Mots dendriques et leurs propriétés

Auteur : Gheeraert, France
Promoteur(s) : Leroy, Julien
Faculté : Faculté des Sciences
Diplôme : Master en sciences mathématiques, à finalité spécialisée en informatique
Année académique : 2019-2020
URI/URL : http://hdl.handle.net/2268.2/9164

Faculté des Sciences
Département de Mathématique

Mots dendriques et leurs propriétés

Mémoire présenté en vue de l'obtention du grade de
Master en Sciences Mathématiques à finalité informatique

France Gheeraert

Promoteur : Julien Leroy

Année 2019-2020

Remerciements

Je tiens avant tout à adresser un merci tout particulier à mon promoteur Julien Leroy pour sa disponibilité, ses relectures attentives et ses nombreux conseils et remarques.

J'aimerais aussi remercier ceux qui, depuis ma plus tendre enfance, m'ont fait découvrir et apprécier les mathématiques ; ceux qui, lors de mes dernières années de secondaire, me les ont fait aimer un peu plus chaque jour et m'ont ainsi encouragée à faire ce choix d'étude ; et ceux qui, au cours de ces cinq dernières années, ont partagé avec moi l'amour des mathématiques et m'ont transmis ne fut-ce qu'une infime partie de leur savoir. J'espère que ce travail pourra leur faire honneur.

Table des matières

Introduction

La combinatoire des mots est une branche des mathématiques discrètes qui apparaît au début du XX$^{\text{ème}}$ avec les travaux d'Axel Thue (1863 – 1922). Elle étudie diverses propriétés de suites (finies ou infinies) sur un ensemble fini (*l'alphabet*). Ces suites sont appelées *mots*. Leur étude a des applications dans de nombreux domaines comme, par exemple, l'informatique théorique, l'algèbre et la théorie des nombres.

Un concept fort étudié en particulier est la notion de complexité factorielle : la *fonction de complexité* d'un mot infini u est la fonction p qui à tout naturel n associe le nombre de facteurs de longueur n différents apparaissant dans u. Les applications de la fonction de complexité ne se limitent pas à la combinatoire des mots puisque Boris Adamczewski et Yann Bugeaud [1] ont notamment montré que dans le cas d'une représentation en base entière d'un nombre irrationnel algébrique, on a la limite suivante

$$\liminf_{n\to+\infty} \frac{p(n)}{n} = +\infty.$$

Parmi tous les mots, on peut en distinguer certains ayant des propriétés particulières. Les plus connus d'entre eux sont probablement les mots *sturmiens* qui sont les mots les plus simples (en terme de complexité), tout en n'étant pas ultimement périodiques. Leur fonction de complexité est donnée par $p(n) = n + 1$. En particulier, ces mots sont construits sur un alphabet de deux lettres.

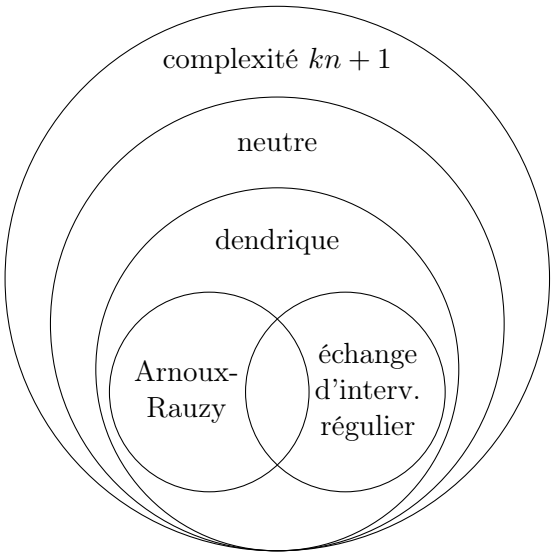
Un des instincts les plus forts du mathématicien est de généraliser, c'est donc tout naturellement que Pierre Arnoux et Gérard Rauzy ont introduits dans [2] les mots actuellement qualifiés *d'épisturmiens stricts* ou *d'Arnoux-Rauzy* qui généralisent la notion de mots sturmiens pour un alphabet de taille quelconque. Ces mots sont définis en considérant les extensions à droite et à gauche de leurs facteurs par des lettres. Plus précisément, si $L(w)$ désigne l'ensemble des lettres a telles que aw soit un facteur du mot infini considéré et si $R(w)$ est défini symétriquement, alors les mots épisturmiens stricts sont les mots infinis tels que, pour toute longueur n, il y ait exactement un facteur w et un facteur w' de longueur n tels que $|L(w)| > 1$ et $|R(w')| > 1$ et dans ce cas w (resp. w') peut être étendu à gauche (resp. à droite) par toutes les lettres de l'alphabet.

Plus récemment, une généralisation des mots épisturmiens stricts a commencée à être étudiée. Il s'agit des *mots dendriques* (appelés à l'origine *tree words*) introduits en 2015 dans [5]. Ils sont toujours définis à partir des extensions à droites et à gauche de leurs facteurs, mais plutôt que de s'intéresser au nombre d'extensions, on étudie le *graphe d'extension* d'un facteur w qui a pour ensemble de sommets l'union disjointe de $L(w)$ et de $R(w)$ et tel que une arrête relie $a \in L(w)$ à $b \in R(w)$ si awb est un facteur du mot infini qu'on considère. Un mot infini est dendrique si, pour tout facteur w, le graphe d'extension de w est un arbre.

Depuis, ces mots ont à nouveau été généralisés dans [11] en restreignant la contrainte aux facteurs suffisamment longs. Ce mots sont appelés *ultimement dendriques* mais nous ne les considérerons pas dans ce travail.

Les mots dendriques conservent de nombreuses propriétés des mots épisturmiens stricts ou, du moins, les résultats sont facilement adaptables. Ce travail s'intéresse à ces propriétés en se basant sur la série d'articles [5, 7, 8] écrits par le même groupe d'auteurs : Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer et Giuseppina Rindone. Plutôt que de travailler avec des mots infinis, nous manipulerons des ensembles de mots finis, tout comme les auteurs l'ont fait dans un premier temps. Ces ensembles correspondront aux ensembles de facteurs, aussi appelés *langages*, de mots infinis. Voici plus précisément la structure de ce travail.

Nous commençons par introduire les concepts de base de combinatoire des mots dans le Chapitre 1 et notamment la notion d'ensemble *factoriel*, i.e. contenant les facteurs de ses éléments, ainsi que les ensembles $L(w)$ et $R(w)$. Nous nous intéressons ensuite à la hiérarchie entourant les mots dendriques. Plus précisément, nous prouvons le diagramme suivant :



où k désigne la taille de l'alphabet moins 1. Un ensemble est *neutre* si, pour chacun de ses mots w, le nombre de couples (a, b) tels que awb soit dans l'ensemble plus 1 est égal à la somme des cardinaux de $L(w)$ et $R(w)$.

Dans le Chapitre 2, nous nous intéressons aux *codes préfixes*, *suffixes* et *bifixes* qui sont inclus dans un ensemble neutre. Nous obtenons deux résultats concernant leur cardinalité. Un premier (Proposition 2.2.3) s'intéressant au nombre d'extensions à droite (resp. à gauche) des préfixes (resp. suffixes) propres d'un code préfixe (resp. suffixe) maximal et un second (Théorème 2.5.10) utilisant la notion de *S-degré*. Parmi ces codes, nous nous intéressons plus particulièrement aux ensembles de *mots de retour* et prouvons notamment que, dans un ensemble neutre récurrent, tout mot possède exactement autant de mots de premier retour qu'il n'y a de lettres dans l'alphabet (Théorème 2.4.2). Ceci permet d'affirmer que tout ensemble neutre récurrent est uniformément récurrent.

Le Chapitre 3 introduit une nouvelle notion qui sera centrale pour tous les chapitres suivants : le *groupe libre*. Grâce aux graphes de Rauzy et aux langages acceptés par des automates particuliers, nous montrons que l'ensemble des mots de premier retour dans un ensemble dendrique récurrent engendre le groupe libre (Théorème 3.5.1). Ce théorème, associé au résultat obtenu au chapitre précédent permet de prouver un des résultats majeurs de ce travail appelé « Théorème de retour » et affirmant que, dans un ensemble dendrique récurrent, l'ensemble des mots de premier retour de n'importe quel mot forme une base du groupe libre.

Nous introduisons ensuite une généralisation des graphes d'extensions dans le Chapitre 4. Plutôt que de se restreindre à des extensions par des lettres, nous considérons des extensions par des mots dans des ensembles fixés. Nous montrons que, si les ensembles choisis vérifient certaines conditions, ces graphes d'extensions généralisés dans un ensemble dendrique sont également des arbres (Théorème 4.1.7). Nous montrons aussi une propriété similaire si l'ensemble considéré est acyclique uniquement (Théorème 4.1.8). Enfin, nous nous intéressons au *décodage bifixe* d'un ensemble qui est l'image inverse de cet ensemble par un morphisme particulier et montrons que cette opération conserve le caractère dendrique (Théorème 4.2.5).

Dans le Chapitre 5, nous prouvons deux propriétés qu'ont les codes bifixes inclus dans un ensemble acyclique. La première, appelée Théorème d'indépendance (Théorème 5.2.6), affirme que ce sont des parties libres. Il s'agit même d'une caractérisation des ensembles acycliques. La seconde (Théorème 5.4.9) dit que le monoïde engendré par un code bifixe est saturé. Ces deux résultats utilisent divers graphes et automates.

Le Chapitre 6 est centré autour de la propriété de base d'indice fini qui lie l'indice du sous-groupe engendré par un code bifixe au S-degré de ce même code bifixe. Tout ensemble dendrique récurrent a cette propriété (Théorème 6.2.5) mais la réciproque est également vraie. Nous nous intéressons ensuite à la stabilité de la famille des ensembles dendriques récurrents pour deux opérations : le passage à l'ensemble *dérivé* (obtenu comme image inverse des mots de retour) et le décodage bifixe (commencé au Chapitre 4).

Enfin, nous abordons le sujet des *représentations $\mathcal{S}$-adiques* dans le Chapitre 7. Certains mots infinis peuvent être vus comme images d'une itération infinie d'un morphisme particulier. Les représentations $\mathcal{S}$-adiques généralisent ce concepts en autorisant des morphismes différents lors de chaque itération. Nous étudions des automorphismes particuliers appelés *tame* et montrons que tout mot dendrique récurrent possède une représentation $\mathcal{S}$-adique par des automorphismes tame (Proposition 7.3.4).

Chapitre 1

Définitions et premiers résultats

Ce chapitre donne une première approche des mots dendriques. Il a pour objectif d'introduire les concepts principaux qui seront réutilisés dans les chapitres suivants mais également, et peut-être surtout, de montrer quelques propriétés que possèdent les mots dendriques (ou, plus généralement, les ensembles dendriques) qui, à elles seules justifient l'étude de ces mots particuliers. Dans un premier temps, nous rappelons certaines définitions générales de combinatoires des mots en se concentrant particulièrement sur la notion de prolongeable. Nous définissons ensuite le principal sujet de ce mémoire : les mots dendriques, et montrons que les ensembles dendriques sont un cas particulier d'ensembles neutres. Nous faisons un rapide passage par la notion de récurrence pour justifier le fait que nous considérerons majoritairement les ensembles dendriques plutôt que les mots dendriques. Nous nous intéressons ensuite à deux familles de mots qui sont des cas particuliers de mots dendriques : les mots épisturmiens stricts (plus communément appelés mots d'Arnoux-Rauzy) et les codages d'échanges d'intervalles réguliers. Enfin, nous montrons que la complexité des ensembles neutres (et donc également celle des ensembles dendriques) est $(k-1)n+1$ où k est le nombre de lettres de l'alphabet.

1.1 Définitions générales

Commençons par introduire les notions élémentaires de combinatoire des mots. Un *alphabet* A est un ensemble fini [(1)] de symboles, appelés *lettres*, et un *mot* sur A est une suite finie ou non d'éléments de A. On note A^* l'ensemble des mots finis sur A et $A^{\mathbb{N}}$ l'ensemble des mots infinis sur ce même alphabet. [(2)]

La *longueur* d'un mot $w \in A^*$, notée $|w|$, est la longueur de la suite correspondante, i.e le nombre de lettres composant w. Le *mot vide*, noté ε, est l'unique mot de longueur 0. La notation A^+ correspond à l'ensemble des mots finis non vides, tandis que A^n est l'ensemble des mots de longueur $n \in \mathbb{N}$. Nous noterons également $A^{\leq n}$ l'ensemble des mots de longueur au plus n.

On munit A^* de l'opération de concaténation. Elle fait de A^* un monoïde, i.e elle est

(1). Dans les Sections 4.2 et 6.3, nous relâchons cette contrainte en manipulant momentanément des alphabets dénombrables.

(2). On adopte des conventions différentes pour la numérotation des lettres suivant le type de mot : s'il est fini, les lettres sont numérotées de 1 à $|w|$ alors que s'il est infini, on commence à 0. Nous utiliserons également la notation $w_{[i,j]}$ pour désigner le mot $w_i \dots w_j$ composé des lettres d'indice i à j dans w.

interne, associative et admet pour neutre ε. Cette opération est représentée en utilisant des notations multiplicatives, autrement dit, la concaténation des mots u et v sera le mot uv et $u^2 = uu$. Cette opération de concaténation peut s'étendre sur $A^* \times A^{\mathbb{N}}$.

Un mot fini x est *facteur* d'un mot w s'il existe un mot fini u et un mot v tels que

$$w = uxv.$$

Si u est vide, on dira que x est un *préfixe* de w. Si w est fini et v est vide, alors x est un *suffixe* de w. Si $x \neq w$, on ajoute le qualificatif *propre*. On note $\mathrm{Fac}(w)$ (resp. $\mathrm{Pref}(w)$, resp. $\mathrm{Suff}(w)$) l'ensemble des facteurs (resp. préfixes, resp. suffixes) de w. L'ensemble des *facteurs internes* d'un mot fini w est

$$\mathrm{IFac}(w) = \{x \in A^* \mid \exists y, z \in A^+ \text{ tq. } w = yxz\}.$$

Nous utiliserons également ces notations dans leurs versions ensemblistes, i.e. si X est un ensemble de mots, $\mathrm{Pref}(X) = \{u \in \mathrm{Pref}(w) \mid w \in X\}$ par exemple.

Dans la suite, nous supposerons que tous les mots sont finis et sur l'alphabet A, sauf mention contraire. De plus, quand nous manipulerons des ensembles de mots, nous supposerons que A est l'alphabet minimal, i.e. que toutes les lettres de A apparaissent dans au moins un mot de l'ensemble.

Définition 1.1.1. Un ensemble de mots est *factoriel* s'il contient tous les facteurs de ses mots.

Remarquons que l'ensemble des facteurs d'un mot est toujours factoriel.

Définition 1.1.2. Pour tout ensemble de mots S et tout mot w, on note

$$w^{-1}S = \{u \in A^* \mid wu \in S\} \quad \text{et} \quad Sw^{-1} = \{u \in A^* \mid uw \in S\}.$$

Il est aisé de voir que, si S est factoriel, le premier ensemble est stable pour les préfixes, i.e. $\mathrm{Pref}(w^{-1}S) \subseteq w^{-1}S$, et que le second est stable pour les suffixes. En particulier, si S est factoriel et si $w \notin S$, ces deux ensembles sont vides.

Passons à présent à des notions plus spécifiques à l'étude des mots dendriques.

Définition 1.1.3. Soient S un ensemble de mots et $w \in S$. On définit

$$\begin{aligned} L_S(w) &= \{a \in A \mid aw \in S\}, \\ R_S(w) &= \{a \in A \mid wa \in S\}, \\ E_S(w) &= \{(a, b) \in A \times A \mid awb \in S\} \end{aligned}$$

ainsi que

$$l_S(w) = |L_S(w)|, \quad r_S(w) = |R_S(w)| \quad \text{et} \quad e_S(w) = |E_S(w)|.$$

Pour alléger les notations, nous nous permettrons de laisser sous-entendre l'indice S quand le contexte le permet.

Exemple 1.1.4. Si $S = \{b, ab, bc\}$, alors on a, entre autres,

$$L(\varepsilon) = \{b\}, \quad R(\varepsilon) = \{b\} \quad \text{et} \quad E(\varepsilon) = \{(a, b), (b, c)\},$$

et

$$L(b) = \{a\} \quad R(b) = \{c\} \quad \text{et} \quad E(b) = \emptyset.$$

On constate notamment que, a priori, on ne peut rien dire de $E(w)$ à partir de $L(w)$ et $R(w)$. Cependant, si S est factoriel, alors

$$E(w) \subseteq L(w) \times R(w).$$

C'est une des raisons pour lesquelles tous les ensembles que nous manipulerons dans ce travail seront factoriels.

Définition 1.1.5. Soit S un ensemble non vide de mots. Un mot $w \in S$ est *prolongeable à droite* (resp. *à gauche*, resp. *biprolongeable*) si $r(w) > 0$ (resp. $l(w) > 0$, resp. $e(w) > 0$). L'ensemble S est *prolongeable à droite* (resp. *à gauche*, resp. *biprolongeable*) si tous ses mots le sont et si S est factoriel.

En particulier, tout ensemble prolongeable à droite, à gauche ou biprolongeable est infini.

Remarquons que si un mot w est prolongeable à gauche et à droite, alors w n'est pas forcément biprolongeable, ce que montre l'exemple ci-dessus pour $w = b$ et ce même si on considère l'ensemble factoriel $\mathrm{Fac}(S)$ plutôt que S. Cependant, on a le résultat suivant.

Proposition 1.1.6. *Un ensemble S est prolongeable à gauche et à droite si, et seulement si, il est biprolongeable.*

Démonstration. Si S est prolongeable à gauche et à droite et si $w \in S$, montrons que w est biprolongeable. Comme S est prolongeable à gauche, il existe $a \in A$ tel que $aw \in S$. À nouveau, S est prolongeable à droite donc il existe $b \in A$ tel que $awb \in S$, d'où la conclusion.

Si S est biprolongeable et si $w \in S$, alors il existe $a, b \in A$ tels que $awb \in S$. Comme S est factoriel, on a $aw, wb \in S$ donc w est prolongeable à gauche et à droite. □

Remarque 1.1.7. L'ensemble des facteurs d'un mot infini u est prolongeable à droite. Par contre, si w est un préfixe de u qui n'a pas d'autre occurrence dans u, alors w n'est pas prolongeable à gauche donc $\mathrm{Fac}(u)$ non plus. Tous les autres facteurs de u sont biprolongeables.

Définition 1.1.8. Si S est un ensemble de mots, alors un mot $w \in S$ est *spécial à droite* (resp. *à gauche*) si $r(w) \geq 2$ (resp. $l(w) \geq 2$) et il est *bispécial* s'il est spécial à droite et à gauche.

La proposition suivante est immédiate.

Proposition 1.1.9. *Soient S un ensemble factoriel et $w \in S$.*

1. *Si w est prolongeable à droite (resp. spécial à droite), alors tous ses suffixes le sont aussi.*
2. *Si w est prolongeable à gauche (resp. spécial à gauche), alors tous ses préfixes le sont aussi.*

Démonstration. Ne montrons que le premier point, le second se prouve similairement. Si v est un suffixe de w, alors

$$\begin{aligned} R(w) &= \{a \in A \mid wa \in S\} \\ &\subseteq \{a \in A \mid va \in S\} \\ &= R(v) \end{aligned}$$

car S est factoriel. Donc $r(v) \geq r(w)$, ce qui permet de conclure. □

Définition 1.1.10. Soit S un ensemble de mots. Un mot $w \in S$ est *neutre* (resp. *faible*, resp. *fort*) si

$$m(w) := e(w) - l(w) - r(w) + 1 = 0$$

(resp. $m(w) < 0$, resp. $m(w) > 0$). L'ensemble S est *neutre* (resp. *faible*, resp. *fort*) si tous ses mots sont neutres (resp. faibles ou neutres, resp. forts ou neutres) et si S est factoriel.

1.2 Graphes d'extensions et mots dendriques

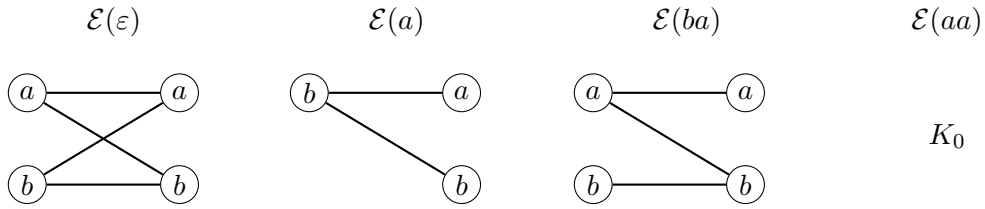
Dans cette section, nous cherchons à définir les mots ou ensembles dendriques. Nous faisons également le lien avec les ensembles neutres. Pour cela, nous considérons un ensemble S factoriel et un mot $w \in S$.

Définition 1.2.1. Le *graphe d'extensions* de w est le graphe biparti $\mathcal{E}_S(w) = (V_S(w), E_S(w))$ tel que [(3)]
- $V_S(w)$ est l'union disjointe de $L(w)$ et $R(w)$,
- $E_S(w)$ est l'ensemble des arêtes (a, b) où $a \in L(w)$, $b \in R(w)$ et $(a, b) \in E(w)$.

Remarquons que, comme nous ne travaillons qu'avec des ensembles factoriels, l'ensemble des arêtes du graphe d'extensions est exactement l'ensemble $E_S(w)$ de la Définition 1.1.3. C'est la raison pour laquelle nous nous permettons ces notations identiques.

Dans la suite, nous favoriserons $\mathcal{E}(w), V(w)$ et $E(w)$ aux notations $\mathcal{E}_S(w), V_S(w)$ et $E_S(w)$ lorsque cela ne prête pas à confusion.

Exemple 1.2.2. Si S est l'ensemble des facteurs du mot $abbbababaa$, alors $S \cap A^2 = \{aa, ab, ba, bb\}$, $S \cap A^3 = \{aba, abb, baa, bab, bba, bbb\}$ et $S \cap A^4 = \{abaa, abab, abbb, baba, bbab, bbba\}$ donc on a les graphes d'extensions suivants [(4)] :



Définition 1.2.3. Un ensemble S de mots finis est *dendrique* s'il est biprolongeable et si, pour tout $w \in S$, le graphe $\mathcal{E}(w)$ est un arbre.

Proposition 1.2.4. *Un ensemble biprolongeable S est dendrique si, et seulement si, $\mathcal{E}(w)$ est un arbre pour tout $w \in S$ bispécial.*

(3). Par union disjointe, on entend que si $a \in L(w) \cap R(w)$, alors deux sommets auront a pour étiquette. Plus rigoureusement, $V_S(w)$ est l'union de $L(w)$ et $\tilde{R}(w)$ où

$$\tilde{R}(w) = \{\tilde{a} \mid a \in R(w)\}.$$

Dans ce cas,

$$E_S(w) = \{(a, \tilde{b}) \in L(w) \times \tilde{R}(w) \mid (a, b) \in E(w)\}.$$

(4). Le symbole K_0 représente le graphe vide (à 0 sommets).

Démonstration. La condition nécessaire est immédiate. Pour la réciproque, montrons que $\mathcal{E}(w)$ est obligatoirement un arbre si w n'est pas bispécial. Comme S est biprolongeable, $L(w)$ et $R(w)$ ne sont pas vides. De plus, w n'est pas bispécial donc un de ces deux ensembles contient exactement un élément. L'autre cas se traitant similairement, supposons que $L(w) = \{a\}$. Comme $E(w) \subseteq \{a\} \times R(w)$, $\mathcal{E}(w)$ est acyclique. Montrons qu'il est connexe. Pour tout $b \in R(w)$, $wb \in S$ donc, comme S est biprolongeable, il existe $c \in A$ tel que $cwb \in S$. Or, S est également factoriel (par définition de biprolongeable) donc $cw \in S$, ce qui implique que $c = a$ par hypothèse. On a donc $(a, b) \in E(w)$ pour tout $b \in R(w)$ et $\mathcal{E}(w)$ est connexe. □

Définition 1.2.5. Un mot infini $u \in A^{\mathbb{N}}$ est *dendrique* si l'ensemble de ses facteurs est dendrique.

Montrons à présent le lien entre le caractère dendrique et le caractère neutre. Commençons par rappeler une propriété concernant les graphes.

Proposition 1.2.6. *Pour tout graphe $G = (V, E)$ ayant c composantes connexes,*

$$|E| \geq |V| - c.$$

L'égalité a lieu si, et seulement si, G est une forêt composée de c arbres disjoints.

Démonstration. Il suffit de montrer le résultat dans le cas où $c = 1$ car le cas général se déduit via une simple addition d'(in)égalités. Supposons donc G connexe et procédons par récurrence sur le nombre d'arêtes.

