)\underline{abb} = (b)(b)(baabb) \tag{1.7}$$
$$(bbba) = (b)(b)(b)\underline{a} \tag{1.8}$$

$$(abbba)\underline{bbba} = (abb)(b)(abbba) \tag{1.9}$$
$$(abbba)\underline{bb} = (abb)(b)(abb) \tag{1.10}$$

$$(abbba)(abb) = (abb)(baabb) \tag{1.11}$$
$$(abbba)(abbba) = (abb)(baabb)\underline{ba} \tag{1.12}$$

  Remarquons que le reste de l'égalité 1.6 est lui-même un mot de X. C'est pourquoi l'égalité 1.11 nous donne une double factorisation du mot $w = abbbaabb$ en mots de X. Puisque nous avons obtenu une double factorisation, nous pouvons en conclure que l'ensemble X n'est pas un code et ainsi arrêter notre recherche.

Observons que dans la méthode présentée ci-dessus, seuls les restes nous importent. Par ailleurs, nous obtenons une double factorisation à l'égalité 1.11 car le reste est le mot vide.

Formalisons cette construction. Étant donné un ensemble $X \subset A^+$, posons :

$$U_1 = X^{-1}X\backslash\{\varepsilon\}$$
$$U_{n+1} = X^{-1}U_n \cup U_n^{-1}X \quad (n \geq 1).$$

Ainsi défini, l'ensemble U_n contient tous les restes calculés à l'étape n. Remarquons que l'ensemble $X^{-1}U_n$ correspond aux restes de U_n ayant un préfixe dans X et $U_n^{-1}X$ correspond aux restes de U_n étant préfixes d'un mot de X.

De manière évidente, l'algorithme se termine lorsqu'un des ensembles U_n contient le mot vide puisque cela signifie que l'on a trouvé un mot qui possède deux factorisations distinctes et par conséquent que X n'est pas un code. Par ailleurs, il est également clair que dès qu'il existe $i \neq j$ tels que $U_i = U_j$, les ensembles U_n seront périodiques : nous pouvons alors arrêter l'algorithme dès que nous sommes dans un tel cas. Si le mot vide n'apparait dans aucun des U_n alors l'ensemble X sera un code.
Nous montrerons dans la suite qu'avec cette deuxième condition l'algorithme s'arrête toujours dans le cas où X est un ensemble régulier.

Remarquons que dans le cas où $X \subset A^+$ est un ensemble préfixe, nous avons directement

$$U_1 = \emptyset.$$

Par conséquent, $U_2 = U_1$, l'algorithme s'arrête et on en conclut que X est un code.

Notons que cette méthode ne nous était pas totalement inconnue : c'est en effet essentiellement la méthode qui nous a permis de conclure que l'ensemble X de l'Exemple 1.2.3 est bien un code (les restes seront toujours égaux à a).

1.3.2 Exemples

Illustrons l'algorithme décrit ci-dessus sur deux exemples.

Exemple 1.3.1. Considérons l'ensemble $X = \{ab, baa, bbab, abbb\}$ sur l'alphabet $A = \{a, b\}$. Calculons la suite $(U_n)_{n\geq 1}$:

$$U_1 = \{bb\}$$
$$U_2 = \{ab\}$$
$$U_3 = \{\varepsilon, bb\}$$

Ainsi $\varepsilon \in U_3$, donc X n'est pas un code.

Deux décompositions d'un même mot peuvent être obtenues en « remontant » l'algorithme : ε apparait dans U_3 car ab est un mot de X, ab apparait dans U_2 parce que bb est préfixe de $bbab$, et bb apparait dans U_1 parce que ab est préfixe de $abbb$. Ceci nous fournit la double décomposition

$$(abbb)(ab) = (ab)(bbab).$$

Exemple 1.3.2. Considérons l'ensemble $X = \{ab, bab, bba, bbba, abaa\}$ sur l'alphabet $A = \{a, b\}$. Calculons la suite $(U_n)_{n\geq 1}$:

$$U_1 = \{aa\}$$
$$U_2 = \emptyset$$
$$U_3 = \emptyset$$

Nous sommes donc dans le cas où $U_2 = U_3$, l'algorithme s'achève et puisque ε n'est contenu dans aucun des ensembles U_n. On en déduit que X est un code.

1.3.3 Résultats

Présentons maintenant les résultats justifiant formellement l'algorithme décrit précédemment.

Lemme 1.3.3. *Soit $X \subset A^+$ et soit la suite $(U_n)_{(n\geq 1)}$ définie comme ci-dessus.*
Pour tout $n \geq 1$, on a $w \in U_n$ si et seulement s'il existe des entiers $p, q \geq 1$ avec $p+q = n+1$ et des mots $x_1, \ldots, x_p, y_1, \ldots, y_q \in X$ avec $x_1 \neq y_1$ et w suffixe de y_q tels que

$$x_1 x_2 \cdots x_p w = y_1 y_2 \cdots y_q$$

Démonstration. Montrons tout d'abord que la condition est nécessaire. Procédons par récurrence sur n.

- Cas de base : $n = 1$
  Si $w \in U_1$, alors par définition on a

  $$xw = y,$$

  avec $x, y \in X$ et $x \neq y$. On a bien que w est suffixe de y.

- Induction : Supposons que la propriété est vraie pour tout $j < n$ et montrons-la pour n.
  Soit $w \in U_n$. On a soit $xw = v$, soit $vw = x$ pour un certain $x \in X$ et un certain $v \in U_{n-1}$. Par hypothèse de récurrence, il existe des entiers $p, q \geq 1$ avec $p + q = n$ et des mots $x_1, \ldots, x_p, y_1, \ldots, y_q \in X$ avec $x_1 \neq y_1$ et v suffixe de y_q tels que

  $$x_1 x_2 \cdots x_p v = y_1 y_2 \cdots y_q,$$

  — Cas 1 : Si $xw = v$, alors

  $$x_1 x_2 \cdots xw = y_1 y_2 \cdots y_q,$$

  et la condition est satisfaite par $x_1, \ldots, x_p, x_{p+1}, y_1, \ldots, y_q$ où $x_{p+1} = x$, puisque w est bien suffixe de y_q.

  — Cas 2 : Si $vw = x$, alors

  $$\begin{aligned} x_1 x_2 \cdots x_p vw &= y_1 y_2 \cdots y_q w \\ x_1 x_2 \cdots x_p x &= y_1 y_2 \cdots y_q w, \end{aligned}$$

  et la condition est vérifiée par $y_1, \ldots, y_q, x_1, \ldots, x_p, x_{p+1}$ où $x_{p+1} = x$, puisque w est suffixe de x.

Inversement, supposons qu'il existe des entiers $p, q \geq 1$ tels que $p + q = n + 1$ et des mots $x_1, \ldots, x_p, y_1, \ldots, y_q \in X$ avec $x_1 \neq y_1$ et w suffixe de y_q tels que

$$x_1 x_2 \cdots x_p w = y_1 y_2 \cdots y_q. \tag{1.13}$$

Procédons par récurrence sur n pour montrer que $w \in U_n$.

- Cas de base : $n = 1$
  On a forcément que $p = q = 1$ et donc

  $$x_1 w = y_1$$

  avec $x_1, y_1 \in X$ et $x_1 \neq y_1$. Ainsi, $w \in U_1 = X^{-1}X \backslash \{\varepsilon\}$.

- Induction : Supposons que la propriété est vraie pour tout $j < n$ et montrons-la pour n.
  Puisque w est suffixe de y_q, on a $y_q = vw$ pour un certain $v \in A^*$. L'égalité 1.13 devient donc
  $$x_1x_2\cdots x_p = y_1y_2\cdots y_{q-1}v.$$
  Posons $v = v'x_{r+1}\cdots x_p$, où v' est suffixe de x_r pour un r tel que $1 \leq r \leq p$. Alors
  $$x_1\cdots x_r = y_1\cdots y_{q-1}v'.$$
  — Cas 1 : $q = 1$. Alors $p > 1$. Dans ce cas, on peut supposer $r = 1$. Il vient alors $x_1 = v' \in X$ et $y_1 = v'x_2\cdots x_pw$. Ainsi, on a $x_2\cdots x_pw \in X^{-1}X\backslash\{\varepsilon\} = U_1 = U_{r+q-1}$.
  — Cas 2 : $q > 1$. Par hypothèse de récurrence, $v' \in U_{r+q-2}$.
  Ensuite, puisque $y_q = v'x_{r+1}\cdots x_pw$, on a $x_{r+1}\cdots x_pw \in U_{r+q-2}^{-1}X \subset U_{r+q-1}$.
  Montrons maintenant par récurrence sur i, $1 \leq i \leq p - r$, que l'on a
  $$x_{r+i}\cdots x_pw \in U_{r+q+i-2}.$$
  — Cas de base : $i = 1$
  On a bien, vu ce qui précède :
  $$x_{r+1}\cdots x_pw \in U_{r+q-1}.$$
  — Induction : Supposons que $x_{r+j}\cdots x_pw \in U_{r+q+j-2}$ pour $j \leq i$. Montrons-le pour $i + 1$.
  On a, par hypothèse de récurrence
  $$x_{r+i}\cdots x_pw \in U_{r+q+i-2}.$$
  Puisque $x_{r+i} \in X$, il vient alors que $x_{r+1+i}\cdots x_pw \in U_{r+q+i-1}$.
  On obtient alors finalement que $x_pw \in U_{p+q-2}$ et par conséquent
  $$w \in X^{-1}U_{p+q-2} \subset U_{p+q-1},$$
  où $p + q - 1 = n$.

□

Théorème 1.3.4. *L'ensemble $X \subset A^+$ est un code si et seulement si aucun des ensembles U_n définis ci-dessus ne contient le mot vide ε.*

Démonstration. Si X n'est pas un code, alors il existe un mot $w \in X^*$, que l'on suppose de longueur minimale, qui possède deux factorisations distinctes en mots de X :
$$\begin{aligned} w &= x_1x_2\cdots x_p \\ &= y_1y_2\cdots y_q \end{aligned}$$

avec $x_1, \ldots, x_p, y_1, \ldots, y_q \in X$ et $x_1 \neq y_1$.
Par le lemme précédent, il vient que $\varepsilon \in U_{p+q+1}$.

Inversement, si $\varepsilon \in U_n$, par le lemme, il existe des entiers $p, q \geq 1$ tels que $p+q = n+1$ et des mots $x_1, \ldots, x_p, y_1, \ldots, y_q \in X$, avec $x_1 \neq y_1$ tels que

$$x_1 x_2 \cdots x_p \ = y_1 y_2 \cdots y_q.$$

Ceci montre que X n'est pas un code.

□

Lorsque que nous considérons le cas où X est fini ou plus généralement où X est régulier, nous allons voir que le nombre d'ensembles U_n est fini, ce qui entraine que l'algorithme s'achève au vu de la deuxième condition d'arrêt. Il en découle alors que le nombre de calculs nécessaires pour prouver que X est un code est fini. Ainsi, on en conclut qu'il est décidable qu'un ensemble fini ou régulier est un code.

Procédons tout d'abord à quelques rappels :

Définition 1.3.5. Soit $X \subset A^*$. On définit la *congruence syntaxique* $\equiv_X$ *de* X de la façon suivante. Soient $x, y \in X$, on pose

$$x \equiv_X y \Leftrightarrow (\forall u, v \in A^* : uxv \in X \Leftrightarrow uyv \in X).$$

Proposition 1.3.6. *Muni de l'opération* $\circ : A^*/\equiv_X \times \ A^*/\equiv_X \to A^*/\equiv_X : ([x], [y]) \mapsto [x] \circ [y]$*, l'ensemble quotient* $A^*/\equiv_X$ *possède une structure de monoïde.*

Définition 1.3.7. Le monoïde $(A^*/\equiv_X, \circ)$ est le *monoïde syntaxique* de X.

Proposition 1.3.8. *Un ensemble* X *est régulier si et seulement si son monoïde syntaxique est fini.*

Avant de démontrer le second résultat qui nous intéresse dans cette sous-section, présentons tout d'abord quelques résultats intermédiaires qui rendront plus abordable la démonstration de celui-ci.

Proposition 1.3.9. *Un ensemble* $X \subset A^*$ *est une union de classes d'équivalence d'une congruence* $\mathcal{R}$ *si et seulement si pour tous* $x \in X$ *et* $y \in A^*$ *tels que* $x\mathcal{R}y$ *, on a* $y \in X$*.*

Démonstration. Supposons d'abord que $X = \bigcup_i [z_i]_{\mathcal{R}}$.
Soient $x \in X, y \in A^*$ tels que $x\mathcal{R}y$. Il existe donc un i tel que $x \in [z_i]_{\mathcal{R}}$. On a alors $x\mathcal{R}z_i$. Puisque $\mathcal{R}$ est une congruence, il vient que $y\mathcal{R}z_i$, i.e. $y \in [z_i]_{\mathcal{R}}$. Par conséquent, $y \in X$.

Supposons ensuite que pour tous $x \in X, y \in A^*$ tels que $x\mathcal{R}y$, on a $y \in X$.
Montrons que $X = \bigcup_{w \in X} [w]_{\mathcal{R}}$ par double inclusion.

$\subset$: Soit $x \in X$. On a forcément $x \in [x]_{\mathcal{R}}$ et donc $x \in \bigcup_{w \in X} [w]_{\mathcal{R}}$.

$\supset$: Soit $x \in \bigcup_{w \in X} [w]_{\mathcal{R}}$. Il existe donc $w \in X$ tel que $x \in [w]_{\mathcal{R}}$. Ainsi $x \mathcal{R} w$ et donc par hypothèse, $x \in X$.

□

Proposition 1.3.10. *Tout ensemble $X \subset A^*$ est une union de classes d'équivalence de sa congruence syntaxique $\equiv_X$.*

Démonstration. Soient $x \in X, y \in A^*$ tels que $x \equiv_X y$. Alors, pour tous $u, v \in A^*$ on a

$$uxv \in X \Leftrightarrow uyv \in X.$$

En particulier, $\varepsilon x \varepsilon = x \in X \Leftrightarrow \varepsilon y \varepsilon = y \in X$. On a donc bien $y \in X$.

□

Lemme 1.3.11. *Si $X \subset A^*$ est une union de classes d'équivalence d'une congruence $\mathcal{R}$ alors pour tout sous-ensemble Y de A^*, $Y^{-1}X$ est une union de classes d'équivalence de $\mathcal{R}$.*

Démonstration. Soient $x \in Y^{-1}X, x' \in A^*$ tels que $x \mathcal{R} x'$. Alors $yx \in X$ pour un certain $y \in Y$. Puisque $\mathcal{R}$ est une congruence il vient que $yx \mathcal{R} yx'$. Comme X est une union de classes d'équivalence, par la Proposition 1.3.9 il vient que $yx' \in X$. Ainsi, $x' \in Y^{-1}X$.

□

Nous avons désormais tous les outils nécessaires à la démonstration du résultat suivant.

Théorème 1.3.12. *Si $X \subset A^+$ est régulier alors l'ensemble de tous les U_n $(n \geq 1)$ est fini.*

Démonstration. Soient $\equiv_X$ la congruence syntaxique de X et μ la congruence[4] sur A^* qui a pour classes $\{\varepsilon\}$ et A^+, i.e. $x \mu y \Leftrightarrow x, y \neq \varepsilon$ ou $x = y = \varepsilon$. Posons $\iota = \equiv_X \cap \mu$.
Nous allons montrer que U_n est une union de classes d'équivalence de ι par récurrence sur $n \geq 1$.

- Cas de base : $n = 1$
  Vu la Proposition 1.3.10, X est une union de classes d'équivalence de $\equiv_X$. Grâce au Lemme 1.3.11, on sait que $X^{-1}X$ est également une union de classes de $\equiv_X$. Vu la définition de ι, $X^{-1}X$ est une union de classes de ι. De plus, comme $\{\varepsilon\}$ est une classe de ι, $X^{-1}X \backslash \{\varepsilon\} = U_1$ est encore une union de classes de ι.
- Induction : Supposons que les ensembles U_k sont des unions de classes d'équivalence pour $k \leq n$ et montrons que U_{n+1} est encore une union de classes d'équivalence.
  Par hypothèse de récurrence nous savons que U_n est une union de classes d'équivalence de ι. De plus, nous savons aussi que X est une union de classes d'équivalence de ι. Ainsi $U_n^{-1}X$ et $X^{-1}U_n$ le sont aussi vu le Lemme 1.3.11. Puisque $U_{n+1} = U_n^{-1}X \cup X^{-1}U_n$, U_{n+1} est encore une union de classes d'équivalence de ι.

4. On vérifie facilement qu'il s'agit bien d'une congruence.

Il vient donc que U_n est une union de classes d'équivalence de ι.
Puisque ι possède un nombre fini de classes d'équivalence [5], il n'y a qu'un nombre fini de U_n.

□

Remarque 1.3.13. La réciproque de ce Théorème n'est pas vraie.
En effet, considérons l'ensemble $X = \{a^n b^n | n \geq 1\}$ qui n'est pas régulier [6]. Puisque X est bifixe nous avons $U_1 = \emptyset$. Ainsi, l'ensemble des U_n est fini.

1.4 Codes et sous-monoïdes libres

Dans cette section, nous considérons le sous-monoïde X^* généré par un code X. Nous verrons notamment qu'il est possible de montrer qu'un ensemble de mots est un code en connaissant uniquement le sous-monoïde qu'il génère.

Proposition 1.4.1. *Tout sous-monoïde M de A^* possède un unique ensemble minimal de générateurs*

$$X = (M\backslash\{\varepsilon\})\backslash(M\backslash\{\varepsilon\})^2$$

Démonstration. Dans un premier temps, nous allons montrer que X engendre M, i.e. $X^* = M$.
Nous savons déjà, par définition de X, que $X \subset M$. Puisque M est un sous-monoïde, il est stable pour le produit et donc $X^* \subset M$. Nous allons montrer la seconde inclusion en procédant par récurrence sur la longueur des mots de M :

- Cas de base :
  On a $\varepsilon \in X^*$.
- Induction : Supposons que la propriété est vérifiée pour les mots de longueur $j < k$ et montrons-la pour les mots de longueur k.
  Soit $m \in M\backslash\{\varepsilon\}$ tel que $|m| = k$.
  - Cas 1 : si $m \notin (M\backslash\{\varepsilon\})^2$, alors on a directement que $m \in X \subset X^*$.
  - Cas 2 : si $m \in (M\backslash\{\varepsilon\})^2$, alors $m = m_1 m_2$ où $m_1, m_2 \in (M\backslash\{\varepsilon\})$ et $|m_1|, |m_2| < |m| = k$. Ainsi par hypothèse de récurrence, $m_1, m_2 \in X^*$. Et par stabilité pour le produit, on obtient $m \in X^*$.

  Nous avons donc bien l'égalité $M = X^*$.
  Montrons maintenant le caractère minimal de X.

Considérons un ensemble Y de générateurs de M. On peut supposer que $\varepsilon \notin Y$. On a donc que chaque $x \in X \subset M$ est dans Y^* et peut donc s'écrire de la manière suivante :

$$x = y_1 y_2 \cdots y_n$$

5. En effet, nous savons que $\equiv_X$ possède un nombre fini de classes d'équivalence vu la Proposition 1.3.8, et que μ possède deux classes d'équivalence. Ainsi puisque les classes de ι sont les intersections des classes de $\equiv_X$ et μ, il y en a un nombre fini.

6. Nous le vérifions aisément grâce au Lemme de la Pompe.

avec $y_i \in Y$ et $n > 0$.
Puisque $x \neq \varepsilon$ et $x \notin (M\backslash\{\varepsilon\})^2$, on a forcément $n = 1$. Autrement dit, $x = y_1 \in Y$. Il vient donc $X \subset Y$.
Ainsi, X est bien un ensemble minimal de générateurs, un tel ensemble est unique.

□

Remarque 1.4.2. L'ensemble $X = (M\backslash\{\varepsilon\})\backslash(M\backslash\{\varepsilon\})^2$ est un ensemble minimum de générateurs de M.

Exemple 1.4.3. Soit $A = \{a, b\}$ et posons $M = \{w \in A^* |\ |w|_a \equiv 0 \mod 2\}$
On voit [7] que $M = (b+ab^*a)^*$. Ainsi, l'ensemble minimal de générateurs de M est l'ensemble $X = b \cup ab^*a$. En effet, montrons que $X = (M\backslash\{\varepsilon\})\backslash(M\backslash\{\varepsilon\})^2$ par double inclusion.
On sait déjà que $X^* = M$. Ainsi X est un ensemble de générateurs de M et par conséquent il contient $(M\backslash\{\varepsilon\})\backslash(M\backslash\{\varepsilon\})^2$. Réciproquement, soit $x \in X$. Il est clair que $x \in M\backslash\{\varepsilon\}$. Par ailleurs, x ne peut se décomposer en deux mots non vides de M. En effet cela est clair si $x = b$, et sinon, chaque mot de la décomposition contiendrait exactement un a.

Nous allons maintenant nous intéresser aux sous-monoïdes engendrés par des codes.

Définition 1.4.4. Un sous-monoïde M de A^* est dit *libre* s'il existe un isomorphisme

$$\alpha : B^* \to M$$

d'un monoïde libre B^* (B éventuellement infini) dans M.

La proposition suivante nous donne un critère d'équivalence au fait d'être libre.

Théorème 1.4.5. *Si M est un sous-monoïde libre de A^*, alors son ensemble minimal de générateurs est un code.*

Inversement, si $X \subset A^*$ est un code, alors le sous-monoïde X^* de A^* est libre et X est son ensemble minimal de générateurs.

Démonstration. Soit $\alpha : B^* \to M$ un isomorphisme. Alors $\alpha' : B^* \to A^* : b \mapsto \alpha(b)$ est un morphisme injectif. Par le Théorème 1.2.7, l'ensemble $X = \alpha(B)$ est un code. On a $M = \alpha(B^*) = (\alpha(B))^* = X^*$. Donc X engendre M.
De plus, $B = B^+\backslash B^+B^+$ et $\alpha(B^+) = M\backslash\{\varepsilon\}$. Il vient donc

$$\begin{aligned} X &= \alpha(B) \\ &= \alpha(B^+\backslash B^+B^+) \\ &= \alpha(B^+)\backslash\alpha(B^+B^+) \\ &= \alpha(B^+)\backslash\alpha(B^+)^2, \end{aligned}$$

ce qui montre que X est l'ensemble minimal de générateurs de M.

7. Il est assez aisé de construire un automate acceptant le language M.

Inversement, supposons que $X \subset A^*$ est un code et considérons un codage pour X

$$\alpha : B^* \to A^*.$$

Par définition, α est injectif et α est une bijection de B dans X. On a alors que α est une bijection entre B^* et $\alpha(B^*) = X^*$ et ainsi X^* est libre.
Il vient :

$$\begin{aligned} X &= \alpha(B) \\ &= \alpha(B^+) \backslash \alpha(B^+)\alpha(B^+) \\ &= \alpha(B)^+ \backslash \alpha(B)^+\alpha(B)^+ \\ &= X^+ \backslash X^+X^+ \\ &= (X^* \backslash \{\varepsilon\}) \backslash (X^* \backslash \{\varepsilon\})^2 \end{aligned}$$

Ainsi, X est bien l'ensemble minimal des générateurs de X^*.

□

Remarque 1.4.6. Le théorème précédent nous permet de construire des sous-monoïdes libres : il suffit simplement de prendre le sous-monoïde engendré par un code.

Si l'on s'intéresse à l'étude des codes non plus dans des monoïdes libres mais dans des groupes libres, cela revient à étudier les sous-groupes d'un groupe libre. Or, nous savons que tout sous-groupe d'un groupe libre est libre, ce qui n'est pas le cas dans les monoïdes libres, i.e. un sous-monoïde d'un monoïde libre n'est pas nécessairement libre. En effet, considérons, par exemple, l'alphabet $A = \{a, b\}$ et le sous-monoïde $M = \{\varepsilon\} \cup A^*bA^*$ de A^*. Montrons par l'absurde que M n'est pas un sous-monoïde libre. Supposons qu'il existe un alphabet C et un isomorphisme φ tel que

$$\varphi : C^* \to M.$$

Puisque φ est en particulier injectif, nous avons $|\varphi(w)| \geq |w|$ pour tout $w \in C^*$. Comme $b \in M$, il existe $c \in C^*$ tel que $\varphi(c) = b$. Ainsi $1 = |b| \geq |c|$, ce qui implique que $c \in C$ ($c \neq \varepsilon$, par injectivité). Les mots ab et ba sont également dans M, il existe donc $d, e \in C^*$ tels que $\varphi(d) = ab$ et $\varphi(e) = ba$. On a donc $|d|, |e| \leq 2$. Cependant puisque $a \notin M$, il vient $|d|, |e| \neq 2$. On a donc forcément $d, e \in C$. On observe alors que

$$\varphi(cd) = bab = \varphi(ec),$$

ce qui contredit l'injectivité de φ puisque $c \neq d, e$.

Le Théorème 1.4.5 implique les deux résultats suivants :

Corollaire 1.4.7. *Un sous-monoïde M de A^* est libre si et seulement s'il existe un code $X \subset M$ tel que $M = X^*$.*

Corollaire 1.4.8. *Soient X et Y des codes sur un même alphabet A. Si $X^* = Y^*$ alors $X = Y$.*

Définition 1.4.9. Le code X qui engendre un sous-monoïde libre M de A^* est appelé *la base de M.*

Remarque 1.4.10. Soit $M = X^*$ le sous-monoïde de A^* engendré par $X \subset A^+$. Si X est un code alors X est la base de M.

Dès lors, si un ensemble X n'est pas un code, nous avons deux possibilités (non exclusives) :

1. X n'est pas l'ensemble minimal de générateurs de X^*.
2. X^* n'est pas libre.

Définition 1.4.11. Un sous-monoïde N d'un monoïde M est *stable* dans M si pour tous $u, v, w \in M$

$$u, v, uw, wv \in N \Rightarrow w \in N,$$

ce qui peut se réécrire

$$N^{-1}N \cap NN^{-1} \subset N.$$

Dans la définition précédente, N étant un sous-monoïde, l'inclusion est en fait une égalité. En effet, si $w \in N$ alors $ww \in N$, ce qui montre que $w \in N^{-1}N \cap NN^{-1}$.

La proposition suivante nous montre que la notion de stabilité coïncide avec celle de liberté dans le cas des monoïdes libres.

Proposition 1.4.12. *Un sous-monoïde N de A^* est stable si et seulement s'il est libre.*

Démonstration. Supposons que N est stable. Posons $X = (N\backslash\{\varepsilon\})\backslash(N\backslash\{\varepsilon\})^2$ et montrons que X est un code car dans ce cas, $X^* = N$ sera libre.
Procédons par l'absurde et supposons que X n'est pas un code.
Il existe donc un mot $z \in X^*$, supposé de longueur minimale, qui possède deux factorisations distinctes en mots de X :

$$z = x_1x_2 \cdots x_m = y_1y_2 \cdots y_n$$

avec $m, n \geq 1$ et $x_i, y_j \in X \subset N$. Puisque l'on a $x_1 \neq y_1$, on peut supposer que x_1 est préfixe propre de y_1. Ainsi, on a

$$y_1 = x_1w$$

pour un certain mot non vide w. On a alors que

$$x_1, \quad y_2 \cdots y_n, \quad x_1w = y_1, \quad wy_2 \cdots y_n = x_2 \cdots x_m$$

sont tous dans N. Puisque N est stable, il vient que $w \in N$. Ainsi on a

$$y_1 = x_1w \notin X$$

puisqu'un mot de X ne peut pas s'écrire comme le produit de deux mots non vides de N. Or, on a supposé $y_1 \in X$. L'ensemble X est donc un code.

Supposons maintenant que N est libre et soit X sa base.
Soient $u, v, w \in A^*$ et supposons que u, v, uw et wv appartiennent à $N = X^*$. Posons

$$u = x_1 \cdots x_k, \quad uw = y_1 \cdots y_l, \quad wv = x_{k+1} \cdots x_r, \quad v = y_{l+1} \cdots y_s$$

avec $x_i, y_j \in X$. L'égalité $u \cdot wv = uw \cdot v$ implique

$$x_1 \cdots x_k \cdot x_{k+1} \cdots x_r = y_1 \cdots y_l \cdot y_{l+1} \cdots y_s.$$

Puisque X est un code, il vient que $r = s$ et $y_i = x_i$ pour $i = 1, \ldots, s$. De plus, on a $l \geq k$ puisque $|u| \leq |uw|$, donc

$$\begin{aligned} uw &= y_1 \cdots y_l \\ &= x_1 \cdots x_l \\ &= u x_{k+1} \cdots x_l. \end{aligned}$$

Ainsi, $w = x_{k+1} \cdots x_l \in X^* = N$. On en conclut donc que N est stable.

□

Définition 1.4.13. Soit M un monoïde.
Le sous-monoïde N de M est *unitaire à droite* dans M si pour tous $u, v \in M$

$$u, uv \in N \Rightarrow v \in N,$$

ce qui peut se réécrire

$$N^{-1}N \subset N. \tag{1.14}$$

Le sous-monoïde N de M est *unitaire à gauche* dans M si pour tous $u, v \in M$

$$u, vu \in N \Rightarrow v \in N,$$

ce qui peut se réécrire

$$NN^{-1} \subset N. \tag{1.15}$$

Le sous-monoïde N de M est *biunitaire* s'il est à la fois unitaire à gauche et à droite.

Dans la définition précédente, N étant un sous-monoïde, les inclusions 1.14 et 1.15 sont en fait des égalités. Ainsi N est biunitaire si et seulement si $N^{-1}N = N = NN^{-1}$.

Remarque 1.4.14. Observons qu'un sous-monoïde unitaire à gauche (resp. à droite) est en particulier stable.

Vu la remarque précédente, un sous-monoïde biunitaire de A^* est en particulier stable. Cependant, la réciproque n'est pas vraie. Le sous-monoïde engendré par un code préfixe (resp. suffixe) qui n'est pas suffixe (resp. préfixe) est libre vu la Proposition 1.4.5 et donc stable par la Proposition 1.4.12. La proposition suivante prouve que ce sous-monoïde est unitaire à droite (resp. à gauche) mais pas à gauche (resp. à droite).

Proposition 1.4.15. *Un sous-monoïde M de A^* est unitaire à droite (unitaire à gauche, biunitaire) si et seulement si son ensemble minimal de générateurs est un code préfixe (suffixe, bifixe).*

Démonstration. Soit $M \subset A^*$ un sous-monoïde. Soit $X = (M\backslash\{\varepsilon\})\backslash(M\backslash\{\varepsilon\})^2$ son ensemble minimal de générateurs.

Supposons que M est unitaire à droite et montrons que X est préfixe.
Soient $x, xu \in X \subset M$ pour un certain $u \in A^*$. Puisque M est unitaire à droite, il vient que $u \in M$. Si $u \neq \varepsilon$, alors $u \in M\backslash\{\varepsilon\}$ et $xu \in (M\backslash\{\varepsilon\})^2$, ce qui contredit le fait que $xu \in X$. Donc $u = \varepsilon$ et X est bien préfixe.

Inversement, supposons que X est préfixe. Soient $u, v \in A^*$ tels que $u, uv \in M = X^*$, montrons que $v \in M$. Posons

$$u = x_1 \cdots x_n \text{ et } uv = y_1 \cdots y_m,$$

avec $x_i, y_j \in X$. Il vient donc que $x_1 \cdots x_n v = y_1 \cdots y_m$. L'ensemble X étant préfixe, x_1 ne peut pas être préfixe propre de y_1 et inversement. Nous avons donc $x_1 = y_1$. En continuant de proche en proche, on obtient finalement $x_i = y_i$ pour $i = 1, \ldots, n$. Il vient alors que

$$v = y_{n+1} \cdots y_m \in M,$$

ce qui montre que M est unitaire à droite.

□

Définition 1.4.16. Un sous-monoïde libre M de A^* est *maximal* si $M \neq A^*$ et si M n'est pas contenu strictement dans un autre sous-monoïde libre, excepté A^*.

Proposition 1.4.17. *Si M est un sous-monoïde libre maximal de A^*, alors sa base X est un code maximal.*

Démonstration. Procédons par l'absurde en supposant qu'il existe un code Y sur A tel que $X \subsetneq Y$.

On a alors $X^* \subset Y^*$ et $X^* \neq Y^*$. En effet, si $X^* = Y^*$, on aurait $X = Y$ vu le Corollaire 1.4.8. Puisque $M = X^*$ est un sous-monoïde libre maximal, on a forcément $Y^* = A^*$ et par conséquent $Y = A$, toujours au vu du Corollaire 1.4.8. Il vient alors que $X \subsetneq A$.
Soit $b \in A\backslash X$ et posons $Z = X \cup \{b^2\}$. L'ensemble Z est un code puisque X ne contient que des lettres de $A\backslash\{b\}$ et qu'en lui ajoutant le mot b^2, l'ensemble Z forme un ensemble bifixe différent de $\{\varepsilon\}$, qui est donc un code. On a

$$M \subsetneq Z^* \subsetneq A^*.$$

La première inclusion est stricte puisque $b^2 \in Z$ et $b^2 \notin X^* = M$. La seconde est également stricte puisque $b \in A^*$ mais $b \notin Z^*$.
Ainsi, M n'est pas un sous-monoïde libre maximal, d'où la contradiction.

□

Remarque 1.4.18. La réciproque de cette proposition est fausse.
En effet, considérons les codes uniformes A^p avec $p \geq 1$. On sait par la Proposition 1.2.23 que ces codes sont maximaux.
Or, si $k, n \geq 2$, on a

$$(A^{kn})^* \subsetneq (A^n)^* \subset A^*$$

ce qui montre que $(A^{kn})^*$ n'est pas un sous-monoïde maximal.

Proposition 1.4.19. *Soient G un groupe et H un sous-groupe de G. Soit $\varphi : A^* \to G$ un morphisme de monoïdes. L'ensemble $M = \varphi^{-1}(H)$ est un sous-monoïde biunitaire.*

Démonstration. Le fait que $\varphi^{-1}(H)$ soit un sous-monoïde est évident puisque $\varphi(\varepsilon) = 1 \in H$, le morphisme envoie le neutre du monoïde libre sur le neutre du groupe et H en tant que sous-groupe possède le neutre du groupe G. De plus, si $m_1, m_2 \in M$, alors $\varphi(m_1), \varphi(m_2) \in H$. Par définition, on a

$$\varphi(m_1 m_2) = \varphi(m_1)\varphi(m_2) \in H.$$

Ainsi, $m_1 m_2 \in M$.
Montrons à présent que M est biunitaire.
Soient $p, pq \in M$. Puisque $M = \varphi^{-1}(H)$, il vient que $\varphi(p), \varphi(pq) \in H$. On a alors

$$(\varphi(p))^{-1}\varphi(pq) = (\varphi(p))^{-1}\varphi(p)\varphi(q) = \varphi(q).$$

Puisque H est un sous-groupe, on sait que $(\varphi(p))^{-1} \in H$, il vient alors que $\varphi(q) \in H$ et donc $q \in M$, ce qui montre que M est unitaire à droite.
On procède de manière analogue pour montrer que M est unitaire à gauche.

□

Remarque 1.4.20. La base de $M = \varphi^{-1}(H)$ est donc un code bifixe.

Définition 1.4.21. Soient G un groupe et H un sous-groupe de G. On appelle *code de groupe* la base X d'un sous-monoïde $M = \varphi^{-1}(H)$, où $\varphi : A^* \to G$ est un morphisme de monoïdes surjectif.

Proposition 1.4.22. *Un code de groupe est un code maximal.*

Démonstration. Soient G un groupe, H un sous-groupe de G et X le code de groupe de $M = \varphi^{-1}(H)$. Si $M = A^*$ alors on a directement que $X = A$ est maximal.
Sinon, prenons un mot $w \in A^* \backslash M$. Posons $Y = X \cup \{w\}$ et montrons que ce n'est pas un code.
Posons $m = \varphi(w) \in G$. On sait que m admet un inverse dans G et φ étant surjectif on sait qu'il existe un mot $w' \in A^*$ tel que $\varphi(w') = m^{-1}$. Les mots $u = ww'$ et $v = w'w$ sont dans M puisque

$$\varphi(ww') = \varphi(w)\varphi(w') = mm^{-1} = 1 \in H$$

et

$$\varphi(w'w) = \varphi(w')\varphi(w) = m^{-1}m = 1 \in H;$$

Ainsi, le mot $ww'w = uw = wv \in Y^*$ possède deux factorisations distinctes [8] en mots de Y.

□

1.4.1 Théorème du défaut

Le but de cette sous-section est de démontrer le Théorème du défaut, qui montre que si X n'est pas un code, alors il existe un code Y, contenant moins d'éléments que X, qui est tel que Y^* est un sous-monoïde libre contenant X.

Proposition 1.4.23. *L'intersection d'une famille arbitraire de sous-monoïdes libres de A^* est un sous-monoïde libre.*

Démonstration. Soit $(M_i)_{i\in I}$ une famille de sous-monoïdes libres de A^* et posons $M = \cap_{i\in I} M_i$.

- M est un sous-monoïde :
  Chaque M_i est un sous-monoïde, donc par définition, $\varepsilon \in M_i\ \forall i \in I$. Ainsi, $\varepsilon \in M$. Soient $m_1, m_2 \in M$. On a en particulier $m_1, m_2 \in M_i\ \forall i \in I$. Par définition on a $m_1m_2 \in M_i\ \ \forall i \in I$. Par conséquent, $m_1m_2 \in M$.
- M est libre :
  Montrons que M est stable. Soient $u, v, uw, wv \in M$. Par définition, ces quatre mots sont dans chacun des M_i. Puisque les M_i sont libres, ils sont stables donc on a bien $w \in M_i\ \forall i \in I$. Ainsi, $w \in M$.

□

Soit $X \subset A^*$. Comme nous venons de le voir, l'intersection de tous les sous-monoïdes libres de A^* contenant X est encore un sous-monoïde. Il s'agit du plus petit [9] sous-monoïde libre de A^* contenant X. Nous l'appellons *l'enveloppe libre de X*.

Remarque 1.4.24. Si X^* est un sous-monoïde libre, alors il coïncide avec son enveloppe libre. Cependant, si X^* n'est pas libre, alors il est inclus strictement dans son enveloppe libre.

Lemme 1.4.25. *Soient $X \subset A^*$ et Y la base de l'enveloppe libre de X. Alors tout élément de Y apparait comme premier (resp. dernier) facteur dans la factorisation d'un mot $x \in X$ en mots de Y, i.e.*

$$Y \subset X(Y^*)^{-1} \cap (Y^*)^{-1}X$$

où $X(Y^*)^{-1} = \{w \in A^* | wY^* \cap X \neq \emptyset\}$ et $(Y^*)^{-1}X = \{w \in A^* | Y^*w \cap X \neq \emptyset\}$.

8. $u = u_1 \cdots u_n$ avec $u_i \in X$ et on a $u_1 \neq w$ puisque $w \notin X^*$.
9. Pour l'inclusion.

Démonstration. Procédons par l'absurde et supposons qu'il existe $y \in Y$ tel que

$$yY^* \cap X = \emptyset,$$

i.e. $y \in Y \backslash X(Y^*)^{-1}$. Montrons que Y^* n'est pas le plus petit sous-monoïde libre contenant X.
Posons $Z = \{zy^i | z \in Y \backslash y, i \geq 0\} = (Y \backslash y)y^*$. Puisque $X \subset Y^*$, tout $x \in X$ s'écrit : $x = y_1 y_2 \cdots y_n$ avec $y_i \in Y$. Comme $yY^* \cap X = \emptyset$, on a forcément que $y_1 \neq y$. Il vient donc que $X \subset \{\varepsilon\} \cup (Y \backslash y)Y^*$. De plus, on a [10]

$$\{\varepsilon\} \cup (Y \backslash y)Y^* \subset Z^*$$

et $y \notin Z^*$. Il vient alors $X \subset Z^* \subsetneq Y^*$.
De plus, Z^* est libre : en effet tout mot $x \in Z^*$ possède une factorisation unique de la forme

$$x = x_1 x_2 \cdots x_n$$

avec $x_1, \ldots, x_n \in Y$ et $x_1 \neq y$, puisque Y est un code. Par conséquent le mot x peut se réécrire de manière unique de la façon suivante :

$$x = z_1 y^{p_1} z_2 y^{p_2} \cdots z_r y^{p_r}$$

avec $z_i \in Y \backslash y$ et $p_i \geq 0$. Ainsi, Z est bien un code.

□

Théorème 1.4.26 (Théorème du défaut)**.** *Soit $X \subset A^*$ et soit Y la base de l'enveloppe libre de X. Si X n'est pas un code, alors*

$$\mathrm{Card}(Y) < \mathrm{Card}(X).$$

Démonstration. Considérons deux cas :

— Cas 1 : Supposons que $\varepsilon \notin X$.
Soit $\alpha : X \to Y$ une application définie par

$$\alpha(x) = y \text{ si } x \in yY^*,$$

i.e. α associe à $x \in X$ le premier facteur de son unique factorisation en mots de Y. Ainsi, l'application est bien et partout définie puisque Y est un code et $X \subset Y^*$. Vu le lemme précédent, l'application α est surjective. Par contre, elle n'est pas injective. En effet, puisque X n'est pas un code, il existe $w \in X^*$ possédant deux factorisations distinctes

$$w = x_1 \cdots x_n = x'_1 \cdots x'_m$$

10. Effectivement, soit $w \in (Y \backslash y)Y^*$, on a $w = y_1 \cdots y_n$, avec $y_1 \neq y$ et $y_i \in Y \ \forall i$. En coupant avant chaque $y_i \neq y$, on obtient des blocs de $(Y \backslash y)y^*$.

avec $x_i, x'_j \in X \subset Y^*$ et $x_1 \neq x'_1$. Chacun des x_i, x'_j se décompose en mots de Y de manière unique. Ainsi,

$$x_1 \cdots x_n = y_{1,1} \cdots y_{1,n_1} \cdots y_{n,1} \cdots y_{n,n_n} = y'_{1,1} \cdots y'_{1,m_1} \cdots y'_{m,1} \cdots y'_{m,m_m} = x'_1 \cdots x'_m,$$

avec $y_{i,j}, y'_{k,l} \in Y$. Vu que Y est un code, on a $y_{1,1} = y'_{1,1}$. Il vient alors $\alpha(x_1) = \alpha(x'_1)$. Nous avons donc bien $\mathrm{Card}(Y) < \mathrm{Card}(X)$.

— Cas 2 : Si $\varepsilon \in X$.
Dans ce cas, X et $X' = X \backslash \{\varepsilon\}$ ont la même enveloppe libre Y^*.

- Si X' est un code, alors le sous-monoïde engendré par X' coïncide avec son enveloppe libre , i.e. $(X')^* = Y^*$. Vu le Corollaire 1.4.8, on a $X' = Y$ et il vient

$$\mathrm{Card}(Y) < \mathrm{Card}(X).$$

- Si X' n'est pas un code, il suffit d'appliquer le Cas 1 à X'. Il vient donc que

$$\mathrm{Card}(Y) < \mathrm{Card}(X') < \mathrm{Card}(X).$$

□

Corollaire 1.4.27. *L'ensemble $X = \{x_1, x_2\}$, $x_1 \neq x_2$, est un code si et seulement si x_1 et x_2 ne sont pas des puissances d'un même mot.*

Démonstration. Supposons que X est un code et procédons par l'absurde en supposant que x_1, x_2 sont des puissances d'un même mot. Soit $w \in A^*$ tel que

$$x_1 = w^r, \ x_2 = w^s,$$

avec $r, s \in \mathbb{N}$. Le mot $x = x_1 x_2 \in X^*$ possède deux factorisations distinctes puisque

$$x = x_1 x_2 = w^r w^s = w^s w^r = x_2 x_1,$$

ce qui contredit le fait que X est un code.

Supposons maintenant que X n'est pas un code. Soit Y la base de l'enveloppe libre de X. Par le théorème précédent, on sait que $\mathrm{Card}(Y) < \mathrm{Card}(X)$. Or, $x_1, x_2 \in Y^*$, i.e. x_1, x_2 sont des concaténations de mots de Y. Étant donné qu'il n'y a qu'un élément dans Y, cela entraine que x_1 et x_2 sont des puissances d'un même mot.

□

Chapitre 2

Description quantitative des codes

Dans ce chapitre, nous allons essentiellement introduire des outils permettant de « mesurer » la taille d'ensembles de mots.

Les lois de Bernoulli constituent le premier de ces outils. Dans le Théorème 2.1.15 nous verrons, par exemple, que les codes ne peuvent pas contenir « trop » de « petits » mots. Avant d'y parvenir, nous introduirons des lois de probabilité appliquées aux ensembles de mots et présenterons certaines de leurs propriétés. D'autres résultats importants seront également démontrés, comme le Théorème de Kraft-McMillan. Par ailleurs, de nombreux résultats de cette section seront utilisés dans des preuves ultérieures.

Le second outil développé dans ce chapitre est la notion de densité d'un ensemble de mots. Tout comme la densité topologique, à laquelle elle est liée, cette notion nous permet d'identifier les ensembles contenant « beaucoup » d'éléments. Ceci nous permettra d'étudier la maximalité des codes.

2.1 Mesure des codes

Les lois de Bernoulli nous permettront d'établir des conditions nécessaires pour qu'un ensemble de mots soit un code. De plus, elles nous permettront de mesurer la taille des codes grâce notamment au Théorème 2.1.15. Les résultats de cette section sont, par ailleurs, également très utiles pour les sections suivantes.

2.1.1 Lois de probabilité

Définition 2.1.1. Soit A un alphabet. Une application $\pi : A^* \to [0; 1]$ telle que

$$\pi(\varepsilon) = 1 \tag{2.1}$$

et

$$\sum_{a \in A} \pi(wa) = \pi(w) \quad \forall\ w \in A^* \tag{2.2}$$

est appelée une *loi de probabilité* sur A^*. La condition 2.2 est appelée *la condition de cohérence*.

Une loi de probabilité est dite *positive* si elle est non nulle, i.e. si $\pi(w) > 0$ pour tout $w \in A^*$.

Proposition 2.1.2. *La condition de cohérence implique que* $\sum_{x \in A^n} \pi(x) = 1$.

Démonstration. Procédons par récurrence sur $n \geq 0$.

- Cas de base : $n = 0$
  On a $A^0 = \{\varepsilon\}$, ainsi $\sum_{x \in A^0} \pi(x) = \pi(\varepsilon) = 1$.
- Induction : Supposons que la propriété est vraie pour $j < n$ et montrons-la pour n. On a

$$\sum_{x \in A^n} \pi(x) = \sum_{y \in A^{n-1}} \sum_{a \in A} \pi(ya) \tag{2.3}$$

$$= \sum_{y \in A^{n-1}} \pi(y) \tag{2.4}$$

$$= 1, \tag{2.5}$$

  où 2.4 est obtenu grâce à la condition de cohérence et 2.5 grâce à l'hypothèse de récurrence.

□

Étant donné une loi de probabilité π sur A^*, on pose pour chaque $X \subset A^*$

$$\pi(X) = \sum_{x \in X} \pi(x)$$

La valeur $\pi(X)$ peut être positive ou infinie.

Remarque 2.1.3. Pour une famille $(X_i)_{i \geq 0}$ de sous-ensembles de A^*, on a

$$\pi\left(\bigcup_{i \geq 0} X_i\right) \leq \sum_{i \geq 0} \pi(X_i).$$

En effet, si les ensembles X_i sont disjoints deux à deux, on a

$$\pi\left(\bigcup_{i \geq 0} X_i\right) = \sum_{w \in \bigcup_{i \geq 0} X_i} \pi(w) \tag{2.6}$$

$$= \sum_{i \geq 0} \sum_{w \in X_i} \pi(w) \tag{2.7}$$

$$= \sum_{i \geq 0} \pi(X_i). \tag{2.8}$$

Dans le cas où les ensembles ne sont pas disjoints, il est clair que nous avons l'inégalité souhaitée puisque certains éléments sont repris plusieurs fois à l'égalité 2.7.