- Si $|E| = 0$, alors le graphe a exactement un sommet. De plus, il s'agit toujours d'un arbre donc le cas de base est démontré.
- Supposons la propriété vraie pour tout graphe connexe possédant au plus n arêtes et montrons-la pour G qui a $n + 1$ arêtes.
  Si G n'est pas un arbre, alors il contient un cycle. Considérons le graphe $G' = (V', E')$ obtenu en retirant une des arêtes de ce cycle. Il est encore connexe et contient n arêtes donc, par hypothèse de récurrence,

  $$|E| = |E'| + 1 > |V'| - 1 = |V| - 1.$$

  Si G est un arbre, montrons l'égalité. Considérons le graphe $G' = (V', E')$ obtenu en supprimant une arête $e = (a, b)$ de G. Le graphe G' possède exactement deux composantes connexes $G_1 = (V_1, E_1)$ et $G_2 = (V_2, E_2)$, l'une contenant a et l'autre contenant b. De fait, G est connexe donc G' a au plus deux composantes connexes et si G' était toujours connexe, alors G posséderait un cycle passant par a et b, ce qui est absurde. Comme G_1 et G_2 sont des arbres ayant n arêtes ou moins, par hypothèse de récurrence,

  $$|E| = 1 + |E_1| + |E_2| = |V_1| + |V_2| - 1 = |V| - 1.$$

□

Proposition 1.2.7. *Si S est un ensemble dendrique, alors il est neutre. Réciproquement, si un ensemble biprolongeable S est neutre et si, pour tout $w \in S$, $\mathcal{E}(w)$ est acyclique ou connexe, alors S est dendrique.*

Démonstration. Pour tout $w \in S$, on a, par définition du graphe d'extensions de w,

$$|E(w)| = e(w) \quad \text{et} \quad |V(w)| = l(w) + r(w).$$

Comme $\mathcal{E}(w)$ est un arbre, par la proposition précédente, on a

$$e(w) = l(w) + r(w) - 1,$$

ce qui revient à dire que w est neutre. On peut conclure car, par définition, S est biprolongeable donc factoriel.

Supposons à présent S neutre. Si $\mathcal{E}(w)$ est acyclique, son nombre de composantes connexes est donné par

$$l(w) + r(w) - e(w) = 1$$

donc $\mathcal{E}(w)$ est connexe. Réciproquement, si $\mathcal{E}(w)$ est connexe, alors

$$e(w) \geq l(w) + r(w) - 1.$$

Comme S est neutre, on a l'égalité donc, par la proposition précédente, $\mathcal{E}(w)$ est acyclique. Dans les deux cas, $\mathcal{E}(w)$ est un arbre donc S est dendrique. □

1.3 Notions de récurrence

Par définition, à tout mot dendrique correspond un ensemble dendrique. La notion d'ensemble dendrique est donc, a priori, plus générale que celle de mot dendrique. Dans cette section, nous montrons que dans le cas d'ensembles/mots récurrents, les deux concepts sont équivalents, ce qui signifie que, dans la suite, tout résultat parlant d'ensembles dendriques récurrents admet une formulation semblable avec des mots dendriques récurrents.

1.3.1 Ensembles récurrents

Dans ce travail, nous étudions et utilisons les notions de récurrence (simple) et de récurrence uniforme. Leurs définitions sont données ci-dessous.

Définition 1.3.1. Un ensemble S non réduit à $\{\varepsilon\}$ est *récurrent* s'il est factoriel et si, pour tous $x, y \in S$, il existe $z \in S$ tel que $xzy \in S$.

En particulier, un ensemble récurrent S est biprolongeable. De fait, comme S n'est pas réduit à $\{\varepsilon\}$, il existe $x \in A^+ \cap S$. Pour tout $w \in S$, il existe alors des mots u et v tels que $xuw \in S$ et $xuwvx \in S$. On en déduit que w est biprolongeable car S est factoriel.

Définition 1.3.2. Un ensemble S est *uniformément récurrent* s'il est prolongeable à droite et si, pour tout $u \in S$, il existe $n \in \mathbb{N}$ tel que u soit facteur de tous les mots de $A^n \cap S$.

L'hypothèse prolongeable à droite permet d'écarter les ensembles finis.

Proposition 1.3.3. *Un ensemble uniformément récurrent est récurrent.*

Démonstration. Soient S un ensemble uniformément récurrent et $x, y \in S$. Vu la définition d'un ensemble prolongeable à droite, S est factoriel. Par définition, il existe $n \in \mathbb{N}$ tel que y soit facteur de tous les mots de $A^n \cap S$. Comme S est prolongeable à droite, il existe un mot $w \in A^n \cap S$ tel que $xw \in S$. Il suffit alors de prendre $z \in S$ tel que zy soit préfixe de w. Un tel z existe car y est facteur de w. □

1.3.2 Mots récurrents

Étudions à présent ces propriétés du point de vue des mots.

Définition 1.3.4. Un mot infini $u \in A^{\mathbb{N}}$ est *récurrent* si tous ses facteurs apparaissent une infinité de fois.

Proposition 1.3.5. *Soit $u \in A^{\mathbb{N}}$. Les assertions suivantes sont équivalentes :*

1. *u est récurrent,*
2. *$\mathrm{Fac}(u)$ est récurrent,*
3. *pour tout $w \in \mathrm{Fac}(u)$, il existe $v \in \mathrm{Fac}(u)$ tel que $wvw \in \mathrm{Fac}(u)$.*

Démonstration. Supposons u récurrent. Soient $x, y \in \mathrm{Fac}(u)$. Considérons une occurrence fixée de x dans u (la première, par exemple). Comme y apparaît une infinité de fois dans u, il apparaît au moins une fois après cette occurrence donc il existe z tel que $xzy \in \mathrm{Fac}(u)$, ce qui signifie que $\mathrm{Fac}(u)$ est récurrent.

Si $\mathrm{Fac}(u)$ est récurrent, la troisième assertion est évidente.

Supposons que, pour tout $w \in \mathrm{Fac}(u)$, il existe $v \in \mathrm{Fac}(u)$ tel que $wvw \in \mathrm{Fac}(u)$ et montrons que tout $x \in \mathrm{Fac}(u)$ apparaît une infinité de fois. Par l'absurde, s'il n'apparaît qu'un nombre fini de fois, notons y le préfixe de u précédant sa dernière occurrence. Autrement dit, yx est un préfixe de u contenant toutes les occurrences de x dans u. Par hypothèse, $yx \in \mathrm{Fac}(u)$ donc il existe z tel que $yxzyx \in \mathrm{Fac}(u)$, ce qui est absurde. □

Nous rappelons ici brièvement la notion de convergence pour les mots mais ne rentrons pas dans les détails.

Définition 1.3.6. Une suite de mots finis $\left(u^{(n)}\right)_{n\in\mathbb{N}}$ *converge* vers un mot infini si

$$\lim_{n\to\infty} |u^{(n)}| = +\infty$$

et si la suite de mots infinis $(u^{(n)}a^{\omega})_{n\in\mathbb{N}}$, où a^{ω} désigne la lettre $a \in A$ répétée un nombre infini de fois, converge dans $A^{\mathbb{N}}$ muni de la distance

$$d(u,v) = 2^{-\min\{i\in\mathbb{N} | u_i \neq v_i\}}$$

si $u \neq v$ et $d(u,v) = 0$ sinon. Dans ce cas,

$$\lim_{n\to\infty} u^{(n)} = \lim_{n\to\infty} u^{(n)}a^{\omega}.$$

Autrement dit, la suite $\left(u^{(n)}\right)_{n\in\mathbb{N}}$ converge si, pour tout $k \in \mathbb{N}$, il existe $v^{(k)} \in A^k$ et $N \in \mathbb{N}$ tels que, pour tout $n \geq N$, $v^{(k)}$ soit un préfixe de $u^{(n)}$. La limite est alors le mot infini u tel que $v^{(k)} \in \mathrm{Pref}(u)$ pour tout $k \in \mathbb{N}$.

Le résultat suivant montre bien que les notions d'ensemble dendrique récurrent ou de mot dendrique récurrent sont équivalentes.

Proposition 1.3.7. *Tout ensemble récurrent est l'ensemble des facteurs d'un mot infini récurrent.*

Démonstration. Soit S un ensemble factoriel récurrent sur l'alphabet A. Supposons disposer d'un ordre sur A qu'on étend en un ordre lexicographique [(5)] sur A^*. On peut alors noter les mots de S $w_0, w_1, w_2, \ldots$ [(6)] selon cet ordre lexicographique. On aura entre autres $w_0 = \varepsilon$ et $|w_1| = 1$. Construisons par induction une suite de mots $(u_i)_{i\in\mathbb{N}}$ telle que

$$u_0 = w_0 \quad \text{et} \quad u_{i+1} = u_i v_i w_{i+1}$$

où v_i est tel que $u_i v_i w_{i+1} \in S$. Un tel v_i existe car $u_i \in S$ par induction et S est récurrent. On a alors

$$|u_{i+1}| > |u_i| \quad \text{et} \quad u_i \in \mathrm{Pref}(u_{i+1})$$

donc la suite $(u_i)_{i\in\mathbb{N}}$ converge vers un mot infini $u \in A^{\mathbb{N}}$.

Par construction, $S \subseteq \mathrm{Fac}(u)$ car $w_n \in \mathrm{Fac}(u_n)$ et $u_n \in \mathrm{Pref}(u)$ pour tout $n \in \mathbb{N}$. Réciproquement, pour tout $w \in \mathrm{Fac}(u)$, il existe $n \in \mathbb{N}$ tel que $w \in \mathrm{Fac}(u_n)$ donc, comme $u_n \in S$ et que S est factoriel, $w \in S$. On a donc bien $S = \mathrm{Fac}(u)$ et le mot u est récurrent par la Proposition 1.3.5. □

Définition 1.3.8. Un mot infini $u \in A^{\mathbb{N}}$ est *uniformément récurrent* si l'ensemble de ses facteurs l'est.

Tout ensemble uniformément récurrent correspond à l'ensemble des facteurs d'un mot infini uniformément récurrent. En effet, un ensemble uniformément récurrent est récurrent donc par la proposition précédente, il correspond aux facteurs d'un mot infini et ce mot est alors uniformément récurrent.

1.4 Mots épisturmiens stricts

Les mots dendriques sont une généralisation d'une famille bien connue de mots : les mots épisturmiens stricts aussi appelés mots d'Arnoux-Rauzy qui est elle-même une généralisation des très célèbres mots sturmiens. Cette section a pour but de définir ces mots et de montrer qu'ils sont effectivement dendriques.

Définition 1.4.1. L'application *miroir* est l'application

$$\cdot^R : A^* \to A^*, \quad w \mapsto w^R$$

telle que $\varepsilon^R = \varepsilon$ et

$$(w_1 \ldots w_n)^R = w_n \ldots w_1$$

pour tous $w_1, \ldots, w_n \in A$.

Définition 1.4.2. Un mot $u \in A^{\mathbb{N}}$ est *épisturmien* si l'ensemble de ses facteurs est stable pour l'application miroir et contient, pour tout $n \in \mathbb{N}$, au plus un mot spécial à droite de longueur n.

(5). L'ordre lexicographique est l'ordre du dictionnaire. Autrement dit, un mot u est plus petit qu'un mot v si u est un préfixe de v ou si, lorsqu'on regarde le premier indice i pour lequel u et v ont des lettres différentes, u_i est plus petit que v_i.

(6). Pour alléger les notations, w_i désigne ici un mot et non la $i^{\text{ème}}$ lettre de w. Nous prenons également cette convention pour u_i et v_i.

Remarquons que la condition sur la stabilité de $\mathrm{Fac}(u)$ pour l'application miroir implique que

$$R(w) = L(w^R)$$

pour tout $w \in \mathrm{Fac}(u)$ car

$$wa \in \mathrm{Fac}(u) \;\Leftrightarrow\; aw^R \in \mathrm{Fac}(u).$$

En particulier, dans la définition, on peut remplacer spécial à droite par spécial à gauche. De plus, comme $\mathrm{Fac}(u)$ est prolongeable à droite, il est également prolongeable à gauche donc biprolongeable.

Définition 1.4.3. Un mot épisturmien $u \in A^{\mathbb{N}}$ est *épisturmien strict* (ou *mot d'Arnoux-Rauzy*) sur A si, pour tout $n \in \mathbb{N}$, il a exactement un facteur w spécial à droite de longueur n et si, de plus, $r(w) = |A|$.

Remarquons que A est alors l'alphabet minimal de u (i.e. toutes les lettres de A apparaissent dans u). Dans la suite, si on suppose $u \in A^{\mathbb{N}}$ épisturmien strict, la locution « sur A » sera sous-entendue.

Abordons à présent un autre type de mot qui nous sera utile pour montrer que les mots épisturmiens stricts sont dendriques : les mots ordinaires.

Définition 1.4.4. Un mot biprolongeable $w \in S$ est *ordinaire* si [(7)] $E(w) \subseteq (a \times A) \cup (A \times b)$ pour $a, b \in A$ tels que $(a, b) \in E(w)$.

Remarquons que si w est ordinaire, alors il est neutre. De fait, S étant biprolongeable, on a alors

$$\begin{aligned} E(w) &= (a \times R(w)) \cup (L(w) \times b) \\ &= (a \times (R(w) \setminus \{b\})) \cup (L(w) \times b) \end{aligned}$$

où l'union est disjointe donc

$$e(w) = r(w) - 1 + l(w)$$

et

$$m(w) = e(w) - r(w) - l(w) + 1 = 0.$$

En réalité, on a même le résultat suivant.

Proposition 1.4.5. *Si w est ordinaire, alors $\mathcal{E}(w)$ est un arbre.*

Démonstration. Comme w est ordinaire, il existe $a, b \in A$ tels que $(a, b) \in E(w)$ et

$$E(w) \subseteq (a \times A) \cup (A \times b).$$

Le graphe $\mathcal{E}(w)$ est connexe car $(a, b) \in E(w)$ et que, pour tout $c \in L(w)$ (resp. $c \in R(w)$), on a $(c, b) \in E(w)$ (resp. $(a, c) \in E(w)$). De plus, ce graphe ne contient pas de cycle car tous les sommets sauf a et b sont de degré 1 et que pour avoir un cycle [(8)], il faut au moins 3 sommets de degré supérieur ou égal à 2. □

Nous pouvons à présent utiliser ce résultat pour montrer que les mots épisturmiens stricts sont bien un cas particulier de mots dendriques.

(7). Il s'agit ici d'une notation abusive pour $E \subseteq (\{a\} \times A) \cup (A \times \{b\})$.
(8). Non dégénéré, i.e. non réduit à $((c, d), (d, c))$ pour $c, d \in V(w)$.

Proposition 1.4.6. *Un mot $u \in A^{\mathbb{N}}$ épisturmien strict est dendrique.*

Démonstration. Par la Proposition 1.2.4, il suffit de montrer que, pour tout $w \in \mathrm{Fac}(u)$ bispécial, w est ordinaire. Supposons donc avoir un tel w. Il existe un unique $v \in \mathrm{Fac}(u) \cap A^{|w|+1}$ spécial à droite car u est épisturmien. En particulier, son suffixe de longueur $|w|$ est spécial à droite. Or, comme u est épisturmien strict, w est le seul mot de cette longueur à être spécial à droite donc w est suffixe de v. Il existe alors un unique $a \in A$ tel que $v = aw$ et

$$R(aw) = A = R(w).$$

De même, il existe un unique $b \in A$ tel que wb soit spécial à gauche et

$$L(wb) = A = L(w).$$

Comme, pour tous $c \in A \setminus \{a\}$, $d \in A \setminus \{b\}$,

$$r(cw) = 1 \quad \text{et} \quad l(wd) = 1,$$

on a

$$R(cw) = \{b\} \quad \text{et} \quad L(wd) = \{a\}.$$

En conséquence,

$$E(w) = (a \times A) \cup (A \times b)$$

donc w est ordinaire et on peut conclure par la proposition précédente. □

Exemple 1.4.7. Le *mot de Tribonacci* est le mot obtenu en itérant le morphisme σ tel que

$$\sigma(a) = ab, \quad \sigma(b) = ac \quad \text{et} \quad \sigma(c) = a$$

un nombre infini de fois à partir de la lettre a, i.e. c'est la limite de la suite $(\sigma^n(a))_{n\in\mathbb{N}}$. Il s'agit donc du mot

$$abacabaabacababacabaabacabac\cdots$$

Ce mot est épisturmien strict sur l'alphabet $\{a, b, c\}$, comme énoncé dans [13].

1.5 Échanges d'intervalles réguliers

Intéressons nous à présent à une autre classe de mots infinis qui est également un cas particulier de mots dendriques : les codages d'échanges d'intervalles réguliers. Les résultats et définitions de cette section sont tirés de [6].

Définition 1.5.1. Soient $<_1$ et $<_2$ deux ordres totaux sur un alphabet A d'au moins deux lettres et soit $([\gamma_a, \mu_a[)_{a\in A}$ une partition de $[0, 1[$ telle que $\mu_a \leq \gamma_b$ si $a <_1 b$. Notons $([\delta_a, \nu_a[)_{a\in A}$ la partition de $[0, 1[$ telle que $\nu_a \leq \delta_b$ si $a <_2 b$ et, pour tout $a \in A$,

$$\mu_a - \gamma_a = \nu_a - \delta_a.$$

On notera $I_a = [\gamma_a, \mu_a[$ et $J_a = [\delta_a, \nu_a[$. L'*échange d'intervalles* correspondant est la fonction

$$T : [0, 1[\to [0, 1[\quad z \mapsto z - \gamma_a + \delta_a \quad \text{si } z \in I_a.$$

Cette fonction est bijective et son inverse est

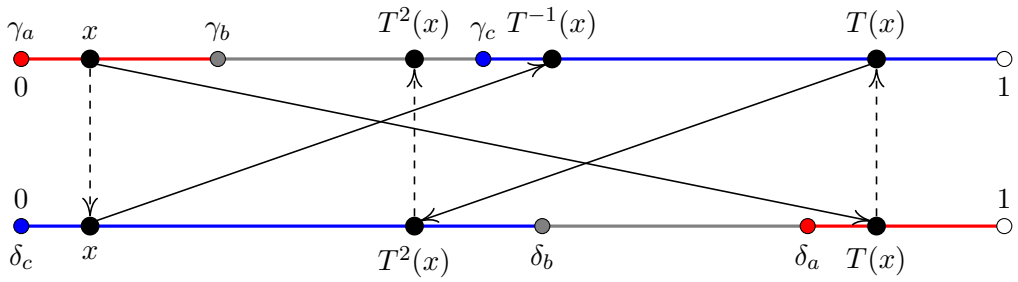
$$T^{-1} : [0,1[\to [0,1[\quad z \mapsto a - \delta_a + \gamma_a \quad \text{si } z \in J_a.$$

Nous utiliserons la notations T^k, $k \in \mathbb{Z}$ pour désigner le résultat de k compositions de T avec lui même si $k > 0$, de $|k|$ compositions de T^{-1} si $k < 0$ et l'identité si $k = 0$.

Exemple 1.5.2. Si $A = \{a, b, c\}$ et

$$a <_1 b <_1 c \quad \text{et} \quad c <_2 b <_2 a,$$

les premières itérations de T et T^{-1} à partir du point $x = 0.07$ sont représentées ci-dessous pour une partition de $[0,1[$ donnée.



Remarquons que T restreint à un intervalle I_a est une simple translation vers J_a. Cependant, ce n'est pas le cas en général car on peut seulement dire de l'image d'un intervalle semi-ouvert $[x, y[$ par T qu'il s'agit d'une union (finie) d'intervalles semi-ouverts dont les bornes sont dans l'ensemble [(9)]

$$\{T(\gamma_a) \mid a \in A\} \cup \{T(x), T(y), 1\}.$$

En conséquence, on peut faire la remarque suivante.

Remarque 1.5.3. Pour tous $n \in \mathbb{N}$ et $a \in A$, $T^n(I_a)$ est une union finie d'intervalles semi-ouverts dont les bornes sont dans l'ensemble

$$\{T^k(\gamma_b) \mid b \in A, 0 \leq k \leq n\} \cup \{1\}.$$

En effet, le résultat est vrai quand $n = 0$ et, pour l'induction,

$$T^{n+1}(I_a) = T(T^n(I_a))$$

donc on peut conclure en appliquant l'observation ci-dessus à chacun des intervalles composant $T^n(I_a)$.

Définition 1.5.4. Soit T un échange d'intervalles. L'*orbite* de $z \in [0,1[$ est

$$Orb(z) = \{T^k(z) \mid k \in \mathbb{Z}\}.$$

(9). Si $y = 1$, $T(y)$ n'est pas défini donc il sera ignoré dans l'ensemble qui suit.

Remarquons que si $x \in Orb(z)$, alors $Orb(x) = Orb(z)$. En effet, il existe alors $k_0 \in \mathbb{Z}$ tel que $x = T^{k_0}(z)$ donc

$$Orb(x) = \{T^k(T^{k_0}(z)) \mid k \in \mathbb{Z}\} = Orb(z).$$

Parmi les échanges d'intervalles, nous ne les considérons pas tous et nous restreignons aux échanges d'intervalles dit réguliers.

Définition 1.5.5. L'ensemble des *points de séparation* d'une partition $([a_i, b_i[)_{i\in I}$ de $[a, b[$ par des semi-ouverts est

$$\{a_i \mid i \in I\} \setminus \{a\} = \{b_i \mid i \in I\} \setminus \{b\}.$$

Définition 1.5.6. L'échange d'intervalles T est *régulier* si les orbites des points de séparation de $([\gamma_a, \mu_a[)_{a\in A}$ sont infinies et disjointes.

Il s'agit donc des orbites des γ_a, $a \in A$, différents de 0. Cette dernière précision est obligatoire, comme le montre la remarque suivante.

Remarque 1.5.7. Soit T un échange d'intervalle régulier.

- Pour tous $m, n \in \mathbb{Z}$ et pour tous $a, b \in A$ différents du minimum de $(A, <_1)$,

  $$T^m(\gamma_a) = T^n(\gamma_b) \Rightarrow a = b \text{ et } m = n.$$

  En effet, si $a \neq b$, les orbites de γ_a et γ_b ne sont pas disjointes et, si $a = b$ mais $m \neq n$, l'orbite de γ_a est finie.
- Remarquons qu'il existe $c \in A$ tel que $T(\gamma_c) = 0$ donc que

  $$Orb(0) = Orb(\gamma_c).$$

  De plus, $\gamma_c \neq 0$ car sinon, si a et b sont les deuxièmes plus petites lettres de A pour $<_1$ et $<_2$ respectivement, alors

  $$T(\gamma_b) = \delta_b = \nu_c = \mu_c = \gamma_a,$$

  ce qui est absurde. L'orbite de 0 ne peut donc pas être disjointe des orbites des autres γ_a, $a \in A$. Par contre, elle sera bien infinie donc

  $$T^m(0) = T^n(0) \Rightarrow m = n.$$

- Pour tout $a \in A$ tel que $\gamma_a \neq 0$, si

  $$T^m(\gamma_a) = T^n(0),$$

  alors, comme $0 = T(\gamma_c)$ pour $\gamma_c \neq 0$, on a par le premier point $a = c$ et $m = n + 1$. En conséquence, pour tous $a, b \in A$, si

  $$T^m(\gamma_a) = T^n(\gamma_b),$$

  alors

  1. si $|m - n| = 1$ et si, sans perte de généralité, $m = n + 1$, alors $\gamma_b = 0$,
  2. sinon, $m = n$ et $a = b$.

Définition 1.5.8. L'échange d'intervalles T est *minimal* si l'orbite de tout $z \in [0, 1[$ est dense dans $[0, 1[$.

Remarquons que si T est minimal, alors les orbites des μ_a sont forcément infinies car elles doivent être denses dans $[0,1[$ mais elles ne sont pas forcément disjointes donc T n'est pas toujours régulier. Par contre, on a le résultat suivant, dû à Michael Keane [14].

Théorème 1.5.9 (Keane)**.** *Si un échange d'intervalles est régulier, alors il est minimal.*

A tout échange d'intervalles régulier, on peut associer un ensemble de mots infinis. Pour ce faire, nous codons par un mot les images successives d'un point de $[0,1[$ de la façon suivante.

Définition 1.5.10. Soient T un échange d'intervalles et $z \in [0,1[$. Le *codage* de z est le mot $\Sigma_T(z) \in A^{\mathbb{N}}$ tel que, pour tout $i \in \mathbb{N}$, sa i-ème lettre [(10)] soit l'unique $a \in A$ tel que

$$T^i(z) \in I_a.$$

Réciproquement, pour $w = w_0 \dots w_n \in A^*$, on note

$$\begin{aligned} I_w &= \{z \in [0,1[\mid w \in \mathrm{Pref}(\Sigma_T(z))\} \\ &= [0,1[\cap I_{w_0} \cap T^{-1}(I_{w_1}) \cap \dots \cap T^{-n}(I_{w_n}). \end{aligned}$$

En particulier, $I_\varepsilon = [0,1[$.

Exemple 1.5.11. En reprenant les données de l'Exemple 1.5.2,

$$x \in I_a, \quad T(x) \in I_c \quad \text{et} \quad T^2(x) \in I_b$$

donc le codage de x commence par acb et

$$x \in I_{acb} \subseteq I_{ac} \subseteq I_a \subseteq I_\varepsilon.$$

Remarque 1.5.12. Montrons par récurrence sur la longueur de w que I_w est soit vide, soit un intervalle semi-ouvert. Si $|w| = 0$, alors $w = \varepsilon$ donc le résultat est immédiat. Supposons le vrai pour les mots de longueur $n-1$ et montrons-le pour w de longueur n. Posons $v = w_2 \dots w_n$. On a

$$I_w = I_{w_1} \cap T^{-1}(I_v).$$

Si I_v est vide, alors $T^{-1}(I_v)$ aussi donc on en déduit directement que I_w aussi. Sinon,

$$T(I_w) = T(I_{w_1}) \cap I_v = J_{w_1} \cap I_v$$

est l'intersection de deux intervalles semi-ouverts. Il s'agit donc soit de l'ensemble vide, soit d'un intervalle semi-ouvert. Or, il s'agit d'un simple translaté de $I_w \subseteq I_{w_1}$, d'où la conclusion.

Proposition 1.5.13. *Soient T un échange d'intervalles régulier et $z \in [0,1[$. Pour tout $w \in A^*$, w est facteur de $\Sigma_T(z)$ si, et seulement si, $I_w \neq \emptyset$. En particulier,* $\mathrm{Fac}(\Sigma_T(z))$ *dépend uniquement de T.*

(10). Rappelons que, pour les mots infinis, nous avons pour convention de commencer à compter les lettres à partir de 0. Dans cette section, pour faciliter les notations, nous prenons la même convention pour les mots finis.

Démonstration. Si $w \in \mathrm{Fac}(\Sigma_T(z))$, alors il existe $n \in \mathbb{N}$ tel que, pour tout $i \in \{0, \dots, |w|-1\}$,

$$w_i = \Sigma_T(z)_{n+i} = \Sigma_T(T^n(z))_i.$$

Par définition, on a alors

$$T^n(z) \in T^{-i}(I_{w_i}), \quad \forall\, i \in \{0, \dots, |w|-1\}$$

donc $T^n(z) \in I_w$ et $I_w \neq \emptyset$.

Réciproquement, supposons I_w non vide donc, par la remarque précédente, I_w est un intervalle semi-ouvert. Comme T est régulier, l'orbite de z est dense dans $[0,1[$ donc il existe $n \in \mathbb{N}$ tel que $T^n(z) \in I_w$. On a alors $w = \Sigma_T(z)_{[n,n+|w|[}$, d'où la conclusion. □

Définition 1.5.14. Soit T un échange d'intervalles régulier. L'*ensemble d'échange d'intervalles régulier* associé à T est

$$\mathrm{Fac}(T) := \mathrm{Fac}(\Sigma_T(z))$$

pour $z \in [0,1[$.

Remarquons d'ores et déjà que cet ensemble est bi-prolongeable. En effet, il est évident qu'il est prolongeable à droite puisqu'il s'agit des facteurs d'un mot infini. De plus, si $w \in \mathrm{Fac}(T)$, alors I_w est non vide donc $T^{-1}(I_w)$ également. Il rencontre donc un intervalle I_a pour $a \in A$. On a alors $I_{aw} = I_a \cap T^{-1}(I_w) \neq \emptyset$ donc $aw \in \mathrm{Fac}(T)$.

Cet ensemble est même dendrique, ce que montrent les résultats qui suivent.

Lemme 1.5.15. *Soient $I = [a,b[$ un intervalle et $(I_i)_{i\leq n}$, $(J_j)_{j\leq m}$ deux partitions de I en intervalles semi-ouverts. Le graphe dont les sommets correspondent aux intervalles de ces deux partitions tel que deux sommets sont reliés si et seulement si les intervalles correspondant s'intersectent est acyclique. De plus, il est connexe si, et seulement si, les points de séparations des deux partitions sont distincts.*

Démonstration. Notons a_i la borne inférieure de I_i et b_j celle de J_j et $a_{n+1} = b_{m+1} = b$. Notons également, pour tout $i \leq n$, $h(i)$ l'indice tel que $a_i \in J_{h(i)}$. On se convainc rapidement que le graphe est biparti. De plus, par construction, pour tout $i \leq n$, le seul voisin de I_i qui puisse éventuellement être relié à des intervalles I_k pour $k < i$ est $J_{h(i)}$. En effet, pour qu'un intervalle J_j intersecte I_i et des intervalles qui le précédent, J_j doit contenir a_i.

Supposons que le graphe contienne un cycle [(11)]. Notons i le plus grand indice tel que le cycle passe par I_i. Notons J_j et I_k les deux sommets suivants dans le cycle. Comme $k < i$, $j = h(i)$. De même, si $J_{j'}$ et $I_{k'}$ sont les deux sommets précédant I_i dans le cycle, $k' < i$ donc $j' = h(i)$, ce qui est absurde car le cycle ne peut pas emprunter la même arête deux fois d'affilée.

Supposons à présent que les points de séparations des partitions soient distincts, i.e. que

$$\{a_2, \dots, a_n\} \cap \{b_2, \dots, b_m\} = \emptyset$$

et montrons que le graphe est connexe. Soient U et V deux sommets (donc deux intervalles parmi ceux des partitions). Supposons, sans perte de généralité, que U commence avant V [(12)].

(11). On ne considère ici que des cycles non dégénérés, i.e. qui n'empruntent pas la même arête deux fois de suite.

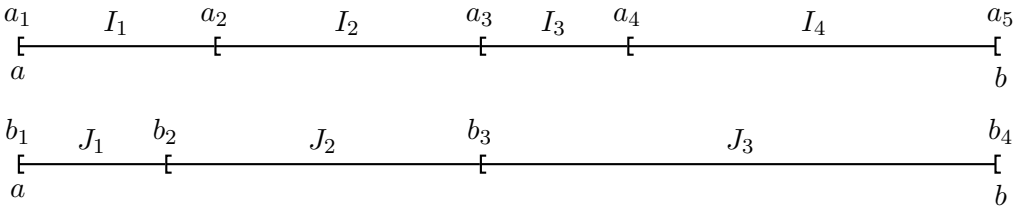
(12). Ce qui signifie que la borne inférieure de U est inférieure à celle de V

Pour simplifier les notations, supposons également que $U = I_i$ et notons $V = [x, y[$. Soit k tel que $x \in I_k$. Par hypothèse, $k \geq i$. Si $k = i$, alors U et V s'intersectent donc il existe un chemin les reliant (soit il s'agit du même sommet, soit ils sont reliés par une arête). Sinon, $x \geq a_{i+1}$. En particulier, $i < n$. Comme $a_{i+1} \notin \{b_2, \ldots, b_n\}$, $J_{h(i+1)}$ intersecte $I_i = U$. Il rencontre également I_{i+1}. On a donc un chemin de U vers $I_{i+1} =: U'$ et on peut conclure par récurrence sur $k - i$.

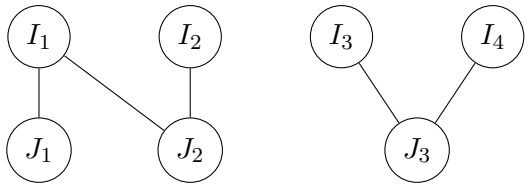
Réciproquement, si $a_i = b_j$, $i, j > 1$, alors $h(i) = j$. Or, $J_j \subseteq [a_i, b[$ donc J_j n'est relié à aucun des intervalles I_k, $k < i$. Il n'existe donc pas de chemin reliant I_i aux intervalles I_k, $k < i$, et le graphe n'est pas connexe. □

Non seulement le graphe est acyclique mais on peut également se convaincre que, si on place, à gauche, les sommets I_i dans l'ordre et, à droite, les sommets J_j également ordonnés, alors le graphe est planaire.

Exemple 1.5.16. Supposons avoir les partitions suivantes :



Le graphe dont parle le lemme précédent est le graphe



Il est bien acyclique. Cependant, il n'est pas connexe car $a_3 = b_3$.

Proposition 1.5.17. *Si T est un échange d'intervalles régulier, alors* $\mathrm{Fac}(T)$ *est dendrique. De façon équivalente, le codage $\Sigma_T(z)$ est dendrique pour tout $z \in [0, 1[$.*

Démonstration. Soit $w \in \mathrm{Fac}(T)$. Comme I_w est non vide, il s'agit d'un intervalle semi-ouvert. Remarquons que

$$\begin{aligned} a \in L(w) &\Leftrightarrow aw \in \mathrm{Fac}(T) \\ &\Leftrightarrow I_{aw} \neq \emptyset \\ &\Leftrightarrow I_a \cap T^{-1}(I_w) \neq \emptyset \\ &\Leftrightarrow T(I_a) \cap I_w \neq \emptyset \\ &\Leftrightarrow J_a \cap I_w \neq \emptyset \end{aligned}$$

donc, comme $\{J_a \mid a \in A\}$ est une partition de $[0, 1[$, $P_1 := \{J_a \cap I_w \mid a \in L(w)\}$ est une partition de I_w. De plus, ses éléments sont de la forme I_{aw} et sont non-vides donc ce sont des intervalles semi-ouverts.

De même,

$$\begin{aligned} b \in R(w) &\Leftrightarrow wb \in \mathrm{Fac}(T) \\ &\Leftrightarrow I_{wb} \neq \emptyset \\ &\Leftrightarrow T^{-|w|}(I_b) \cap I_w \neq \emptyset \end{aligned}$$

donc, comme $\{T^{-|w|}(I_b) \mid b \in A\}$ est une partition de $[0,1[$, $P_2 := \{I_{wb} \mid b \in R(w)\}$ est une partition de I_w par des intervalles semi-ouverts.

De plus,

$$\begin{aligned} (a,b) \in E(w) &\Leftrightarrow awb \in \mathrm{Fac}(T) \\ &\Leftrightarrow T(I_{awb}) \neq \emptyset \\ &\Leftrightarrow J_a \cap I_{wb} \neq \emptyset \end{aligned}$$

donc le graphe d'extensions de w est le graphe du Lemme 1.5.15 pour les partitions P_1 et P_2 de I_w. Pour conclure qu'il s'agit d'un arbre, montrons que les points de séparation de ces deux partitions sont distincts.

D'une part, les points de séparation de P_1 sont de la forme $T(\gamma_c)$, $c \in A$, car ce sont des bornes d'intervalles J_a, $a \in L(w)$. D'autre part, les points de séparation de P_2 sont dans l'ensemble

$$\{T^{-k}(\gamma_d) \mid d \in A, 0 \leq k \leq |w|\} \setminus \{0\}$$

par la Remarque 1.5.3 appliquée à T^{-1}. L'échange d'intervalles T étant régulier, les points de séparation sont bien distincts. En effet, par la Remarque 1.5.7, la seule façon d'avoir $T(\gamma_c) = T^{-k}(\gamma_d)$, $0 \leq k \leq |w|$ est d'avoir $k = 0$ et $\gamma_d = 0$ mais dans ce cas, $T^{-k}(\gamma_d) = 0$ n'est pas un point de séparation de P_2. □

On peut montrer (voir [6] pour les détails) que l'ordre de la partition P_1 correspond à une restriction de $<_2$ et que l'ordre de P_2 est une restriction de $<_1$. Ceci nous permet d'affirmer que, si un mot $u \in A^{\mathbb{N}}$ est obtenu par codage d'échange d'intervalles régulier, alors il est non seulement dendrique mais il existe deux ordres sur A tels que, pour tout $w \in \mathrm{Fac}(u)$, si les éléments de $L(w)$ sont placés à gauche et ordonnés selon le premier ordre et ceux de $R(w)$ à droite et selon le second ordre, alors le graphe $\mathcal{E}(w)$ est planaire. Nous pouvons donc à présent exhiber un mot dendrique qui n'est ni épisturmien strict, ni codage d'un échange d'intervalles régulier.

Exemple 1.5.18. Si σ est le morphisme défini par

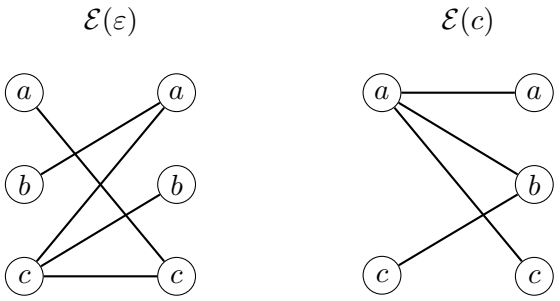
$$\sigma(a) = ac, \quad \sigma(b) = bac \quad \text{et} \quad \sigma(c) = cbac$$

et si u est le mot infini obtenu en itérant σ à partir de a, alors

$$u = accbaccbacbacaccbaccbacbacaccbacbac\cdots$$

Pour chaque longueur, ce mot a exactement un facteur spécial à droite w tel que $r(w) = 3$, il s'agit du suffixe de $\sigma^n(c)$ pour n assez grand. Cependant, il n'est pas épisturmien strict car on peut notamment remarquer que a et c sont spéciaux à gauche. De plus, on a les graphes d'extensions suivants :

Pour que u soit obtenu comme codage d'un échange d'intervalles régulier, il faut donc trouver deux ordres sur $\{a,b,c\}$ rendant ces graphes planaires. Or, le premier nous impose



que, pour l'ordre de droite, b soit au milieu tandis que le deuxième nous impose le contraire. Le mot u n'est donc pas non plus codage d'un échange d'intervalles régulier. Cependant, on connaît la forme exacte des facteurs bispéciaux de u puisque ce sont des suffixes de $\sigma^n(c)$, $n \in \mathbb{N}$. On se convainc alors que tout facteur bispécial v suffisamment long s'écrit de la forme $ac\sigma(w)$ ou $cbac\sigma(w)$ où w est lui-même un facteur bispécial. Le graphe d'extensions de v se déduit à partir de celui de w donc on montre par récurrence sur la longueur de v que u est dendrique.

1.6 Ensembles neutres et complexité

La notion de complexité est essentielle en combinatoire des mots. Nous montrons ici que la complexité des mots dendriques et, plus généralement, celle des ensembles neutres est connue et qu'elle est linéaire.

Définition 1.6.1. La *complexité d'un ensemble* $S \subseteq A^*$ est la fonction

$$p_S \ : \ \mathbb{N} \to \mathbb{N} \quad n \mapsto |S \cap A^n|.$$

La *complexité d'un mot* est la complexité de l'ensemble de ses facteurs. Autrement dit, $p_w(n)$ est le nombre de facteurs de longueur n différents apparaissant dans le mot w.

Exemple 1.6.2. Les premières valeurs de la fonction de complexité du mot de Tribonacci x (Exemple 1.4.7) sont données par $p_x(0) = 1$, $p_x(1) = 3$, $p_x(2) = 5$ et $p_x(3) = 7$ car

$$\mathrm{Fac}(x) \cap A^0 = \{\varepsilon\}, \quad \mathrm{Fac}(x) \cap A = \{a, b, c\}, \quad \mathrm{Fac}(x) \cap A^2 = \{aa, ab, ac, ba, ca\}$$

et

$$\mathrm{Fac}(x) \cap A^3 = \{aab, aba, aca, baa, bab, bac, cab\}.$$

Remarque 1.6.3. Si S est factoriel, alors tout mot de longueur $n+1$ de S correspond à un unique mot de longueur n de S auquel on a rajouté une lettre à gauche (resp. à droite). On a donc

$$p_S(n+1) = \sum_{w \in S \cap A^n} l(w) = \sum_{w \in S \cap A^n} r(w).$$

En particulier, si S est prolongeable à droite (resp. à gauche), alors sa fonction de complexité est croissante. De même, tout mot de longueur $n+2$ est un mot de longueur n auquel on a ajouté une lettre à gauche et une lettre à droite. Cette décomposition étant unique, on a

$$p_S(n+2) = \sum_{w \in S \cap A^n} e(w).$$

Vu leur définition, la complexité des mots épisturmiens stricts est très facile à calculer, comme le montre la proposition suivante.

Proposition 1.6.4. *Si $u \in A^{\mathbb{N}}$ est épisturmien strict, alors*

$$p_u(n) = (|A| - 1) \cdot n + 1.$$

Démonstration. Procédons par récurrence sur $n \in \mathbb{N}$. Si $n = 0$, la propriété est immédiate, le seul facteur de longueur 0 étant ε. Supposons la propriété vraie pour n et montrons-la pour $n + 1$. Vu la Remarque 1.6.3, on a

$$p_u(n+1) = \sum_{w \in \mathrm{Fac}(u) \cap A^n} r(w).$$

Notons v l'unique facteur de u de longueur n qui soit spécial à droite. On a alors, pour tout $w \in \mathrm{Fac}(u) \cap A^n$,

$$r(w) = \begin{cases} |A| & \text{si } w = v \\ 1 & \text{sinon} \end{cases}$$

donc

$$p_u(n+1) = p_u(n) - 1 + |A| = (|A| - 1) \cdot (n+1) + 1.$$

□

Il est peut-être moins évident de calculer la complexité des échanges d'intervalles réguliers ou des mots dendriques à partir de leurs définitions. Cependant, il est en réalité assez aisé d'obtenir la complexité des ensembles neutres en général. Cela nécessite juste un lemme rendu très rapide par des observations faites précédemment.

Lemme 1.6.5. *Soit S un ensemble factoriel. Si, pour tout $n \in \mathbb{N}$, on note*

$$s_n = p_S(n+1) - p_S(n) \quad \textit{et} \quad b_n = s_{n+1} - s_n,$$

alors

$$s_n = \sum_{w \in S \cap A^n} (r(w) - 1) \quad \textit{et} \quad b_n = \sum_{w \in S \cap A^n} m(w)$$

pour tout $n \in \mathbb{N}$.

Démonstration. Le résultat découle directement de la remarque 1.6.3. De fait, on a alors

$$\sum_{w \in S \cap A^n} (r(w) - 1) = \left(\sum_{w \in S \cap A^n} r(w) \right) - p_S(n) = p_S(n+1) - p_S(n) = s_n$$

et

$$\begin{aligned} \sum_{w \in S \cap A^n} m(w) &= \left(\sum_{w \in S \cap A^n} e(w) \right) - \left(\sum_{w \in S \cap A^n} r(w) \right) - \left(\sum_{w \in S \cap A^n} l(w) \right) + p_S(n) \\ &= p_S(n+2) - p_S(n+1) - p_S(n+1) + p_S(n) \\ &= s_{n+1} - s_n \\ &= b_n. \end{aligned}$$

□

Proposition 1.6.6. *Si S est neutre, alors, pour tout $n \in \mathbb{N}$, $p_S(n) = (|A| - 1)n + 1$* [13].

Démonstration. Si S est neutre, alors, pour tout $w \in S$, $m(w) = 0$ donc

$$b_n = 0$$

et

$$s_{n+1} = s_n$$

pour tout $n \in \mathbb{N}$. Or, par définition,

$$s_0 = p_S(1) - p_S(0) = |A| - 1$$

donc

$$p_S(n+1) = p_S(n) + s_n = p_S(n) + s_0 = p_S(n) + |A| - 1$$

pour tout $n \in \mathbb{N}$. Une simple récurrence permet de conclure. □

(13). Rappelons qu'on suppose l'alphabet A minimal donc, comme S est factoriel, cela signifie que $A \subseteq S$.

Chapitre 2

Cardinalité des codes bifixes dans un ensemble neutre

Nous allons dans ce chapitre momentanément quitter le cadre des ensembles et mots dendriques pour considérer plus généralement des ensembles neutres. Ce chapitre introduit de nouvelles notions générales de combinatoire des mots qui seront primordiales pour la suite telles que les codes bifixes, les mots de retour et la notion de S-degré. Nous démontrons ainsi divers résultats concernant la cardinalité des codes préfixes et bifixes dans un ensemble neutre récurrent. En particulier, nous montrons que tout ensemble neutre récurrent est uniformément récurrent.

2.1 Codes préfixes, suffixes et bifixes

Dans cette section, nous introduisons les notions de codes préfixes, suffixes et bifixes qui sont des notions fondamentales de combinatoire des mots. Nous étudions également quelques unes de leurs propriétés élémentaires.

Définition 2.1.1. Un ensemble $X \subseteq A^*$ est un *code* si, pour tous $x_1, \dots, x_n, y_1, \dots, y_m \in X$, on a

$$x_1 \dots x_n = y_1 \dots y_m \Rightarrow n = m \text{ et } x_i = y_i \quad \forall\, i \leq n.$$

Définition 2.1.2. Un ensemble X de mots est un *code préfixe* si, pour tous $x, y \in X$, $x \neq y$, x n'est pas un préfixe de y. (1)

Un code préfixe $X \subseteq S$ est *S-maximal* s'il n'est inclus dans aucun autre code préfixe $Y \subseteq S$.

On définit symétriquement un *code suffixe* et un *code suffixe S-maximal*.

Un code *bifixe* est un ensemble qui est à la fois un code préfixe et un code suffixe. Il sera *S-maximal* s'il est maximal parmi les codes bifixes sur S.

Exemple 2.1.3. L'ensemble $ab^* = \{ab^n \mid n \in \mathbb{N}\}$ est un code suffixe mais pas un code préfixe. L'ensemble $ab^*a = \{ab^na \mid n \in \mathbb{N}\}$, lui, est un code bifixe mais il n'est pas maximal dans $\{a, b\}^*$ car $ab^*a \cup \{b\}$ est un code bifixe de $\{a, b\}^*$ strictement plus grand.

(1). On exclut également les cas $X = \emptyset$ et $X = \{\varepsilon\}$.

Tous les résultats de cette section seront exprimés pour les codes préfixes mais il est évident qu'il existe des résultats symétriques pour les codes suffixes.

On peut faire quelques remarques concernant ces concepts que nous manipulerons abondamment dans la suite.

Remarque 2.1.4. Soit X un code préfixe.

1. Tout $Y \subseteq X$ non vide est également un code préfixe.
2. Pour tout ensemble non vide Y tel que $\varepsilon \notin Y$, l'ensemble
$$Z = Y \setminus YA^+$$
est un code préfixe. En effet, si $x \in Y$ est de longueur minimale, alors $x \in Z$ donc $Z \neq \emptyset$ et $Z \neq \{\varepsilon\}$. De plus, si $x, y \in Z$ et $x \in \mathrm{Pref}(y)$, alors $y \in xA^* \subseteq YA^*$ et la seule possibilité est d'avoir $x = y$.
3. Tout mot w a au plus un préfixe dans X. En effet, si $x, y \in X$ étaient deux préfixes distincts de w, alors le plus court serait un préfixe de l'autre, ce qui est absurde. La réciproque est également vraie. En effet, si tout mot a au plus un préfixe dans X alors c'est valable pour les éléments de X donc X est un code préfixe.
4. Tout w peut s'écrire de façon unique comme $w = x_1 \dots x_n u$ où u n'a pas de préfixe dans X et $x_1, \dots, x_n \in X$. En conséquence, X est un code. Ce résultat peut se montrer par récurrence sur la longueur de w. Il est évident quand $w = \varepsilon$ et sinon, soit w n'a pas de préfixe dans X auquel cas $u = w$ et $n = 0$ est la seule possibilité, soit il en a un unique donc $w = yv$ où $y \in X$ et on peut conclure par récurrence sur v.
5. En conséquence du point précédent, si $u, uv \in X^*$, alors $v \in X^*$. On dit alors que X^* est *unitaire à droite*. En effet, si on note $x_1 \dots x_n$ la factorisation de uv par des mots de X et si $k \leq n$ est tel que $u = x_1 \dots x_k p$ où p est préfixe propre de x_{k+1} alors, par le point précédent, on doit avoir $p = \varepsilon$ car $x \in X^*$. En conséquence, $v = x_{k+1} \dots x_n \in X^*$.
6. Pour tout $x \in X$ et pour tout $w \in A^*$,
$$(X \setminus \{x\}) \cup \{xw\}$$
est encore un code préfixe. En effet, comme x est un préfixe de xw, xw n'est préfixe d'aucun mot de $X \setminus \{x\}$ et il n'a aucun préfixe dans $X \setminus \{x\}$ car cela signifierait qu'il en a 2 dans X. Par un raisonnement similaire, pour tout $x \in \mathrm{Pref}(X)$,
$$(X \setminus xA^*) \cup \{x\}$$
est un code préfixe.

Proposition 2.1.5. *Soient S un ensemble factoriel et $X \subseteq S$. Les affirmations suivantes sont équivalentes.*

1. *Tout mot $w \in S$ est préfixe propre d'un mot de X ou a un préfixe dans X.*
2. *XA^* est S-*dense à droite*, i.e. $S \subseteq \mathrm{Pref}(XA^*)$.*
3. *X est S-*complet à droite*, i.e. $S \subseteq \mathrm{Pref}(X^*)$.*

Si, de plus, X est un code préfixe, alors il est S-maximal si, et seulement si, une des trois propositions ci-dessus est vraie.