Définition 2.1.4. La *série génératrice* de $X \subset A^*$ est la série

$$f_X(t) = \sum_{n \geq 0} u_n t^n$$

où $u_n = \mathrm{Card}(X \cap A^n)$.

Définition 2.1.5. La *série génératrice des probabilités* d'un ensemble $X \subset A^*$ est la série

$$F_X(t) = \sum_{n \geq 0} \pi(X \cap A^n) t^n.$$

En particulier, on a

$$\begin{aligned} F_X(1) &= \sum_{n \geq 0} \pi(X \cap A^n) && (2.9)\\ &= \sum_{n \geq 0} \sum_{w \in X \cap A^n} \pi(w) && (2.10)\\ &= \sum_{w \in X} \pi(w) && (2.11)\\ &= \pi(X). && (2.12) \end{aligned}$$

Remarque 2.1.6. Grâce à la Proposition 2.1.2, on sait que $\pi(A^n) = 1$. Ainsi on a $\pi(X \cap A^n) \leq 1$ et par conséquent $|\pi(X \cap A^n)t^n| \leq |t^n|$. Or on sait que $\sum_{n=0}^{+\infty} t^n$ converge sur $]-1;1[$. On en déduit alors que $F_X(t) = \sum_{n \geq 0} \pi(X \cap A^n)t^n$ converge au moins sur $]-1;1[$, i.e. le rayon de convergence de la série $F_X(t)$ est au moins 1.

2.1.2 Lois de Bernoulli

Intéressons-nous maintenant à une classe de lois de probabilité en particulier.

Définition 2.1.7. Une *loi de Bernoulli* est un morphisme π de A^* dans $[0,1]$ muni de la multiplication tel que

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

Proposition 2.1.8. *Une loi de Bernoulli est une loi de probabilité.*

Démonstration. Soit π une loi de Bernoulli sur A^*. Étant donné que π est un morphisme, nous avons, d'une part, $\pi(\varepsilon) = 1$ et, d'autre part, soit $w \in A^*$, on a

$$\begin{aligned}\sum_{a\in A} \pi(wa) &= \sum_{a\in A} \pi(w)\pi(a) \\ &= \pi(w) \underbrace{\sum_{a\in A} \pi(a)}_{=1} \\ &= \pi(w)\end{aligned}$$

□

Définition 2.1.9. La *loi uniforme de Bernoulli* sur A est définie par $\pi(a) = \dfrac{1}{\operatorname{Card} A}$ pour tout $a \in A$.

Dans le cas d'une loi uniforme de Bernoulli, la série génératrice des probabilités est liée à la série génératrice par la relation suivante :

$$f_X(t) = F_X(kt), \tag{2.13}$$

où $k = \operatorname{Card} A$.
En effet, dans ce cas on a

$$\begin{aligned}k^n \pi(X \cap A^n) &= k^n \sum_{w\in X\cap A^n} \pi(w) && (2.14)\\ &= k^n \sum_{w\in X\cap A^n} \frac{1}{k^n} && (2.15)\\ &= \sum_{w\in X\cap A^n} 1 && (2.16)\\ &= \operatorname{Card}(X \cap A^n). && (2.17)\end{aligned}$$

L'égalité 2.15 provient du fait que si $w \in X \cap A^n$, alors $w = w_1 \cdots w_n$ avec $w_i \in A$. Ainsi $\pi(w) = \pi(w_1) \cdots \pi(w_n)$. Puisque π est uniforme chaque $\pi(w_i) = \dfrac{1}{k}$.

Par ailleurs, puisque $\pi(X) = F_X(1)$, il vient

$$\pi(X) = f_X\left(\frac{1}{k}\right) = \sum_{n\geq 1} u_n k^{-n}. \tag{2.18}$$

où $u_n = \operatorname{Card}(X \cap A^n)$.

Proposition 2.1.10. *Soient $X, Y \subset A^*$ et π une loi de Bernoulli sur A^*. On a*

$$\pi(XY) \leq \pi(X)\pi(Y).$$

En particulier, on a l'égalité si le produit XY est non ambigu.

Démonstration. On a

$$\begin{aligned}\pi(XY) &= \pi\left(\bigcup_{x\in X}\bigcup_{y\in Y}\{xy\}\right)\\ &\leq \sum_{x\in X}\sum_{y\in Y}\pi(xy)\\ &= \pi(X)\pi(Y).\end{aligned}$$

En particulier, si le produit XY est non ambigu, alors l'union $\bigcup_{\substack{x\in X\\ y\in Y}}\{xy\}$ est disjointe, d'où l'égalité.

□

2.1.3 Codes et lois de Bernoulli

De manière intuitive, si nous cherchons à construire un code sur un alphabet donné, nous nous rendons vite compte qu'il ne faut pas lui ajouter « trop » de « trop petits » mots. Grâce aux lois de Bernoulli, nous allons pouvoir « quantifier » ces notions de « trop » et « trop petits ». Cette quantification sera réalisée au moyen des Théorèmes 2.1.15 et 2.1.23.

Lemme 2.1.11. *Soit π une loi de Bernoulli sur A^*. Pour des ensembles $X, Y \subset A^+$, on a*

1. *$F_{X\cup Y}(t) = F_X(t) + F_Y(t)$ si $X \cap Y \neq \emptyset$.*
2. *$F_{XY}(t) = F_X(t)F_Y(t)$ si le produit XY est non ambigu.*

Démonstration. 1. On a

$$\begin{aligned}F_{X\cup Y}(t) &= \sum_{n\geq 0}\pi((X\cup Y)\cap A^n)t^n && (2.19)\\ &= \sum_{n\geq 0}\pi((X\cap A^n)\cup(Y\cap A^n))t^n && (2.20)\\ &= \sum_{n\geq 0}(\pi(X\cap A^n)+\pi(Y\cap A^n))t^n && (2.21)\\ &= \sum_{n\geq 0}\pi(X\cap A^n)t^n + \sum_{n\geq 0}\pi(Y\cap A^n)t^n && (2.22)\\ &= F_X(t)+F_Y(t), && (2.23)\end{aligned}$$

l'égalité 2.21 découlant de la Remarque 2.1.3.

2. Pour tout n, on a

$$XY\cap A^n = \bigcup_{i+j=n}(X\cap A^i)(Y\cap A^j)$$

Puisque le produit XY est non ambigu, cette union est disjointe.
En effet supposons le contraire : soit $w \in (X \cap A^i)(Y \cap A^j) \cap (X \cap A^k)(Y \cap A^l)$ avec $i \neq k, j \neq l$. Il vient alors

$$w = x_1 \cdots x_i y_1 \cdots y_j = x'_1 \cdots x'_k y'_1 \cdots y'_l,$$

avec $x_1 \cdots x_i, x'_1 \cdots x'_k \in X$ et $y_1 \cdots y_j, y'_1 \cdots y'_l \in Y$ tels que $x_1 \cdots x_i \neq x'_1 \cdots x'_k$ et $y_1 \cdots y_j \neq y'_1 \cdots y'_l$, ce qui contredit le fait que le produit XY est non ambigu.
On obtient alors

$$\begin{aligned}\pi(XY \cap A^n) &= \pi\left(\bigcup_{i+j=n} (X \cap A^i)(Y \cap A^j)\right) \\ &= \sum_{i+j=n} \pi\left((X \cap A^i)(Y \cap A^j)\right).\end{aligned}$$

Ensuite, π étant un morphisme, il vient :

$$\begin{aligned}\pi\left((X \cap A^i)(Y \cap A^j)\right) &= \sum_{w \in (X \cap A^i)(Y \cap A^j)} \pi(w) \\ &= \sum_{\substack{x \in X \cap A^i \\ y \in X \cap A^j}} \pi(x)\pi(y) \\ &= \sum_{x \in X \cap A^i} \pi(x) \sum_{y \in X \cap A^j} \pi(y) \\ &= \pi(X \cap A^i)\pi(Y \cap A^j).\end{aligned}$$

Ainsi, d'une part nous avons

$$\begin{aligned}F_{XY}(t) &= \sum_{n \geq 0} \pi(XY \cap A^n) t^n \\ &= \sum_{n \geq 0} \sum_{i+j=n} (\pi(X \cap A^i)\pi(Y \cap A^j)) t^n.\end{aligned}$$

D'autre part, nous avons

$$\begin{aligned}F_X(t)F_Y(t) &= \sum_{n \geq 0} \pi(X \cap A^n) t^n \sum_{n \geq 0} \pi(Y \cap A^n) t^n \\ &= \sum_{n \geq 0} \left(\sum_{i+j=n} \pi(X \cap A^i)\pi(X \cap A^j)\right) t^n.\end{aligned}$$

L'égalité de l'énoncé est ainsi démontrée.

□

Remarque 2.1.12. Si chaque mot de $X_1 \cdots X_m$ possède une factorisation unique en mots de $X_1, \ldots, X_m$, alors on a [1]

$$F_{X_1 \cdots X_m}(t) = F_{X_1}(t) \cdots F_{X_m}(t).$$

Proposition 2.1.13. *Soit $X \subset A^+$ un code et soit π une loi de Bernoulli sur A^*. Alors on a*

$$F_{X^*}(t) = \frac{1}{1 - F_X(t)}.$$

Démonstration. Puisque $F_X(0) = 0$, nous savons que le terme indépendant de la série $1 - F_X(t)$ est 1, qui est inversible. Ainsi, vu la Proposition 1.1.1, nous savons que la série $1 - F_X(t)$ est inversible. Grâce à la Proposition 1.1.2, il vient $(F_X(t))^* = \sum_{n \geq 0} F_X(t)^n = \frac{1}{1 - F_X(t)}$. L'ensemble X étant un code, les produits X^n sont non ambigus. En effet, tout mot de X^n possède une factorisation unique en mots de X et par la Remarque 2.1.12 il vient que

$$(F_X(t))^n = F_{X^n}(t).$$

De plus les ensembles X^n sont deux-à-deux disjoints. En effet, s'il existait un mot $w \in X^m \cap X^n$, alors $w = x_1 \cdots x_n = y_1 \cdots y_m$ avec $x_i, y_j \in X$, ce qui contredit le fait que X est un code. Ainsi il vient

$$\begin{aligned} F_{\bigcup_{n \geq 0} X^n}(t) &= \sum_{m \geq 0} \pi \left(\left(\bigcup_{n \geq 0} X^n \right) \cap A^m \right) t^m \\ &= \sum_{m \geq 0} \pi \left(\bigcup_{n \geq 0} (X^n \cap A^m) \right) t^m \\ &= \sum_{m \geq 0} \sum_{n \geq 0} \pi(X^n \cap A^m) t^m \\ &= \sum_{n \geq 0} F_{X^n}(t). \end{aligned}$$

Finalement, on a donc

$$\begin{aligned} F_{X^*}(t) &= F_{\bigcup_{n \geq 0} X^n}(t) \\ &= \sum_{n \geq 0} F_{X^n}(t) \\ &= \sum_{n \geq 0} (F_X(t))^n \\ &= \frac{1}{1 - F_X(t)}. \end{aligned}$$

□

1. On peut le démontrer grâce à un raisonnement analogue à celui du point 2 du lemme précédent.

Dans le cas des lois uniformes de Bernoulli, nous avons le corollaire suivant :

Corollaire 2.1.14. *Soit X un code sur un alphabet fini A. On a alors*

$$f_{X^*}(t) = \frac{1}{1 - f_X(t)}.$$

Vu la définition des lois de Bernoulli, les mots « plus longs » auront une probabilité plus petite que les mots « plus courts ». La condition $\pi(X) \leq 1$ exprime alors le fait que X ne contienne « pas trop » de petits mots. Cela peut se voir encore plus facilement dans le cas de la loi uniforme de Bernoulli vu la relation 2.18.

Théorème 2.1.15. *Si X est un code sur A, alors $\pi(X) \leq 1$ pour toute loi de Bernoulli π sur A^*.*

Démonstration. Supposons d'abord que X est fini.
Alors $\pi(X) = \sum_{x \in X} \pi(x)$ est également fini. Procédons par l'absurde et supposons que $\pi(X) > 1$.
On sait alors que $F_X(1) > 1$. Puisque $F_X(0) = 0$ et que $F_X(t)$ est continu sur $[0; 1]$, il existe un nombre $r < 1$ tel que

$$F_X(r) = 1.$$

Puisque X est un code, on a $F_{X^*}(t) = \frac{1}{1 - F_X(t)}$. Ainsi, $F_{X^*}(t)$ diverge si $t = r$. On en tire alors que le rayon de convergence R de la série $F_{X^*}(t)$ est tel que $R \leq r < 1$. En effet, si le rayon de convergence R était tel que $R > r$, la série convergerait en r. Or, vu la Remarque 2.1.6, le rayon de convergence doit valoir au moins 1, d'où la contradiction.

Considérons maintenant le cas où X est infini.
Nous savons que A^* est dénombrable, ce qui implique que X l'est également. Nous pouvons alors réécrire l'ensemble X comme étant $\{x_0, x_1, x_2, x_3 \ldots\}$. Il vient alors que

$$\pi(X) = \sum_{x \in X} \pi(x) \tag{2.24}$$

$$= \sum_{n \geq 0} \pi(x_n) \tag{2.25}$$

$$= \sup_{N \in \mathbb{N}} \sum_{n=0}^{N} \pi(x_n). \tag{2.26}$$

Vu ce qui précède, nous savons que $\sum_{n=0}^{N} \pi(x_n) = \pi(\{x_0, \ldots, x_n\}) \leq 1$. Il en va donc de même pour sa borne supérieure. Ceci conclut la preuve.

□

En appliquant le théorème précédent au cas des lois uniformes de Bernoulli, nous obtenons le corollaire suivant :

Corollaire 2.1.16. *Soit X un code sur un alphabet A de k lettres. On a alors*

$$\sum_{x\in X} k^{-|x|} \leq 1. \tag{2.27}$$

Démonstration. Soit π la loi uniforme de Bernoulli sur A. Puisque X est un code, on a $\pi(X) = \sum_{x\in X} \pi(x) \leq 1$.

Or $\pi(x) = \pi(x_1)\cdots\pi(x_{|x|})$, avec $x_i \in A$. La loi étant uniforme chacun des $\pi(x_i)$ vaut $\frac{1}{k}$. Ainsi $\pi(x) = \frac{1}{k^{|x|}}$, d'où la conclusion.

□

Remarque 2.1.17. L'inégalité 2.27 peut se réécrire

$$\sum_{n\geq 1} \mathrm{Card}(X \cap A^n)k^{-n} \leq 1. \tag{2.28}$$

Pour montrer qu'un ensemble X n'est pas un code, il suffit de trouver une loi de Bernoulli telle que $\pi(X) > 1$.

Exemple 2.1.18. Soient $A = \{a, b\}$ et $X = \{a, ab, ba, baa\}$. Soit π la loi uniforme de Bernoulli sur A. Il vient

$$\begin{aligned}\pi(X) &= \sum_{x\in X} \pi(x)\\ &= \pi(a) + \pi(ab) + \pi(ba) + \pi(bba)\\ &= \frac{1}{2} + \frac{2}{4} + \frac{1}{8}\\ &= \frac{9}{8} > 1.\end{aligned}$$

On a donc trouvé une loi de Bernoulli qui est telle que $\pi(X) > 1$, par conséquent l'ensemble X n'est pas un code.

La réciproque du Théorème 2.1.15 n'est pas vraie. Comme le montre l'exemple suivant il est possible qu'un ensemble X qui n'est pas un code soit tel que $\pi(X) \leq 1$ pour toute loi de Bernoulli π.

Exemple 2.1.19. Considérons l'ensemble $X = \{ab, aba, aab\}$ sur l'alphabet $A = \{a, b\}$. Étant donné que le mot $w = aba \cdot ab = ab \cdot aab$ possède deux factorisations distinctes en mots de X, il est clair que X n'est pas un code. Cependant, n'importe quelle loi de Bernoulli π sur A^* nous donne $\pi(X) \leq 1$.
En effet, posons $p = \pi(a)$ et $q = \pi(b)$. On a alors

$$\begin{aligned}\pi(X) &= \pi(ab) + \pi(aba) + \pi(aab)\\ &= pq + p^2q + p^2q\\ &= pq + 2p^2q\end{aligned}$$

Remarquons que l'on a[2] $pq \leq \frac{1}{4}$ et $p^2q \leq \frac{4}{27}$. Il vient donc que $\pi(X) \leq \frac{1}{4} + \frac{8}{27} < 1$.

Jusqu'à présent, nous avons démontré quelques conditions nécessaires pour être un code, la proposition suivante va, quant à elle, nous fournir une condition suffisante.

Proposition 2.1.20. *Soient $X \subset A^+$ et π une loi de Bernoulli sur A^*.*

1. *Si X est un code, alors*
$$\pi(X^n) = \pi(X)^n \ \forall \ n \geq 0.$$
2. *Si π est positif, si $\pi(X)$ est fini et si $\pi(X^n) = \pi(X)^n$ pour tout $n \geq 1$, alors X est un code.*

Démonstration. 1. Si X est un code, les mots de X^n possèdent une factorisation unique en mots de X. Ainsi vu la Remarque 2.1.12, nous avons en particulier $\pi(X^n) = F_{X^n}(1) = F_X(1) \cdots F_X(1) = (\pi(X))^n$.

2. Procédons par l'absurde et supposons que X n'est pas un code. Il existe donc un mot $w \in X^*$, supposé de longueur minimale, tel que
$$w = x_1x_2 \cdots x_n = y_1y_2 \cdots y_m$$
avec $n, m \geq 1$, $x_i, y_j \in X$ et $x_1 \neq y_1$. Le mot $u = ww$ possède alors deux factorisations de $k = m + n$ facteurs
$$u = x_1x_2 \cdots x_ny_1y_2 \cdots y_m = y_1y_2 \cdots y_mx_1x_2 \cdots x_n$$
On a donc
$$\begin{aligned}(\pi(X))^k &= \left(\sum_{x\in X} \pi(x)\right)^k \\ &= \left(\sum_{z_1\in X} \pi(z_1)\right)\left(\sum_{z_2\in X} \pi(z_2)\right) \cdots \left(\sum_{z_k\in X} \pi(z_k)\right) \\ &= \sum_{z_1\in X}\sum_{z_2\in X} \cdots \sum_{z_k\in X} \pi(z_1z_2 \cdots z_k)\end{aligned}$$
et
$$\pi(X^k) = \sum_{v\in X^k} \pi(v).$$
Il vient alors que
$$(\pi(X))^k \geq \pi(X^k) + \pi(u).$$
Or, par hypothèse $(\pi(X))^k = \pi(X^k)$, ainsi $\pi(u) = 0$, ce qui contredit le fait que π est positif.

□

2. Il suffit d'étudier le signe des polynômes $p(1-p) - \frac{1}{4}$ et $p^2(1-p) - \frac{4}{27}$ pour s'en convaincre.

Les lois de Bernoulli nous permettent également de trouver des codes maximaux.

Proposition 2.1.21. *Soit X un code sur A. S'il existe une loi de Bernoulli positive π sur A^* telle que $\pi(X) = 1$, alors le code X est maximal.*

Démonstration. Procédons par l'absurde et supposons que X n'est pas maximal.
Il existe un certain mot $w \notin X$ de A^* tel que $Y = X \cup \{w\}$ est un code. Par le Théorème 2.1.15, on a $\pi(Y) \leq 1$. Or,

$$\pi(Y) = \sum_{y \in Y} \pi(y) = \sum_{x \in X} \pi(x) + \pi(w) = \pi(X) + \pi(w) = 1 + \pi(w).$$

Il s'ensuit alors que $\pi(w) = 0$, ce qui contredit le fait que π est positif.

□

Exemple 2.1.22. Soient $A = \{a, b\}$ et $X = \{a, ba, bb\}$ un ensemble préfixe. Soit π une distribution de Bernoulli sur A^* telle que $\pi(a) = p$ et $\pi(b) = q$. On a

$$\begin{aligned}
\pi(X) &= \sum_{x \in X} \pi(x) \\
&= \pi(a) + \pi(ba) + \pi(bb) \\
&= \pi(a) + \pi(b)\pi(a) + \pi(b)\pi(b) \\
&= p + pq + q^2 \\
&= p + q.\underbrace{(p+q)}_{=1} \\
&= 1
\end{aligned}$$

L'ensemble est donc un code maximal.

Comme montré à l'Exemple 2.1.19, la réciproque du Théorème 2.1.15 (et donc du Corollaire 2.1.16) n'est pas vraie. En particulier, si $X \subset A^*$ et si on pose $u_n = \mathrm{Card}(X \cap A^n)$, la condition $\sum_{n \geq 1} u_n k^{-n} \leq 1$ n'implique pas que X soit un code. Par contre, le théorème suivant montre que cette condition implique l'existence d'un code X' tel que $u_n = \mathrm{Card}(X' \cap A^n)$.

Théorème 2.1.23 (Kraft-McMillan)**.** *Étant donné une suite $(u_n)_{n \geq 1}$ de naturels, il existe un code X sur un alphabet A de k symboles tel que $u_n = \mathrm{Card}(X \cap A^n)$ si et seulement si*

$$\sum_{n \geq 1} u_n k^{-n} \leq 1.$$

De plus, le code X peut être choisi préfixe.

Démonstration. La condition est nécessaire vu le Corollaire 2.1.16. En effet, si π est la loi uniforme de Bernoulli sur A, on a

$$\begin{aligned}\sum_{n\geq 1} u_n k^{-n} &= \sum_{n\geq 1} \mathrm{Card}(X\cap A^n)k^{-n}\\ &= \sum_{n\geq 1} k^n \pi(X\cap A^n)k^{-n}\\ &= \sum_{n\geq 1}\sum_{w\in X\cap A^n} \pi(w)\\ &= \sum_{w\in X} k^{-|w|}.\end{aligned}$$

D'autre part, observons tout d'abord que $\sum_{n\geq 1} u_n k^{-n} \leq 1$ entraine $\sum_{1\leq i\leq n} u_i k^{-i} \leq 1$ et par conséquent $\sum_{1\leq i\leq n} u_i k^{n-i} \leq k^n$.
Nous allons construire, par récurrence sur $n \geq 1$, une suite $X_1, X_2, X_3, \ldots$ de codes préfixes qui seront tels que

$$\forall n\ X_n \subset X_{n+1} \tag{2.29}$$
$$\forall n\ \forall 1 \leq i \leq n, \mathrm{Card}(X_n \cap A^i) = u_i \tag{2.30}$$
$$\forall n\ \forall i > n, \mathrm{Card}(X_n \cap A^i) = 0. \tag{2.31}$$

- Cas de base : $n = 1$
  On a $u_1 \leq k$. Il suffit de choisir u_1 lettres de A pour former X_1. Alors X_1 est bien un code préfixe tel que $\mathrm{Card}(X_1 \cap A) = u_1$.
- Induction : Supposons avoir construit des codes préfixes $X_1, \ldots, X_n$ vérifiant les conditions 2.29, 2.30 et 2.31, et construisons X_{n+1}.
  L'ensemble des mots de A^* de longueur $n+1$ et qui ont un préfixe dans X_n s'écrit

  $$\bigcup_{i=1}^{n} (X_n \cap A^i)A^{n+1-i}.$$

  Cette union est disjointe car X_n est un code préfixe.
  Le nombre de mots de longueur $n+1$ ayant un préfixe dans X_n est donc

  $$\begin{aligned}S &= \sum_{i=1}^{n} \mathrm{Card}\left((X_n \cap A^i)A^{n+1-i}\right)\\ &= \sum_{i=1}^{n} \mathrm{Card}(X_n \cap A^i)\,\mathrm{Card}(A^{n+1-i})\\ &= \sum_{i=1}^{n} u_i k^{n+1-i}.\end{aligned}$$

Par hypothèse,

$$\sum_{i=1}^{n+1} u_i k^{n+1-i} \leq k^{n+1}$$
$$\Leftrightarrow S + u_{n+1} \leq k^{n+1}.$$

Il y a donc au moins u_{n+1} mots de longueur $n+1$ qui n'ont pas de préfixe dans X_n. Soit Y un ensemble de u_{n+1} mots de longueur $n+1$ qui n'ont pas de préfixe dans X_n. Posons

$$X_{n+1} = X_n \cup Y.$$

L'ensemble X_{n+1} est un code préfixe tel que $\forall\ 1 \leq i \leq n+1, \mathrm{Card}(X_{n+1} \cap A^i) = u_i$.