Démonstration. L'équivalence entre les deux premiers points est immédiate car le premier peut se réécrire

$$S \subseteq XA^+ \cup \mathrm{Pref}(X) = \mathrm{Pref}(XA^*).$$

Montrons que les points 2. et 3. sont équivalents. Il est immédiat que

$$S \subseteq \mathrm{Pref}(X^*) \Rightarrow S \subseteq \mathrm{Pref}(XA^*).$$

Pour la réciproque, procédons par récurrence sur la longueur de $w \in S$. Par hypothèse, $w \in \mathrm{Pref}(XA^*)$. Si $w \in \mathrm{Pref}(X)$, le résultat est direct. Sinon, $w = xu$ avec $x \in X$. L'ensemble X étant un code préfixe, $x \neq \varepsilon$ et, par hypothèse de récurrence sur $u \in S$, on a $u \in \mathrm{Pref}(X^*)$. On a donc bien $w = xu \in \mathrm{Pref}(X^*)$.

Si X est un code préfixe, il est S-maximal si, et seulement si, pour tout $w \in S \setminus X$, $X \cup \{w\}$ n'est pas un code préfixe, i.e. w a un préfixe dans X ou est préfixe propre d'un élément de X donc on a l'équivalence avec la première assertion. □

C'est ce résultat que nous utilisons pour généraliser la notion de code préfixe S-maximal dans le cas où $X \not\subseteq S$. On dira qu'un code préfixe X quelconque est *S-maximal* si tout mot de S est préfixe d'un mot de X ou a un mot de X pour préfixe.

Remarquons que si $S' \subseteq S$ et si X est un code préfixe S-maximal, alors X est également S'-maximal.

Il est évident qu'un code bifixe X maximal parmi les codes préfixes ou les codes suffixes est également maximal parmi les codes bifixes mais la réciproque est également vraie quand S est récurrent et X est S-fin. C'est ce que nous allons montrer dans cette section.

Définition 2.1.6. Un ensemble $X \subseteq S$ est S-fin si $S \setminus \mathrm{Fac}(X)$ est non vide.

Si l'ensemble S est infini, alors tout ensemble fini X est S-fin donc il s'agit a priori d'une généralisation de la notion d'ensemble fini. Cependant, dans le cas d'un ensemble S uniformément récurrent, tout ensemble S-fin est fini. En effet, si X est S-fin, prenons $w \in S \setminus \mathrm{Fac}(X)$. L'ensemble S étant uniformément récurrent, il existe $n \in \mathbb{N}$ tel que w soit facteur de tous les mots de S de longueur n. Montrons que

$$X \subseteq A^{\leq n-1}$$

et donc que X est fini. Supposons qu'il existe $x \in X \cap A^k$ pour $k \geq n$. Par hypothèse, $w \in \mathrm{Fac}(x) \subseteq \mathrm{Fac}(X)$, ce qui est absurde.

Nous montrons à présent toute une série de lemme issus de [3] qui permettront de prouver le résultat principal de cette section. Pour cela, nous devons également introduire quelques notations.

Définition 2.1.7. Soient X un ensemble, $u, v \in A^*$ et $y \in \mathrm{Pref}(u)$. Si $z \in A^*$ est tel que $yz = u$, alors $P_{u,v}(y)$ est l'ensemble des préfixes p de u tels que l'une des deux conditions suivantes soit vérifiée

1. $p \in yX$,
2. $zvp = x_1 \dots x_n$ où $x_i \in X$, $z \in \mathrm{Pref}(x_1) \setminus \{x_1\}$ et $p \in \mathrm{Suff}(x_n) \setminus \{x_n\}$.

Exemple 2.1.8. Si $X = \{a^n b^m \mid n, m \in \mathbb{N}\}$ et si $u = baaba$ et $y = ba$, alors les mots p dans $\mathrm{Pref}(u) \cap yX$ sont $ba, baa, baab$. Cependant, $z = aba$ n'est préfixe propre d'aucun mot de X donc aucun préfixe p de u ne vérifie la seconde condition.

Si $X = A^2$ et si $u = aab$, $y = aa$, alors aucun préfixe p de u ne vérifie la première condition. Pour la seconde condition, le mot $z = b$ est de longueur 1 donc il vérifiera toujours $z \in \mathrm{Pref}(x_1) \setminus \{x_1\}$. Etant donné que p doit être suffixe propre d'un élément de X, on doit avoir $p \in \{\varepsilon, a\}$. Selon la parité de $|v|$, seule une des deux possibilité sera telle que $zvp \in X^*$. On a donc

$$P_{u,v}(y) = \begin{cases} \{a\} & \text{si } |v| \equiv 0 \mod 2, \\ \{\varepsilon\} & \text{si } |v| \equiv 1 \mod 2. \end{cases}$$

Ceci est cohérent avec le résultat suivant.

Lemme 2.1.9. *Soient X un code préfixe et $u, v \in A^*$. Pour tout $y \in \mathrm{Pref}(u)$,*

$$|P_{u,v}(y)| \leq 1.$$

Démonstration. Supposons avoir $p, p' \in P_{u,v}(y)$ et montrons qu'ils sont égaux. Traitons différents cas suivant les conditions vérifiées par p et p' dans la Définition 2.1.7.

1. Si p et p' vérifient la première condition, alors $p, p' \in yX$ donc il existe $x, x' \in X$ tels que $p = yx$ et $p' = yx'$. Supposons sans perte de généralité que $|p| \geq |p'|$. Dans ce cas, comme ils sont tous deux préfixes de u, $p' \in \mathrm{Pref}(p)$ et $x' \in \mathrm{Pref}(x)$. L'ensemble X étant un code préfixe, on en conclut que $x = x'$ et $p = p'$.
2. Si p et p' vérifient la seconde condition, notons $u = yz$, $zvp = x_1 \dots x_n$ et $zvp' = x'_1 \dots x'_m$. À nouveau, nous pouvons supposer que $p' \in \mathrm{Pref}(p)$. Notons alors $p = p'q$ et $q = y_1 \dots y_k q'$ où $y_i \in X$ et q' n'a pas de préfixe dans X. On a alors

   $$x_1 \dots x_n = zvp = zvp'q = x'_1 \dots x'_m y_1 \dots y_k q'.$$

   Or, X est un code préfixe donc les deux décompositions sont identiques par la Remarque 2.1.4. Cela signifie donc que

   $$q' = \varepsilon, \quad n = m + k \quad \text{et} \quad y_i = x_{m+i}.$$

   En particulier,

   $$p = p'q = p'x_{m+1} \dots x_n.$$

   Étant donné que p est un suffixe propre de x_n, la seule possibilité est d'avoir $m = n$ donc $p = p'$.
3. Supposons à présent que p vérifie la première condition et p' la deuxième. L'autre cas est symétrique. Notons $x \in X$ tel que $p = yx$. Comme p est un préfixe de $u = yz$, on en déduit que $x \in \mathrm{Pref}(z)$. Or, par hypothèse sur p', z est préfixe propre d'un élément $x_1 \in X$. Cette configuration ne peut donc pas se produire car X est un code préfixe et $x, x_1 \in X$.

□

Ce résultat admet une réciproque partielle pour laquelle il suffit de considérer des couples (u, v) particuliers.

Définition 2.1.10. Soient S un ensemble récurrent et $X \subseteq S$. On note $C(X, S)$ l'ensemble des couples (u, v) de mots tels que

- $v \neq \varepsilon$,

- $u \notin \mathrm{IFac}(X)$,
- $uvu \in S$.

Lemme 2.1.11. *Soient S un ensemble récurrent et $X \subseteq S$ un code suffixe S-maximal et S-fin. Si $|P_{u,v}(y)| \leq 1$ pour tous $(u,v) \in C(X,S)$, $y \in \mathrm{Pref}(u)$, alors X est un code préfixe.*

Démonstration. L'ensemble X étant S-fin, notons $w \in S \setminus \mathrm{Fac}(X)$. Procédons par contraposition. Supposons donc que X n'est pas un code préfixe et montrons qu'on peut alors construire u, v, y tels que $(u,v) \in C(X,S)$, $y \in \mathrm{Pref}(u)$ et $|P_{u,v}(y)| \geq 2$.

Si X n'est pas un code préfixe, il existe $x, x' \in X$, $x \neq x'$, tels que $x \in \mathrm{Pref}(x')$. Notons alors q tel que $x' = xq$. Comme S est récurrent, il existe q' tel que $xqq'w = x'q'w \in S$. Notons $u = qq'w$. On a donc $xu \in S$ et, à nouveau, il existe q'' tel que $uq''xu \in S$. Posons $v = q''x$. On a alors

- $v \neq \varepsilon$ car $x \in X$ est non vide étant donné que X est un code suffixe,
- $u \notin \mathrm{IFac}(X)$ car, par hypothèse, $w \notin \mathrm{Fac}(X)$ et $w \in \mathrm{Fac}(u)$,
- $uvu = uq''xu \in S$

donc $(u,v) \in C(X,S)$. Cherchons $y \in \mathrm{Pref}(u)$ tel que

$$|P_{u,v}(y)| \geq 2.$$

Il existe $n \in \mathbb{N}$, $t_1, \ldots, t_n \in X$ et $r \in S$ n'ayant pas de suffixe dans X tels que

$$uq'' = rt_1 \ldots t_n.$$

Par la Proposition 2.1.5, étant donné que X est un code suffixe S-maximal, r doit être suffixe propre d'un mot de X. En particulier, u ne peut pas être un préfixe propre de r car $u \notin \mathrm{IFac}(X)$. On en déduit que $r \in \mathrm{Pref}(u)$. On a alors deux possibilités :

1. S'il existe $k \leq n$ tel que $u = rt_1 \ldots t_{k-1}$, alors $q'' = t_k \ldots t_n \in X^*$. Posons $y = u$ et $z = \varepsilon$. Dans ce cas,
$$zv = q''x = t_k \ldots t_n x.$$
2. Sinon, il existe $k \leq n$ et $s, s' \in A^+$ tels que
$$t_k = ss', \quad u = rt_1 \ldots t_{k-1}s \quad \text{et} \quad q'' = s't_{k+1} \ldots t_n.$$
Posons $y = rt_1 \ldots t_{k-1}$ et $z = s$, de sorte que $yz = u$. On a alors également
$$zv = sq''x = t_k \ldots t_n x$$
et z est un préfixe propre de t_k.

Montrons que ε et q sont deux éléments distincts de $P_{u,v}(y)$ dans les deux cas, ce qui contredira l'hypothèse. Comme $x \neq x'$, q est non vide. Rappelons aussi que $u = qq'w$ donc q et ε sont bien deux préfixes distincts de u. Pour $p = \varepsilon$,

$$zvp = t_k \ldots t_n x \in X^+$$

donc ε vérifie la seconde condition de la Définition 2.1.7. De même, pour $p = q$, on a

$$zvp = t_k \ldots t_n xq = t_k \ldots t_n x' \in X^+.$$

□

Etudions à présent des conditions nécessaires ou suffisantes pour avoir $P_{u,v}(y) \geq 1$.

Lemme 2.1.12. *Soient S un ensemble récurrent et $X \subseteq S$ S-complet à droite. Pour tous $(u, v) \in C(X, S)$ et $y \in \mathrm{Pref}(u)$,*

$$P_{u,v}(y) \neq \emptyset.$$

Démonstration. Notons $u = yz$. Comme S est factoriel et que $uvu \in S$, on a $zvu \in S$. De plus, X est S-complet à droite donc $zvu \in \mathrm{Pref}(X^*)$. Notons $w \in A^*$ et $x_1, \ldots, x_n \in X$ tels que

$$zvuw = x_1 \ldots x_n.$$

Si $x_1 \in \mathrm{Pref}(z)$, alors $yx_1 \in \mathrm{Pref}(u) \cap yX$ donc $yx_1 \in P_{u,v}(y)$. Sinon, posons $m \leq n$ minimal tel que $zv \in \mathrm{Pref}(x_1 \ldots x_m)$ et notons

$$zvp = x_1 \ldots x_m.$$

Par construction, comme nous ne sommes pas dans le cas dégénéré où $zv = \varepsilon$, p est un suffixe propre de x_m. En conséquence, u ne peut pas être un préfixe propre de p car cela signifierait que $u \in \mathrm{IFac}(X)$, ce qui contredit le fait que $(u, v) \in C(X, S)$. Étant donné que $p \in \mathrm{Pref}(uw)$, on a donc $p \in \mathrm{Pref}(u)$. Par hypothèse, z est un préfixe propre de x_1 donc p vérifie la seconde condition de la Définition 2.1.7 et $P_{u,v}(y) \neq \emptyset$. □

Lemme 2.1.13. *Soient S un ensemble récurrent et $X \subseteq S$ S-fin. Si pour tous $(u, v) \in C(X, S)$ et $y \in \mathrm{Pref}(u)$,*

$$P_{u,v}(y) \neq \emptyset,$$

alors X est S-complet à droite.

Démonstration. Par la Proposition 2.1.5, montrons que tout $x \in S$ est préfixe d'un mot de X ou a un mot de X pour préfixe. Notons $w \in S \setminus \mathrm{Fac}(X)$. L'ensemble S étant récurrent, il existe $q \in A^*$ tel que $xqw \in S$. Posons $u = xqw$. À nouveau, il existe $v \in A^+$ tel que $uvu \in S$. [(2)] Comme $w \notin \mathrm{Fac}(X)$, $u \notin \mathrm{IFac}(X)$ donc $(u, v) \in C(X, S)$. Par hypothèse, pour $y = \varepsilon$, il existe $p \in P_{u,v}(y)$.

1. Si $p \in \varepsilon X = X$, alors la conclusion est immédiate car x et p sont tous deux des préfixes de u.
2. Si $uvp = x_1 \ldots x_n$ où u est un préfixe propre de x_1, alors x est également un préfixe de x_1 donc on peut conclure.

□

Nous énonçons ici simplement les résultats équivalents à ces quatre lemmes mais concernant les suffixes.

Définition 2.1.14. Soient X un ensemble, $u, v \in A^*$ et $s \in \mathrm{Suff}(u)$. Si $p \in A^*$ est tel que $ps = u$, alors $P'_{u,v}(s)$ est l'ensemble des suffixes z de u tels que l'une des deux conditions suivantes soit vérifiée

1. $z \in Xs$,
2. $zvp = x_1 \ldots x_n$ où $x_i \in X$, $z \in \mathrm{Pref}(x_1) \setminus \{x_1\}$ et $p \in \mathrm{Suff}(x_n) \setminus \{x_n\}$.

(2). Si $uu \in S$, il suffit de considérer $u' = uu_1$ pour s'assurer que v soit non vide.

Remarque 2.1.15. Si $u = yz = ps$, alors

$$p \in P_{u,v}(y) \;\Leftrightarrow\; z \in P'_{u,v}(s).$$

En effet, les secondes conditions des deux définitions sont identiques et

$$p = yx \;\Leftrightarrow\; u = yxs \;\Leftrightarrow\; z = xs.$$

Lemme 2.1.16. *Soient X un code suffixe et $u, v \in A^*$. Pour tout $s \in \mathrm{Suff}(u)$,*

$$|P'_{u,v}(s)| \leq 1.$$

Lemme 2.1.17. *Soient S un ensemble récurrent et $X \subseteq S$ un code préfixe S-maximal et S-fin. Si $|P'_{u,v}(s)| \leq 1$ pour tous $(u, v) \in C(X, S)$, $s \in \mathrm{Suff}(u)$, alors X est un code suffixe.*

Lemme 2.1.18. *Soient S un ensemble récurrent et $X \subseteq S$ S-complet à gauche. Pour tous $(u, v) \in C(X, S)$ et $s \in \mathrm{Suff}(u)$,*

$$P'_{u,v}(s) \neq \emptyset.$$

Lemme 2.1.19. *Soient S un ensemble récurrent et $X \subseteq S$ S-fin. Si pour tous $(u, v) \in C(X, S)$ et $s \in \mathrm{Suff}(u)$,*

$$P'_{u,v}(s) \neq \emptyset,$$

alors X est S-complet à gauche.

Ces huit lemmes permettent de montrer le résultat suivant.

Proposition 2.1.20. *Soient S un ensemble récurrent et $X \subseteq S$ un code préfixe S-maximal et S-fin. L'ensemble X est un code suffixe si, et seulement si, X est S-complet à gauche.*

Démonstration. Soient $(u, v) \in C(X, S)$. Par les Lemmes 2.1.9 et 2.1.12, étant donné que X est S-complet à droite par la Proposition 2.1.5, pour tout $y \in \mathrm{Pref}(u)$,

$$|P_{u,v}(y)| = 1.$$

Notons alors

$$\varphi_{u,v} : \mathrm{Pref}(u) \to \mathrm{Pref}(u) \quad y \mapsto p \text{ où } P_{u,v}(y) = \{p\}.$$

Si X est un code suffixe, par le Lemme 2.1.16 et par la Remarque 2.1.15, pour tout $p \in \mathrm{Pref}(u)$, il existe au plus un $y \in \mathrm{Pref}(u)$ tel que

$$p = \varphi_{u,v}(y).$$

Autrement dit la fonction $\varphi_{u,v}$ est injective. L'ensemble $\mathrm{Pref}(u)$ étant fini, il s'agit alors d'une bijection. Pour tout $p \in \mathrm{Pref}(u)$, il existe donc $y \in \mathrm{Pref}(u)$ tel que $p = \varphi_{u,v}(y)$, ce qui se traduit par

$$\forall\, s \in \mathrm{Suff}(u) \quad P'_{u,v}(s) \neq \emptyset$$

via la Remarque 2.1.15. C'est valable pour tous $(u, v) \in C(X, S)$ donc, par le Lemme 2.1.19, X est S-complet à gauche.

Réciproquement, par le Lemme 2.1.18, si X est S-complet à gauche, la fonction $\varphi_{u,v}$ est surjective donc elle est injective. On en déduit, par le Lemme 2.1.17, que X est un code suffixe. □

Nous pouvons à présent montrer qu'un code bifixe S-maximal X est également maximal parmi les codes préfixes et les codes suffixes, à condition que S soit récurrent et que X soit S-fin.

Proposition 2.1.21. *Soient S un ensemble récurrent et $X \subseteq S$ un ensemble S-fin. Les affirmations suivantes sont équivalentes.*

1. *L'ensemble X est un code bifixe S-maximal.*
2. *L'ensemble X est un code préfixe S-complet à gauche.*
3. *L'ensemble X est un code suffixe S-complet à droite.*
4. *L'ensemble X est un code préfixe S-maximal et un code suffixe S-maximal.*
5. *L'ensemble X est un code bifixe S-maximal parmi les codes préfixes.*
6. *L'ensemble X est un code bifixe S-maximal parmi les codes suffixes.*

Démonstration. Les implications $5 \Rightarrow 1$ et $6 \Rightarrow 1$ sont immédiates, de même que $4 \Rightarrow 5$ et $4 \Rightarrow 6$. De plus, les affirmations 2 et 3 sont symétriques donc montrons uniquement $1 \Rightarrow 2 \Rightarrow 4$.

$1 \Rightarrow 2$: Si X est S-maximal parmi les codes suffixes, la conclusion découle directement de la Proposition 2.1.5. S'il est S-maximal parmi les codes préfixes, on peut conclure directement grâce à la Proposition 2.1.20. Si X n'est ni maximal parmi les codes préfixes, ni parmi les codes suffixes, alors il existe $x, y \in S \setminus X$ tels que $X \cup \{x\}$ soit un code préfixe et $X \cup \{y\}$ un code suffixe. L'ensemble S étant récurrent, il existe $w \in S$ tel que $xwy \in S$. Dans ce cas, $X \cup \{xwy\}$ est un code bifixe par la Remarque 2.1.4, ce qui est absurde.

$2 \Rightarrow 4$: Posons

$$Y = X \setminus A^+X.$$

Comme $Y \subseteq X$, il s'agit d'un code préfixe par hypothèse. De plus, par la Remarque 2.1.4, c'est un code suffixe. Montrons qu'il est S-maximal parmi les codes suffixes. On a

$$A^*Y = A^*X$$

donc on peut conclure par la Proposition 2.1.5 car X est S-complet à gauche donc A^*X est S-dense à gauche. Montrons à présent qu'il est S-maximal parmi les codes préfixes. Pour cela, il suffit que X soit S-complet à droite, ce qui est effectivement le cas par la Proposition 2.1.20. Or, $Y \subseteq X$ et X est un code préfixe. On a donc $Y = X$, ce qui permet de conclure que X est S-maximal parmi les codes préfixes et les codes suffixes. □

2.2 Premiers résultats sur la cardinalité

Nous montrons ici un résultat concernant la cardinalité d'un code préfixe en fonction des prolongements à droite de ses préfixes propres. Un résultat symétrique peut être obtenu concernant les codes suffixes et les prolongements à gauche de leurs suffixes propres.

Définition 2.2.1. Soit $f : S \to [0, +\infty[$. On notera, pour tout $X \subseteq S$ fini,

$$f(X) = \sum_{x \in X} f(x).$$

On étend cette définition à tout $X \subseteq S$ par [3]

$$f(X) = \lim_{n\to\infty} f(X \cap A^{\leq n}),$$

ce que nous noterons parfois abusivement

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

même dans le cas où X est infini. Remarquons que cette limite existe toujours mais qu'elle n'est pas forcément finie. En effet, f est à valeurs positives donc

$$f(X \cap A^{\leq n+1}) \geq f(X \cap A^{\leq n})$$

et $f(X)$ est la limite d'une suite croissante.

Par exemple, dans le cas de la fonction $|\cdot| : x \mapsto 1$, $|X|$ correspond bien au cardinal de X.

Lemme 2.2.2. *Soit X un code préfixe. Notons P l'ensemble des préfixes propres de X et, pour tout $q \in P$, $\alpha(q) = r_{X\cup P}(q) - 1$. Si X est non vide, alors*

$$|X| = 1 + \sum_{q\in P} \alpha(q).$$

Démonstration. Notons

$$C_k = X \cap A^k, \quad P_k = P \cap A^k, \quad C_{\leq k} = \bigcup_{i\leq k} C_i \quad \text{et} \quad P_{\leq k} = \bigcup_{i\leq k} P_i.$$

On a alors

$$C_{k+1} \cup P_{k+1} = \{qa \mid q \in P_k, a \in R_{X\cup P}(q)\}$$

donc, comme C_{k+1} et P_{k+1} sont disjoints (car X est un code préfixe),

$$|C_{k+1}| + |P_{k+1}| = \sum_{q\in P_k} r_{X\cup P}(q).$$

Par définition de α, on obtient

$$|C_{k+1}| + |P_{k+1}| - |P_k| = \sum_{q\in P_k} \alpha(q)$$

pour tout $k \in \mathbb{N}$. On a alors, pour tout $K \in \mathbb{N}$,

$$\begin{aligned}\sum_{q\in P_{\leq K}} \alpha(q) &= \sum_{k=0}^{K} (|C_{k+1}| + |P_{k+1}| - |P_k|) \\ &= |C_{\leq K+1}| + |P_{K+1}| - |P_0| \\ &= |C_{\leq K+1}| + |P_{K+1}| - 1\end{aligned}$$

car X est non vide. Donc [4]

$$|C_{\leq K+1}| + |P_{K+1}| = 1 + \sum_{q\in P_{\leq K}} \alpha(q).$$

(3). On note $A^{\leq n} = \cup_{i\leq n} A^i$.

(4). Remarquons que cette égalité peut aussi être obtenue en considérant la restriction à la profondeur K de l'arbre ayant pour sommets les mots de X et P et pour arêtes les couples (u, v) pour lesquels il existe $a \in A$ tel que $v = ua$.

- Si X est fini, alors, pour $K = \max\{|w| \mid w \in X\} - 1$, on a

$$X = C_{\leq K+1}, \quad P_{K+1} = \emptyset \quad \text{et} \quad P = P_{\leq K}$$

donc

$$|X| = 1 + \sum_{q \in P} \alpha(q).$$

- Si X est infini, par la remarque suivant la Définition 2.2.1, on a

$$|X| = \lim_{K \to \infty} |C_{\leq K+1}|.$$

De plus, $|P_{K+1}|$ est positif donc [(5)]

$$\begin{aligned} |X| &= \lim_{K \to \infty} |C_{\leq K+1}| + |P_{K+1}| \\ &= \lim_{K \to \infty} 1 + \sum_{q \in P_{\leq K}} \alpha(q) \\ &= 1 + \sum_{q \in P} \alpha(q). \end{aligned}$$

□

Proposition 2.2.3. *Soient S un ensemble factoriel (non vide) et $X \subseteq S$ un code préfixe S-maximal. Si P est l'ensemble des préfixes propres de X, alors*

$$|X| = 1 + \sum_{q \in P} (r_S(q) - 1).$$

Démonstration. Vu le lemme précédent, montrer que $R_{X \cup P}(q) = R_S(q)$ pour tout $q \in P$ est suffisant pour conclure. Comme S est factoriel, $X \cup P \subseteq S$ donc $R_{X \cup P}(q) \subseteq R_S(q)$. Montrons l'autre inclusion. Soit $a \in R_S(q)$. Si $qa \notin X \cup P$, alors $X \cup \{qa\}$ est un code préfixe. En effet, les préfixes propres de qa sont préfixes de $q \in P$ donc ils sont également dans P. Comme X est un code préfixe, ils ne peuvent pas être dans X. De plus, $qa \notin P$ donc il ne peut pas être préfixe propre d'un élément de X. C'est donc absurde car X est maximal. □

Remarquons qu'il suffisait que S soit stable pour les préfixes et pas factoriel pour avoir ce résultat.

2.3 Mots de retour

Nous introduisons à présent le deuxième nouveau concept de ce chapitre : les mots de retour d'un mot w dans un ensemble factoriel.

Définition 2.3.1. Soit S un ensemble factoriel. Pour tout $w \in S \setminus \{\varepsilon\}$, l'ensemble des *mots de retour* de w est

$$\Gamma_S(w) = \{u \in A^+ \mid wu \in S, w \in \mathrm{Suff}(wu)\},$$

(5). Comme $\alpha(q) \geq 0$ pour tout $q \in P$, on peut utiliser les notations introduites dans la Définition 2.2.1.

l'ensemble des *mots de premier retour* de w est

$$\mathcal{R}_S(w) = \{u \in A^+ \mid wu \in S, w \in \mathrm{Suff}(wu) \setminus \mathrm{IFac}(wu)\}$$

et l'ensemble des *mots de premier retour complets* de w est

$$\mathcal{CR}_S(w) = \{wu \in S \mid u \in A^+, w \in \mathrm{Suff}(wu) \setminus \mathrm{IFac}(wu)\}.$$

Les deux premiers termes sont parfois accompagnés de la locution « à droite » pour les distinguer des notions symétriques

$$\Gamma'_S(w) := \{u \in A^+ \mid uw \in S, w \in \mathrm{Pref}(uw)\} = \{u \in A^+ \mid uw \in w\Gamma_S(w)\}$$

et

$$\mathcal{R}'_S(w) = \{u \in A^+ \mid uw \in S, w \in \mathrm{Pref}(uw) \setminus \mathrm{IFac}(uw)\} = \{u \in A^+ \mid uw \in w\mathcal{R}_S(w)\}$$

qui, elles, sont qualifiées de « à gauche ». Dans ce travail, nous travaillerons majoritairement avec les mots de (premier) retour à droite mais, moyennant l'adaptation des preuves, les résultats seront également vrais pour les mots de (premier) retour à gauche.

Ces notions peuvent également être définie pour un mot infini u en considérant l'ensemble $\mathrm{Fac}(u)$.

Exemple 2.3.2. Pour rappel, le mot de Tribonacci (Exemple 1.4.7) est le mot

$$x = abacabaabacababacabaabacabac\cdots$$

Deux occurrences successives de a étant soit consécutives, soit séparées par un b ou un c, on a

$$\mathcal{R}_x(a) = \{a, ba, ca\}, \quad \mathcal{CR}_x(a) = \{aa, aba, aca\} \quad \text{et} \quad \mathcal{R}'_x(a) = \{a, ab, ac\}.$$

On peut aussi remarquer que les mots

$$acab, aab, ab, acabaab, abacab, aabacab, aabacabab, \ldots$$

sont des mots de retour (à droite) pour ab. Les trois premiers sont mêmes des mots de premier retour. Nous montrons plus loin dans ce chapitre que ce sont les seuls mots de premier retour pour ab dans le mot de Tribonacci.

Faisons à présent quelques observations plus ou moins immédiates concernant ces concepts.

Remarque 2.3.3.

1. Si S est récurrent, aucun des trois ensembles de la Définition 2.3.1 n'est vide.
2. Il y a évidemment une bijection entre $\mathcal{R}_S(w)$ et $\mathcal{CR}_S(w)$ puisque

   $$\mathcal{CR}_S(w) = w\mathcal{R}_S(w)$$

   mais les deux notations seront utilisées dans la suite.
3. Remarquons que $\mathcal{R}_S(w)$ et $\mathcal{CR}_S(w)$ sont des codes préfixes. De fait, si $x, y \in \mathcal{R}_S(w)$ sont tels que x est préfixe propre de y, alors w est suffixe propre de wx qui est préfixe propre de wy donc w est facteur interne de wy, ce qui est absurde. Le raisonnement pour $\mathcal{CR}_S(w)$ est le même.

4. Tout mot de retour se factorise en mots de premier retour, i.e $\Gamma_S(w) \subseteq (\mathcal{R}_S(w))^*$. De fait, si $u \in \Gamma_S(w)$, alors notons $x_0 < \cdots < x_n$ les indices de fin des différentes occurrences de w dans wu. Par définition, $x_0 = |w|$ et $x_n = |wu|$. Si $u_i = (wu)_{[x_i+1,x_{i+1}]}$ pour tout $i < n$, alors
$$u = u_0 \dots u_{n-1}.$$
De plus, par définition, $w \in \mathrm{Suff}(wu_i) \setminus \mathrm{IFac}(wu_i)$ donc $u_i \in \mathcal{R}_S(w)$ et $u \in (\mathcal{R}_S(w))^*$.
5. Si S est récurrent, l'ensemble $\mathcal{R}_S(w)$ est même un code préfixe $w^{-1}S$-maximal car, pour tout $u \in w^{-1}S$, il existe $v \in S$ tel que $wuvw \in S$ donc u est préfixe d'un mot de retour pour w et, par le point précédent, est soit préfixe d'un mot de premier retour, soit a un mot de premier retour pour préfixe.
6. On a
$$\mathcal{CR}_S(w) = \mathcal{CR}_{A^*}(w) \cap S$$
et
$$u \in \mathcal{R}_S(w) \;\Leftrightarrow\; u \in \mathcal{R}_{A^*}(w) \text{ et } wu \in S.$$

Les mots de premier retour sont liés au concept de récurrence de la façon suivante.

Proposition 2.3.4. *Un ensemble récurrent S est uniformément récurrent si et seulement si $\mathcal{R}_S(w)$ est fini pour tout $w \in S \setminus \{\varepsilon\}$.*

Démonstration. Supposons S uniformément récurrent. Soit $w \in S$. Par définition, il existe $n \in \mathbb{N}$ tel que w soit facteur de tout mot de $S \cap A^n$. Pour tout $u \in A^{>n}$, w est facteur non suffixe [(6)] de u donc u n'est pas un mot de premier retour. On a alors
$$\mathcal{R}_S(w) \subseteq A^{\leq n}$$
qui est fini. Réciproquement, si $\mathcal{R}_S(w)$ est fini, alors posons
$$n = \max\{|u| \mid u \in \mathcal{R}_S(w)\}.$$
Montrons que w est facteur de tous les mots de $S \cap A^{n+|w|}$. Soit $u \in S \cap A^{n+|w|}$. Comme S est récurrent, il existe $v, v' \in S$ tel que $wvuv'w \in S$. Par définition de n, deux occurrences consécutives de w dans $wvuv'w$ sont le décalage l'une de l'autre d'au plus n lettres. Vu la longueur de u, w est bien facteur de u. □

Il existe bien évidemment un résultat identique pour $\mathcal{CR}_S(w)$.

2.4 Récurrence dans les ensembles neutres

Dans le cadre d'un ensemble neutre récurrent, le cardinal de l'ensemble des mots de premier retour est connu, comme nous le montrons dans cette section. Ceci nous permet alors d'affirmer que tout ensemble neutre récurrent est uniformément récurrent.

Commençons par montrer un résultat tiré de [4]. Ce résultat concerne à l'origine les distributions de probabilité. Nous n'abordons pas ce sujet ici mais nous invitons le lecteur intéressé à consulter [4] pour plus d'informations sur les distributions de probabilité.

(6). On entend par là que w peut éventuellement être suffixe de u mais qu'il doit apparaître à un autre endroit donc soit comme préfixe, soit comme facteur interne.

Lemme 2.4.1. *Soit S un ensemble factoriel et soit $\pi : S \to [0, +\infty[$ une application telle que, pour tout $w \in S$,*

$$\sum_{a \in L(w)} \pi(aw) = \pi(w).$$

Si $X \subseteq S$ est un code suffixe, alors

$$\pi(X) \leq \pi(\varepsilon).$$

Si, de plus, X est fini et S-maximal, alors on a l'égalité.

Démonstration. Commençons par faire remarquer que, pour tout $w \in S$,

$$Aw \cap S = \bigcup_{a_1 \in L(w)} \{a_1 w\}$$

donc, comme S est factoriel, on a

$$\begin{aligned} A^2 w \cap S &= \bigcup_{a_1 \in L(w)} A a_1 w \cap S \\ &= \bigcup_{a_1 \in L(w)} \bigcup_{a_2 \in L(a_1 w)} \{a_2 a_1 w\} \end{aligned}$$

où l'union est disjointe. On peut étendre ce résultat à $A^k w$ pour tout $k \in \mathbb{N}$. On a alors

$$\begin{aligned} \pi(A^k w \cap S) &= \sum_{x \in A^k w \cap S} \pi(x) \\ &= \sum_{a_1 \in L(w)} \sum_{a_2 \in L(a_1 w)} \cdots \sum_{a_k \in L(a_{k-1} \ldots a_2 a_1 w)} \pi(a_k \ldots a_2 a_1 w) \\ &= \sum_{a_1 \in L(w)} \sum_{a_2 \in L(a_1 w)} \cdots \sum_{a_{k-1} \in L(a_{k-2} \ldots a_2 a_1 w)} \pi(a_{k-1} \ldots a_2 a_1 w) \\ &= \cdots \\ &= \sum_{a_1 \in L(w)} \pi(a_1 w) \\ &= \pi(w) \end{aligned} \tag{2.1}$$

par hypothèse sur π.

Passons maintenant à la preuve proprement dite et supposons X fini dans un premier temps. Dans ce cas, notons

$$n = \max\{|w| \mid w \in X\}.$$

Pour tous $x, y \in X$ tels que $x \neq y$, on a

$$A^{n-|x|} x \cap A^{n-|y|} y = \emptyset.$$

De fait, sinon, il existerait un mot w ayant x et y comme suffixes, ce qui contredit le fait que X est un code suffixe. On a donc

$$\bigcup_{x \in X} (A^{n-|x|} x \cap S) \subseteq A^n \cap S$$

où l'union est disjointe, ce qui implique

$$\begin{aligned}
\pi(X) &= \sum_{x\in X} \pi(x) \\
&= \sum_{x\in X} \pi(A^{n-|x|}x \cap S) && \text{par (2.1)} \\
&= \pi\left(\bigcup_{x\in X} (A^{n-|x|}x \cap S)\right) && \text{car l'union est disjointe} \\
&\leq \pi(A^n \cap S) \\
&= \pi(\varepsilon) && \text{par (2.1).}
\end{aligned}$$

Pour avoir l'égalité, il suffit d'avoir

$$\bigcup_{x\in X} (A^{n-|x|}x \cap S) = A^n \cap S. \tag{2.2}$$

Montrons que c'est le cas quand X est S-maximal. L'égalité (2.2) est équivalente à demander que tout mot de $A^n \cap S$ ait un suffixe dans X. Or, les mots de $A^n \cap S$ ne peuvent pas être suffixes propres d'éléments de X par définition de n. On a donc la conclusion par la Proposition 2.1.5.

Dans le cas où X est infini, par la Définition 2.2.1, on a

$$\begin{aligned}
\pi(X) &= \lim_{n\to\infty} \pi(X \cap A^{\leq n}) \\
&\leq \pi(\varepsilon)
\end{aligned}$$

car $X \cap A^{\leq n} \subseteq X$ est soit vide, soit un code suffixe fini. □

Nous pouvons à présent démontrer le résultat principal de cette section.

Théorème 2.4.2. *Soit S un ensemble neutre récurrent. Pour tout $x \in S \setminus \{\varepsilon\}$,*

$$|\mathcal{CR}_S(x)| = |A|.$$

Démonstration. Soit $x \in S$. Reprenons les notations du lemme 2.2.2 pour $X = \mathcal{CR}_S(x)$. Remarquons que, comme S est récurrent, $X \neq \emptyset$. Si P' désigne l'ensemble des préfixes propres de x, alors $P' \subseteq P$ et, pour tout $q \in P'$,

$$\alpha(q) = r_{X\cup P}(q) - 1 = 0 \tag{2.3}$$

car, pour tout $a \in A$, $qa \in X \cup P$ si et seulement si qa est préfixe de x donc il n'y a qu'une valeur possible pour a. Notons $Y = P \setminus P'$. On a alors

$$\begin{aligned}
Y &= \{xw \in S \mid \exists\ v \neq \varepsilon \text{ tq. } xwv \in X\} \\
&= \{xw \in S \mid \exists\ v \neq \varepsilon \text{ tq. } x \in \mathrm{Suff}(xwv) \text{ et } x \notin \mathrm{IFac}(xwv)\} \\
&= \{xw \in S \mid x \notin \mathrm{Fac}(x_{[2,|x|]}w)\}
\end{aligned}$$

car S est récurrent. (7) Montrons que, pour tout $q \in Y$,

$$R_{X\cup P}(q) = R_S(q). \tag{2.4}$$

(7). Plus précisément, pour tout $w \in S$ tel que $xw \in S$, il existe $v' \in S$ tel que $xwv'x \in S$. Il suffit alors de prendre comme v le préfixe de $v'x$ de longueur minimale tel que $x \in \mathrm{Suff}(xwv)$.

Pour cela, procédons par double inclusion. Si $a \in R_{X\cup P}(q)$, alors $qa \in S$ car S est factoriel donc $X \cup P \subseteq S$. On a donc bien $a \in R_S(q)$. Réciproquement, si $a \in R_S(q)$, alors $qa \in S$ donc, comme S est récurrent, il existe $u \in S$ tel que

$$qauqa \in S.$$

Étant donné que $q \in Y$, il existe $w \in S$ tel que $q = xw$. On a donc

$$xwaux \in S$$

car S est factoriel. Autrement dit, $waux$ est un mot de retour pour x. Le mot x n'est pas facteur de

$$x_{[2,|x|]}w$$

car $q \in Y$. Donc, si u' est le plus petit préfixe (éventuellement vide) de ux tel que

$$x \in \mathrm{Suff}(xwau'),$$

alors $xwau'$ est un mot de premier retour complet pour x donc

$$qa = xwa \in X \cup P$$

et $a \in R_{X\cup P}(q)$.

En conséquence, par le Lemme 2.2.2, on a [8]

$$\begin{aligned} |X| &= 1 + \sum_{q\in P} \alpha(q) \\ &= 1 + \sum_{q\in P'} \alpha(q) + \sum_{q\in Y} \alpha(q) \\ &= 1 + \sum_{q\in Y} (r_{X\cup P}(q) - 1) \\ &= 1 + \sum_{q\in Y} (r_S(q) - 1) \end{aligned}$$

vu (2.3) et (2.4).

Notons

$$\pi(q) := r_S(q) - 1 \geq 0$$

pour tout $q \in S$. L'ensemble S étant neutre, on a

$$\begin{aligned} \pi(q) &= r_S(q) - 1 \\ &= e_S(q) - l_S(q) \\ &= \sum_{a\in L_S(q)} (r_S(aq) - 1) \\ &= \sum_{a\in L_S(q)} \pi(aq). \end{aligned}$$

(8). Même remarque que dans la note (5).

Montrons que Y est un code suffixe. Par l'absurde, supposons avoir $u, v \in Y$ tels que u soit un suffixe propre de v. Par définition de Y, x est un préfixe de u donc un facteur de $v_{[2,|v|]}$, ce qui est absurde car $v \in Y$.

Les hypothèses du Lemme 2.4.1 sont vérifiées pour π et Y donc

$$\sum_{q \in Y} (r_S(q) - 1) \leq \pi(\varepsilon) = |A| - 1,$$

ce qui implique

$$|X| \leq |A|.$$

En particulier, X est fini donc P et Y aussi. Montrons que Y est S-maximal pour pouvoir en déduire l'égalité en appliquant la deuxième partie du Lemme 2.4.1. Soit $w \in S \setminus Y$. Montrons que $Y \cup \{w\}$ n'est pas un code suffixe. Comme S est récurrent, il existe $u \in S$ tel que $xuw \in S$. Si y est le plus court suffixe de xuw ayant x comme préfixe, alors $y \in Y$ par définition de Y. De plus, soit il est suffixe de w, soit il a w pour suffixe. Dans tous les cas, $Y \cup \{w\}$ n'est pas un code suffixe si $y \neq w$. Or, si $y = w$, alors $w \in Y$, ce qui est absurde. On a donc, par le Lemme 2.4.1,

$$|X| = \sum_{q \in Y} (r_S(q) - 1) + 1 = |A|.$$

□

Corollaire 2.4.3. *Si S est neutre et récurrent, alors S est uniformément récurrent.*

Démonstration. Pour tout $x \in S \setminus \{\varepsilon\}$, par le théorème précédent, $\mathcal{CR}_S(x)$ est fini donc S est uniformément récurrent vu la Proposition 2.3.4. □

Exemple 2.4.4. Le mot de Tribonacci étant neutre et récurrent, ce résultat nous dit que tout facteur du mot de Tribonacci x aura exactement 3 mots de premier retour dans $\mathrm{Fac}(x)$, ce qui justifie ce que nous avons affirmé dans l'Exemple 2.3.2.

2.5 S-degré et Théorème de cardinalité

Nous continuons à nous intéresser à la cardinalité de codes inclus dans une ensemble neutre récurrent mais en se penchant cette fois-ci sur les codes bifixes. Pour cela, nous définissons les notions de découpages et de S-degré en montrons certaines de leurs propriétés avant de montrer le résultat principal.

Définition 2.5.1. Un *découpage* d'un mot w par rapport à un ensemble X est un triplet (u, x, v) de mots tels que

- $w = uxv$,
- u n'a pas de suffixe dans X,
- $x \in X^*$,
- v n'a pas de préfixe dans X.

On note $\delta_X(w)$ le nombre de tels découpages.

Exemple 2.5.2. Le seul découpage possible pour le mot ε est donné par $(\varepsilon, \varepsilon, \varepsilon)$. Ce n'est un découpage que si $\varepsilon \notin X$. En réalité, si $\varepsilon \in X$, alors il est impossible de trouver un mot n'ayant pas de suffixe dans X donc

$$\delta_X(w) = 0, \quad \forall\, w \in A^*.$$

On considérera donc que $\varepsilon \notin X$ (ce qui est immédiat si X est un code préfixe ou suffixe).

Dans ce cas, pour tout mot w, il existe $x \in X^*$ et v n'ayant pas de préfixe dans X tels que $w = xv$. Un découpage trivial de w est donc donné par (ε, x, v). De façon symétrique, w a également un découpage de la forme (u, x', ε). On a donc

$$\delta_X(w) \geq 2, \quad \forall\, w \notin X^*.$$

Proposition 2.5.3. *Soit X un code préfixe. Pour tout mot w, il existe une bijection entre les découpages de w par rapport à X et les préfixes de w n'ayant pas de suffixes dans X. En particulier, pour tout $a \in A$,*

$$\delta_X(wa) = \begin{cases} \delta_X(w) & \text{si } wa \in A^*X, \\ \delta_X(w) + 1 & \text{sinon.} \end{cases}$$

De plus, si X est un code bifixe,

$$\delta_X(w) = |w| + 1 - |\{(x, y, z) \mid y \in X, xyz = w\}|.$$

Démonstration. À un découpage (u, x, v), on peut associer u qui est un préfixe de w n'ayant pas de suffixes dans X. Réciproquement, si u est un tel préfixe, notons $w = uw'$. Par la Remarque 2.1.4, w' se décompose de façon unique en $w' = xv$ où $x \in X^*$ et v n'a pas de préfixe dans X. On peut donc associer le triplet (u, x, v) à u. On a alors

$$\begin{aligned} \delta_X(wa) &= |\{u \in \mathrm{Pref}(wa) \mid u \notin A^*X\}| \\ &= \begin{cases} |\{u \in \mathrm{Pref}(w) \mid u \notin A^*X\}| & \text{si } wa \in A^*X \\ |\{u \in \mathrm{Pref}(w) \mid u \notin A^*X\}| + 1 & \text{sinon} \end{cases} \\ &= \begin{cases} \delta_X(w) & \text{si } wa \in A^*X \\ \delta_X(w) + 1 & \text{sinon.} \end{cases} \end{aligned}$$

De plus, à toute factorisation xyz de w telle que $y \in X$, on peut faire correspondre le préfixe $u = xy$ de w ayant un suffixe dans X. Réciproquement, si X est un code suffixe, à tout $u \in \mathrm{Pref}(w) \cap A^*X$ correspond un unique $y \in X \cap \mathrm{Suff}(u)$ donc une unique factorisation xyz de w telle que $u = xy$ et $y \in Y$. On a alors

$$\begin{aligned} |w| + 1 &= |\mathrm{Pref}(w)| \\ &= \delta_X(w) + |\{u \in \mathrm{Pref}(w) \cap A^*X\}| \\ &= \delta_X(w) + |\{(x, y, z) \mid y \in X, xyz = w\}|, \end{aligned}$$

ce qui permet de conclure dans le cas où X est un code bifixe. □

On a un résultat symétrique en inversant les notions de préfixe et suffixe.

Proposition 2.5.4. *Soit X un code bifixe. Pour tous mots w, u et v,*

$$\delta_X(w) \leq \delta_X(uwv).$$

De plus, si $u, v \neq \varepsilon$ et $uwv \in X$, l'inégalité est stricte.

Démonstration. Soit (u', x, v') un découpage de w. Par la Remarque 2.1.4, il existe d'uniques u'' n'ayant pas de suffixe dans X et $y \in X^*$ tels que $uu' = u''y$. De même, il existe d'uniques v'' n'ayant pas de préfixe dans X et $z \in X^*$ tels que $v'v = zv''$. On peut donc associer à (u', x, v') le découpage (u'', yxz, v'') de uwv, d'où l'inégalité.

Si u, v sont non vides et si $uwv \in X$, alors le découpage $(\varepsilon, uwv, \varepsilon)$ de uwv ne peut pas être obtenu par le procédé décrit ci-dessus à partir d'un découpage de w donc il y a strictement plus de découpages de uwv que pour w. □

Définition 2.5.5. Soit S un ensemble factoriel. Le *S-degré* d'un ensemble X est

$$d_S(X) := \max_{w \in S} \delta_X(w).$$

Exemple 2.5.6. Soit S un ensemble factoriel infini. Pour tout $n \in \mathbb{N}$, l'ensemble $S \cap A^n$ est de S-degré n. En effet, pour tout $w \in S$, $\delta_{S \cap A^n}(w)$ est le nombre de préfixes de w n'ayant pas de suffixes dans $S \cap A^n$. Or, comme S est factoriel, on a les égalités suivantes

$$\begin{aligned}\{u \in \mathrm{Pref}(w) \mid \mathrm{Suff}(u) \cap S \cap A^n = \emptyset\} &= \{u \in \mathrm{Pref}(w) \mid \mathrm{Suff}(u) \cap A^n = \emptyset\} \\ &= \{u \in \mathrm{Pref}(w) \mid |u| < n\}\end{aligned}$$

donc

$$\delta_{S \cap A^n}(w) = \min(n, |w| + 1)$$

et

$$d_S(S \cap A^n) = \max_{w \in S} \min(n, |w| + 1) = n$$

car S est infini.

Exemple 2.5.7. Si A est l'alphabet minimal de S, l'ensemble A est l'unique code bifixe dont le S-degré vaut 1. En effet, si X est un code bifixe de S-degré 1, étant donné que $(\varepsilon, \varepsilon, \varepsilon)$ est l'unique découpage de ε, on doit avoir, pour tout $a \in A$,

$$1 \geq \delta_X(a) \geq \delta_X(\varepsilon) = 1,$$

ce qui, par la Proposition 2.5.3, implique que $a \in A^*X$ donc que $A \subseteq A^*X$. Comme $\varepsilon \notin X$, on en déduit que $A \subseteq X$. On trouve donc bien $X = A$ car X est un code bifixe.

On peut se convaincre rapidement que, si $n \in \mathbb{N}_0$ l'ensemble $S \cap A^n$ est un code bifixe S-maximal. L'Exemple 2.5.6 a permis de constater que son S-degré était fini et que, pour tout $w \in S$,

$$\begin{aligned}\delta_{S \cap A^n}(w) < d_S(S \cap A^n) &\Leftrightarrow |w| + 1 < n \\ &\Leftrightarrow |w| \leq n - 2 \\ &\Leftrightarrow w \in \mathrm{IFac}(S \cap A^n).\end{aligned}$$

Ce résultat n'est pas uniquement vrai pour $S \cap A^n$, comme le montre la proposition suivante.