Posons $X = \bigcup_{n \geq 1} X_n$.

— L'ensemble X est préfixe :
Si $v, w \in X$, il existe n, m tels que $v \in X_n$ et $w \in X_m$. Supposons par exemple que $m \geq n$, alors $v, w \in X_m$ qui est un code préfixe. Ainsi v, w ne peuvent pas être préfixes propres l'un de l'autre.

— On a $\mathrm{Card}(X \cap A^n) = u_n$:
En effet, nous avons $\mathrm{Card}(X \cap A^n) = \mathrm{Card}(X_n \cap A^n) = u_n$.

Ainsi l'ensemble X satisfait aux conditions de l'énoncé.

□

2.2 Ensembles complets

Nous savons, jusqu'à présent, que tout sous-ensemble d'un code est encore un code. Il peut donc être intéressant d'étudier les codes maximaux. Par ailleurs, dans cette section, nous définirons les ensembles complets, possédant la propriété contraire : tout ensemble contenant un ensemble complet est complet. Nous verrons alors dans le Théorème 2.2.11 que les codes maximaux se trouvent à l'intersection entre les codes et les ensembles complets.

Définition 2.2.1. Soient M un monoïde et P un sous-ensemble de M. Un élément $m \in M$ est dit *complétable* dans P s'il existe des éléments $u, v \in M$ tels que $umv \in P$.

Cette définition est équivalente au fait que P rencontre l'idéal MmM :

$$MmM \cap P \neq \emptyset.$$

L'ensemble des éléments complétables dans P est $F(P) = M^{-1}PM^{-1}$.
Un élément qui n'est pas complétable dans P est dit *incomplétable* dans P.

Définition 2.2.2. Un sous-ensemble P d'un monoïde M est *dense* dans M si tous les éléments de M sont complétables dans P, i.e. $F(P) = M$.

Remarque 2.2.3. Cette définition est équivalente au fait que P rencontre tous les idéaux de M. En effet, P est dense dans M si et seulement si pour tout $m \in M$, on a $MmM \cap P \neq \emptyset$. Soit I un idéal de M. Par définition, on a $MIM \subset I$. Soit $i \in I$, on a $MiM \subset I$. En particulier $\emptyset \neq MiM \cap P \subset I \cap P$.

La remarque précédente nous permet de justifier l'utilisation du terme « dense ». En effet, nous pouvons vérifier que $\mathcal{T} = \{\emptyset\} \cup \{I \mid I \text{ idéal de } M\}$ est une topologie sur M. Puisqu'un ensemble est dense, au sens topologique, si et seulement s'il rencontre tout ouvert non vide, la notion de densité définie ci-dessus est équivalente à celle de densité topologique.

Intuitivement, un ensemble dense contient « beaucoup » d'éléments.

Soient P, Q deux sous-ensembles d'un monoïde M tels que $P \subset Q$. Si P est dense alors Q l'est aussi.

Remarque 2.2.4. Dans le cas du monoïde libre A^*, $X \subset A^*$ est dense si et seulement tout mot de A^* est facteur d'un mot de X.

Définition 2.2.5. Un sous-ensemble P d'un monoïde M est *complet* dans M si le sous-monoïde engendré par P est dense, i.e. $F(P^*) = M$.

De manière équivalente, P est complet si et seulement si, pour tout élément $m \in M$,

$$MmM \cap P^* \neq \emptyset.$$

Nous avons trivialement le résultat suivant.

Proposition 2.2.6. *1. Tout ensemble dense est complet.*

2. Tout ensemble contenant un ensemble complet est encore complet.

Définition 2.2.7. Un mot $w \in A^+$ est dit *sans bords* si aucun préfixe propre non vide de w n'est un suffixe de w.

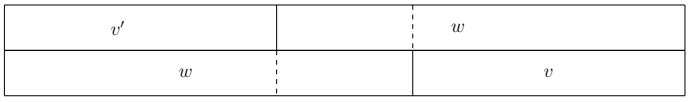
En d'autres termes, w est sans bords si et seulement si $w \in uA^+ \cap A^+u$ implique $u = \varepsilon$.

Proposition 2.2.8. *Si un mot $w \in A^*$ est sans bords alors $wA^* \cap A^*w = wA^*w \cup \{w\}$.*

Démonstration. Procédons par double inclusion.

$\subset$: Si $u \in wA^* \cap A^*w$, alors $u = wv = v'w$ pour des mots $v, v' \in A^*$.

- Si $v = v' = \varepsilon$ alors $u \in \{w\}$.
- Sinon, $|v|, |v'| \geq |w|$. En effet, si $|v|, |v'| < |w|$,



alors il existe un suffixe de w qui est également préfixe propre de w, ce qui contredit le fait que w est sans bords. Ainsi, $v = tw$ et $v' = wt'$ pour des mots $t, t' \in A^*$. Il vient alors que $u = wv = wtw \in wA^*w$.

$\supset$: Soit $u \in wA^*w \cup \{w\}$.

- Si $u \in \{w\}$ alors $u \in wA^* \cap A^*w$.
- Si $u \in wA^*w$, alors il existe $v \in A^*$ tel que $u = wvw = wv' = v''w$ avec $v' = vw \in A^*$ et $v'' = wv \in A^*$. Ainsi, $u \in wA^* \cap A^*w$.

□

Lemme 2.2.9. *Soit A un alphabet d'au moins deux lettres. Pour tout mot $u \in A^+$, il existe un mot $v \in A^+$ tel que uv est sans bords.*

Démonstration. Soit $a \in A$ la première lettre de u et soit $b \in A\backslash\{a\}$. Montrons que le mot $w = uab^{|u|}$ est sans bords, autrement dit, le mot $v = ab^{|u|}$ convient.
Procédons par l'absurde et supposons qu'il existe un mot t qui est préfixe propre non vide et suffixe de w. On a alors nécessairement $|t| > |u|$. Ainsi, t étant suffixe, on a $t = sab^{|u|}$ pour un certain $s \in A^*$ et puisqu'il est également préfixe, on a $t = uab^{|s|}$. Il vient donc $|s| = |u|$. On obtient finalement $t = w$, ce qui contredit le fait que t est préfixe propre de w.

□

Proposition 2.2.10. *Soit $X \subset A^+$ un code maximal. Pour tout mot $w \in A^*$ on a*

$$X^*wA^* \cap X^* \neq \emptyset. \tag{2.32}$$

Démonstration. Traitons tout d'abord le cas où $\operatorname{Card} A = 1$. Supposons $A = \{a\}$, alors $\operatorname{Card} X = 1$ avec $X = \{a^n\}$ pour un n fixé. Soit $w \in A^*$, alors $w = a^m$ pour un certain m. Alors le mot $a^m a^{kn-m}$ appartient à $X^*wA^* \cap X^*$ pour un k tel que $kn > m$. Ensuite, traitons le cas où $w = \varepsilon$. Nous avons directement $w \in X^*wA^* \cap X^*$. Maintenant, nous pouvons donc supposer que $\operatorname{Card} A \geq 2$ et $w \in A^+$.
Par le lemme précédent, il existe un mot $w' \in A^+$ tel que $y = ww'$ est sans bords. Si $y \in X$, alors il vient directement que $y = ww' \in X^*wA^* \cap X^*$. Supposons donc que $y \notin X$ et posons $Y = X \cup \{y\}$. Nous allons alors montrer que $X^*yA^* \cap X^* \neq \emptyset$. L'ensemble Y ne peut pas être un code vu la maximalité de X. Il existe donc un mot $t \in Y^*$, supposé de longueur minimale, qui possède deux factorisations en mots de Y :

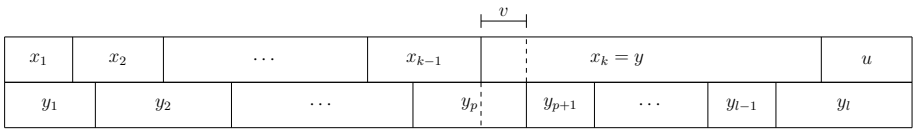
$$\begin{aligned} t &= x_1x_2\cdots x_n \\ &= y_1y_2\cdots y_m, \end{aligned}$$

avec $m, n \geq 1$, $x_i, y_j \in Y$ et $x_1 \neq y_1$. L'ensemble X étant un code, il y a forcément un des x_i, y_j qui est égal à y. Considérons l'occurence la plus à gauche de y parmi les x_i, y_j, on peut supposer $x_k = y$. Autrement dit, le mot $x_1 \cdots x_{k-1}$ est le plus petit préfixe de t qui ne contient pas y. On a donc $x_1, \ldots, x_{k-1} \in X$. Posons l le plus petit indice tel que $x_1 \cdots x_k$ est préfixe de $y_1 \cdots y_l$. Posons

$$z = \underbrace{x_1x_2\cdots x_{k-1}}_{\in X^*}\underbrace{x_k}_{=y} u = y_1y_2\cdots y_l,$$

pour $u \in A^*$. On a donc $z \in X^*yA^*$. Montrons que $z \in X^*$ en montrant que $y_1, \ldots, y_l \in X$. Soit p le plus petit indice tel que $x_1x_2\cdots x_{k-1}$ soit préfixe de $y_1y_2\cdots y_p$.

- Cas 1 : $p = l$
  Si $y_l = y_p$, alors $x_1 \cdots x_{k-1}$ est préfixe de $y_1 \cdots y_l$, ce qui implique que $y_1 \cdots y_l \in X^*$. En effet, sinon il y a un des $y_1, \ldots, y_l$ qui est égal à y ce qui contredit le fait que x_k est l'occurence la plus à gauche de y.
- Cas 2 : Posons
  $$x_1 x_2 \cdots x_{k-1} v = y_1 y_2 \cdots y_p$$
  avec $v \neq \varepsilon$ puisque X est un code.



  Il vient $x_k u = v y_{p+1} \cdots y_l$. On a alors que $y_{p+1}, \ldots, y_{l-1}$ sont des facteurs propres de $x_k = y$, i.e. $|y_{p+1}|, \ldots, |y_{l-1}| < |y|$. Ainsi ils sont aussi dans X. De plus, puisque $y_l \neq y$, y étant sans bords, on a $y_l \in X$.

Finalement, z est bien dans X^*.

□

Le théorème suivant découle simplement de la proposition précédente.

Théorème 2.2.11. *Tout code maximal est complet.*

Grâce à ce résultat, nous allons désormais pouvoir montrer qu'il existe des codes finis qui ne sont pas inclus dans un code maximal fini, comme stipulé dans la Remarque 1.2.25.

Exemple 2.2.12. Soit $X = \{a^5, ba^2, ab, b\}$ un code[3] sur l'alphabet $A = \{a, b\}$. Chaque code maximal contenant X est infini. Nous allons le montrer par l'absurde.
Soit Y un code maximal sur A contenant X et supposons que Y est fini. Posons $m = \max\{|y| \mid y \in Y\}$ et $u = b^m a^{4+5m} b^m \in A^*$. L'ensemble Y étant maximal, il est complet, ainsi u est facteur d'un mot de Y^*. Cependant, b^m et a^{4+5m} ne peuvent être des facteurs propres d'un mot de Y vu leurs longueurs, il existe donc des mots $y, y' \in Y \cup \{\varepsilon\}$ et des naturels $p, q, r \geq 0$ tels que
$$u = b^p y a^q y' b^r$$
avec $a^q \in Y^*$.
Le mot a^5 est le seul mot de Y qui ne contient pas de b. En effet, nous allons montrer que si une autre puissance de a appartient à Y, nous trouverons un mot possédant deux factorisations différentes, ce qui contredit le fait que Y est un code.

— Si $a^{5i} \in Y$, pour $i > 1$[4], alors $a^{5i} = (a^5)^i$.
— Si $a^{5i+1} \in Y$, pour $i \geq 0$, alors $(a^5)^i(ab) = (a^{5i+1})(b)$.

3. Nous le vérifions grâce à l'algorithme développé dans la Section 1.3 : $U_1 = \{aa\}, U_2 = \{aaa\}, U_3 = U_1$ et $\varepsilon \notin U_n$ pour $n = 1, 2, 3$.
4. Si $i = 0$ alors $a^{5i} = \varepsilon \notin Y$ et on sait déjà que $a^5 \in Y$.

— Si $a^{5i+2} \in Y$, pour $i \geq 0$, alors $(ba^2)(a^5)^i = (b)(a^{5i+2})$.
— Si $a^{5i+3} \in Y$, pour $i \geq 0$, alors $(ba^2)(a^5)^i(ab) = (b)(a^{5i+3})(b)$.
— Si $a^{5i+4} \in Y$, pour $i \geq 0$, alors $(a^5)^{i+1}(b) = (a^{5i+4})(ab)$.

Ainsi, q est forcément un multiple de 5, ce qui implique que $|y|_a + |y'|_a \equiv 4 \mod 5$. Soient

$$y = b^h a^{5s+i} \text{ et } y' = a^{j+5t} b^k$$

avec $0 \leq i, j \leq 4$. On a $i + j \equiv 4 \mod 5$. Il vient donc $i + j = 4$. Nous allons alors montrer que pour tout choix pour i, j, nous arrivons à la conclusion que Y n'est pas un code en trouvant des mots qui possèdent des décompositions distinctes en mots de Y.

— $i = 0$, $j = 4$, alors $k \geq 1$ et on a $(ba^2)(a^{5t+4}b^k) = (b)(a^5)^{t+1}(ab)(b)^{k-1}$.
— $i = 1$, $j = 3$, alors on a $(b^h a^{5s+1})(b) = (b)^h(a^5)^s(ab)$.
— $i = 2$, $j = 2$, alors on a $(b)(a^{5t+2}b^k) = (ba^2)(a^5)^t(b)^k$.
— $i = 3$, $j = 1$, alors $h \geq 1$ et on a $(b^h a^{5s+3})(b) = (b)^{h-1}(ba^2)(a^5)^s(ab)$.
— $i = 4$, $j = 0$, alors on a $(b^h a^{5s+4})(ab) = (b)^h(a^5)^{s+1}(b)$.

Ainsi, Y n'est pas fini.

Nous allons maintenant nous intéresser plus spécifiquement à la notion de densité et à ses liens avec la maximalité des codes.

Définition 2.2.13. Un sous-ensemble P d'un monoïde M est dit *fin* lorsqu'il n'est pas dense. Autrement dit, il existe au moins un élément $m \in M$ qui n'est pas complétable dans P, i.e. $MmM \cap P = \emptyset$.

Proposition 2.2.14. *Tout sous-ensemble d'un ensemble fin est fin.*

Proposition 2.2.15. *Soient M un monoïde et P, Q, $R \subset M$.*

1. *L'ensemble $P \cup Q$ est fin si et seulement si P et Q sont fins.*
2. *Si R est dense et P est fin, alors $R \backslash P$ est dense.*

Démonstration. 1. Supposons tout d'abord que P et Q sont fins. Ainsi il existe des mots $m, n \in M$ tels que

$$MmM \cap P = \emptyset \text{ et } MnM \cap Q = \emptyset.$$

Le mot mn est incomplétable dans $P \cup Q$:

$$\begin{aligned} M(mn)M \cap (P \cup Q) &= (M(mn)M \cap P) \cup (M(mn)M \cap Q) \\ &\subset (MmM \cap P) \cup (MnM \cap Q) = \emptyset. \end{aligned}$$

L'ensemble $P \cup Q$ est donc fin.

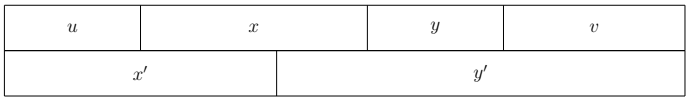
La réciproque découle directement de la proposition précédente.

2. Procédons par l'absurde et supposons que $R\backslash P$ est fin. Alors, vu le point précédent, $(R\backslash P) \cup P$ est fin. Or, $R \subset (R\backslash P) \cup P$, ce qui contredit notre hypothèse.

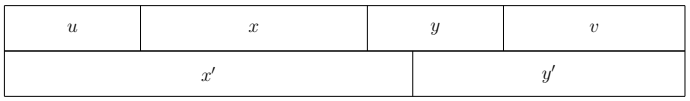
□

Remarque 2.2.16. Les ensembles fins d'un monoïde libre ont des propriétés supplémentaires :

1. Tout sous-ensemble fini $X \subset A^*$ est fin.
   En effet, si on prend un élément $w \in A^*\backslash X$ tel que $|w| > \max\{|x| \mid x \in X\}$, alors $A^*wA^* \cap X = \emptyset$.
2. Si $X, Y \subset A^*$ sont des ensembles fins, alors XY est fin.
   Comme X et Y sont fins, il existe $x, y \in A^*$ tels que x et y sont incomplétables dans X et Y, respectivement. Montrons par l'absurde que $w = xy$ est incomplétable dans XY. Supposons qu'il existe $u, v \in A^*$ tels que $uxyv \in XY$. Il existe alors $x' \in X$ et $y' \in Y$ tels que $uxyv = x'y'$. Posons $l = |ux|$.
   — Si $|x'| \leq l$:



   Alors $y' \in A^*yA^* \cap Y$, ce qui contredit le fait que y est incomplétable dans Y.
   — Si $|x'| \geq l$:



   Alors $x' \in A^*xA^* \cap X$, ce qui contredit le fait que x est incomplétable dans X.

Puisqu'un ensemble est fin s'il n'est pas dense, intuitivement nous savons qu'il ne contient que « peu » d'éléments. La proposition suivante va confirmer cette intuition.

Proposition 2.2.17. *Soit $X \subset A^*$ un ensemble fin. Pour toute loi de Bernoulli positive π sur A^*, on a*

$$\pi(X) < \infty.$$

Démonstration. Soit w un mot qui n'est pas facteur d'un mot de X, i.e. w est incomplétable dans X. Posons $n = |w|$. On a $n \geq 1$. Pour $0 \leq i \leq n-1$, considérons

$$X_i = \{x \in X \mid |x| \equiv i \mod n\}.$$

Puisque $\pi(X) = \pi\left(\bigcup_{i=0}^{n-1} X_i\right) = \sum_{i=0}^{n-1} \pi(X_i)$, l'union étant disjointe, il suffit de montrer que $\pi(X_i)$ est fini pour $i = 0, \ldots, n-1$.

On a $X_i \subset A^i(A^n\backslash\{w\})^*$. Puisque $A^n\backslash\{w\}$ est un code [5], nous avons

$$\pi\left((A^n\backslash\{w\})^*\right) = \pi\left(\bigcup_{k\geq 0}(A^n\backslash\{w\})^k\right) \tag{2.33}$$

$$= \sum_{k\geq 0}\pi\left((A^n\backslash\{w\})^k\right) \tag{2.34}$$

$$= \sum_{k\geq 0}\left(\pi\left(A^n\backslash\{w\}\right)\right)^k \tag{2.35}$$

$$= \sum_{k\geq 0}(1-\pi(w))^k \tag{2.36}$$

puisque le produit $(A^n\backslash\{w\})^k$ est non ambigu et $\pi(A^n) = 1$ d'après la Proposition 2.1.2. Étant donné que π est positif par hypothèse, nous savons que $\pi(w) > 0$, et par conséquent $(1-\pi(w)) < 1$. Nous obtenons donc

$$\pi\left((A^n\backslash\{w\})^*\right) = \frac{1}{\pi(w)}.$$

Finalement, il vient

$$\begin{aligned}\pi(X_i) &\leq \pi(A^i(A^n\backslash\{w\})^*)\\ &= \underbrace{\pi(A^i)}_{=1}\pi((A^n\backslash\{w\})^*)\\ &= \frac{1}{\pi(w)}\end{aligned}$$

car le produit $A^i(A^n\backslash\{w\})^*$ est non ambigu.

□

Proposition 2.2.18. *Soit $X \subset A^*$ un ensemble fin et complet. Soit w un mot incomplétable dans X. Alors*

$$A^* = \bigcup_{d\in D, g\in G} d^{-1}X^*g^{-1} = D^{-1}X^*G^{-1}, \tag{2.37}$$

où D est l'ensemble des suffixes de w et G est l'ensemble des préfixes de w.

Démonstration. Soit $z \in A^*$. Puisque X est complet, X^* est dense. Le mot $wzw \in A^*$ est donc complétable dans X^*, il existe alors $u, v \in A^*$ tels que

$$uwzwv \in X^*.$$

Puisque w est incomplétable dans X, il n'est facteur d'aucun mot de X. Par conséquent, il existe deux factorisations $w = g_1d = gd_1$ telles que

$$ug_1, dzg, d_1v \in X^*.$$

5. C'est un sous-ensemble du code uniforme A^n.

Ainsi, $z \in d^{-1}X^*g^{-1}$, ce qui permet de conclure étant donné que la seconde inclusion est triviale.

□

Proposition 2.2.19. *Soit $X \subset A^*$ un ensemble fin et complet. Pour toute loi de Bernoulli positive π sur A^*, on a*

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

Démonstration. On a $\pi(A^*) = \pi\left(\bigcup_{n\geq 0} A^n\right) = \sum_{n\geq 0} \underbrace{\pi(A^n)}_{=1} = \infty$. Vu la proposition précédente, on a

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

Ainsi il existe une paire $(g, d) \in G \times D$ telle que $\pi(d^{-1}X^*g^{-1}) = \infty$, la somme étant finie. Puisque $d(d^{-1}X^*g^{-1})g \subset X^*$, il vient

$$\begin{aligned}\pi(d(d^{-1}X^*g^{-1})g) &= \pi(d)\pi(d^{-1}X^*g^{-1})\pi(g)\\ &\leq \pi(X^*)\end{aligned}$$

étant donné que le produit $d(d^{-1}X^*g^{-1})g$ est non ambigu, d et g étant fixés. Puisque π est supposée positif, $\pi(X^*) = \infty$. De plus,

$$\begin{aligned}\pi(X^*) &= \pi\left(\bigcup_{n\geq 0} X^n\right)\\ &\leq \sum_{n\geq 0} \pi(X^n)\\ &\leq \sum_{n\geq 0} (\pi(X))^n.\end{aligned}$$

Procédons alors par l'absurde en supposant que $\pi(X) < 1$. Vu ce qui précède, on a $\pi(X^*) < \infty$, ce qui mène à une contradiction. Ainsi $\pi(X) \geq 1$.

□

Théorème 2.2.20. *Tout code fin et complet est maximal.*

Démonstration. Soit X un code fin et complet. Par le Théorème 2.1.15, nous savons que $\pi(X) \leq 1$ pour toute loi de Bernoulli π sur A^*. Par la proposition précédente, nous savons que $\pi(X) \geq 1$ pour toute loi de Bernoulli positive π sur A^*. Ainsi, il vient que $\pi(X) = 1$ pour toute loi de Bernoulli positive π sur A^*. Finalement, par la Proposition 2.1.21, on en tire que X est maximal. □

Les Théorèmes 2.2.11 et 2.2.20 nous permettent d'obtenir une condition équivalente au fait d'être complet.

Théorème 2.2.21. *Soit X un code sur A. L'ensemble X est complet si et seulement si X est dense ou maximal.*

Démonstration. La condition est nécessaire. En effet, supposons que X est complet mais pas dense. Ainsi, X est fin et vu le Théorème 2.2.20, on en tire que X est maximal.

La condition est suffisante puisque nous savons que tout ensemble dense est complet vu la Proposition 2.2.6 et que tout code maximal est complet vu le Théorème 2.2.11.

□

Proposition 2.2.22. *Soit $X \subset A^*$ un code maximal fini. Pour tout sous-ensemble non vide B de A, le code $X \cap B^*$ est un code maximal sur B.*

Démonstration. Soit $n = \max\{|x| \mid x \in X\}$. Puisque $Y = X \cap B^* \subset X$ et que X est fini, par la Remarque 2.2.16, nous savons que Y est fin. Ainsi pour montrer que Y est un code maximal sur B, vu le Théorème 2.2.20, il suffit de montrer que Y est complet dans B^*. Soit $w \in B^*$ et $b \in B$. Considérons le mot $w' = b^{n+1}wb^{n+1} \in B^* \subset A^*$. Nous savons que X est un code maximal sur A, ainsi il est complet sur A. Il existe donc des mots $u, v \in A^*$ tels que

$$uw'v = x_1x_2 \cdots x_k,$$

avec $x_1, \ldots, x_k \in X$, i.e. $uw'v \in X^*$. Vu la définition de n, il existe deux naturels $1 < i \leq j < k$ tels que

$$x_ix_{i+1} \cdots x_j = b^rwb^s$$

pour certains $r, s \in \{1, \ldots, n\}$. On a alors $x_i, x_{i+1}, \ldots, x_j \in X \cap B^* = Y$, et donc w est complétable dans Y^*.

□

Corollaire 2.2.23. *Soit $X \subset A^*$ un code maximal fini. Pour chaque lettre $a \in A$, il existe un naturel n tel que $a^n \in X$.*

Démonstration. Soit $B = \{a\}$. Vu la proposition précédente, $X \cap \{a\}^*$ est un code maximal, et donc non vide. Il existe alors $w \in X \cap \{a\}^*$, i.e. il existe n tel que $w = a^n \in X$.

□

Définition 2.2.24. Soit $X \subset A^*$ un code maximal fini et $a \in A$. Le naturel n tel que $a^n \in X$ est appelé *l'ordre de a relatif* à X.

Théorème 2.2.25. *Soit X un code fin sur A. Les propositions suivantes sont équivalentes.*

1. *L'ensemble X est un code maximal.*
2. *Il existe une loi de Bernoulli positive π sur A^* telle que $\pi(X) = 1$.*
3. *Pour toute loi de Bernoulli positive π sur A^*, on a $\pi(X) = 1$.*
4. *L'ensemble X est complet.*

Démonstration. On a les implications suivantes :

$1 \Rightarrow 4$: Immédiat vu le Théorème 2.2.11.

$4 \Rightarrow 3$: On sait, vu le Théorème 2.1.15 et la Proposition 2.2.19, que pour toute loi de Bernoulli positive π sur A^*, on a $\pi(X) = 1$.

$3 \Rightarrow 2$: Évident vu qu'il existe des lois de Bernoulli positives π sur A^*.

$2 \Rightarrow 1$: Direct vu la Proposition 2.1.21. □

Ce théorème nous permet de vérifier plus aisément si un code fin est maximal. En effet, il nous suffit maintenant de considérer n'importe quelle loi de Bernoulli positive et de voir si on a bien $\pi(X) = 1$.

Remarque 2.2.26. Considérons l'ensemble préfixe $X = \bigcup_{n\geq 0} a^n b A^n$ sur l'alphabet $A = \{a, b\}$.

- L'ensemble X est complet puisqu'il est dense :
  En effet, pour tout $w \in A^*$, $a^{|w|}bw \in X$.
- Toute loi de Bernoulli positive π sur A^* est telle que $\pi(X) = 1$.
  En effet, soit π une loi de Bernoulli positive sur A^* telle que $\pi(a) = p$ et $\pi(b) = 1-p$, avec $0 < p < 1$. On a alors

$$\begin{aligned}
\pi(X) &= \pi\left(\bigcup_{n\geq 0} a^n b A^n\right) \\
&= \sum_{n\geq 0} \pi(a^n b A^n) \\
&= \sum_{n\geq 0} \left(\sum_{w\in a^n b A^n} \pi(w)\right) \\
&= \sum_{n\geq 0} \pi(a^n)\pi(b) \underbrace{\left(\sum_{u\in A^n} \pi(u)\right)}_{=1} \\
&= \sum_{n\geq 0} p^n(1-p) \\
&= (1-p)\frac{1}{1-p} \\
&= 1.
\end{aligned}$$

Ainsi, X satisfait les quatre conditions du Théorème 2.2.25 alors qu'il n'est pas fin.

Théorème 2.2.27. *Soit $X \subset A^+$ un ensemble fin et soit π une loi de Bernoulli positive sur A^*. Chaque combinaison de deux des trois assertions suivantes implique la troisième.*

1. *L'ensemble X est un code.*

2. *On a* $\pi(X) = 1$.
3. *L'ensemble* X *est complet.*

Démonstration. On a les implications suivantes :

$1 + 2 \Rightarrow 3$: Si $\pi(X) = 1$, avec X un code, nous savons que X est un code maximal par la Proposition 2.1.21. Vu le Théorème 2.2.11, il vient que X est complet.

$1 + 3 \Rightarrow 2$: Puisque X est un code fin et complet, vu le Théorème 2.2.25, on sait que $\pi(X) = 1$.

$2 + 3 \Rightarrow 1$: Soit $n \geq 1$ un naturel. Commençons par montrer que l'ensemble X^n est fin et complet.

- Complet : Soit $u \in A^*$. Puisque X est complet par hypothèse, il existe $v, w \in A^*$ tels que $vuw \in X^*$. Il existe donc $k \geq 0$ tel que $vuw \in X^k$. Ainsi $(vuw)^n \in (X^n)^k \subset (X^n)^*$, ce qui montre que u est complétable dans $(X^n)^*$.
- Fin : Nous savons que X est fin et par la Remarque 2.2.16 nous savons que le produit de deux ensembles fins est fin. Ainsi il vient que X^n est un ensemble fin.

Grâce à la Proposition 2.2.19, nous savons que $\pi(X^n) \geq 1$. Or, vu la Proposition 2.1.10, nous savons que $\pi(X^n) \leq \pi(X)^n$, i.e. $\pi(X^n) \leq 1$. Il vient alors que $\pi(X^n) = 1$. Ainsi, pour tout $n \geq 1$ nous avons $\pi(X^n) = \pi(X)^n$. On en conclut, par la Proposition 2.1.20, que X est un code.

□

Nous terminons cette section en nous intéressant aux codes réguliers. Comme nous l'avons vu précédemment, un code fini n'est pas nécessairement inclus dans un code fini maximal. Cependant, nous pouvons nous en sortir en considérant des codes réguliers.

Proposition 2.2.28. *Tout code régulier est fin.*

Démonstration. Soit $X \subset A^*$ un code régulier. Soit $\mathcal{A} = \{Q, q_0, F, A, \delta\}$ un automate fini déterministe acceptant X. À tout mot $w \in A^*$, nous allons associer le nombre suivant :

$$\rho(w) = \mathrm{Card}(\delta(Q, w)) = \mathrm{Card}\{\delta(q, w) \mid q \in Q\}.$$

On a alors $\rho(w) \leq \mathrm{Card}(Q)$ et $\rho(uwv) \leq \rho(w)$ pour tous mots $u, v \in A^*$. Soit J l'ensemble des mots $w \in A^*$ tels que $\rho(w)$ est minimal. L'ensemble J est un idéal de A^*. En effet,

- $J \subset A^*$
- J est non vide
- Soit $w \in J$, puisque $\rho(uwv) \leq \rho(w)$ avec $\rho(w)$ minimal, on a $\rho(uwv) = \rho(w)$ et donc $uwv \in J$ pour $u, v \in A^*$.

Soit $w \in J$, posons $P = \delta(Q, w)$. Montrons que l'on a $P = \delta(P, w)$.
D'une part $\delta(P, w) \subset \delta(Q, w) = P$ et d'autre part $\delta(P, w) = \delta(Q, w^2)$. Il vient alors que

$\mathrm{Card}(\delta(P,w)) = \mathrm{Card}(\delta(Q,w^2)) = \rho(w^2)$. Puisque $\rho(w^2) \leq \rho(w)$, avec $\rho(w)$ minimal, il vient

$$\mathrm{Card}(\delta(P,w)) = \rho(w) = \mathrm{Card}(P).$$

L'égalité souhaitée en découle.
Ceci nous montre alors que l'application $g : P \to Q, p \mapsto \delta(p,w)$ a pour image P, et est donc une bijection de P dans lui-même. Il existe donc un naturel $n \geq 1$ tel que l'application $p \mapsto \delta(p,w^n)$ est l'application identité sur P. Puisque $P = \delta(Q,w)$, on a $\delta(q,w) = \delta(q,w^{n+1})$ pour tout $q \in Q$. Pour montrer que X est fin, il suffit de montrer qu'il ne rencontre pas l'idéal J, vu la Remarque 2.2.3. Procédons par l'absurde en supposant que $J \cap X \neq \emptyset$.
Soit $x \in J \cap X$. Nous avons alors $\delta(q_0,x) = f \in F$ puisque $\mathcal{A}$ accepte X. De plus, puisque $x \in J$, nous savons que $\delta(q_0,x^{n+1}) = \delta(q_0,x) = f \in F$. Il vient alors que $x^{n+1} \in X$, ce qui contredit le fait que X est un code puisque $x^{n+1} = (x)^{n+1}$.

□

Remarque 2.2.29. La réciproque est fausse. En effet, l'ensemble $X = \{a^n b^n \mid n \geq 1\}$ est fin, puisque par exemple le mot ba est incomplétable dans X. Par contre nous savons que X n'est pas régulier.

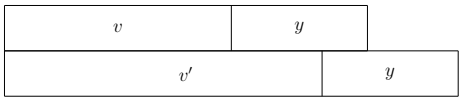
Proposition 2.2.30. *Soit $X \subset A^+$ un code. Soit $y \in A^*$ un mot sans bords tel que $A^*yA^* \cap X^* = \emptyset$. Soit*

$$U = A^* \backslash (X^* \cup A^*yA^*).$$

L'ensemble $Y = X \cup y(Uy)^*$ est un code complet.

Démonstration. Posons $V = A^* \backslash A^*yA^*$. Par hypothèse, nous avons $X^* \subset V$ et $U = V \backslash X^*$.

- L'ensemble $Z = Vy$ est un code préfixe :
  Soient $v, v' \in V$ tels que vy soit préfixe propre de $v'y$. Plus précisément vy doit être préfixe de v'. En effet, si ce n'était pas le cas, cela contredirait le fait que y est sans bords :



  Ainsi, il existe un mot $u \in A^*$ tel que $vyu = v'$. Il vient alors que $v' \in A^*yA^*$, ce qui contredit le fait que $v' \in V$.

- L'ensemble Y est un code :
  Procédons par l'absurde et supposons que Y n'est pas un code. Il existe donc un mot $w \in Y^*$, supposé de longueur minimale, qui possède deux factorisations distinctes en mots de Y :

$$w = y_1 y_2 \cdots y_n = y'_1 y'_2 \cdots y'_m, \tag{2.38}$$

  avec $n, m \geq 1$, $y_i, y'_j \in Y$ et $y_1 \neq y'_1$. L'ensemble X étant un code, il doit forcément y avoir un des y_i, y'_j qui appartient à X. Supposons que ce soit l'un des $y_1, \ldots, y_n$ qui soit dans $Y \backslash X$ et soit p le plus petit indice tel que $y_p \in y(Uy)^*$. Par hypothèse,

nous savons que y n'est pas complétable dans X^*, i.e. $y \notin F(X^*)$. Par conséquent, $y_p \notin F(X^*)$ et donc $w \notin X^*$. Ainsi, nous savons qu'un des $y'_1, \ldots, y'_m$ est également dans $y(Uy)^*$. Soit q le plus petit indice tel que $y'_q \in y(Uy)^*$. Nous avons donc

$$\underbrace{y_1 \cdots y_{p-1}}_{\in X^* \subset V} y, \underbrace{y'_1 \cdots y'_{q-1}}_{\in X^* \subset V} y \in Z$$

Vu 2.38, on a que $y_1 \cdots y_{p-1}$ est préfixe de $y'_1 \cdots y'_{q-1}$ ou inversement. S'il s'agit d'un préfixe propre, alors $y_1 \cdots y_{p-1}y$ reste un préfixe propre de $y'_1 \cdots y'_{q-1}y$, ce qui contredit le fait que Z est préfixe. Ainsi, on a $y_1 \cdots y_{p-1} = y'_1 \cdots y'_{q-1}$. Or, X est un code et $y_1 \neq y'_1$, on a alors forcément que $p = q = 1$, i.e. $y_1, y'_1 \in y(Uy)^*$. Posons donc

$$y_1 = yu_1y \cdots yu_ky \text{ et } y'_1 = yu'_1y \cdots yu'_ly$$

avec $u_1, \ldots u_k, u'_1, \ldots u'_l \in U$. Supposons que y_1 soit préfixe de y'_1. On sait que $U \subset V$ et donc $u_iy, u'_jy \in Z$. Il s'ensuit que

$$u_1 = u'_1, \ldots, u_k = u'_k.$$

Posons $t = u'_{k+1}y \cdots yu'_ly$. On a

$$y_2y_3 \cdots y_n = ty'_2 \cdots y'_m.$$

Le mot y, étant facteur de t, apparait dans $y_2 \cdots y_n$. Ainsi, l'un des $y_2, \ldots, y_n$ est dans $y(Uy)^*$. Soit r le plus petit indice tel que $y_r \in y(Uy)^*$. Donc $y_2y_3 \cdots y_{r-1}y \in Z$ et $u'_{k+1}y \in Z$ sont préfixes d'un même mot, par conséquent

$$\underbrace{y_2y_3 \cdots y_{r-1}}_{\in X^*} = u'_{k+1}.$$

Ainsi $u'_{k+1} \in X^*$, ce qui contredit le fait que $u'_{k+1} \in U$. Ainsi, Y est un code.

- L'ensemble Y est complet :
  Soit $w \in A^*$ et décomposons-le de la façon suivante :

$$w = v_1yv_2y \cdots yv_{n-1}yv_n$$

avec $n \geq 1$ et $v_i \in V$. Puisque $V = U \cup X^*$, notons $v_{i_1}, \ldots, v_{i_k}$ les v_i qui sont dans X^*, il vient alors

$$ywy = (yv_1y \cdots yv_{i_1-1})v_{i_1}(yv_{i_1+1}y \cdots yv_{i_2-1}y) \cdots v_{i_k}(yv_{i_k+1}y \cdots yv_ny).$$

Chaque parenthèse appartient à $y(Uy)^* \subset Y$. Ainsi, le mot $ywy \in Y^*$, i.e. w est complétable dans Y^*.

□

Théorème 2.2.31. *Tout code régulier est inclus dans un code régulier maximal.*

Démonstration. Soit X un code régulier. Si X est complet, alors il est maximal puisqu'il est fin par la Proposition 2.2.28. Sinon, il existe $y \in A^*$ tel que $A^*yA^* \cap X^* = \emptyset$. Au vu du Lemme 2.2.9 [6], on peut le supposer sans bords. Puisque l'ensemble X est régulier, nous savons que $U = A^* \backslash (X^* \cup A^*yA^*)$ est également régulier. Il en va de même pour $Y = X \cup y(Uy)^*$. Vu les Propositions 2.2.28 et 2.2.30, l'ensemble Y est un code fin et complet. Finalement, Y est maximal vu le Théorème 2.2.20.

□

En particulier, tout code fini est inclus dans un code régulier maximal.

6. Si Card$(A) = 1$, le théorème est évident. On peut donc supposer travailler sur un alphabet d'au moins deux lettres.

Chapitre 3

Codes préfixes

Dans ce chapitre, nous allons nous focaliser sur une famille particulière de codes : les codes préfixes. Lorsque l'on s'intéresse aux codes, les codes préfixes occupent rapidement une place privilégiée puisqu'il s'agit des codes les plus faciles à construire et à manipuler. Il est donc naturel de les étudier plus en profondeur.

Nous allons tout d'abord donner quelques résultats généraux concernant les ensembles préfixes avant de nous intéresser aux liens entre les codes préfixes et les séries formelles.

Une partie de ce chapitre sera consacrée à l'étude des automates acceptant les ensembles préfixes. Pour cela, nous aurons besoin d'introduire préalablement la notion de représentation littérale d'un ensemble $X \subset A^*$.

Ensuite, nous adapterons les notions de densité vues dans la Section 2.2 au cas particulier des ensembles préfixes dans le but d'en étudier la maximalité.

Finalement, nous nous concentrerons sur un sous-ensemble spécifique des codes préfixes : les codes sémaphores. Nous établirons une série de résultats concernant cette nouvelle famille avant de déboucher sur le Théorème 3.7.15 qui nous montrera que celle-ci possède une propriété intéressante qui n'est pas partagée par les codes préfixes en toute généralité.

Évidemment, tous les résultats présentés dans ce chapitre peuvent aisément être adaptés aux ensembles suffixes.

3.1 Premiers résultats

Rappelons que pour des mots $x, y \in A^*$, nous notons $x \preceq y$ (resp. $x \prec y$) le fait que x est préfixe (resp. préfixe propre) de y. L'ordre $\preceq$ est l'ordre préfixe.

Définition 3.1.1. Deux mots $x, y \in A^*$ sont dits *incomparables* pour l'ordre préfixe si x n'est pas préfixe de y et si y n'est pas préfixe de x.

Naturellement, un sous-ensemble $X \subset A^*$ est *préfixe* si les mots de X sont incomparables deux-à-deux pour l'ordre préfixe.

Posons

$$XA^- = X(A^+)^{-1} = \{w \in A^* \mid \exists u \in A^+ : wu \in X\},$$

l'ensemble des préfixes propres des mots de X. Ainsi, $w \in XA^-$ si et seulement si $w \prec x$ pour un $x \in X$.

Proposition 3.1.2. *Soit $X \subset A^*$. Les assertions suivantes sont équivalentes :*

1. *L'ensemble X est préfixe.*
2. *$X \cap XA^- = \emptyset$.*
3. *$X \cap XA^+ = \emptyset$.*
4. *Les ensembles XA^-, X, XA^+ sont deux-à-deux disjoints.*
5. *Si $xu = x'u'$ avec $x, x' \in X$ et $u, u' \in A^*$, alors $x = x'$ et $u = u'$.*
6. *Si $x, xu \in X$ alors $u = \varepsilon$.*

Démonstration. Les implications $1 \Rightarrow 2$, $2 \Rightarrow 3$, $3 \Rightarrow 4$, $5 \Rightarrow 6$ et $6 \Rightarrow 1$ sont immédiates. Nous allons démontrer que 4 implique 5, afin de compléter le cycle d'implications.
Soient $x, x' \in X$ et $u, u' \in A^*$ tels que $xu = x'u'$.

- Cas 1 : $u = u' = \varepsilon$
  On a directement $x = x'$ et $u = u'$.
- Cas 2 : $u = \varepsilon$ et $u' \in A^+$.
  Alors $x = x'u'$. Ainsi, $x \in X \cap XA^+$, ce qui contredit l'hypothèse.
- Cas 3 : $u' = \varepsilon$ et $u \in A^+$.
  Alors $x' = xu$. Ainsi, $x' \in X \cap XA^+$, ce qui contredit l'hypothèse.
- Cas 4 : $u, u' \in A^+$.
  - Supposons $x \prec x'$. Il existe alors $v \in A^+$ tel que $xv = x'$, i.e. $xv \in X$. Ainsi, $x \in X \cap XA^-$, ce qui contredit l'hypothèse.
  - Supposons $x' \prec x$. Il existe alors $v' \in A^+$ tel que $x'v' = x$, i.e. $x'v' \in X$. Ainsi, $x' \in X \cap XA^-$, ce qui contredit l'hypothèse.

  Il vient donc $x = x'$ et par conséquent $u = u'$.

□

La proposition suivante nous permet de lier les notions de code préfixe et d'idéal à droite.

Proposition 3.1.3. *Pour tout ensemble $Y \subset A^*$, l'ensemble $X = Y \setminus YA^+$ est préfixe. De plus,*

$$XA^* = YA^* \tag{3.1}$$

et X est l'ensemble minimum avec cette propriété.

Démonstration. Montrons tout d'abord que l'ensemble X est préfixe.
Puisque $X \subset Y$, il vient $XA^+ \subset YA^+$. Ainsi,

$$X \cap XA^+ \subset X \cap YA^+ = \emptyset.$$

Vu les équivalences de la Proposition 3.1.2, on en tire que X est bien préfixe.

Montrons ensuite l'égalité 3.1.

- $\subset$: On a trivialement $XA^* \subset YA^*$.
- $\supset$: Soient $y \in Y$ et u son plus petit préfixe dans Y. On a alors $u \in X$. Ainsi, $y \in XA^*$, et donc $Y \subset XA^*$.

Montrons finalement que X est minimum pour cette propriété.
Soit Z un ensemble de générateurs de l'idéal à droite YA^*, i.e. $ZA^* = YA^*$. Soit $x \in X \subset Y \subset YA^*$, alors $x = zu$ pour $u \in A^*$ et $z \in Z$. Puisque $XA^* = YA^*$, X génère également l'idéal YA^*, et donc $z = x'u'$ pour $x' \in X$ et $u' \in A^*$. Il vient donc

$$x = zu = x'u'u.$$

L'ensemble X étant préfixe, il vient que $u'u = \varepsilon$, ce qui montre que $X \subset Z$. □

Définition 3.1.4. L'ensemble $X = Y \backslash YA^+$ est appelé *la partie initiale* de Y ou *la base de l'idéal à droite* YA^*.

La notion d'ensemble préfixe est intimement liée à celles d'idéal à droite et d'ensemble fermé par préfixes. En effet, la proposition suivante nous fournit des bijections naturelles entre les familles d'ensembles suivantes :

1. La famille $\mathcal{X}$ des sous-ensembles préfixes de A^* ;
2. La famille $\mathcal{I}$ composée de l'ensemble vide et des idéaux à droite de A^* ;
3. La famille $\mathcal{R}$ des sous-ensembles fermés par préfixe de A^*.

Proposition 3.1.5. *1. L'application $f : X \mapsto XA^*$ est une bijection de $\mathcal{X}$ dans $\mathcal{I}$. L'application $g : I \mapsto I \backslash IA^+$ est son inverse de $\mathcal{I}$ dans $\mathcal{X}$.*

2. L'application $h : R \mapsto A^* \backslash R$ est une bijection de $\mathcal{R}$ dans $\mathcal{I}$. L'application $i : I \mapsto A^* \backslash I$ est son inverse de $\mathcal{I}$ dans $\mathcal{R}$.

3. L'application $j : X \mapsto A^* \backslash XA^*$ est une bijection de $\mathcal{X}$ dans $\mathcal{R}$. L'application $k : R \mapsto (\{\varepsilon\} \cup RA) \backslash R$ est son inverse de $\mathcal{R}$ dans $\mathcal{X}$.

Démonstration. 1. Pour tout sous-ensemble non vide $X \subset A^*$, l'ensemble XA^* est un idéal à droite de A^*.
Pour tout sous-ensemble I de A^*, l'ensemble $I \backslash IA^+$ est préfixe vu la Proposition 3.1.3.
Ainsi, les applications f et g sont bien définies. Nous allons montrer que ces applications sont inverses l'une de l'autre.