Proposition 2.5.8. *Soit S un ensemble récurrent. Un code bifixe $X \subseteq S$ est S-maximal S-fin si, et seulement si, le S-degré $d_S(X)$ est fini. Dans ce cas, on a*

$$\mathrm{IFac}(X) = \{w \in S \mid \delta_X(w) < d_S(X)\}.$$

Démonstration. Supposons que X est un code bifixe S-maximal et S-fin. L'ensemble X étant S-fin, il existe $u \in S \setminus \mathrm{IFac}(X)$. Soit $w \in S$. Montrons que $\delta_X(w) \leq \delta_X(u)$ car on pourra en déduire que $d_S(X)$ est fini. L'ensemble S étant récurrent, il existe v tel que $uvw \in S$. Par la Proposition 2.1.5, on a alors

$$F := S \setminus A^*X = \mathrm{Suff}(X) \setminus X$$

car X est également S-maximal parmi les codes suffixes par la Proposition 2.1.21. Comme u n'est pas un facteur interne de X, il ne peut pas être préfixe propre d'un mot de F. On a donc, par la Proposition 2.5.4 pour les mots w, uv et ε,

$$\begin{aligned}\delta_X(w) &\leq \delta_X(uvw)\\ &= |F \cap \mathrm{Pref}(uvw)|\\ &= |F \cap \mathrm{Pref}(u)|\\ &= \delta_X(u).\end{aligned}$$

En conséquence, $d_S(X) = \delta_X(u)$. Ce résultat étant vrai pour tout $u \in S \setminus \mathrm{IFac}(X)$, on a

$$S \setminus \mathrm{IFac}(X) \subseteq \{w \in S \mid \delta_X(w) = d_S(X)\}$$

ou, de façon équivalente,

$$\{w \in S \mid \delta_X(w) < d_S(X)\} \subseteq \mathrm{IFac}(X).$$

Montrons l'autre inclusion. Soit $w \in \mathrm{IFac}(X)$. Notons $u, v \in A^+$ tels que $uwv \in X$. On a

$$\delta_X(w) < \delta_X(uwv) \leq d_S(X)$$

par la Proposition 2.5.4.

Réciproquement, supposons $d_S(X)$ fini. Il existe alors $w \in S$ tel que $\delta_X(w) = d_S(X)$. Montrons que $w \notin \mathrm{IFac}(X)$. Par l'absurde, s'il existe $u, v \in A^+$ tels que $uwv \in X$, alors, par la Proposition 2.5.4,

$$\delta_X(uwv) > \delta_X(w) = d_S(X),$$

ce qui est absurde. Comme S est biprolongeable, il existe $a, b \in A$ tels que $awb \in S$ mais $awb \notin \mathrm{Fac}(X)$, ce qui permet d'en déduire que X est S-fin. Montrons que X est S-maximal. Par les Propositions 2.1.5 et 2.1.21, il suffit de montrer que $S \subseteq \mathrm{Pref}(XA^*)$. Soit $u \in S$. On peut supposer $u \neq \varepsilon$ car, dans le cas contraire, la conclusion est immédiate. L'ensemble S étant récurrent, il existe $v \in S$ tel que $uvw \in S$. On a alors, si $u = au'$, $a \in A$,

$$\delta_X(au'vw) \geq \delta_X(u'vw) \geq \delta_X(w)$$

donc

$$\delta_X(au'vw) = \delta_X(u'vw) = \delta_X(w).$$

Par la Proposition 2.5.3, $uvw \in XA^*$ donc $u \in \mathrm{Pref}(XA^*)$. □

Lemme 2.5.9. *Soient S un ensemble factoriel biprolongeable (non vide) et $N \in \mathbb{N}$. Si*

$$\alpha, \pi : S \cap A^{\leq N} \to [0, +\infty[$$

sont tels que, pour tout $w \in S \cap A^{\leq N}$,

- *si* $|w| < N$,

$$\pi(w) = \sum_{a \in L(w)} \pi(aw),$$

- *si* $|w| = N$,

$$d = \sum_{u \in \mathrm{Suff}(w)} \alpha(u)$$

pour $d \in [0, +\infty[$ *fixé,*

alors

$$\sum_{w \in S \cap A^{\leq N}} \alpha(w)\pi(w) = d\pi(\varepsilon).$$

Démonstration. Procédons par récurrence sur la valeur de N. Si $N = 0$, alors $S \cap A^{\leq N} = \{\varepsilon\}$ et $\alpha(\varepsilon) = d$ donc la conclusion est immédiate. Supposons le résultat vrai pour N et montrons-le pour $N+1$. Si $aw \in S \cap A^{N+1}$, remarquons que

$$\alpha(aw) = d - \sum_{u \in \mathrm{Suff}(w)} \alpha(u)$$

ne dépend pas de a. On peut donc définir, pour tout $w \in S \cap A^N$, $m(w) = \alpha(aw)$ pour $a \in L(w)$ quelconque et, pour tout $w \in S \cap A^{\leq N}$,

$$\alpha'(w) = \begin{cases} \alpha(w) + m(w) & \text{si } |w| = N, \\ \alpha(w) & \text{sinon.} \end{cases}$$

Cherchons à appliquer l'hypothèse de récurrence. Il est évident que π restreint à $S \cap A^N$ vérifie la première condition et α' vérifie la seconde pour N. On a donc

$$\begin{aligned} d\pi(\varepsilon) &= \sum_{w \in S \cap A^{\leq N}} \alpha'(w)\pi(w) \\ &= \sum_{w \in S \cap A^{\leq N}} \alpha(w)\pi(w) + \sum_{w \in S \cap A^N} m(w)\pi(w) \\ &= \sum_{w \in S \cap A^{\leq N}} \alpha(w)\pi(w) + \sum_{w \in S \cap A^N} \sum_{a \in L(w)} \alpha(aw)\pi(aw) \\ &= \sum_{w \in S \cap A^{\leq N+1}} \alpha(w)\pi(w). \end{aligned}$$

□

Nous pouvons à présent démontrer ce que nous appelons le Théorème de cardinalité. Il permet d'exprimer le cardinal d'un code bifixe S-maximal en fonction de son S-degré dans le cadre d'un ensemble neutre et récurrent S.

Théorème 2.5.10 (Théorème de cardinalité)**.** *Soit S un ensemble neutre récurrent. Si $X \subseteq S$ est un code bifixe S-maximal et S-fin, alors*

$$|X| = 1 + d_S(X)(|A| - 1).$$

Démonstration. Comme S est uniformément récurrent par le Corollaire 2.4.3, tout ensemble S-fin est fini par la remarque suivant la Définition 2.1.6 donc X est fini. Posons

$$N = \max\{|w| \mid w \in X\},$$

$$\pi(w) = r_S(w) - 1,$$

et

$$\alpha(w) = \mathbb{1}_P(w)$$

où P est l'ensemble des préfixes propres de X. Par un raisonnement fait dans la démonstration du Théorème 2.4.2, π vérifie les hypothèses du Lemme 2.5.9. Montrons que c'est également le cas de α. Soit $w \in A^N$. Par définition de N, $w \notin \mathrm{IFac}(X)$. On a alors

$$\begin{aligned} d := d_S(X) &= \delta_X(w) \\ &= |\{u \in \mathrm{Suff}(w) \mid u \notin XA^*\}| \\ &= |\{u \in \mathrm{Suff}(w) \mid u \in P\}| \\ &= \sum_{u \in \mathrm{Suff}(w)} \alpha(u) \end{aligned}$$

car S est un code bifixe S-maximal donc également un code préfixe S-maximal. On a alors

$$\begin{aligned} d_S(X)(|A| - 1) &= d_S(X)\pi(\varepsilon) \\ &= \sum_{w \in S \cap A^{\leq N}} \alpha(w)\pi(w) \\ &= \sum_{w \in P} (r_S(w) - 1) \\ &= |X| - 1. \end{aligned}$$

par le Lemme 2.5.9 et la Proposition 2.2.3, □

Chapitre 3

Groupe libre et mots de retour dans un ensemble dendrique

Ce chapitre introduit de nouvelles notions telles que le groupe libre associé à un alphabet, le graphe de Rauzy d'un ensemble et le groupe décrit par un graphe. Nous utilisons ici ces concepts pour montrer que, dans le cas d'un ensemble dendrique récurrent, l'ensemble des mots de premier retour pour un mot quelconque est une base du groupe libre. Ce résultat, appelé Théorème de retour, est capital pour la suite du travail.

3.1 Groupe libre sur un alphabet

Il est immédiat que l'ensemble des mots sur un alphabet muni de l'opération de concaténation est un monoïde. On peut étendre l'alphabet et la notion de concaténation pour obtenir un groupe. C'est ce qu'on appelle le groupe libre sur un alphabet donné.

Définition 3.1.1. Soit A un alphabet. On peut lui associer un autre alphabet noté A^{-1} qui contient le symbole a^{-1} pour tout $a \in A$. On définit alors la relation d'équivalence $\equiv$ sur $(A \cup A^{-1})^*$ qui respecte la concaténation, i.e

$$u \equiv x \text{ et } v \equiv y \Rightarrow uv \equiv xy$$

et telle que

$$aa^{-1} \equiv \varepsilon \equiv a^{-1}a$$

pour tout $a \in A$. Un mot est *réduit* si, pour tout $a \in A$, il ne contient aucun facteur de la forme aa^{-1} ou $a^{-1}a$. En particulier, les mots de A^* et de $(A^{-1})^*$ sont réduits.

Dans chaque classe d'équivalence, il existe un unique mot réduit w. On dira alors que ce mot est la *réduction* de tous les mots de sa classe, ce qu'on notera $w = \rho(u)$ pour tout $u \in [w]_{\equiv}$.

Définition 3.1.2. Le *groupe libre* F_A sur l'alphabet A est l'ensemble des mots réduits de $(A \cup A^{-1})^*$ muni de l'opération binaire

$$\cdot : (u, v) \mapsto u \cdot v = \rho(uv).$$

Le neutre est alors ε et l'inverse de $u = a_1 \dots a_n$, $a_i \in A \cup A^{-1}$ pour tout i, est le mot

$$u^{-1} = a_n^{-1} \dots a_1^{-1},$$

à condition d'avoir imposé $(a^{-1})^{-1} = a$ pour tout $a \in A$. Remarquons que u est réduit si, et seulement si, u^{-1} l'est. Il s'agit donc bien d'un groupe. De plus, si $u, v \in A^*$,

$$u \cdot v = uv$$

donc l'opération $\cdot$ est bien une extension de l'opération de concaténation.

Les deux définitions qui suivent sont des notions habituelles d'algèbre et ne devraient donc pas surprendre.

Définition 3.1.3. Soit $X \subseteq F_A$. Le *sous-groupe engendré* par X, noté $\langle X \rangle$, est

$$\langle X \rangle = \{\rho(x) \mid x \in (X \cup X^{-1})^*\}$$

où

$$X^{-1} = \{x^{-1} \mid x \in X\}.$$

Définition 3.1.4. Un ensemble $X \subseteq F_A$ est *libre* si, pour tous $x_1, \dots, x_n \in X \cup X^{-1}$ tels que $x_i \neq x_{i+1}^{-1}$,

$$x_1 \cdot \dots \cdot x_n = \varepsilon \Rightarrow n = 0.$$

3.2 Généralités sur les groupes libres

Il existe une approche plus abstraite des groupes libres en algèbre universelle. Nous ne rentrons ici pas dans les détails et ne faisons que donner un bref aperçu de cette autre approche. Le lecteur intéressé est invité à consulter [15].

Définition 3.2.1. Le groupe F est *libre* de *base* $X \subseteq F$ si pour toute fonction φ de X dans un groupe H, il existe une unique façon d'étendre φ en un homomorphisme de F dans H.

L'utilisation du terme *base* n'est pas anodine puisque ce concept est en réalité similaire à ce qu'on pourrait trouver dans les espaces vectoriels, comme le montre le résultat suivant.

Proposition 3.2.2. *Soient $(F, \cdot, \varepsilon)$ un groupe et $X \subseteq F$. Le groupe F est libre de base X si, et seulement si, X engendre F et il n'existe pas de produits triviaux, i.e. pour tous $x_1, \dots, x_n \in X \cup X^{-1}$ tels que $x_i \neq x_{i+1}^{-1}$,*

$$x_1 \cdot \dots \cdot x_n = \varepsilon \Rightarrow n = 0.$$

Démonstration.

$\Rightarrow$ Remarquons que si X n'engendre pas F, alors il existe $w \in F \setminus (X \cup X^{-1})^*$. L'extension de $\varphi : X \to H$ n'est alors pas unique car on peut obtenir une extension pour n'importe quelle valeur de $\varphi(w) \in H$.
Si X engendre F, montrons qu'il n'existe pas de produit trivial. Procédons par l'absurde et supposons qu'il existe $x_1, \dots, x_n \in X \cup X^{-1}, n > 0$, tels que $x_i \neq x_{i+1}^{-1}$ et $x_1 \cdot \dots \cdot x_n = \varepsilon$. On a alors

$$e_H = \varphi(\varepsilon) = \varphi(x_1) \cdot \dots \cdot \varphi(x_n)$$

si φ est un homomorphisme entre F et H. Or, cette égalité n'est pas vraie pour toute fonction $\varphi : X \to H$ donc X n'est pas une base de F.

$\Leftarrow$ Soient H un groupe et $\varphi : X \to H$. Montrons que si X engendre F et n'a pas de produits triviaux, alors on ne peut étendre φ que d'une seule façon sur F. Pour tout $w \in F$, il existe $x_1, \ldots, x_n \in X \cup X^{-1}$ tels que $w = x_1 \cdot \ldots \cdot x_n$. On peut supposer que $x_i \neq (x_{i+1})^{-1}$ (sinon on retire $x_i x_{i+1}$ de cette factorisation). Montrons que cette décomposition est unique pour tout $w \in F$. Si $w = \varepsilon$, alors $n = 0$ par hypothèse donc c'est bien unique. Si $w \neq \varepsilon$, supposons qu'il existe $x_1, \ldots, x_n, y_1, \ldots, y_m \in X \cup X^{-1}$ tels que $x_i \neq (x_{i+1})^{-1}$, $y_j \neq (y_{j+1})^{-1}$ et

$$x_1 \cdot \ldots \cdot x_n = w = y_1 \cdot \ldots \cdot y_m.$$

On a alors

$$x_1 \cdot \ldots \cdot x_n \cdot (y_m)^{-1} \cdot \ldots \cdot (y_1)^{-1} = \varepsilon.$$

Or, $n + m > 0$ car $w \neq \varepsilon$, cela signifie donc que deux élément consécutifs sont inverses l'un de l'autre. Vu les hypothèses, la seule possibilité est $x_n = y_m$ et on peut retirer $x_n \cdot (y_m)^{-1}$. En répétant le raisonnement, on trouve que $n = m$ et $x_i = y_i$ pour tout i donc la décomposition est bien unique.
Le seul prolongement possible de φ en un homomorphisme est alors tels que

$$\varphi(w) = \varphi(x_1) \cdot \ldots \cdot \varphi(x_n)$$

pour tous $x_1, \ldots, x_n \in X \cup X^{-1}$ tels que $w = x_1 \cdot \ldots \cdot x_n$. Il est correctement défini car toutes les décompositions se ramènent à l'unique décomposition réduite en ajoutant ou retirant des facteurs de la forme $x_i \cdot (x_i)^{-1}$, ce qui ne modifie pas $\varphi(x_1) \cdot \ldots \cdot \varphi(x_n)$.
□

Il faut cependant rester prudent car, bien que le résultat précédent évoque le résultat de théorie des espaces vectoriels « Une partie X d'un espace vectoriel est une base si elle est génératrice et libre », il n'y a pas d'équivalents aux résultats « Toute partie génératrice contient une base » et « Toute partie libre est incluse dans une base ». En effet, $(\mathbb{Z}, +)$ est un groupe libre dont les seules bases sont $\{1\}$ et $\{-1\}$ donc $\{2\}$, bien que « libre » (sans produit trivial) n'est inclus dans aucune base. De façon similaire, $\{2, 3\}$ engendre $\mathbb{Z}$ mais ne contient pas de base.

Certains résultats sont tout de même conservés (après une éventuelle adaptation). Entre autres,

Proposition 3.2.3. *Toutes les bases d'un groupe libre ont le même cardinal qui est alors appelé* rang *du groupe libre.*

Proposition 3.2.4. *Soit F un groupe libre de base X. Si $Y \subseteq F$ engendre F et si $|X| = |Y|$, alors Y est une base de F.*

Ces deux résultats ne sont pas démontrés ici car cela sortirait du cadre de ce travail (à nouveau, nous vous invitons à consulter [15]) mais ils nous seront utiles par la suite.

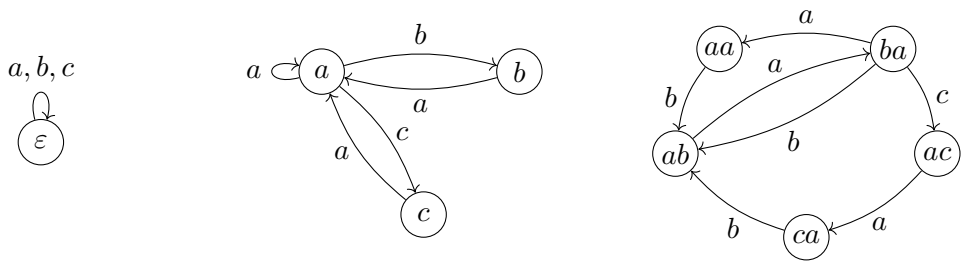
3.3 Graphes de Rauzy

Les graphes de Rauzy d'un ensemble (ou d'un mot infini) ont été introduits par Gérard Rauzy en 1983. Il s'agit d'un outil important pour décrire les mots (ou facteurs) d'une longueur donnée.

Définition 3.3.1. Soit S un ensemble factoriel. Le *graphe de Rauzy* d'ordre $n \in \mathbb{N}$ de S est le graphe étiqueté $G_n(S) = (V, E)$ dont les sommets sont les mots de $S \cap A^n$ et tel que (u, v) est une arête d'étiquette $a \in A$ (ce qu'on notera $(u, a, v) \in E$) si $ua \in S \cap Av$.

Exemple 3.3.2. Par l'Exemple 1.6.2, les graphes de Rauzy de petits ordres pour l'ensemble T des facteurs du mot de Tribonacci sont

$G_0(T)$ $\qquad$ $G_1(T)$ $\qquad$ $G_2(T)$



Nous allons à présent définir une relation d'équivalence sur les mots de longueur fixée dans un ensemble factoriel et montrer que cette relation donne un lien entre les graphes de Rauzy d'ordres successifs.

Définition 3.3.3. Soit S un ensemble factoriel. On définit la relation θ_n sur $S \cap A^n$ par $u \sim_{\theta_n} v$ s'il existe $w \in S$ et $a, b \in L(w)$ tels que $u = aw$, $v = bw$ et tels que a et b [(1)] soient reliés par un chemin dans le graphe d'extensions $\mathcal{E}(w)$.

On peut vérifier aisément qu'il s'agit d'une relation d'équivalence.

Définition 3.3.4. Soit $G = (V, E)$ un graphe étiqueté et σ une relation d'équivalence sur V. Le *quotient* de G par σ est le graphe étiqueté G/σ dont les sommets sont les classes d'équivalence pour σ et qui contient l'arête (x, a, y) pour x, y des classes d'équivalence et $a \in A$ si, et seulement si, il existe $u \in x$, $v \in y$ tels que $(u, a, v) \in E$.

Définition 3.3.5. Soient deux graphes étiquetés $G = (V, E)$ et $H = (V', E')$. Un *morphisme* φ de G dans H est une application de V dans V' telle que

$$(x, a, y) \in E' \quad \Leftrightarrow \quad \exists\, u, v \in V \text{ tq. } \varphi(u) = x, \varphi(v) = y \text{ et } (u, a, v) \in E.$$

Si φ est une bijection entre V et V', on parle *d'isomorphisme*.

Proposition 3.3.6. *Si S est dendrique, alors, pour tout $n \in \mathbb{N}_0$, le graphe $G_n(S)/\theta_n = (V, E)$ est isomorphe à $G_{n-1}(S) = (V', E')$.*

Démonstration. Considérons

$$\varphi : S \cap A^n \to S \cap A^{n-1} \quad u \mapsto u_{[2,n]}.$$

Comme S est factoriel, cette opération est correctement définie. Montrons qu'elle détermine un isomorphisme entre les deux graphes. Dans la suite, pour alléger les notations, on écrira $[x]$ pour désigner la classe d'équivalence de x pour la relation θ_n.

(1). Il s'agit ici des sommets correspondant à a et b vus comme éléments de $L(w)$.

- Si $u \sim_{\theta_n} v$, alors $\varphi(u) = \varphi(v)$ donc on peut définir φ pour les classes d'équivalence pour θ_n en posant
$$\varphi([u]) = \varphi(u) \in V'.$$
- Par définition, pour tous $u, v \in S \cap A^{n-1}$, $a \in A$, on a
$$\begin{aligned}(u,a,v) \in E' &\Leftrightarrow ua \in S \cap Av \\ &\Leftrightarrow \exists\, u', v' \in S \cap A^n \text{ tq. } \varphi(u') = u, \varphi(v') = v, u'a \in S \cap Av' \text{ (2)} \\ &\Leftrightarrow \exists\, u', v' \in S \cap A^n \text{ tq. } \varphi([u']) = u, \varphi([v']) = v, ([u'], a, [v']) \in E \\ &\Leftrightarrow \exists\, x, y \in V \text{ tq. } \varphi(x) = u, \varphi(y) = v, (x, a, y) \in E,\end{aligned}$$
ce qui revient à dire que φ est un morphisme de graphes étiquetés.
- Comme S est biprolongeable, pour tout $u \in S \cap A^{n-1}$, il existe $v \in S \cap A^n$ tel que u soit suffixe de v. On a alors
$$u = \varphi(v) = \varphi([v])$$
donc φ est surjectif.
- Si $[u], [v] \in V$ sont tels que $\varphi([u]) = \varphi([v])$, alors u et v ont le même suffixe w de longueur $n-1$ donc il existe $a, b \in L(w)$ tels que $u = aw$ et $v = bw$. Or, S est dendrique donc $\mathcal{E}(w)$ est connexe. Par définition de θ_n, on a donc
$$u \sim_{\theta_n} v,$$
ce qui signifie que $[u] = [v]$ et donc que φ est injectif.

L'application φ est alors un isomorphisme entre les graphes $G_n(S)/\theta_n$ et $G_{n-1}(S)$. □

Remarquons que seul le caractère connexe de $\mathcal{E}(w)$ a été utilisé dans cette démonstration et pas le caractère acyclique.

3.4 Groupes décrits par un automate et par un graphe

Nous allons maintenant nous intéresser à des sous-groupes particuliers du groupe libre. Ils sont définis à partir d'automates ou de graphes étiquetés. En particulier, nous allons montrer que, dans le cas d'un ensemble dendrique récurrent, le sous-groupe décrit par le graphe de Rauzy est le groupe F_A lui-même.

3.4.1 Brève introduction aux automates

Nous nous contentons ici d'introduire les notions fondamentales des automates et ne faisons qu'effleurer le sujet. Le lecteur désireux d'en savoir plus est invité à consulter [16].

Définition 3.4.1. Un *automate (déterministe)* est un quintuple $\mathcal{A} = (Q, q_0, F, A, \delta)$ où
- Q est un ensemble fini d'éléments appelés *états*,
- $q_0 \in Q$ est *l'état initial*,
- $F \subseteq Q$ est l'ensemble des *états accepteurs*,
- A est *l'alphabet* de l'automate,

(2). Il suffit de prendre $v' = ua$ et $u' = bu$ pour $b \in L(ua)$, ce qui est possible car S est biprolongeable.

- $\delta \subseteq Q \times A \times Q$ tel que

$$|\{q \in Q \mid (p, a, q) \in \delta\}| \leq 1 \qquad \forall\ a \in A,\ \forall\ p \in Q$$

est la *relation de transition*. Si, pour $p \in Q$ et $a \in A$, un tel q existe, on notera alors $q = \delta(p, a)$.

Un automate est, en général, représenté par le graphe orienté étiqueté $G = (Q, \delta)$ dont les sommets sont des cercles et auquel on ajoute une flèche entrante à l'état initial et des flèches sortantes aux états accepteurs [(3)].

Remarque 3.4.2. On peut étendre la notation $\delta(\,\cdot\,,\,\cdot\,)$ par

$$\delta(q, w) = \begin{cases} q & \text{si } w = \varepsilon, \\ \delta(\delta(q, v), a) & \text{si } w = va,\ a \in A. \end{cases}$$

Définition 3.4.3. Un mot w est *accepté* par l'automate $\mathcal{A} = (Q, q_0, F, A, \delta)$ s'il existe un chemin étiqueté par w entre q_0 et un état accepteur dans le graphe correspondant à l'automate. L'ensemble des mots acceptés par $\mathcal{A}$ forme le *langage de l'automate*, noté $L(\mathcal{A})$.

En d'autres mots, un mot w est accepté si

$$\delta(q_0, w) \in F.$$

Définition 3.4.4. Un *automate non déterministe* est un quintuple $\mathcal{A} = (Q, I, F, A, \Delta)$ où
- Q est un ensemble fini d'éléments appelés *états*,
- $I \subseteq Q$ est l'ensemble des *états initiaux*,
- $F \subseteq Q$ est l'ensemble des *états accepteurs*,
- A est *l'alphabet* de l'automate,
- $\Delta \subseteq Q \times A^* \times Q$ est la *relation de transition* (finie).

On peut étendre de façon naturelle la définition précédente aux automates non déterministes mais la notation $\delta(\,\cdot\,,\,\cdot\,)$ n'est plus correctement définie.

Définition 3.4.5. Le *sous-groupe décrit* par l'automate $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ est l'ensemble des réductions des mots de $L(\mathcal{B})$ où $\mathcal{B} = (Q, \{q_0\}, \{q_0\}, A \cup A^{-1}, \Delta)$ est l'automate non déterministe tel que

$$\Delta = \{(x, a, y) \in Q \times (A \cup A^{-1}) \times Q \mid (x, a, y) \in \delta \text{ ou } (y, a^{-1}, x) \in \delta\}.$$

À nouveau, on peut étendre ce concept au cas où $\mathcal{A}$ est un automate non déterministe ayant un unique état initial qui est également l'unique état accepteur.

Proposition 3.4.6. *Le sous-groupe H décrit par l'automate $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ est un sous-groupe de F_A.*

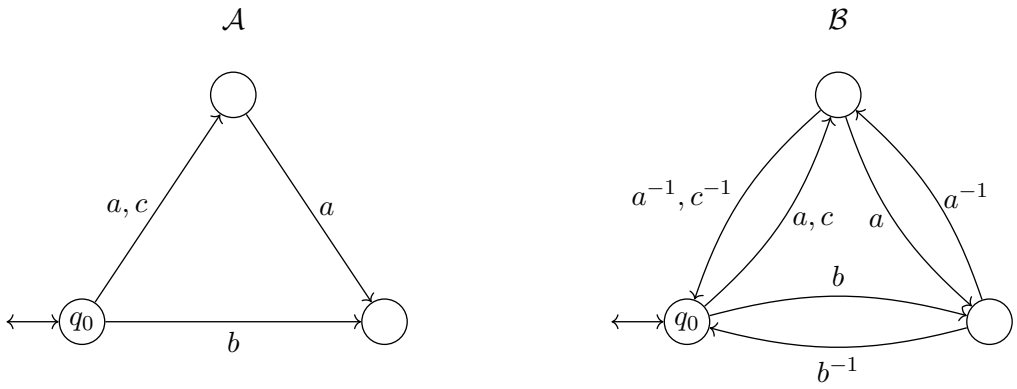
Démonstration. Tout élément de H est un élément de F_A puisqu'on considère les réductions. Montrons donc que H est non vide, stable pour la concaténation de F_A et pour l'inverse.
- Le mot ε correspond au chemin « vide » qui part et arrive en q_0 donc $\varepsilon \in H$.

(3). Une autre convention est de doubler les cercles correspondant aux états accepteurs.

- Si $x, x' \in H$, alors il existe des mots $y, y' \in L(\mathcal{B})$ tels que x (resp. x') soit la réduction de y (resp. y') et qu'il existe des chemins correspondant à y et y' démarrant et aboutissant en q_0. En concaténant ces chemins, on trouve que $yy' \in L(\mathcal{B})$ donc $x \cdot x' = \rho(xx') = \rho(yy') \in H$.
- Si $x \in H$, alors x est la réduction d'un mot y correspondant à un chemin de $\mathcal{B}$ commençant et finissant en q_0. Le chemin inverse aura pour étiquette y^{-1} qui sera donc également accepté par $\mathcal{B}$. On vérifie facilement que $x^{-1} = \rho(y^{-1}) \in H$.

□

Exemple 3.4.7. Voici, par exemple, un automate $\mathcal{A}$ et l'automate $\mathcal{B}$ correspondant.



Les mots $\varepsilon, ca^{-1}, cab^{-1}, ba^{-1}a^{-1}, \ldots$ sont dans le sous-groupe décrit par $\mathcal{A}$. Ce groupe est en réalité donné par $\langle\{ac^{-1}, aab^{-1}\}\rangle$.

3.4.2 Groupe décrit par un graphe

Nous pouvons à présent définir une notion similaire mais en partant d'un graphe.

Définition 3.4.8. Si $G = (V, E)$ est un graphe orienté, étiqueté par des lettres de A, alors le *groupe décrit* par G par rapport à un sommet $v \in V$ est le sous-groupe décrit par l'automate non déterministe $\mathcal{B} = (V, \{v\}, \{v\}, A, E)$.

Il s'agit donc de l'ensemble des réductions des mots de $L(\mathcal{A}_G)$ où

$$\mathcal{A}_G = (V, \{v\}, \{v\}, A \cup A^{-1}, \Delta)$$

avec

$$\Delta = E \cup \{(x, a^{-1}, y) \mid (y, a, x) \in E\}.$$

Proposition 3.4.9. *Si $G = (V, E)$ est fortement connexe, alors, pour tout $v \in V$, le langage accepté par l'automate $\mathcal{A} = (V, \{v\}, \{v\}, A, E)$ engendre le groupe H décrit par G par rapport à v.*

Démonstration. Soit $v \in V$. Notons $\mathcal{A}_G = (V, \{v\}, \{v\}, A \cup A^{-1}, \Delta)$ l'automate non déterministe correspondant au groupe décrit par G par rapport à v (voir Définition 3.4.8). Remarquons que $L(\mathcal{A}) \subseteq A^*$ donc $L(\mathcal{A}) \subseteq F_A$. De plus, tout mot accepté par $\mathcal{A}$ l'est aussi par $\mathcal{A}_G$. On a donc immédiatement que le groupe engendré par $L(\mathcal{A})$ est un sous-groupe de H. Montrons l'autre inclusion.

Soit $x \in H$. Notons $y \in L(\mathcal{A}_G)$ tel que

$$x = \rho(y).$$

Comme $y \in (A \cup A^{-1})^*$, il existe $u_1, \dots, u_n \in A^*$, $v_1, \dots, v_n \in (A^{-1})^*$ tels que

$$y = u_1 v_1 \dots u_n v_n.$$

Par définition, y correspond à un chemin de $\mathcal{A}_G$ commençant et arrivant en v. Notons $p_i \in V$, pour $i \in \{1, \dots, 2n+1\}$, l'état atteint en parcourant ce chemin et en ne lisant que $u_1 v_1 \dots v_{\frac{i-1}{2}}$ si i est impair ou que $u_1 v_1 \dots u_{\frac{i}{2}}$ si i est pair. Le mot u_i (resp v_i) correspond donc à un chemin de p_{2i-1} à p_{2i} dans $\mathcal{A}_G$ (resp. de p_{2i} à p_{2i+1}).

Comme G est fortement connexe, il existe des chemins de p_i vers v et inversement pour tout $i \in \{1, \dots, 2n+1\}$. Notons $x_k \in A^*$ l'étiquette d'un chemin de p_k vers v (resp. de v vers p_k) si k est pair (resp. impair). Comme $p_1 = v = p_{2n+1}$, on peut supposer $x_1 = x_{2n+1} = \varepsilon$. Par construction de $\mathcal{A}$, il existe des chemins commençant en v, étiquetés par $x_{2i-1} u_i x_{2i} \in A^*$ et arrivant en v pour tout $i \leq n$. Autrement dit, $x_{2i-1} u_i x_{2i} \in L(\mathcal{A})$. Comme $v_i \in (A^{-1})^*$, $(v_i)^{-1} \in A^*$ correspond à un chemin de $\mathcal{A}$ de p_{2i+1} vers p_{2i}. On a donc un chemin commençant en v, étiqueté par $x_{2i+1}(v_i)^{-1}x_{2i} \in A^*$ et arrivant en v, ce qui implique que

$$x_{2i+1}(v_i)^{-1}x_{2i} \in L(\mathcal{A}).$$

Si

$$w = (x_1 u_1 x_2)(x_3 v_1^{-1} x_2)^{-1}(x_3 u_2 x_4) \dots (x_{2n+1} v_n^{-1} x_{2n})^{-1},$$

alors $w \in (L(\mathcal{A}) \cup L(\mathcal{A})^{-1})^*$ et

$$\rho(w) = \rho(u_1 v_1 \dots u_n v_n) = \rho(y) = x.$$

En conclusion, x est dans le sous-groupe engendré par $L(\mathcal{A})$ donc on a bien l'autre inclusion. □

Dans le cas d'un graphe de Rauzy d'un ensemble dendrique récurrent, le groupe décrit est F_A tout entier, ce que nous montrons maintenant.

Définition 3.4.10. Soit $G = (V, E)$ un graphe étiqueté. S'il existe $p_1, p_2, q \in V$ et $a \in A$ tels que $(p_1, a, q), (p_2, a, q) \in E$, alors l'opération qui consiste à remplacer p_1 et p_2 par un nouveau sommet p_0 et les arêtes partant ou arrivant en p_1 ou p_2 par des arêtes partant ou arrivant en p_0 avec la même étiquette est appelée un *pliage de Stallings*. Plus formellement, le graphe G est remplacé par le graphe $G' = (V', E')$ où

$$V' = \{p_0\} \cup V \setminus \{p_1, p_2\}, \qquad p_0 \notin V$$

et

$$\begin{aligned} E' = &\{(p_0, b, p_0) \mid b \in A, \exists\, r, s \in \{p_1, p_2\} \text{ tq. } (r, b, s) \in E\} \\ &\cup \{(p, b, p_0) \mid p \in V',\ b \in A,\ (p, b, p_1) \in E \text{ ou } (p, b, p_2) \in E\} \\ &\cup \{(p_0, b, p) \mid p \in V',\ b \in A,\ (p_1, b, p) \in E \text{ ou } (p_2, b, p) \in E\} \\ &\cup \{(p, b, p') \in E \mid b \in A,\ p, p' \in V'\}. \end{aligned}$$

On peut généraliser cette opération pour le pliage de plus de 2 sommets.

Lemme 3.4.11. *Soit $G = (V, E)$ un graphe étiqueté. S'il existe $p_1, \dots, p_n, q \in V$ et $a \in A$ tels que $(p_1, a, q), \dots, (p_n, a, q) \in E$, alors le groupe décrit par G par rapport à un sommet v est le même que celui décrit par le graphe $G' = (V', E')$ par rapport au sommet v' où G' est le graphe obtenu après pliage de Stallings de $p_1, \dots, p_n$ et où*

$$v' = \begin{cases} v & \text{si } v \notin \{p_1, \dots, p_n\}, \\ p_0 & \text{sinon.} \end{cases}$$

Démonstration. Notons $\mathcal{A}_G$ et $\mathcal{A}_{G'}$ les automates non déterministes correspondant respectivement à G et G' (voir Définition 3.4.8). Si y est accepté par $\mathcal{A}_G$, alors tout chemin permettant d'accepter y dans $\mathcal{A}_G$ peut aisément être vu comme un chemin permettant d'accepter y dans $\mathcal{A}_{G'}$ (au lieu de passer par p_i, on passera par p_0). Réciproquement, supposons que y est accepté par $\mathcal{A}_{G'}$ et considérons un chemin permettant d'accepter y dans $\mathcal{A}_{G'}$. Si ce chemin ne passe pas par p_0, alors y est trivialement accepté par $\mathcal{A}_G$. Sinon, notons ce chemin $v' = q_0, \dots, q_m = v'$. Pour tout $j < m$, par construction de E', on sait qu'il existe $k, l \in \{1, \dots, n\}$ tels que l'arête $(q_j, y_{j+1}, q_{j+1}) \in \Delta'$ corresponde à l'arête

$$\begin{cases} (q_j, y_{j+1}, q_{j+1}) & \text{si } q_j, q_{j+1} \neq p_0 \\ (p_k, y_{j+1}, q_{j+1}) & \text{si } q_j = p_0, q_{j+1} \neq p_0 \\ (q_j, y_{j+1}, p_l) & \text{si } q_j \neq p_0, q_{j+1} = p_0 \\ (p_k, y_{j+1}, p_l) & \text{si } q_j = q_{j+1} = p_0 \end{cases}$$

dans Δ. Or, dans $\mathcal{A}_G$, on peut passer de p_l à p_k en lisant aa^{-1} (on passe par le sommet q). Donc, quitte à rajouter des transitions par le sommet q, y est dans la même classe qu'un mot accepté par $\mathcal{A}_G$, ce qui permet de conclure l'égalité des groupes décrits par G et G' par rapport aux sommets v et v' respectivement. □

Proposition 3.4.12. *Soit S un ensemble dendrique récurrent. Pour tout $n \in \mathbb{N}$, le groupe décrit par le graphe de Rauzy d'ordre n de S par rapport à n'importe quel sommet est le groupe libre F_A.*

Démonstration. Montrons qu'on peut passer de $G_n(S) = (V, E)$ à $G_n(S)/\theta_n$ en effectuant un nombre fini de pliages de Stallings. Notons $G' = (V', E')$ le graphe intermédiaire provisoire. Au début, on a donc $V' = V$ et $E' = E$. Comme S est dendrique, on a

$$\begin{aligned} u \sim_{\theta_n} v &\Leftrightarrow u_{[2,n]} = v_{[2,n]} \\ &\Leftrightarrow \forall a \in A, w \in V, (ua \in Aw \Leftrightarrow va \in Aw) \\ &\Leftrightarrow \forall a \in A, w \in V, ((u, a, w) \in E \Leftrightarrow (v, a, w) \in E)\,. \end{aligned}$$

Soient $u, v \in V$ tels que $u \sim_{\theta_n} v$. Comme S est biprolongeable, il existera toujours $a \in A$ et $w \in S$ tels que $ua, va \in Aw$. Pour chacun de ces couples, on a deux possibilités. Soit w n'a pas été modifié précédemment et donc $(u, a, w), (v, a, w) \in E'$, soit $(u, a, [w]_{\theta_n}), (v, a, [w]_{\theta_n}) \in E'$. Dans les deux cas, on peut fusionner tous les sommets de la classe $[u]_{\theta_n}$ et adapter V' et E' conformément à ce qui est fait dans le lemme précédent. En répétant le processus, on groupe ainsi tous les sommets équivalents pour obtenir le graphe $G_n(S)/\theta_n$. Par le lemme, le groupe décrit par $G_n(S)$ par rapport à un sommet $v_n \in V$ est égal au groupe décrit par $G_n(S)/\theta_n$ par

rapport au sommet correspondant $[v_n]_{\theta_n}$. Vu la Proposition 3.3.6, il est alors égal au groupe décrit par $G_{n-1}(S)$ par rapport à un sommet v_{n-1}, et ainsi de suite. On se ramène donc au groupe décrit par $G_0(S)$ par rapport à un de ses sommets v_0. Or, $G_0(S)$ a ε pour unique sommet et des boucles étiquetées par toutes les lettres de A comme arêtes. Le groupe décrit est alors F_A tout entier. □

3.5 Théorème de retour

Le théorème principal de ce chapitre, appelé Théorème de retour, est une conséquence du théorème suivant.

Théorème 3.5.1. *Soit S un ensemble dendrique récurrent. Pour tout $w \in S$, l'ensemble $\mathcal{R}_S(w)$ engendre le groupe libre F_A.*

Démonstration. Soit $w \in S$. Par le Théorème 2.4.2, $\mathcal{CR}_S(w)$ est fini non vide donc posons

$$n = \max\{|u| \mid u \in \mathcal{CR}_S(w)\}.$$

Par définition de $\mathcal{CR}_S(w)$, on a alors $n > |w|$. Comme S est biprolongeable, il existe $x \in S \cap A^n$ ayant w pour suffixe. Considérons l'automate $\mathcal{A} = (V, \{x\}, \{x\}, A, E)$ où $G_n(S) = (V, E)$. Le graphe $G_n(S)$ est fortement connexe. En effet, S est récurrent et $V \subseteq S$ donc, pour tous $y, z \in V$, il existe v tel que $yvz \in S$. L'ensemble S étant factoriel, on a donc un chemin de y vers z étiqueté par vz.

Par la Proposition 3.4.12, le groupe décrit par $G_n(S)$ est F_A et par la Proposition 3.4.9, il est engendré par $L(\mathcal{A})$. Il suffit donc de montrer que $L(\mathcal{A}) \subseteq (\mathcal{R}_S(w))^*$ pour pouvoir conclure que $F_A \subseteq \langle R_S(w)\rangle$ et donc avoir l'égalité.

Soit $y \in L(\mathcal{A})$. Si $y = \varepsilon$, la conclusion est immédiate. Sinon, ce mot correspond à un chemin $\mathcal{C}$ partant de x et arrivant en x dans $G_n(S)$ donc xy a x pour suffixe par définition des graphes de Rauzy. Comme w est suffixe de x, on a donc

$$wy \in A^+w$$

et y est un mot de retour pour w dans A^*. Il existe alors $u_1, \ldots, u_m \in \mathcal{R}_{A^*}(w)$ tels que

$$y = u_1 \ldots u_m.$$

Si on montre que, pour tout $i \leq m$, $u_i \in \mathcal{R}_S(w)$, alors on pourra conclure. Pour cela, il suffit de montrer que $wu_i \in S$ pour tout $i \leq m$.

Soit $i \leq m$. Notons z le sommet atteint en lisant $u_1 \ldots u_{i-1}$ sur le chemin $\mathcal{C}$. Comme $z \in S$ et que S est biprolongeable, il existe $v \in S \cap A^n$ tel que $vz \in S$. Il existe alors un chemin de v vers z étiqueté par z. Comme $n > |w|$ et que z est un suffixe de $xu_1 \ldots u_{i-1}$, par définition des u_j, w est suffixe de z. Autrement dit, on peut trouver un chemin étiqueté par wu_i dans $G_n(S)$. Montrons que $|wu_i| \leq n$ car alors ce chemin arrive dans un sommet ayant wu_i comme suffixe. Or, par définition de $G_n(S)$, ce sommet sera un élément de S et on pourra conclure car S est factoriel.

Par l'absurde, si $|wu_i| > n$, notons p le préfixe de longueur n de $|wu_i|$. Comme $|w| < n$, il existe v' tel que $p = wv'$. Si on ne considère que les n premières étapes du chemin étiqueté

par wu_i, on arrive dans un sommet étiqueté par p donc $p \in S$. Comme S est récurrent, v' est préfixe d'un mot de retour pour w. De plus,

$$|v'| = n - |w| = \max\{|u| \mid u \in \mathcal{R}_S(w)\}$$

donc v' a un préfixe u dans $\mathcal{R}_S(w)$. En particulier, w est suffixe propre de wu donc $w \in \mathrm{IFac}(wu_i)$, ce qui est absurde car $u_i \in \mathcal{R}_{A^*}(w)$. □

Remarquons que si S est uniformément récurrent, seul le caractère connexe de la définition de dendrique est nécessaire pour que la preuve soit valide.

Le Théorème de retour n'est alors qu'un simple corollaire.

Théorème 3.5.2 (Théorème de retour)**.** *Soit S un ensemble dendrique récurrent. Pour tout $w \in S$, l'ensemble $\mathcal{R}_S(w)$ est une base de F_A.*

Démonstration. Vu le théorème précédent, $\mathcal{R}_S(w)$ engendre F_A. De plus, par le Théorème 2.4.2,

$$|\mathcal{R}_S(w)| = |A|$$

donc, comme A est une base de F_A, on en conclut que $\mathcal{R}_S(w)$ est également une base de F_A, pour tout $w \in S$, par la Proposition 3.2.4. □

Chapitre 4

Stabilité par décodage bifixe

Dans ce chapitre, nous montrons que l'image inverse d'un ensemble dendrique S par un morphisme codant pour un code bifixe S-maximal est également dendrique. Pour cela, nous introduisons une généralisation des graphes d'extensions et nous intéressons à leurs propriétés.

4.1 Graphes d'extensions généralisés

Plutôt que de considérer les extensions à gauche et à droite d'un mot par des lettres, nous considérons les extensions par des mots dans des ensembles donnés. Cela permet de généraliser le concept de graphe d'extensions.

Définition 4.1.1. Soient S un ensemble factoriel et $w \in S$. Si $U, V \subseteq S$, alors on note

$$L^U(w) = \{l \in U \mid lw \in S\}, \quad R^V(w) = \{r \in V \mid wr \in S\}$$

et

$$E^{U,V}(w) = \{(l, r) \in L^U(w) \times R^V(w) \mid lwr \in S\}.$$

Le *graphe d'extensions généralisé* de w est le graphe non orienté biparti

$$\mathcal{E}^{U,V}(w) = (V^{U,V}(w), E^{U,V}(w))$$

où $V^{U,V}(w)$ est l'union disjointe de $L^U(w)$ et $R^V(w)$.

Remarque 4.1.2. Il s'agit bien d'une généralisation car on a

$$L(w) = L^A(w), \quad R(w) = R^A(w), \quad E(w) = E^{A,A}(w).$$

De plus, comme $L^{L^U(w)}(w) = L^U(w)$ et $R^{R^V(w)}(w) = R^V(w)$, on a

$$\mathcal{E}^{U,V}(w) = \mathcal{E}^{L^U(w),R^V(w)}(w)$$

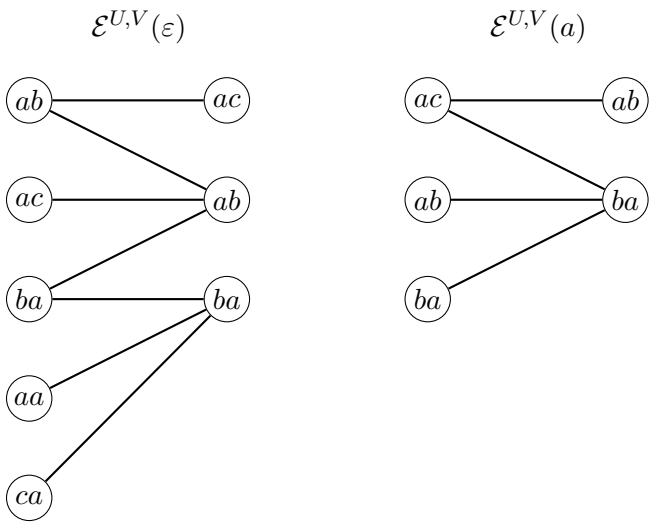
donc

$$\mathcal{E}(w) = \mathcal{E}^{A,A}(w) = \mathcal{E}^{L(w),R(w)}(w).$$

Exemple 4.1.3. Pour l'ensemble des facteurs du mot de Tribonacci (Exemple 1.4.7), si

$$U = A^2 \quad \text{et} \quad V = \{ab, ac, ba\},$$

alors on a, entre autres, les graphes d'extensions généralisés suivants :



Nous montrons ici que, sous certaines conditions, si l'ensemble S est dendrique, ces nouveaux graphes d'extensions sont également des arbres. Nous nous intéressons également à un résultat similaire mais concernant uniquement le caractère acyclique. Ces résultats nécessitent plusieurs lemmes.

Lemme 4.1.4. *Soient S un ensemble biprolongeable, $w \in S$, U un code suffixe fini Sw^{-1}-maximal et $v \in S$ tel que $vw \in S$. Si $\mathrm{Suff}(v) \cap U = \emptyset$, alors il existe $u \in L^U(w)$ tel que v soit un suffixe propre de u. En particulier, pour tout $w \in S$, $L^U(w)$ n'est pas vide.*

Symétriquement, si V est un code préfixe fini $w^{-1}S$-maximal, si $u \in S$ est tel que $wu \in S$ et si $\mathrm{Pref}(u) \cap V = \emptyset$, alors il existe $v \in R^V(w)$ tel que u soit un préfixe propre de v. En particulier, pour tout $w \in S$, $R^V(w)$ n'est pas vide.

Démonstration. Notons $n = \max\{|u| \mid u \in U\}$. Comme S est biprolongeable, il existe $a_0, \ldots, a_n \in A$ tels que

$$a_n \ldots a_0 vw \in S.$$

On a donc $a_n \ldots a_0 v \in Sw^{-1}$ et, de par sa longueur, $a_n \ldots a_0 v$ n'est suffixe d'aucun mot de U. Or, U est un code suffixe Sw^{-1}-maximal donc ce mot possède un suffixe dans U. Comme v n'a pas de suffixe dans U par hypothèse, il existe donc $i \in \{0, \ldots, n\}$ tel que

$$a_i \ldots a_0 v \in U.$$

L'ensemble S étant factoriel, on a alors

$$a_i \ldots a_0 vw \in S,$$

ce qui implique que $a_i \ldots a_0 v \in L^U(w)$ et prouve la première partie de l'énoncé. Un raisonnement symétrique permet de conclure pour l'autre cas. □

Lemme 4.1.5. *Soient S un ensemble biprolongeable, $w \in S$, $U, V \subseteq S$ et $T \subseteq A$. Soit $l \in S \setminus U$ (resp. $r \in S \setminus V$) tel que*

$$lw \in S \quad (\textit{resp. } wr \in S)$$

et soit

$$U' = (U \setminus Tl) \cup \{l\} \quad (\textit{resp. } V' = (V \setminus rT) \cup \{r\}).$$

Si

$$\mathcal{E}^{U',V}(w) \textit{ et } \mathcal{E}^{T,V}(lw)$$
$$(\textit{resp. } \mathcal{E}^{U,V'}(w) \textit{ et } \mathcal{E}^{U,T}(wr))$$

sont acycliques, alors $\mathcal{E}^{U,V}(w)$ *aussi.*

Démonstration. Les deux cas se traitent de façon symétrique donc ne prouvons que le cas pour $l \in S \setminus U$. Procédons par l'absurde et supposons que $\mathcal{E}^{U,V}(w)$ contienne un cycle $\mathcal{C}$. Comme $\mathcal{E}^{U,V}(w)$ est biparti, notons ce cycle $u_1, v_1, u_2, \ldots, v_n, u_1$, avec $u_i \in L^U(w)$, $v_i \in R^V(w)$ pour tout $i \leq n$.