Montrons que $f \circ g = \mathrm{id}$, i.e. pour tout idéal à droite I de A^*, on a $(I\backslash IA^+)A^* = I$.
Soit I un idéal à droite de A^*. Posons $X = I\backslash IA^+$. Par la Proposition 3.1.3, nous avons $XA^* = IA^*$. De plus, nous savons que $IA^* = I$, vu la définition d'un idéal à droite. Il vient alors $XA^* = I$.
Montrons que $g \circ f = \mathrm{id}$, i.e. pour tout ensemble préfixe X de A^*, on a $XA^*\backslash XA^+ = X$.
Soit $X \subset A^*$ un ensemble préfixe. Vu la Proposition 3.1.2, nous avons $X \cap XA^+ = \emptyset$. Il s'ensuit que $X = X\backslash XA^+$. On a alors

$$X = X\backslash XA^+ = (X \cup XA^+)\backslash XA^+ = XA^*\backslash XA^+.$$

2. Soit I un idéal à droite de A^*. Si $w \notin I$, alors aucun de ses préfixes n'est dans I, i.e. tous ses préfixes sont dans $A^*\backslash I$. En effet, si $w = vu$ avec $v \in I$ et $u \in A^*$, puisque I est un idéal à droite on aurait que $vu \in I$, ce qui contredirait le fait que $w \notin I$. L'ensemble $R = A^*\backslash I$ est donc bien un ensemble fermé par préfixe.
   Soit R un ensemble fermé par préfixe. L'ensemble $A^*\backslash R$ est soit vide soit un idéal à droite de A^*. En effet si $w \notin R$ alors $wu \notin R$ pour tout $u \in A^*$. Autrement dit, si $w \in A^*\backslash R$ alors $wu \in A^*\backslash R$.
   Les applications h et i sont bien définies. De plus, il est évident que ces applications sont bien inverses l'une de l'autre.
3. Vu les points précédents, on a $i \circ f : \mathcal{X} \to \mathcal{R}, X \mapsto A^*\backslash XA^*$. Ainsi, $j = i \circ f$ est bien une bijection de $\mathcal{X}$ dans $\mathcal{R}$.
   Soit R un ensemble fermé par préfixe. On a

$$g \circ h : \mathcal{R} \to \mathcal{X}, R \mapsto (A^*\backslash R)\backslash(A^*\backslash R)A^+.$$

   Nous allons montrer que $(A^*\backslash R)\backslash(A^*\backslash R)A^+ = (\{\varepsilon\} \cup RA)\backslash R$, ainsi nous aurons $k = g \circ h$, qui est une bijection.
   Remarquons tout d'abord que dans le cas où $R = \emptyset$, l'égalité est évidente puisque $(\{\varepsilon\} \cup RA)\backslash R = \{\varepsilon\}$ et $(A^*\backslash R)\backslash(A^*\backslash R)A^+ = A^*\backslash A^+ = \{\varepsilon\}$. Supposons donc que $R \neq \emptyset$. Dans ce cas, $\varepsilon \in R$, ainsi il suffit de montrer l'égalité $(A^*\backslash R)\backslash(A^*\backslash R)A^+ = RA\backslash R$.

   $\subset$: Soit $x \in (A^*\backslash R)\backslash(A^*\backslash R)A^+ \subset A^*\backslash R$. Posons $x = ua$ avec $u \in A^*$ et $a \in A$. On a forcément que $u \in R$, sinon $ua \in (A^*\backslash R)A$, ce qui contredit le fait que $x = ua \notin (A^*\backslash R)A^+$. Ainsi, $x = ua \in RA\backslash R$.

   $\supset$: Soit $x \in RA\backslash R \subset A^*\backslash R$. Il existe donc $r \in R$ et $a \in A$ tels que $x = ra$. Puisque R est fermé par préfixe, tous les préfixes propres de x sont dans R. Ainsi, x n'a aucun préfixe propre dans $A^*\backslash R$, i.e. $x \notin (A^*\backslash R)A^+$. Donc $x \in (A^*\backslash R)\backslash(A^*\backslash R)A^+$.

□

Corollaire 3.1.6. *Soient $X, Y \subset A^*$ des ensembles préfixes. Si $XA^* = YA^*$, alors $X = Y$.*

3.2 Lien avec les séries formelles

Proposition 3.2.1. *Soit X un sous-ensemble de A^+ et soit X^* le sous-monoïde engendré par X. L'ensemble X est un code si et seulement si $\underline{X^*} = (\underline{X})^*$ ou, de manière équivalente, $\underline{X^*} = (\underline{\varepsilon} - \underline{X})^{-1}$.*

Démonstration. Vu la Proposition 1.1.5, on sait que le coefficient $((\underline{X})^*, w)$ d'un mot $w \in A^*$ est le nombre de factorisations distinctes de w en mots de X. La condition est nécessaire puisque si X est un code alors

$$\begin{aligned}((\underline{X})^*, w) &= \mathrm{Card}\{(x_1, x_2, \ldots, x_n) \mid n \geq 0, x_i \in X, w = x_1 \cdots x_n\} \\ &= \begin{cases} 1 \text{ si } w \in X^* \\ 0 \text{ sinon} \end{cases} \\ &= (\underline{X^*}, w).\end{aligned}$$

La condition est aussi suffisante. Procédons par l'absurde et supposons que X n'est pas un code. Il existe alors un mot $w \in X^*$ tel que $((\underline{X})^*, w) \geq 2$, ce qui contredit le fait que $\underline{X^*} = (\underline{X})^*$. □

Proposition 3.2.2. *Soient X un code préfixe sur A et $R = A^* \backslash XA^*$. Alors*

$$\underline{X} - \underline{\varepsilon} = \underline{R}(\underline{A} - \underline{\varepsilon}) \tag{3.2}$$

et

$$\underline{A^*} = \underline{X^*}\ \underline{R}. \tag{3.3}$$

Démonstration. Le produit XA^* est non ambigu vu le point 5 de la Proposition 3.1.2. Ainsi, par la Proposition 1.1.4, on a $\underline{XA^*} = \underline{X}\ \underline{A^*}$. Il vient donc

$$\underline{R} = \underline{A^* \backslash XA^*} = \underline{A^*} - \underline{X}\ \underline{A^*} = (\underline{\varepsilon} - \underline{X})\underline{A^*}. \tag{3.4}$$

On a $\underline{A^*} = (\underline{\varepsilon} - \underline{A})^{-1}$ grâce à la proposition précédente. Ainsi, en multipliant 3.4 par $(\underline{\varepsilon} - \underline{A})$ à droite, on obtient 3.2 :

$$\begin{aligned}\underline{R} &= (\underline{\varepsilon} - \underline{X})\underline{A^*} \\ \Rightarrow \underline{R}(\underline{\varepsilon} - \underline{A}) &= (\underline{\varepsilon} - \underline{X})\underline{A^*}(\underline{\varepsilon} - \underline{A}) \\ \Rightarrow \underline{R}(\underline{A} - \underline{\varepsilon}) &= (\underline{X} - \underline{\varepsilon})\underbrace{(\underline{\varepsilon} - \underline{A})^{-1}(\underline{\varepsilon} - \underline{A})}_{=\underline{\varepsilon}} \\ \Rightarrow \underline{R}(\underline{A} - \underline{\varepsilon}) &= (\underline{X} - \underline{\varepsilon}).\end{aligned}$$

De plus, les égalités 3.2 et 3.3 sont équivalentes. Par la proposition précédente, puisque X est un code, nous savons que $\underline{X^*} = (\underline{\varepsilon} - \underline{X})^{-1}$. Ainsi, en multipliant 3.2 par $\underline{X^*}$ à gauche et $\underline{A^*}$ à droite, on retombe sur 3.3. De manière analogue, en multipliant 3.3 à gauche par $(\underline{\varepsilon} - \underline{X})$ et à droite par $(\underline{\varepsilon} - \underline{A})$, nous obtenons 3.2. □

Remarque 3.2.3. 1. L'égalité 3.2 peut être réécrite

$$\underline{X} - \underline{\varepsilon} = \underline{R}(\underline{A} - \underline{\varepsilon})$$
$$\underline{X} - \underline{\varepsilon} = \underline{RA} - \underline{R}$$
$$\underline{X} + \underline{R} = \underline{RA} + \underline{\varepsilon}.$$

Ainsi, on voit qu'un mot de R concaténé à une lettre est soit dans X soit dans R. De plus tout mot de X se décompose en un mot de R concaténé à une lettre.

2. L'égalité 3.3 nous apprend, quant à elle, que tout mot $w \in A^*$ possède une unique factorisation de la forme

$$w = x_1 \cdots x_n u$$

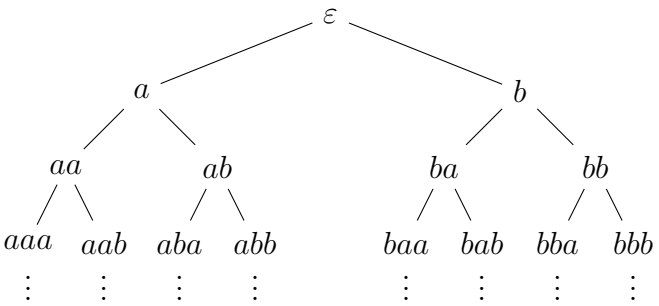
avec $x_i \in X$ et $u \in R$.

3.3 Représentation graphique des codes préfixes

Nous allons maintenant représenter les codes préfixes au moyen d'arbres dont les feuilles correspondront aux mots du code.

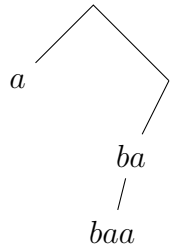
Tout d'abord, nous associons un arbre infini à l'ensemble A^* de la manière suivante : L'alphabet est totalement ordonné et les mots de A^* de même longueur sont ordonnés lexicographiquement. Chaque sommet de l'arbre représente un mot de A^*, les mots les plus courts se trouvant en haut de l'arbre et les mots les plus longs sont en bas. Dans le cas de mots de même longueur, ils sont disposés horizontalement selon l'ordre lexicographique de gauche à droite. On dessine une arête d'un mot u vers un mot v si et seulement si $v = ua$ pour un $a \in A$.
L'arbre obtenu de cette manière est appelé la *représentation littérale de* A^*.

Exemple 3.3.1. Considérons l'alphabet $A = \{a, b\}$. Sa représentation littérale débute comme suit :



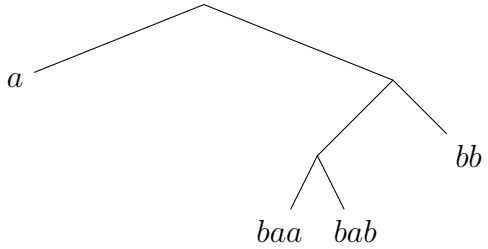
Maintenant nous considérons un sous-ensemble X de A^*. Nous lui associons un « sous-arbre » de la représentation littérale de A^* :
On ne conserve que les sommets qui correspondent aux préfixes des mots de X. Parmi ceux-ci nous ne nommons que ceux appartenant à X. L'arbre ainsi obtenu est la *représentation littérale de* X.

Exemple 3.3.2. Considérons l'ensemble $X = \{a, ba, baa\}$ sur l'alphabet $A = \{a, b\}$. La représentation littérale de X est



On se convainc alors qu'un ensemble X est préfixe si et seulement si, dans sa représentation littérale, les sommets correspondant aux mots de X sont exactement les feuilles de l'arbre.

Exemple 3.3.3. Considérons l'ensemble préfixe $X = \{a, baa, bab, bb\}$ sur l'alphabet $A = \{a, b\}$. La représentation littérale de X est



Nous pouvons remarquer que nous avons en effet une correspondance entre les mots de X et les feuilles de sa représentation littérale.

3.4 Ensembles préfixes et automates

Comme nous venons de le voir, la représentation littérale d'un sous-ensemble X de A^* nous permet de vérifier s'il est ou non préfixe. De plus, elle nous fournit une méthode pour déterminer si un mot w est dans X^* pour un certain code préfixe X fixé. En effet, il nous suffit de suivre le chemin, partant de la racine et suivant chaque lettre successivement de w. À chaque fois qu'une feuille est atteinte, nous continuons la lecture des lettres en repartant de la racine. Si après avoir lu la dernière lettre de w nous atteignons une feuille, cela signifie que le mot w appartient bien à l'ensemble X^*.

Dans cette section, nous allons nous intéresser à quelques automates dérivés des représentations littérales.

Nous commençons par présenter une caractérisation des ensembles préfixes en termes d'automates.

Proposition 3.4.1. *Soit $X \subset A^*$. Les assertions suivantes sont équivalentes :*

1. *L'ensemble X est préfixe.*
2. *L'automate minimal $\mathcal{A}_X$ n'a soit qu'un état et celui-ci est non accepteur, soit qu'un état final f. On a de plus $\delta_X(f, w) \notin F_X$ pour tout $w \in A^+$.*
3. *Il existe un automate déterministe $\mathcal{A} = (Q, q_0, F, A, \delta)$ acceptant X tel que $\delta(f, w) \notin F$ pour tous $f \in F$ et $w \in A^+$.*

Démonstration. L'implication $2 \Rightarrow 3$ est évidente. Montrons alors les deux implications suivantes afin de compléter le cycle.

$1 \Rightarrow 2$: Si $X = \emptyset$, il est évident que l'automate minimal qui l'accepte ne possède qu'un état et que celui-ci n'est pas final. Supposons alors que $X \neq \emptyset$. Soit $\mathcal{A}_X$ l'automate minimal de X. Pour tout $f \in F_X$, on a

$$\{w \in A^* \mid \delta_X(f, w) \in F_X\} = \{\varepsilon\}.$$

En effet, soit $x \in X$ et $w \in A^*$ tels que $\delta_X(q_{0,X}, x) = f$ et $\delta_X(f, w) \in F_X$. Alors $\delta_X(q_{0,X}, xw) \in F_X$, i.e. $xw \in X$, d'où $w = \varepsilon$ puisque X est préfixe. Ainsi, on a bien $\delta_X(f, w) \notin F$ pour tout $w \in A^+$.
Puisque $\mathcal{A}_X$ est minimal, il est réduit. Donc si $p, q \in F_X$, on a

$$\{w \in A^* \mid \delta_X(p, w) \in F_X\} = \{\varepsilon\} = \{w \in A^* \mid \delta_X(q, w) \in F_X\},$$

ce qui implique $p = q$, ce qui montre que l'état final est unique.

$3 \Rightarrow 1$: On sait que $\delta(f, w) \notin F$ pour tous $f \in F$ et $w \in A^+$. Si $x \in X$ et $w \in A^+$ alors $\delta(q_0, xw) = \delta(\underbrace{\delta(q_0, x)}_{\in F}, w) \notin F$. Donc $xw \notin X$. Ainsi, X est préfixe.

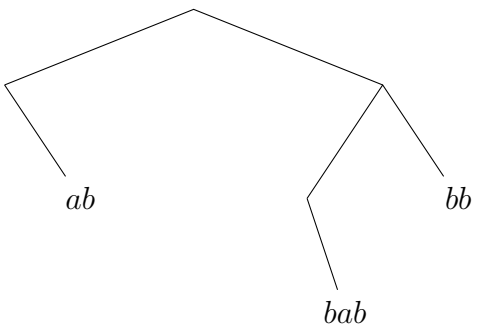
□

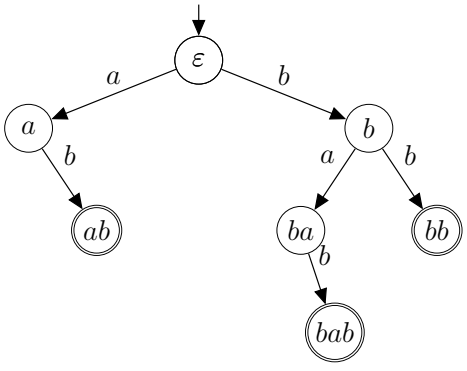
Soit X un code préfixe. Nous allons construire un automate $\mathcal{A}$ acceptant X à partir de la représentation littérale de X. Nous appellons cet automate *l'automate littéral* de X :

$$\mathcal{A} = (XA^- \cup X, \varepsilon, X, A, \Delta)$$

où la relation de transition est définie par $\Delta = \{(u, a, ua) \mid ua \in XA^- \cup X\}$.

Exemple 3.4.2. Soit l'ensemble $X = \{ab, bab, bb\}$ sur l'alphabet $A = \{a, b\}$.
La représentation littérale de X est : L'automate littéral de X est :



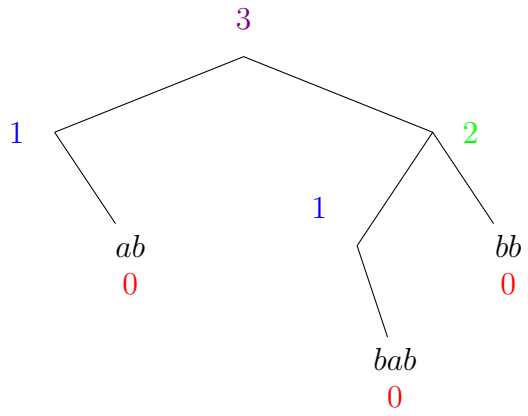


Remarque 3.4.3. Pour rendre l'automate littéral déterministe, il suffit de lui ajouter un puits vers lequel sont dirigées toutes les transitions manquantes.

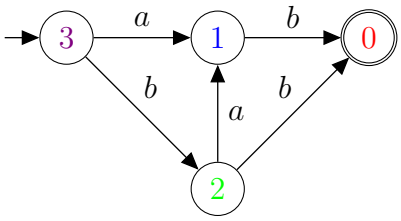
Considérons deux états de l'automate littéral. Cela revient à considérer deux préfixes de mots de X, disons u et v. Ces deux états sont non distingués si et seulement si $u^{-1}X = v^{-1}X$. Étant donné la construction de la représentation littérale, cette dernière égalité traduit le fait que le sous-arbre de la représentation littérale de X partant de la racine u est isomorphe à celui partant de la racine v.

Nous construisons alors l'automate minimal de X de la façon suivante : nous étiquetons d'abord par 0 tous les états finals, c'est-à-dire les feuilles. Supposons que nous avons déjà défini i étiquettes. Considérons les sous-arbres dont tous les nœuds sont étiquetés sauf la racine. Nous étiquetons alors les racines de sous-arbres isomorphes de la même manière. Finalement, ces étiquettes correspondront aux états de notre automate minimal.

Exemple 3.4.4. Considérons l'ensemble X défini à l'Exemple 3.4.2. Commençons par étiqueter les nœuds de notre représentation littérale.



Nous pouvons maintenant construire notre automate minimal.



Nous présentons maintenant une caractérisation des sous-monoïdes unitaires à droite en termes d'automates.

Proposition 3.4.5. *Soit M un sous-ensemble de A^*. Les assertions suivantes sont équivalentes :*

1. *M est un sous-monoïde unitaire à droite.*

2. *L'automate minimal $\mathcal{A}_M$ a un unique état final qui est l'état initial.*
3. *Il existe un automate déterministe acceptant M ayant l'état initial comme unique état final.*

Démonstration. L'implication $2 \Rightarrow 3$ est évidente. Montrons alors les deux implications suivantes afin de compléter le cycle.

$1 \Rightarrow 2$: Les états de $\mathcal{A}_M$ sont les ensembles $u^{-1}M$ pour $u \in A^*$. Si $u \in M$ alors $u^{-1}M = M$. En effet, si $u \in M$, on a $uv \in M$ si et seulement si $v \in M$. Ainsi, il y a un unique état final qui correspond à l'état initial.

$3 \Rightarrow 1$: Soit $\mathcal{A} = \{Q, q_0, \{q_0\}, A, \delta\}$ un automate acceptant M. L'ensemble M est un sous-monoïde puisque $\varepsilon \in M$, étant donné que l'état initial est accepteur, et si $u, v \in M$, alors $\delta(q_0, uv) = \delta(\underbrace{\delta(q_0, u)}_{=q_0}, v) = q_0$. Ensuite, l'ensemble M est unitaire à droite. En effet, soient $u, uv \in M$. On a

$$\delta(q_0, v) = \delta(\delta(q_0, u), v) = \delta(q_0, uv) = q_0.$$

Ainsi, $v \in M$.

□

Définition 3.4.6. Soit $\mathcal{A} = \{Q, q_0, F, A, \delta\}$ un automate déterministe. Le *stabilisateur* d'un état $q \in Q$ est

$$\mathrm{Stab}(q) = \{w \in A^* \mid \delta(q, w) = q\}.$$

Remarquons que le stabilisateur d'un état d'un automate déterministe est un sous-monoïde unitaire à droite.

Proposition 3.4.7. *Tout sous-monoïde unitaire à droite est le stabilisateur d'un état d'un automate déterministe.*

Démonstration. Soit M un sous-monoïde unitaire à droite. Vu la proposition précédente, on sait qu'il existe un automate déterministe acceptant M qui possède comme unique état final l'état initial. Ainsi, $\mathrm{Stab}(q_0) = M$. □

Soit X un code préfixe. Vu la Proposition 1.4.15, nous savons que X^* est unitaire à droite.

La proposition suivante nous fournit une méthode pour construire l'automate minimal $\mathcal{A}_{X^*}$ de X^* à partir de l'automate minimal $\mathcal{A}_X$ de X.

Proposition 3.4.8. *Soient X un code préfixe non vide sur A et $\mathcal{A}_X = \{Q, q_0, \{f\}, A, \delta\}$ l'automate minimal de X. Alors l'automate minimal de X^* est*

$$\mathcal{A}_{X^*} = \begin{cases} (Q, f, \{f\}, A, \circ) \text{ si } \mathrm{Stab}(q_0) \neq \{\varepsilon\} \\ (Q \backslash \{q_0\}, f, \{f\}, A, \circ) \text{ si } \mathrm{Stab}(q_0) = \{\varepsilon\} \end{cases}$$

où la fonction de transition, notée $\circ$, est définie par

$$q \circ a = \delta(q, a) \text{ si } q \neq f$$

$$f \circ a = \delta(q_0, a).$$

Démonstration. Soit $\mathcal{B}$ l'automate $(Q, f, \{f\}, A, \circ)$. On a $\mathcal{L}(\mathcal{B}) = \{w \in A^* \mid f \circ w = f\} = X^*$. Vérifions que $\mathcal{B}$ est réduit. Considérons des états distincts p et q. Puisque $\mathcal{A}_X$ est réduit, il existe un mot $u \in A^+$ qui distingue p et q. Supposons par exemple que

$$\delta(p, u) = f \text{ et } \delta(q, u) \neq f.$$

On peut remarquer que, vu la Proposition 3.4.1, l'état p est distinct de l'état final f, puisque $\delta(f, u) \neq f$ et que de plus, tous les états atteints en lisant successivement une lettre de u sont également différents de l'état final. On a alors $p \circ u = \delta(p, u) = f$ et $p \circ v \neq f$ pour tout $v \prec u$.

- Si $q \circ u \neq f$, alors le mot u distingue p et q dans l'automate $\mathcal{B}$.
- Sinon, si $q \circ u = f$, il existe un préfixe v de u, supposons-le être le plus petit, tel que $q \circ v = f$. Montrons alors que $\delta(q, v) = f$. Procédons par récurrence pour montrer que $\forall\ v' \preceq v, q \circ v' = \delta(q, v')$.
  - — Cas de base : $v' = \varepsilon$.
    Par convention, on a $q \circ \varepsilon = q = \delta(q, \varepsilon)$.
  - — Induction : Supposons que l'on a $q \circ v'' = \delta(q, v'')$ pour tout $v'' \prec v'$.
    Supposons que $v' = v''a$. On a alors

$$\begin{aligned} q \circ v' &= (q \circ v'') \circ a \\ &= \delta(q, v'') \circ a. \end{aligned}$$

    Comme $v'' \prec v' \preceq v$, on sait que $\delta(q, v'') = q \circ v'' \neq f$, donc

$$\delta(q, v'') \circ a = \delta(\delta(q, v''), a) = \delta(q, v').$$

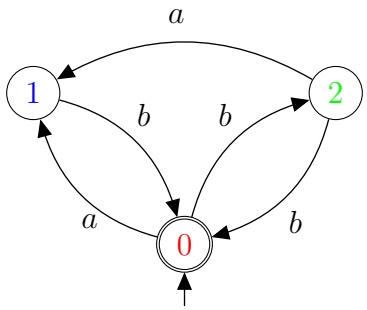
  Donc $\delta(q, v) = q \circ v = f$. Il vient alors que $u \neq v$ et donc $v \prec u$. On a alors $p \circ v \neq f$ et $q \circ v = f$. Ainsi, p et q sont distingués par v.

On sait que $\varepsilon \notin X$ puisque X est un code préfixe. On a donc $q_0 \neq f$. Ainsi, l'état q_0 est accessible dans $\mathcal{B}$ si et seulement si $\{w \in A^* \mid f \circ w = q_0\} \neq \emptyset$, autrement dit si et seulement $\mathrm{Stab}(q_0) \neq \{\varepsilon\}$.

Ceci montre que $\mathcal{B}$ est l'automate minimal de X^* si $\mathrm{Stab}(q_0) \neq \{\varepsilon\}$. Sinon la partie accessible de $\mathcal{B}$ est sa restriction à $Q \backslash \{q_0\}$. □

Remarque 3.4.9. Dans le cas où X est un code préfixe fini, son automate minimal a la forme $\mathcal{A}_{X^*} = (Q \backslash \{q_0\}, f, \{f\}, A, \circ)$.

Exemple 3.4.10. Considérons l'ensemble X défini à l'Exemple 3.4.2. L'automate minimal de X^* est



Nous pouvons également construire l'automate littéral de X^* pour un code préfixe X. Il s'agit de l'automate

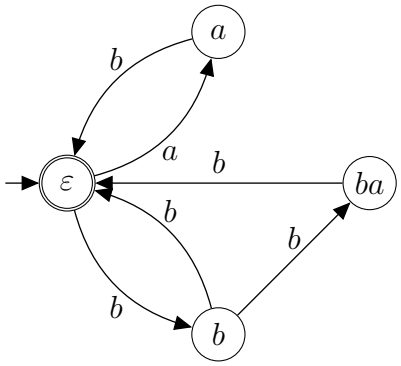
$$\mathcal{A} = (XA^-, \varepsilon, \{\varepsilon\}, A, \Delta)$$

où les états sont les préfixes propres des mots de X et où la relation de transition est définie par

$$\Delta = \{(u, a, ua) \mid ua \in XA^-\} \cup \{(u, a, \varepsilon) \mid ua \in X\}.$$

Il est donc obtenu depuis l'automate littéral de X en identifiant les états finals à l'état initial.

Exemple 3.4.11. Considérons l'ensemble X défini à l'Exemple 3.4.2. L'automate littéral de X^* est



3.5 Codes préfixes maximaux

Tout comme nous l'avons fait dans le cadre des codes généraux, nous allons maintenant nous intéresser à la structure des codes préfixes maximaux.

Un ensemble préfixe $X \subset A^*$ est *maximal* s'il n'est pas strictement inclus dans un autre ensemble préfixe de A^*, i.e. $X \subset Y \subset A^*$ et Y préfixe implique $X = Y$.

L'ensemble $\{\varepsilon\}$ est un ensemble préfixe maximal. Tout autre ensemble préfixe maximal est un code.

Un code maximal qui est préfixe est en particulier un ensemble préfixe maximal. Notons que la réciproque n'est pas vraie puisqu'il existe des codes préfixes maximaux qui ne sont pas des codes maximaux. Cependant, dans le cas des ensembles fins, nous verrons que les codes préfixes maximaux sont des codes maximaux.

Définition 3.5.1. Soient M un monoïde et N un sous-ensemble de M. Un élément $m \in M$ est dit *complétable à droite dans* N si $mw \in N$ pour un certain $w \in M$, ce qui est équivalent au fait que N rencontre l'idéal à droite mM, i.e. $N \cap mM \neq \emptyset$.

L'ensemble N est *dense à droite* si pour tout $m \in M$, m est complétable à droite dans N, ce qui est équivalent au fait que N rencontre tous les idéaux à droite de M.

L'ensemble N est *complet à droite* si le sous-monoïde engendré par N est dense à droite.

L'ensemble N est *fin à droite* s'il n'est pas dense à droite.

Soit N un sous-ensemble d'un monoïde M, on a trivialement les implications suivantes :

$$\begin{aligned} &N \text{ dense à droite } \Rightarrow N \text{ dense} \\ &N \text{ complet à droite} \Rightarrow N \text{ complet} \\ &N \text{ fin } \Rightarrow N \text{ fin à droite.} \end{aligned}$$

Remarque 3.5.2. Dans le cas du monoïde libre A^*, un sous-ensemble X de A^* est dense à droite si et seulement si tout mot de A^* est préfixe d'un mot de X. Ainsi, tout idéal à gauche de A^* est dense à droite.

L'ensemble X est complet à droite si tout mot $w \in A^*$ peut s'écrire

$$w = x_1 x_2 \cdots x_n p$$

pour $x_1, \ldots, x_n \in X$ $(n \geq 0)$ et p un préfixe d'un mot de X.

Proposition 3.5.3. *Soit $X \subset A^*$. Les conditions suivantes sont équivalentes :*

1. *L'ensemble XA^* est dense à droite.*
2. *$A^* = XA^- \cup X \cup XA^+$.*
3. *Pour tout $w \in A^*$, il existe $u, v \in A^*$ et $x \in X$ tels que $wu = xv$.*

Démonstration. On a les implications suivantes :

$1 \Rightarrow 3$: Soit $w \in A^*$. L'ensemble XA^* étant dense à droite, nous avons

$$XA^* \cap wA^* \neq \emptyset.$$

Il existe donc un mot $w' \in XA^* \cap wA^*$. Ainsi, $w' = xv = wu$ pour des mots $x \in X$ et $u, v \in A^*$.

$3 \Rightarrow 2$: L'inclusion $XA^- \cup X \cup XA^+ \subset A^*$ est évidente. Soit $w \in A^*$. Il existe $u, v \in A^*$ et $x \in X$ tels que $wu = xv$.

Cas 1 : Si $w \prec x$, alors il existe $w' \in A^+$ tel que $ww' = x$. Ainsi, $w \in XA^-$.

Cas 2 : Si $w = x$, alors $w \in X$.

Cas 3 : Si $x \prec w$, alors il existe $v' \in A^+$ tel que $xv' = w$. Ainsi, $w \in XA^+$.

Ainsi, $w \in XA^- \cup X \cup XA^+$.

$2 \Rightarrow 1$: L'ensemble des préfixes de XA^* est $XA^- \cup X \cup XA^+ = A^*$. Ainsi, soit $w \in A^*$, il existe $u \in A^*$ tel que $wu \in XA^*$. Le mot w est donc complétable à droite dans XA^*.

□

Proposition 3.5.4. *Soit $X \subset A^+$. L'ensemble XA^* est dense à droite si et seulement si X est complet à droite.*

Démonstration. Supposons que XA^* est dense à droite et considérons un mot $w \in A^* = XA^- \cup X \cup XA^+$. Montrons par récurrence sur la longueur de w que w est complétable à droite dans X^*.

- Cas de base : $|w| = 0$, i.e. $w = \varepsilon$.
  Alors $w \in X^*$.
- Induction : Supposons que pour tout $v \in A^*$ tel que $|v| < |w|$, v est complétable à droite dans X^*.
  - Cas 1 : $w \in XA^- \cup X$. Alors, on a directement qu'il existe $u \in A^*$ tel que $wu \in X \in X^*$.
  - Cas 2 : $w \in XA^+$. Alors il existe $x \in X$ et $v \in A^+$ tels que $w = xv$. On a $x \neq \varepsilon$, donc $|v| < |w|$. Par hypothèse de récurrence, il existe $u \in A^*$ tel que $vu \in X^*$. Il vient alors $wu = \underbrace{x}_{\in X}\underbrace{vu}_{\in X^*} \in X^*$.

Réciproquement, supposons que $w \in A^*$. Par hypothèse, $wu \in X^*$ pour un certain $u \in A^*$. Quitte à concaténer ce dernier avec un mot de X, on a $wu \neq \varepsilon$. Ainsi, $wu \in X^+ \subset XA^*$. □

Remarque 3.5.5. Si $X = \{\varepsilon\}$, alors $XA^* = A^*$. Or A^* est dense à droite alors que X n'est pas complet à droite puisque $X^* = \{\varepsilon\}$ n'est pas dense à droite. Ceci montre donc que la proposition précédente ne se généralise pas au cas $X \subset A^*$.

Nous allons démontrer l'analogue de la Proposition 3.1.5 dans le cas des ensembles préfixes maximaux. Considérons les familles suivantes :

1. La famille $\mathcal{M}$ des ensembles préfixes maximaux ;
2. La famille $\mathcal{D}$ des idéaux à droite qui sont denses à droite ;
3. La famille $\mathcal{P}$ des ensembles fermés par préfixe qui ne contiennent pas d'idéaux à droite.

Proposition 3.5.6. *Nous avons les bijections suivantes :*

1. *L'application $f' : X \mapsto XA^*$ est une bijection de $\mathcal{M}$ dans $\mathcal{D}$. L'application $g' : I \mapsto I \backslash IA^+$ est son inverse de $\mathcal{D}$ dans $\mathcal{M}$.*

2. *L'application* $h' : P \mapsto A^*\backslash P$ *est une bijection de* $\mathcal{P}$ *dans* $\mathcal{D}$. *L'application* $i' : I \mapsto A^*\backslash I$ *est son inverse de* $\mathcal{D}$ *dans* $\mathcal{P}$.
3. *L'application* $j' : X \mapsto XA^-$ *est une bijection de* $\mathcal{M}$ *dans* $\mathcal{P}$. *L'application* $k' : P \mapsto (PA \cup \{\varepsilon\})\backslash P$ *est son inverse de* $\mathcal{P}$ *dans* $\mathcal{M}$.

Démonstration. Remarquons que ces applications ne sont que les restrictions des applications définies dans la Proposition 3.1.5. C'est évident pour celles des points 1 et 2, nous le vérifierons ci-dessous pour le point 3.

1. Soit X un ensemble préfixe maximal. Tout mot $u \in A^*$ est comparable pour l'ordre préfixe avec un mot de X. En effet, supposons que u et x sont incomparables pour tout $x \in X$, i.e. x n'est pas préfixe de u et u n'est pas préfixe de x. Alors $X \cup \{u\}$ est encore un ensemble préfixe, ce qui contredit la maximalité de X. On a alors

Cas 1 : Si $u \prec x$ alors il existe $u' \in A^*$ tel que $uu' = x \in X \subset XA^*$.

Cas 2 : Si $u = x$ alors $u \in X \subset XA^*$.

Cas 3 : Si $x \prec u$ alors il existe $v \in A^*$ tel que $xv = u \in XA^*$.

On en conclut donc que XA^* est dense à droite.

Soit I un idéal à droite qui est dense à droite. Vu la Proposition 3.1.3 nous savons déjà que l'ensemble $I\backslash IA^+$ est préfixe. Montrons que pour tout $w \in A^*$, il existe $i \in I\backslash IA^+$ tel que w est comparable avec i.

Soit $w \in A^*$. Puisque I est dense à droite, il existe $u \in A^*$ tel que $wu \in I$. Soit i le plus petit préfixe de wu tel que $i \in I$. Alors $i \notin IA^+$. On a donc soit $i \prec w$ soit $i = w$ soit $w \prec i$.
Ainsi, $I\backslash IA^+$ est un ensemble préfixe maximal.

2. Soit I un idéal à droite qui est dense à droite. On sait déjà, vu la Proposition 3.1.5, que $A^*\backslash I$ est fermé par préfixe. Montrons alors que $A^*\backslash I$ ne contient pas d'idéal à droite.
Procédons par l'absurde et supposons que $I' \subset A^*\backslash I$ est un idéal à droite. Soit $i' \in I'$. On a alors $i'w \in I'$ pour tout $w \in A^*$, ce qui contredit le fait que I est dense à droite puisque i' n'est pas complétable à droite dans I.

Soit P un ensemble fermé par préfixe ne contenant pas d'idéal à droite. On sait déjà, vu la Proposition 3.1.5, que $A^*\backslash P$ est un idéal à droite. Montrons alors que $A^*\backslash P$ est dense à droite.
Soit $w \in A^*$. Puisque wA^* est un idéal à droite, on sait que $wA^* \not\subset P$. Il existe donc $u \in wA^*\backslash P$. Dès lors, $u = wv$ pour un certain $v \in A^*$. Donc $wv \in A^*\backslash P$, ce qui montre que w est complétable dans $A^*\backslash P$.

3. Soit X un ensemble préfixe maximal. Vu le point 1, on sait que XA^* est dense à droite. Par la Proposition 3.1.2, on sait que X, XA^+ et XA^- sont des ensembles disjoints deux-à-deux et par la Proposition 3.5.3, on sait que $A^* = XA^- \cup X \cup XA^+$. Il vient alors :
$$A^*\backslash XA^* = A^*\backslash (X \cup XA^+) = XA^-.$$

Ainsi, $j' = i' \circ f'$ est bien une bijection de $\mathcal{M}$ dans $\mathcal{P}$.
Soit P un ensemble fermé par préfixe qui ne contient pas d'idéal droite. On a

$$g' \circ h' : \mathcal{P} \to \mathcal{M}, P \mapsto (A^*\backslash P)\backslash(A^*\backslash P)A^+.$$

Vu la Proposition 3.1.5, on sait que $(PA \cup \{\varepsilon\})\backslash P = (A^*\backslash P)\backslash(A^*\backslash P)A^+$. Ainsi, $k' = g' \circ h'$ est bien une bijection de $\mathcal{P}$ dans $\mathcal{M}$.

□

Corollaire 3.5.7. *Soient $Y \subset A^+$ et $X = Y\backslash YA^+$. L'ensemble Y est complet à droite si et seulement si X est un code préfixe maximal.*

Démonstration. Vu la Proposition 3.5.4, Y est complet à droite si et seulement si YA^* est dense à droite. Vu la Proposition 3.1.3, on sait que X est un ensemble préfixe et $YA^* = XA^*$. Ainsi, XA^* est un idéal à droite qui est dense à droite. Par conséquent $XA^*\backslash(XA^*)A^+$ est un ensemble préfixe maximal vu la proposition précédente. Or $XA^*\backslash(XA^*)A^+ = XA^*\backslash XA^+ = (X \cup XA^+)\backslash XA^+$. De plus, nous savons que $X \cup XA^+$ est une union disjointe puisque X est un ensemble préfixe. Ainsi, $XA^*\backslash(XA^*)A^+ = X$. □

Ce corollaire nous fournit trivialement le résultat suivant :

Théorème 3.5.8. *Soit X un code préfixe. L'ensemble X est complet à droite si et seulement si X est un code préfixe maximal.*

Nous pouvons exprimer la maximalité d'un code préfixe en termes de séries formelles comme le montre le théorème suivant.

Théorème 3.5.9. *Soient X un code préfixe sur A et $P = XA^-$ l'ensemble des préfixes propres des mots de X. L'ensemble X est préfixe maximal si et seulement si on a les conditions équivalentes suivantes :*

$$\underline{X} - \underline{\varepsilon} = \underline{P}(\underline{A} - \underline{\varepsilon}) \tag{3.5}$$

et

$$\underline{A^*} = \underline{X^*}\ \underline{P}. \tag{3.6}$$

Démonstration. Posons $R = A^*\backslash XA^*$. Si X est préfixe maximal alors XA^* est dense à droite. Par la Proposition 3.1.2, nous savons que les ensembles XA^-, X, XA^+ sont disjoints deux à deux. Ainsi, grâce à la Proposition 3.5.3, il vient

$$R = A^*\backslash XA^* = A^*\backslash(XA^+ \cup X) = XA^- = P.$$

Ainsi, par la Proposition 3.2.2, il vient directement que $\underline{X} - \underline{\varepsilon} = \underline{P}(\underline{A} - \underline{\varepsilon})$.
Inversement, si $\underline{X} - \underline{\varepsilon} = \underline{P}(\underline{A} - \underline{\varepsilon})$, vu la Proposition 3.2.2, il vient

$$\underline{P}(\underline{A} - \underline{\varepsilon}) = \underline{R}(\underline{A} - \underline{\varepsilon}).$$

De plus, on sait que $(\underline{A} - \underline{\varepsilon})$ est inversible. On a alors $P = R$. Il s'ensuit que

$$\begin{aligned} & A^*\backslash(XA^+ \cup X) = XA^- \\ \Leftrightarrow\ & A^* = XA^- \cup X \cup XA^+. \end{aligned}$$

Ainsi, XA^* est dense à droite. □

Observons que lorsque X est un code préfixe maximal fini, alors P est également fini et donc $\underline{P}$ est un polynôme. Ainsi, l' équation 3.5 nous donne une factorisation de $\underline{X} - \underline{\varepsilon}$ en deux polynômes.

Remarque 3.5.10. L'équation 3.5 peut se réécrire $\underline{X} + \underline{P} = \underline{\varepsilon} + \underline{P}\ \underline{A}$. Sous cette forme, la maximalité d'un code préfixe X peut se vérifier aisément à partir de sa représentation littérale. En effet, nous allons montrer que X est un code préfixe maximal si et seulement si pour tout nœud p qui n'est pas dans X, on a pour tout $a \in A$, pa qui est un nœud dans la représentation littérale de X, i.e. X est un code préfixe maximal si et seulement si $\forall p \in P$ et $a \in A, pa \in X \cup P$.

Supposons que X est un code préfixe maximal. On sait alors que $\underline{X} + \underline{P} = \underline{\varepsilon} + \underline{P}\ \underline{A}$. Soient $p \in P$ et $a \in A$. Le mot $pa \in PA$, il vient alors que $(\underline{P}\ \underline{A}, pa) = 1$. Puisque $pa \neq \varepsilon$, on a même $(\underline{P}\ \underline{A} + \underline{\varepsilon}, pa) = 1$. Puisque les ensembles X et P sont disjoints, vu que X est préfixe, il s'ensuit que $(\underline{X} + \underline{P}, pa) = (\underline{X \cup P}, pa) = 1$. Ainsi, $pa \in X \cup P$.

Réciproquement, supposons que pour tous $p \in P$ et $a \in A$, $pa \in X \cup P$. Montrons que $w \in X \cup P$ si et seulement $w \in PA \cup \{\varepsilon\}$.
Soit $w \in X \cup P$. Si $w = \varepsilon$, alors $w \in PA \cup \{\varepsilon\}$. Sinon, $w = pa$ où p est un préfixe de $w \in X \cup P$. Donc $p \in P$. On en tire $w \in PA$.
Soit $w \in PA \cup \{\varepsilon\}$. Si $w = \varepsilon$, alors $w \in P$. Sinon, si $w \in PA$, alors $w = pa$ pour $p \in P$ et $a \in A$. Par hypothèse, on a alors que $w \in X \cup P$.

Nous allons maintenant montrer que dans le cas particulier d'ensembles fins, un code préfixe maximal est aussi un code maximal.

Théorème 3.5.11. *Soit X un sous-ensemble fin de A^+. Les assertions suivantes sont équivalentes :*

1. *L'ensemble X est un code préfixe maximal.*
2. *L'ensemble X est préfixe et un code maximal.*
3. *L'ensemble X est complet à droite et un code.*

Démonstration. L'implication $2 \Rightarrow 1$ est évidente.

$1 \Rightarrow 3$: Soit X un code préfixe maximal. Alors, vu la Proposition 3.5.6, XA^* est dense à droite. Par la Proposition 3.5.4, on en tire que X est complet à droite.

$3 \Rightarrow 2$: Soit $Y = X\backslash XA^+$. Vu la Proposition 3.1.3, on sait que Y est préfixe et que $XA^* = YA^*$. Par la Proposition 3.5.4, XA^* est dense à droite. Il s'ensuit, par la même proposition, que Y est complet à droite et par conséquent, Y est complet. Puisque $Y \subset X$, avec X fin, l'ensemble Y est fin également. Par le Théorème 2.2.20, Y est un code maximal. Il vient alors $Y = X$.

□

La proposition suivante découle directement des Théorèmes 3.5.11 et 2.2.25.

Proposition 3.5.12. *Soit X un sous-ensemble fin de A^+. Les assertions suivantes sont équivalentes :*

1. *L'ensemble X est un code préfixe maximal.*
2. *L'ensemble X est préfixe et il existe une loi de Bernoulli positive π telle que $\pi(X) = 1$.*
3. *L'ensemble X est préfixe et $\pi(X) = 1$ pour toute loi de Bernoulli positive π.*

Pour terminer cette section, nous allons caractériser les codes préfixes maximaux au moyen des automates acceptant les sous-monoïdes qu'ils engendrent.

Définition 3.5.13. Un état q d'un automate déterministe $\mathcal{A} = \{Q, q_0, F, A, \delta\}$ est *récurrent* si pour tout $u \in A^*$ il existe un mot $v \in A^*$ tel que $\delta(q, uv) = q$.

Proposition 3.5.14. *Soit X un code préfixe sur A. Les assertions suivantes sont équivalentes :*

1. *L'ensemble X est préfixe maximal.*
2. *L'automate minimal de X^* est coaccessible.*
3. *Tous les états de l'automate minimal de X^* sont récurrents.*
4. *L'état initial de l'automate minimal de X^* est récurrent.*
5. *L'ensemble X^* est le stabilisateur d'un état récurrent d'un automate déterministe.*

Démonstration.

$1 \Rightarrow 2$: Soit $\mathcal{A}_{X^*} = (Q_{X^*}, q_{0,X^*}, q_{0,X^*}, A, \delta_{X^*})$ l'automate minimal de X^*. Soit $q \in Q_{X^*}$. Puisque $\mathcal{A}_{X^*}$ est accessible, il existe un mot $u \in A^*$ tel que $\delta_{X^*}(q_{0,X^*}, u) = q$. Par le Théorème 3.5.8, X est complet à droite. Ainsi u est complétable à droite dans X^*, donc il existe un mot $v \in A^*$ tel que $uv \in X^*$. Il vient que $\delta_{X^*}(q, v) = \delta_{X^*}(\delta_{X^*}(q_{0,X^*}, u), v) = \delta_{X^*}(q_{0,X^*}, uv) = q_{0,X^*}$. L'état q est donc coaccessible.

$2 \Rightarrow 3$: Soit $q \in Q_{X^*}$ et $u \in A^*$. Puisque $\mathcal{A}_{X^*}$ est accessible, il existe $v \in A^*$ tel que $\delta_{X^*}(q_{0,X^*}, v) = q$. Par hypothèse, l'automate minimal est coaccessible, il existe donc $w \in A^*$ tel que $\delta_{X^*}(\delta_{X^*}(q, u), w) = q_{0,X^*}$. Ainsi $\delta_{X^*}(q, uwv) = \delta_{X^*}(q_{0,X^*}, v) = q$.

$3 \Rightarrow 4$: Évident.

$4 \Rightarrow 5$: On a $X^* = \mathcal{L}(\mathcal{A}_{X^*}) = \{w \in A^* | \delta_{X^*}(q_{0,X^*}, w) = q_{0,X^*}\} = \mathrm{Stab}(q_{0,X^*})$ et q_{0,X^*} est récurrent par hypothèse.

$5 \Rightarrow 1$: Soient $\mathcal{A} = (Q, i, F, A, \delta)$ un automate déterministe et $q \in Q$ un état récurrent tel que $X^* = \mathrm{Stab}(q) = \{w \in A^* | \delta(q, w) = q\}$. Soit $u \in A^*$, il existe $v \in A^*$ tel que $\delta(q, uv) = q$, puisque q est récurrent. Ainsi, $uv \in \mathrm{Stab}(q) = X^*$. Il vient alors que X est complet à droite. Ainsi, par le Théorème 3.5.8, l'ensemble X est un code préfixe maximal. □

3.6 Quelques opérations sur les ensembles préfixes

Proposition 3.6.1. *Soient X et $(Y_i)_{i\in I}$ des sous-ensembles de A^*. Soit $(X_i)_{i\in I}$ une partition de X. Posons*

$$Z = \bigcup_{i\in I} X_i Y_i.$$

1. *Si X et les Y_i sont préfixes (resp. préfixes maximaux) alors Z est préfixe (resp. préfixe maximal).*
2. *Si Z est préfixe alors tous les Y_i sont préfixes.*
3. *Si X est préfixe et Z est préfixe maximal, alors X et les Y_i sont préfixes maximaux.*

Démonstration. 1. Supposons que $z, zu \in Z$. Il existe alors $x \in X_i$, $y \in Y_i$ et $x' \in X_j$, $y' \in Y_j$, avec $i, j \in I$, tels que $z = xy$ et $zu = x'y'$. On a alors $zu = xyu = x'y'$. Puisque X est préfixe, il vient que $x = x'$. Il en découle $i = j$. On a alors $y, y' \in Y_i$ avec $y \preceq y'$. Étant donné que Y_i est préfixe, il vient $y = y'$. Par conséquent $u = \varepsilon$. L'ensemble Z est donc préfixe.

Supposons maintenant que X et les Y_i sont préfixes maximaux. Par la Proposition 3.5.6, les ensembles XA^* et Y_iA^* sont des ensembles denses à droite. Soit $w \in A^*$. Par la Proposition 3.5.3, il existe $w', v \in A^*$ et $x \in X$ tels que

$$ww' = xv.$$

Soit $i \in I$ tel que $x \in X_i$. Puisque Y_iA^* est dense à droite, il existe $v' \in A^*$ tel que $vv' \in Y_iA^*$. Donc $ww'v' = xvv' \in X_iY_iA^* \subset ZA^*$, ce qui montre que ZA^* est dense à droite. Par la Proposition 3.5.6, l'ensemble Z est préfixe maximal.

2. Soient $y, yu \in Y_i$ et $x \in X_i$ avec $i \in I$. On a $xy, xyu \in Z$. Or Z est préfixe donc $u = \varepsilon$, ce qui montre que les ensembles Y_i sont donc bien préfixes.
3. Puisque Z est préfixe maximal, on sait que ZA^* est dense à droite. Étant donné que $ZA^* \subset XA^*$, on en tire que XA^* est également dense à droite. Ainsi, X est préfixe maximal. Pour montrer que les Y_i sont préfixes maximaux, nous allons montrer que Y_iA^* est dense à droite.

   Soit $w \in A^*$. Pour tout $x \in X_i$, xw est complétable à droite dans ZA^*. Il existe alors $t \in A^*$ tel que $xwt \in ZA^*$. Donc il existe $z \in Z$ et $u \in A^*$ tels que $xwt = zu$. Il existe alors $x' \in X_j$ et $y' \in Y_j$ tels que $z = x'y'$. Il vient alors $xwt = x'y'u$. Puisque X est préfixe on a $x = x'$. Il s'ensuit que $i = j$ et donc $wt = y'u \in Y_iA^*$. Ainsi, w est complétable à droite dans Y_iA^*.

□

Corollaire 3.6.2. *Si X et Y sont des codes préfixes (resp. codes préfixes maximaux) alors XY est un code préfixe (resp. code préfixe maximal).*

Corollaire 3.6.3. *Soient $X \subset A^+$ et $n \geq 1$. L'ensemble X est préfixe (resp. préfixe maximal) si et seulement si X^n est préfixe (resp. préfixe maximal).*

Démonstration. Supposons que X est un code préfixe (resp. code préfixe maximal). Par le corollaire précédent, le produit X^n est un code préfixe (resp. code préfixe maximal).

Inversement, supposons que X^n est préfixe. Puisque $X^n = X^{n-1}X$, par le point 2 de la Proposition 3.6.1, il vient que X est préfixe. Dans le cas où X^n est préfixe maximal, alors on conclut par le point 3 que X est préfixe maximal. □

3.7 Codes sémaphores

Cette dernière section est dédiée à un sous-ensemble particulier des codes préfixes : les codes sémaphores. Après avoir défini cette nouvelle notion et donné quelques caractérisations et propriétés, nous pourrons finalement démontrer le Théorème 3.7.15 propre à cette nouvelle famille.

Proposition 3.7.1. *Pour tout ensemble non vide $S \subset A^+$, l'ensemble*

$$X = (A^*S)\backslash(A^*S)A^+ \tag{3.7}$$

est un code préfixe maximal.

Démonstration. L'ensemble A^*S est le plus petit idéal à gauche de A^* engendré par S. Vu la Remarque 3.5.2, on sait que A^*S est dense à droite et par conséquent complet à droite. On conclut alors que X est un code préfixe maximal par le Corollaire 3.5.7.

□

Définition 3.7.2. Un code X de la forme 3.7 est appelé un *code sémaphore*. L'ensemble $S \subset A^+$ est l'ensemble des *sémaphores* de X.

Soit $X = (A^*S)\backslash(A^*S)A^+$ un code sémaphore. Un mot x est dans X si et seulement s'il se termine par un sémaphore mais qu'aucun de ses préfixes propres ne se finit par un sémaphore.

On justifie l'emploi du terme « sémaphore » de la façon suivante : lorsqu'on lit un mot de X^*, chaque occurrence d'un sémaphore dans ce mot est un « signal » indiquant la fin d'un facteur dans la décomposition en mots de X [1].

Les quelques propositions suivantes nous donnent des caractérisations des codes sémaphores. Nous rappelons au lecteur que dans le cas où $X \subset A^+$, l'ensemble X est préfixe si et seulement si X est un code préfixe.

Proposition 3.7.3. *Soit $X \subset A^+$ non vide. L'ensemble X est un code sémaphore si et seulement si X est préfixe et $A^*X \subset XA^*$.*

Démonstration. Supposons tout d'abord que X est un code sémaphore. Il existe alors $S \subset A^+$ non vide tel que $X = (A^*S)\backslash(A^*S)A^+$. Par définition, on sait directement que X est un code préfixe. Montrons donc que $A^*X \subset XA^*$.

1. Tout comme un sémaphore ferroviaire indique la fin d'un canton.

Soit $w \in A^*X \subset A^*S$. Le mot w possède donc un facteur dans S. Soit w' le plus petit préfixe de w qui est dans A^*S. On a alors $w' \in X$. Ainsi, il existe $u \in A^*$ tel que $w = w'u \in XA^*$.

Supposons maintenant que X est un code préfixe et que $A^*X \subset XA^*$. Posons $M = XA^*$. Par la Proposition 3.1.5, puisque X est préfixe, on sait que $X = XA^*\backslash(XA^*)A^+ = M\backslash MA^+$. On a alors

$$A^*M = A^*XA^* \subset XA^* = M.$$

Donc $M = A^*M$ et $X = (A^*M)\backslash(A^*M)A^+$, ce qui montre que X est un code sémaphore. □

L'exemple suivant nous montre qu'être un code préfixe maximal n'est pas suffisant pour être un code sémaphore.

Exemple 3.7.4. Considérons l'ensemble $X = \{a^2, aba, ab^2, b\}$ sur l'alphabet $A = \{a, b\}$. Il s'agit d'un code préfixe. On sait aussi que l'ensemble X étant fini, il est fin. De plus, la loi uniforme de Bernoulli sur A est telle que

$$\begin{aligned}\pi(X) &= \pi(a)\pi(a) + \pi(a)\pi(b)\pi(a) + \pi(a)\pi(b)\pi(b) + \pi(b)\\ &= \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{2}\\ &= 1.\end{aligned}$$

Ainsi par la Proposition 3.5.12, X est un code préfixe maximal. Cependant, X n'est pas un code sémaphore puisque $ab \in A^*X$ mais $ab \notin XA^*$.

Proposition 3.7.5. *Soit $X \subset A^+$. L'ensemble X est un code sémaphore si et seulement si X est complet à droite et $X \cap A^*XA^+ = \emptyset$.*

Démonstration. Supposons que X est un code sémaphore. S'agissant d'un code préfixe maximal, il est complet à droite vu le Théorème 3.5.8. Montrons alors que $X \cap A^*XA^+ = \emptyset$. Vu la proposition précédente, on sait que $A^*X \subset XA^*$. On a donc $A^*XA^+ \subset XA^+$. Ainsi, il vient

$$X \cap A^*XA^+ \subset X \cap XA^+.$$

L'ensemble X étant préfixe, nous savons, grâce à la Proposition 3.1.2, que $X \cap XA^+ = \emptyset$, d'où la conclusion.

Supposons maintenant que X est complet à droite et que $X \cap A^*XA^+ = \emptyset$. Puisque $X \cap XA^+ \subset X \cap A^*XA^+ = \emptyset$, on en tire que X est préfixe par la Proposition 3.1.2. Pour vérifier que X est bien un code sémaphore, nous allons montrer que $A^*X \subset XA^*$. Soit $w = ux \in A^*X$ avec $u \in A^*$ et $x \in X$. L'ensemble X étant complet à droite, il existe un mot $v \in A^*$ tel que $uxv \in X^*$. Il existe donc $x' \in X$ et $y' \in X^*$ tels que $uxv = x'y'$. Puisque $X \cap A^*XA^+ = \emptyset$, ux n'est pas un préfixe propre de x'. Ainsi, $ux \in x'A^* \subset XA^*$. □

Corollaire 3.7.6. *Soient $X \subset A^+$ un code sémaphore et $P = XA^-$. On a alors $PX \subset XP \cup X^2$.*

Démonstration. Soient $p \in P$ et $x \in X$. On a $px \in PX \subset A^*X \subset XA^*$. Il existe donc $y \in X$ et $u \in A^*$ tels que $px = yu$. L'ensemble X étant un code préfixe, on a $|p| < |y|$. En effet, si $|p| \geq |y|$, alors y est préfixe de p qui est lui-même un préfixe d'un mot de X. Ainsi, y est préfixe d'un mot de X, ce qui contredit le fait que X est préfixe. Dès lors, u est un suffixe propre de x et vu que $X \cap A^*XA^+ = \emptyset$, $u \notin XA^+$. Étant donné que X est un code préfixe maximal, nous savons, par la Proposition 3.5.6, que XA^* est dense à droite. Ainsi, par la Proposition 3.5.3, on sait alors que $A^* = XA^- \cup X \cup XA^+$. Il vient alors $u \in XA^- \cup X = P \cup X$. On obtient finalement $px = yu \in X(P \cup X)$. □

Soit X un code sémaphore. La condition $X \cap A^*XA^+ = \emptyset$ impose non seulement que X est préfixe, c'est-à-dire qu'aucun mot de X n'est préfixe d'un autre mot de X, mais également qu'aucun mot de X n'est facteur d'un autre mot de X sauf s'il en est un suffixe. Autrement dit, soient x, x' deux éléments de X, la seule possibilité pour que x apparaisse comme facteur de x' est que x soit suffixe de x'.

Proposition 3.7.7. *Soient $X \subset A^+$ et $P = XA^-$. L'ensemble X est un code sémaphore si et seulement si X est un code préfixe maximal et P est fermé par suffixe.*

Démonstration. Supposons que X est un code sémaphore. On sait donc que c'est un code préfixe maximal. Montrons que P est fermé par suffixe.
Soit $p = uq \in P$ avec $u, q \in A^*$. Par définition de P, il existe un mot $v \in A^+$ tel que $pv \in X$. Le mot q ne peut pas être dans XA^* sinon $pv = uqv \in X \cap A^*XA^+ = \emptyset$. Il s'ensuit que $q \in A^* \backslash XA^*$. Or, on sait par la Proposition 3.7.5 que X est complet à droite. Ainsi, XA^* est dense à droite et $A^* = XA^- \cup X \cup XA^+$. Il vient donc $q \in XA^- = P$.

Inversement, supposons que X est un code préfixe maximal et que P est fermé par suffixe. Par le Théorème 3.5.8, X est complet à droite. Nous allons montrer que $X \cap A^*XA^+ = \emptyset$ et grâce à la Proposition 3.7.5 nous pourrons conclure que X est un code sémaphore.
Procédons par l'absurde. Supposons qu'il existe $x \in X \cap A^*XA^+$. Il existe donc $u \in A^*, x' \in X$ et $v \in A^+$ tels que $x = ux'v$. On a alors que $ux' \in P$. L'ensemble P étant fermé par suffixe, il vient que $x' \in P$, ce qui contredit le fait que X est préfixe. □

Proposition 3.7.8. *Tout code sémaphore est fin.*

Démonstration. Soit X un code sémaphore. On a alors $X \cap A^*XA^+ = \emptyset$. Ainsi, aucun mot de XA^+ n'est facteur d'aucun mot de X. On conclut donc par la Remarque 2.2.4 que X n'est pas dense. □

Corollaire 3.7.9. *Tout code sémaphore est un code maximal.*

Proposition 3.7.10. *Deux sous-ensembles non vides S et T de A^+ définissent le même code sémaphore si et seulement si $A^*SA^* = A^*TA^*$.*

Pour tout code sémaphore X, il existe un unique ensemble minimal de sémaphores : $T = X \backslash A^+X$.

Démonstration. Soient $X = A^*S\backslash A^*SA^+$ et $Y = A^*T\backslash A^*TA^+$ deux codes sémaphores. Par la Proposition 3.1.3, il vient $XA^* = A^*SA^*$ et $YA^* = A^*TA^*$. Par le Corollaire 3.1.6, on sait que $XA^* = YA^*$ si et seulement si $X = Y$.

Montrons maintenant que $T = X\backslash A^+X$ est un ensemble de sémaphores de X. Soit $X = A^*S\backslash A^*SA^+$ un code sémaphore. En utilisant la Proposition 3.1.3 adaptée aux ensembles suffixes [2], nous avons $A^*T = A^*X$. Il vient alors $A^*TA^* = A^*XA^* = A^*SA^*$, ce qui implique que T et S définissent le même code sémaphore X, i.e. $X = A^*T\backslash A^*TA^+$.

Finalement, montrons que T est bien minimal. Soit $S \subset A^+$ tel que $X = A^*S\backslash A^*SA^+ = A^*T\backslash A^*TA^+$. Soit $t \in T$. Puisque $A^*TA^* = A^*SA^*$, il existe $u, v \in A^*$ et $s \in S$ tels que $t = usv$. De plus, il existe $u', v' \in A^*$ et $t' \in T$ tels que $s = u't'v'$. On a alors $t = uu't'v'v$. Par définition $T \subset X$ donc $v'v = \varepsilon$ puisque $X \cap A^*XA^+ = \emptyset$. De plus, on sait que T est un code suffixe[2], donc $uu' = \varepsilon$. Il vient donc $t = t' = s \in S$. □

Étudions maintenant quelques opérations sur les codes sémaphores.

Proposition 3.7.11. *Si X et Y sont des codes sémaphores, alors XY est un code sémaphore.*

Inversement, si XY est un code sémaphore et si X est un code préfixe, alors X est un code sémaphore.

Démonstration. Supposons que X et Y sont des codes sémaphores. Par le Corollaire 3.6.2, XY est un code préfixe. Vu la Proposition 3.7.3, pour montrer que XY est un code sémaphore, il nous reste à montrer que $A^*XY \subset XYA^*$. Puisque X et Y sont des codes sémaphores, on a $A^*X \subset XA^*$ et $A^*Y \subset YA^*$. Il vient donc

$$A^*XY \subset XA^*Y \subset XYA^*.$$

Inversement, supposons que XY est un code sémaphore et que X est un code préfixe. Pour montrer que X est un code sémaphore, il nous reste à montrer que $A^*X \subset XA^*$.
Soit $w = ux \in A^*X$ avec $u \in A^*$ et $x \in X$. Soit y un mot de Y supposé de longueur minimale. Puisque XY est sémaphore, on a $A^*XY \subset XYA^*$. Il vient alors $wy = uxy = x'y'u'$, pour $x' \in X, y' \in Y$ et $u' \in A^*$. Étant donné que y est de longueur minimale, on a $|y| \leq |y'| \leq |y'u'|$ et par conséquent $|ux| \geq |x'|$, ce qui montre que $ux \in XA^*$. □

Si XY est un code sémaphore, alors Y n'est, quant à lui, pas nécessairement sémaphore, comme nous le montre l'exemple suivant.

Exemple 3.7.12. Soient $X = a^*b$ et $Y = \{a^2, aba, ab^2, b\}$ des ensembles définis sur l'alphabet $A = \{a, b\}$. Comme nous l'avons vu dans l'Exemple 3.7.4, Y est un code préfixe maximal mais n'est pas un code sémaphore. Par contre, l'ensemble X est un code sémaphore. En effet, nous voyons que X est préfixe. De plus, soit $w \in A^*X$. Il existe $u \in A^*$ et $n \in \mathbb{N}$ tels que

$$w = ua^nb.$$

2. Pour tout ensemble $Y \subset A^*$, l'ensemble $X = Y\backslash A^+Y$ est suffixe. De plus, $A^*X = A^*Y$ et X est l'ensemble minimum avec cette propriété.

- Si u commence par a, après avoir lu une première occurrence de b dans w (dans u ou non), nous avons un préfixe de w dans X. On a donc $w \in XA^*$.
- Si u commence par b, cette première lettre est notre préfixe dans X. Ainsi, $w \in XA^*$.
- Si $u = \varepsilon$, on a directement $w \in X \subset XA^*$.

Ainsi, par la Proposition 3.7.3, X est bien un code sémaphore.

L'ensemble $Z = XY$ est un code sémaphore. En effet, grâce au Corollaire 3.6.2, on sait que Z est un ensemble préfixe maximal. De plus, l'ensemble $ZA^- = a^* \cup a^*b \cup a^*ba \cup a^*bab$ est fermé par suffixe. Par la Proposition 3.7.7, l'ensemble Z est bien sémaphore.

Corollaire 3.7.13. *Pour tous $X \subset A^+$ et $n \geq 1$, l'ensemble X est un code sémaphore si et seulement si X^n est un code sémaphore.*

Démonstration. Si X est un code sémaphore, alors il découle directement de la proposition précédente que X^n est un code sémaphore.

Inversement, si X^n est un code sémaphore, alors par le Corollaire 3.6.3 l'ensemble X est un code préfixe maximal. Par la proposition précédente, on en tire que X est un code sémaphore. □

Comme nous l'avons déjà mentionné, tous les résultats s'appliquant aux codes préfixes sont adaptables au cas des codes suffixes. Il est clair que nous pouvons passer d'un code préfixe à un code suffixe au moyen d'une bijection qui associe à un mot son miroir. Cependant, dans le cas des codes sémaphores, nous obtenons le résultat suivant, plus surprenant, nous permettant de passer d'un code sémaphore à son dual, qui est alors suffixe, au moyen d'une bijection associant à un mot l'un de ses conjugués.

Définition 3.7.14. Deux mots x et y sont *conjugués* s'il existe des mots $u, v \in A^*$ tels que $x = uv$ et $y = vu$. On dit que y est un *conjugué* de x.

Théorème 3.7.15. *Soit $S \subset A^+$. Il existe une bijection β de $X = A^*S \backslash A^*SA^+$ dans $Y = SA^* \backslash A^+SA^*$ telle que pour tout $x \in X$, $\beta(x)$ est un conjugué de x.*

Démonstration. Premièrement, considérons l'idéal $J = A^*SA^*$. On a alors

$$X = J \backslash JA^+ \text{ et } Y = J \backslash A^+J.$$

En effet, on a $A^*JA^* = A^*SA^*$ et par la Proposition 3.7.10, il vient $X = A^*J \backslash A^*JA^+$. Puisque $A^*J = J$, il vient que $X = J \backslash JA^+$. On montre que $Y = J \backslash A^+J$ par un raisonnement analogue.

Définissons ensuite, pour chaque $x \in X$,

$$D(x) = \{d \in A^+ | \exists g \in A^* \text{ tel que } x = gd \text{ et } dg \in J\}.$$

L'ensemble $D(x)$ ne contient donc que des suffixes de x. Remarquons que $D(x) \neq \emptyset$ puisque $x \in D(x)$. Ainsi, chaque $D(x)$ contient un plus petit élément.

Nous allons définir l'application β comme suit : pour $x \in X$

$$\beta(x) = dg,$$

où d est le plus petit mot de $D(x)$ et g est tel que $x = gd$. De cette manière $\beta(x)$ est bien un conjugué de x et $\beta(x) \in J$. Montrons alors que $\beta(x) \in J \backslash A^+J = Y$.
Supposons le contraire, supposons qu'il existe $x \in X$ tel que

$$\beta(x) = dg = uj,$$

avec $u \in A^+$ et $j \in J$. Puisque g est un préfixe propre de x, $g \notin J$. En effet, si $g \in J$ alors $x = gd \in JA^+$, ce qui contredit le fait que $x \in X$. Ainsi, $|g| < |j|$. Sinon, si $|j| < |g|$, il existe $v \in A^*$ tel que $g = vj$. Or, J est un idéal donc $vj \in J$, ce qui contredit le fait que $g \notin J$. Donc on a forcément $|d| > |u|$. Il existe donc $d' \in A^+$ tel que $d = ud'$. Cependant, $d' \in D(x)$ puisque $d'gu = ju \in J$ et $gud' = gd = x$. Ceci contredit donc le fait que d est le plus petit mot de $D(x)$. Donc $\beta(x) \in Y$, ce qui montre que l'application β est bien définie.

Considérons maintenant, pour $y \in Y$

$$G(y) = \{e \in A^+ | \exists h \text{ tel que } y = eh \text{ et } he \in J\}.$$

Définissons l'application inverse γ de Y dans X en posant

$$\gamma(y) = he$$

avec $e \in G(y)$ supposé de longueur minimale et h tel que $y = eh$. Si $y = \beta(x) = dg$ et si $\gamma(y) = he$ alors

$$dg = \beta(x) = eh.$$

Puisque $x = gd \in J$, on a $d \in G(y)$. On en tire alors que $|d| \geq |e|$. Mais e n'est pas un préfixe propre de d. Sinon il existe $u \in A^+$ tel que $d = eu$ et $ug = h$. Il vient alors

$$geu = gd = x \text{ et } uge = he \in J,$$

ce qui montre que $u \in D(x)$ et contredit le fait que d est le plus petit mot de $D(x)$. On en déduit alors que $d = e$ et $g = h$. On obtient donc

$$\gamma(\beta(x)) = \gamma(eh) = he = gd = x.$$

Par un raisonnement similaire, nous obtenons

$$\beta(\gamma(y)) = \beta(he) = \beta(gd) = dg = y.$$

Ainsi, les applications β et γ sont bien inverses l'une de l'autre, ce qui montre que β est une bijection. □

Exemple 3.7.16. Illustrons la construction de la preuve précédente. Soient $A = \{a, b\}$ et $S = \{a^2, ba, b^2\}$ un ensemble de sémaphores. On a

$$X = A^*S \backslash A^*SA^+ = \{a^2, ba, b^2, aba, ab^2\}$$

et

$$Y = SA^* \backslash A^+SA^* = \{a^2, a^2b, ba, bab, b^2\}.$$

Le tableau suivant nous fournit, pour chaque mot x de X, ses suffixes, l'ensemble $D(x)$ et son image par β.

x	Suffixes $\neq \varepsilon$ de x	$D(x)$	d	g	$\beta(x)$
aa	a, aa	a, aa	a	a	aa
ba	a, ba	ba	ba	ε	ba
bb	b, bb	b, bb	b	b	bb
aba	a, ba, aba	a, ba, aba	a	ab	aab
abb	b, bb, abb	b, bb, abb	b	ab	bab

On constate effectivement que les mots de Y sont bien les images par β des mots de X.

Les codes sémaphores ont la propriété d'être en bijection avec un code suffixe par une application qui associe à un mot l'un de ses conjugués. Ceci n'est pas vrai pour les codes préfixes quelconques, comme nous le montre l'exemple suivant.

Exemple 3.7.17. Considérons l'ensemble $X = \{ab, ba, c, ac, bca\}$ sur l'alphabet $A = \{a, b, c\}$. Il est évident que c'est un code préfixe. Supposons qu'il existe une bijection β de X dans un code suffixe Y qui à x associe un de ses conjugués.

x	Conjugués
ab	ab , ba
ba	ba, ab
c	c
ac	ac, ca
bca	bca, cab, abc

Il est clair, vu le tableau ci-dessus, que Y contient forcément c, ab et ba. De plus, il doit contenir ca. Si ce n'est pas le cas, il contient alors ac et c contredisant le fait qu'il est suffixe. Cependant, les conjugués de bca ont des suffixes égaux à c, ab, ca, ce qui montre qu'une telle bijection β ne peut exister.

Bibliographie

[1] J. Berstel, D. Perrin, J. F. Perrot, et A. Restivo, *Sur le théorème du défaut*, J. Algebra **60** (1979), 169–180.

[2] Jean Berstel et Dominique Perrin, *Theory of codes*, 2002, disponible via l'URL `<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.851&rep=rep1&type=pdf>`.

[3] Jean Berstel, Dominique Perrin, et Christophe Reutenauer, *Codes and automata*, vol. 129, Cambridge : Cambridge University Press, 2010.

[4] Jean Berstel et Christophe Reutenauer, *Noncommutative rational series with applications*, vol. 137, Cambridge : Cambridge University Press, 2011.

[5] M. Lothaire, *Algebraic combinatorics on words*, vol. 90, Cambridge : Cambridge University Press, 2002.

[6] Michel Rigo, *Automates et langages formels*, 2009–2010, disponible via l'URL `<http://www.discmath.ulg.ac.be/cours/main_autom.pdf>`.