- Si, pour tout i, $u_i \in Tl$ [(1)] alors notons $a_i \in T$ tel que $u_i = a_i l$. Dans ce cas, $a_1, v_1, a_2, \ldots, v_n, a_1$ est un cycle de $\mathcal{E}^{T,V}(lw)$, ce qui est absurde.
- Si, pour tout i, $u_i \notin Tl$, alors $\mathcal{C}$ est un cycle de $\mathcal{E}^{U',V}(w)$, ce qui est absurde.
- Sinon, supposons, sans perte de généralité que $u_1 \notin Tl$. Pour tous $i, j \geq 1$ tels que $u_i, u_{i+j+1} \notin Tl$ et $u_{i+1}, \ldots, u_{i+j} \in Tl$, remplaçons le tronçon $v_i, u_{i+1}, \ldots, u_{i+j}, v_{i+j}$ par v_i, l, v_{i+j} dans $\mathcal{E}^{U',V}(w)$. Il s'agit bien d'un chemin de $\mathcal{E}^{U',V}(w)$ car, si $u_{i+1} = bl$, alors [(2)]
$$(u_{i+1}, v_i) \in E^{U,V}(w) \Rightarrow blwv_i \in S \Rightarrow (l, v_i) \in E^{U',V}(w)$$
On procède de même pour prouver l'existence de l'arête (l, v_{i+j}) dans $\mathcal{E}^{U',V}(w)$. Après avoir fait toutes les substitutions, on obtient, dans $\mathcal{E}^{U',V}(w)$, un cycle non trivial car il passe par u_1 et l, ce qui est absurde.

□

En ajoutant quelques hypothèses, on obtient un résultat similaire mais concernant également le caractère connexe des graphes d'extensions généralisés.

Lemme 4.1.6. *Soient S un ensemble biprolongeable, $w \in S$ et $U, V \subseteq S$ tels que $L^U(w)$ et $R^V(w)$ soient non vides. Soit $l \in S \setminus U$ (resp. $r \in S \setminus V$) tel que*

$$lw \in S \textit{ et } Al \cap S \subseteq U$$
$$(\textit{resp. } wr \in S \textit{ et } rA \cap S \subseteq V)$$

et soit

$$U' = (U \setminus Al) \cup \{l\} \quad (\textit{resp. } V' = (V \setminus rA) \cup \{r\}).$$

Si

$$\mathcal{E}^{U',V}(w) \textit{ et } \mathcal{E}^{A,V}(lw)$$
$$(\textit{resp. } \mathcal{E}^{U,V'}(w) \textit{ et } \mathcal{E}^{U,A}(wr))$$

sont des arbres, alors $\mathcal{E}^{U,V}(w)$ *aussi.*

Démonstration. Par le résultat précédent, $\mathcal{E}^{U,V}(w)$ est acyclique. Pour le caractère connexe, procédons par étapes : [(3)]

(1). On prendra la convention $u_{n+1} = u_1$.
(2). Rappellons que nous considérons ici des graphes non orientés donc considérer l'arête (u, v) ou (v, u) revient au même.
(3). Nous traitons ici le cas $l \in S \setminus U$, l'autre cas étant symétrique.

1. Pour tous $al \in L^U(w)$, $v \in R^V(lw)$, il existe un chemin les reliant dans $\mathcal{E}^{U,V}(w)$. De fait, a et v sont connectés dans $\mathcal{E}^{A,V}(lw)$ et
$$\begin{aligned}(b,u) \in E^{A,V}(lw) &\Leftrightarrow u \in V \text{ et } blwu \in S\\ &\Leftrightarrow u \in V,\ bl \in U \text{ et } blwu \in S\\ &\Leftrightarrow (bl,u) \in E^{U,V}(w)\end{aligned}$$
donc en remplaçant tous les sommets $b \in L^A(lw)$ par $bl \in L^U(w)$ dans le chemin reliant a et v, on obtient un chemin de al vers v dans $\mathcal{E}^{U,V}(w)$.
2. Pour tous $al \in L^U(w)$, $v \in R^V(w)$, il existe un chemin les reliant dans $\mathcal{E}^{U,V}(w)$. De fait, comme $\mathcal{E}^{U',V}(w)$ est connexe, il contient un chemin reliant l à v. On peut supposer que ce chemin ne repasse pas par l donc que, si (l,u) est la première arête, la partie du chemin reliant u à v se trouve également dans le graphe $\mathcal{E}^{U,V}(w)$. Par le point 1., comme $u \in R^V(lw)$, il existe un chemin reliant al à u dans $\mathcal{E}^{U,V}(w)$ qui, combiné avec le chemin de u vers v, permet de conclure.
3. Pour tous $u \in L^{U'}(w) \setminus \{l\}$, $v \in R^V(w)$, il existe un chemin les reliant dans $\mathcal{E}^{U,V}(w)$. En effet, $\mathcal{E}^{U',V}(w)$ est connexe donc il contient un chemin reliant u et v. Les arêtes ne faisant pas intervenir le sommet l sont également dans $\mathcal{E}^{U,V}(w)$. Il suffit donc de s'intéresser aux éventuels tronçons v_1', l, v_2'. Comme S est biprolongeable, il existe $a \in A$ tel que $alw \in S$. On a alors $al \in L^U(w)$ donc, par le point 2., il existe un chemin reliant v_1' à v_2' dans $\mathcal{E}^{U,V}(w)$ en passant par le sommet al.

Toute paire $\{u,v\}$, $u \in L^U(w)$, $v \in R^V(w)$ peut ainsi être reliée, soit par le point 2., soit par le point 3. Comme $L^U(w)$ et $R^V(w)$ sont non vides, on en déduit que $\mathcal{E}^{U,V}(w)$ est connexe. □

Nous pouvons à présent montrer que, dans un ensemble dendrique, sous certaines hypothèses, les graphes d'extensions généralisés sont des arbres.

Théorème 4.1.7. *Soient S dendrique et $w \in S$. Si $U \subseteq S$ est un code suffixe fini Sw^{-1}-maximal et si $V \subseteq S$ est un code préfixe fini $w^{-1}S$-maximal, alors le graphe d'extensions généralisé $\mathcal{E}^{U,V}(w)$ est un arbre.*

En particulier, si S est dendrique, pour tout code suffixe fini $U \subseteq S$ S-maximal et tout code préfixe fini $V \subseteq S$ S-maximal, $\mathcal{E}^{U,V}(w)$ est un arbre pour tout $w \in S$.

Démonstration. Montrons ce résultat par récurrence sur
$$n(U,V) = \sum_{u\in U}|u| + \sum_{v\in V}|v|$$
pour tout $w \in S$ et tous U, V vérifiant les hypothèses. Le cas de base de la récurrence est inclus dans le cas où $L^U(w), R^V(w) \subseteq A$. Pour tout $a \in L(w)$, on a alors
$$A^+a \cap L^U(w) = \emptyset$$
donc, par contraposition du Lemme 4.1.4, $\mathrm{Suff}(a) \cap U \neq \emptyset$, ce qui signifie que $a \in U$. Symétriquement, on obtient $R(w) \subseteq V$ donc
$$L^U(w) = L(w) \quad \text{et} \quad R^V(w) = R(w).$$
En particulier, on a
$$\mathcal{E}^{U,V}(w) = \mathcal{E}^{L(w),R(w)}(w) = \mathcal{E}(w)$$

qui est un arbre car S est dendrique.

Passons maintenant à la récurrence proprement dite. Supposons le résultat vrai pour tout mot w' et tout couple (U', V') vérifiant les hypothèses et tels que

$$n(U', V') < n(U, V)$$

et où U est un code suffixe Sw^{-1}-maximal et V un code préfixe $w^{-1}S$-maximal pour $w \in S$. On peut donc supposer que, soit $L^U(w)$, soit $R^V(w)$ contient au moins un mot de longueur supérieure ou égale à 2. Supposons dans un premier temps qu'il s'agisse de $L^U(w)$. La démonstration est symétrique dans l'autre cas.

Notons $v \in L^U(w)$ de longueur maximale. Il existe alors $a \in A$ et $l \in A^+$ tels que $v = al$. On a donc $alw \in S$ et $lw \in S$ car S est factoriel. De plus, $l \notin U$ car U est un code suffixe.

Posons

$$W' = (U \setminus A^*l) \cup (Al \cap \mathrm{Suff}(U)).$$

Cette opération consiste à « raccourcir » les mots de U ayant l pour suffixe. Montrons que ça ne modifie pas le graphe d'extensions. Par définition de v, on a

$$L^U(w) \cap A^*l = L^U(w) \cap Al = L^{W'}(w) \cap Al = L^{W'}(w) \cap A^*l$$

donc

$$L^U(w) = L^{W'}(w).$$

Si $W = W' \cup (Al \setminus Sw^{-1})$, le graphe d'extensions n'est pas non plus modifié.

Posons

$$U' = (W \setminus Al) \cup \{l\} = (U \setminus A^*l) \cup \{l\}.$$

et montrons que U' est un code suffixe Sw^{-1}-maximal. Comme $l \in \mathrm{Suff}(U)$, par la Remarque 2.1.4, U' est un code suffixe. Regardons le caractère Sw^{-1}-maximal. Soit $u \in Sw^{-1} \setminus U'$. Si $u \in A^*l$ ou si u est un suffixe propre de l, alors $\{l\} \cup \{u\}$ ne saurait être un code suffixe donc $U' \cup \{u\}$ non plus. Si ce n'est pas le cas, alors $u \notin U$ donc, comme U est un code suffixe Sw^{-1}-maximal, il existe $u' \in U$ qui soit suffixe propre de u ou qui ait u pour suffixe propre. Par définition de u, $u' \notin A^*l$ donc $u' \in U'$ et $U' \cup \{u\}$ n'est pas un code suffixe.

Cherchons à appliquer le Lemme 4.1.6 pour W et V.

- Par hypothèse sur U et V et par le Lemme 4.1.4, $L^U(w) = L^W(w)$ et $R^V(w)$ ne sont pas vides. De plus, $lw \in S$ et $l \notin W$.
- Soit $bl \in Al \cap S$. Si $bl \notin Sw^{-1}$, alors on a directement $bl \in W$. Si $bl \in Sw^{-1}$, alors, l'ensemble U étant un code suffixe Sw^{-1}-maximal ne contenant aucun suffixe de l, il existe v tel que $vbl \in U$. Par construction de W, on a alors $bl \in W$ donc

$$Al \cap S \subseteq W.$$

- Par construction,

$$\begin{aligned}
\sum_{u \in U'} |u| &= |l| + \sum_{u \in U \setminus A^*l} |u| \\
&\leq |l| + \sum_{u \in U \setminus \{al\}} |u| \\
&< |al| + \sum_{u \in U \setminus \{al\}} |u| \\
&= \sum_{u \in U} |u|
\end{aligned}$$

donc

$$n(U', V) < n(U, V).$$

Par hypothèse de récurrence, $\mathcal{E}^{U',V}(w)$ est alors un arbre.

- Comme U est Sw^{-1}-maximal et qu'il ne contient aucun suffixe de l, il contient au moins un mot se terminant par al pour tout $a \in L(lw)$ donc

$$n(L(lw), V) < n(U, V).$$

De plus, $L(lw)$ est $S(lw)^{-1}$-maximal et V est $w^{-1}S$-maximal donc il est $(lw)^{-1}S$-maximal car $(lw)^{-1}S \subseteq w^{-1}S$. Par hypothèse de récurrence, $\mathcal{E}^{A,V}(lw) = \mathcal{E}^{L(lw),V}(lw)$ est un arbre.

Par le Lemme 4.1.6 appliqué à W et V, on conclut alors que

$$\mathcal{E}^{U,V}(w) = \mathcal{E}^{W,V}(w)$$

est un arbre. □

Si on souhaite uniquement s'intéresser au caractère acyclique des graphes d'extensions, certaines hypothèses deviennent superflues. Nous dirons qu'un ensemble biprolongeable S est *acyclique* si, pour tout $w \in S$, $\mathcal{E}(w)$ est acyclique. Nous avons alors le résultat suivant :

Théorème 4.1.8. *Soit S acyclique. Si $U \subseteq S$ est un code suffixe et $V \subseteq S$ un code préfixe, alors le graphe d'extensions généralisé $\mathcal{E}^{U,V}(w)$ est acyclique pour tout $w \in S$.*

Démonstration. La démonstration n'est qu'une version simplifiée de celle du théorème précédent. Le cas de base est immédiat et ne nécessite plus l'appel au Lemme 4.1.4 car tout chemin dans $\mathcal{E}^{U,V}(w)$ est un chemin de $\mathcal{E}(w)$ si $L^U(w), R^V(w) \subseteq A$. Pour l'induction, on peut supposer $U = L^U(w)$ et $V = R^V(w)$ car sinon on conclut immédiatement par hypothèse de récurrence vu que $\mathcal{E}^{U,V}(w) = \mathcal{E}^{L^U(w),R^V(w)}(w)$. Utilisons les notations de la démonstration précédente et posons $T = \{b \in A \mid bl \in U\}$. On a alors

$$U' = (U \setminus A^*l) \cup \{l\} = (U \setminus Al) \cup \{l\} = (U \setminus Tl) \cup \{l\}$$

car $U = L^U(w)$ donc $\max\{|u| \mid u \in U\} = |v| = 1 + |l|$. De plus, $n(T, V) < n(U, V)$. En procédant comme pour la démonstration précédente, on constate que les hypothèses du Lemme 4.1.5 sont vérifiées pour U et V donc $\mathcal{E}^{U,V}(w)$ est acyclique. □

4.2 Décodage bifixe

Dans cette section, nous nous intéressons ici à des morphismes particuliers et, plus spécifiquement, à l'image inverse d'un ensemble par un tel morphisme. C'est ce qu'on appelle un décodage bifixe. Nous montrons que le caractère dendrique est stable pour cette opération sous certaines hypothèses.

Définition 4.2.1. Soient A et B deux alphabets et $X \subseteq A^*$. Un *morphisme codant* pour X est un morphisme $f : B^* \to A^*$ tel que la restriction de f à B soit une bijection entre B et X.

Remarquons que nous autorisons ici l'alphabet B à être infini dénombrable pour définir le concept de morphisme codant dans un cadre plus général. Les notions fondamentales de combinatoire des mots sont alors étendues de façon naturelle. Cependant, dans cette section, seule la Proposition 4.2.2 peut réellement utiliser cette généralisation. Les autres résultats supposent l'ensemble X fini, ce qui nous permet de revenir au contexte plus classique des alphabets finis.

Rappelons qu'un ensemble X est un code si, pour tous $x_1, \ldots, x_n, y_1, \ldots, y_m \in X$, on a

$$x_1 \ldots x_n = y_1 \ldots y_m \Rightarrow n = m \text{ et } x_i = y_i \quad \forall\, i \leq n.$$

Proposition 4.2.2. *Si $X \subseteq A^*$ est un code et $f : B^* \to A^*$ est un morphisme codant pour X, alors f est une bijection entre B^* et X^*.*

Démonstration. La surjectivité est immédiate car f est un morphisme. Pour l'injectivité, supposons avoir $a_1, \ldots, a_n, b_1, \ldots, b_m \in B$ tels que

$$f(a_1 \ldots a_n) = f(b_1 \ldots b_m).$$

Notons $x_i = f(a_i)$ et $y_j = f(b_j)$ pour tous $i \leq n$, $j \leq m$. Comme f est un morphisme, on a alors

$$x_1 \ldots x_n = y_1 \ldots y_m$$

donc, par hypothèse sur X, $n = m$ et $x_i = y_i$ pour tout $i \leq n$. Comme f est injectif sur B, on a également $a_i = b_i$ pour tout $i \leq n$, ce qui permet de conclure. □

Définition 4.2.3. Soit S un ensemble factoriel et soit $X \subseteq S$ un code bifixe fini. Si f est un morphisme codant pour X, alors $f^{-1}(S)$ est un *décodage bifixe* de S. Si X est un code bifixe S-maximal, on parle de *décodage bifixe maximal*.

Exemple 4.2.4. Soit T l'ensemble des facteurs du mot de Tribonacci (Exemple 1.4.7) et $X = T \cap A^2 = \{aa, ab, ac, ba, ca\}$ qui est un code bifixe T-maximal. Si f est le morphisme $f : \{0, 1, 2, 3, 4\}^* \to \{a, b, c\}^*$ tel que

$$f(0) = aa, \quad f(1) = ab, \quad f(2) = ac, \quad f(3) = ba \quad \text{et} \quad f(4) = ca,$$

alors les mots

$$03, 103, 343, 1121, 4312112, \ldots$$

sont dans l'ensemble $f^{-1}(T)$ car les mots

$$aaba, abaaba, bacaba, cabaabacababac, \ldots$$

respectivement sont des facteurs du mot de Tribonacci.

Théorème 4.2.5. *Soient S un ensemble récurrent, $X \subseteq S$ un code bifixe fini S-maximal et f un morphisme codant pour X. On a les résultats suivants :*

1. *$f^{-1}(S)$ est biprolongeable,*
2. *pour tout $v \in f^{-1}(S)$, $\mathcal{E}^{X,X}(f(v))$ est isomorphe à $\mathcal{E}_{f^{-1}(S)}(v)$,*
3. *si S est dendrique, $f^{-1}(S)$ aussi.*

Démonstration.

1. Montrons que $f^{-1}(S)$ est factoriel. Soient $w \in f^{-1}(S)$ et $u \in \mathrm{Fac}(w)$. Montrons que $f(u) \in S$. Comme f est un morphisme, $f(u) \in \mathrm{Fac}(f(w))$ donc on peut conclure car $f(w) \in S$ et car S est factoriel.
   Soit $w \in f^{-1}(S)$. Montrons qu'il existe $a, b \in B$ tels que $awb \in f^{-1}(S)$. Notons
   $$n = \max\{|x| \mid x \in X\}$$
   qui existe car X est fini. Comme S est biprolongeable et $f(w) \in S$, il existe $u, v \in S$ tels que
   $$|u| > n, \quad |v| > n, \quad uf(w)v \in S.$$
   L'ensemble X étant un code suffixe S-maximal, u possède un unique suffixe $u' \in X$. De même, v a un unique préfixe $v' \in X$. Notons $a = f^{-1}(u') \in B$ et $b = f^{-1}(v') \in B$. On a alors
   $$f(awb) = u'f(w)v' \in S$$
   d'où la conclusion.
2. Considérons la restriction [(4)]
   $$f' : V_{f^{-1}(S)}(v) \to V^{X,X}(f(v)) \quad a \mapsto f(a)$$
   Montrons qu'elle est correctement définie et qu'il s'agit d'une bijection. Pour tout $x \in A^*$, on a
   $$\begin{aligned} x \in f'(L_{f^{-1}(S)}(v)) &\Leftrightarrow \exists\, a \in L_{f^{-1}(S)}(v) \text{ tq. } x = f(a) \\ &\Leftrightarrow \exists\, a \in B \text{ tq. } x = f(a), av \in f^{-1}(S) \\ &\Leftrightarrow \exists\, a \in B \text{ tq. } x = f(a), f(a)f(v) \in S \\ &\Leftrightarrow x \in X, xf(v) \in S \\ &\Leftrightarrow x \in L^X(f(v)) \end{aligned}$$
   car f définit une bijection entre B et X. De même,
   $$f'(R_{f^{-1}(S)}(v)) = R^X(f(v))$$
   donc f' est correctement défini et surjectif. Comme la fonction f' est une restriction d'une fonction injective, elle est également injective.
   Montrons maintenant que f' est un isomorphisme de graphes entre les graphes d'extensions $\mathcal{E}_{f^{-1}(S)}(v)$ et $\mathcal{E}^{X,X}(f(v))$. Autrement dit, montrons que
   $$(x, y) \in E^{X,X}(f(v)) \Leftrightarrow (f'^{-1}(x), f'^{-1}(y)) \in E_{f^{-1}(S)}(v).$$
   On a
   $$\begin{aligned} (x, y) \in E^{X,X}(f(v)) &\Leftrightarrow x, y \in V^{X,X}(f(v)), xf(v)y \in S \\ &\Leftrightarrow f'^{-1}(x), f'^{-1}(y) \in V_{f^{-1}(S)}(v), f'^{-1}(x)vf'^{-1}(y) \in f^{-1}(S) \\ &\Leftrightarrow (f'^{-1}(x), f'^{-1}(y)) \in E_{f^{-1}(S)}(v), \end{aligned}$$
   ce qui permet de conclure.

(4). Rappelons que la notation $V_S(w)$ représente l'union disjointe des ensemble $L_S(w)$ et $R_S(w)$.

3. Par le Théorème 4.1.7 et la Proposition 2.1.21, $\mathcal{E}^{X,X}(w)$ est un arbre pour tout $w \in S$. Donc, pour tout $v \in f^{-1}(S)$, $\mathcal{E}_{f^{-1}(S)}(v)$ est également un arbre. En conséquence, $f^{-1}(S)$ est dendrique.

□

Remarquons que nous n'avons pas montré ici la stabilité de la famille des ensembles dendriques récurrents pour les décodages bifixes maximaux mais bien uniquement la stabilité du caractère dendrique sous certaines conditions. Pour montrer que le caractère récurrent est également conservé, il faudra attendre le Chapitre 6 et plus particulièrement le Théorème 6.4.6.

Chapitre 5

Codes bifixes dans un ensemble acyclique

Dans ce chapitre, nous prouvons deux résultats importants qui ne sont pas uniquement valables pour les ensembles dendriques mais plus généralement pour les ensembles acycliques. Rappelons qu'un ensemble S est acyclique s'il est biprolongeable et si, pour tout $w \in S$, $\mathcal{E}(w)$ est acyclique.

Ces deux résultats concernent les propriétés qu'ont les codes bifixes inclus dans un ensemble S acyclique. Le premier, appelé Théorème d'indépendance, s'intéresse à leur caractère libre (dans le groupe F_A). Le second affirme que les codes bifixes dans un ensemble acyclique sont saturés dans ce même ensemble.

5.1 Graphes d'incidence

On peut représenter les factorisations des mots d'un ensemble X dans un graphe appelé graphe d'incidence. Nous nous intéressons ici aux propriétés de ce graphe lorsque X est un code bifixe dans un ensemble acyclique. Nous montrons en particulier que ce graphe est toujours acyclique.

Définition 5.1.1. Soit X un ensemble. Notons P_X l'ensemble des préfixes propres non vides de X et S_X l'ensemble des suffixes propres non vides de X. Le *graphe d'incidence* de X est le graphe non orienté $G_X = (V, E)$ où V est l'union disjointe de P_X et S_X et où

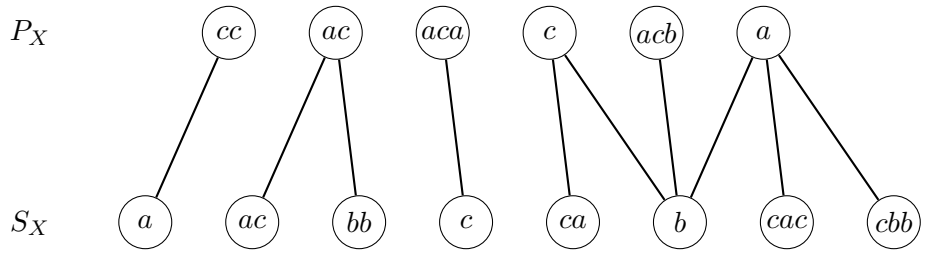
$$E = \{(p, s) \in P_X \times S_X \mid ps \in X\}.$$

Remarque 5.1.2. Si, pour tous $u, v \in X$ distincts, leur plus long préfixe commun est préfixe propre de u et de v, alors X est un code préfixe. En effet, il est alors impossible d'avoir u préfixe de v et $u \neq v$, car ça impliquerait que u est leur plus long préfixe commun et devrait donc être préfixe propre de lui-même, ce qui est impossible. On a bien évidemment un résultat similaire avec les suffixes.

Exemple 5.1.3. Si $X = \{ab, acac, acbb, cb, cca\}$, on a

$$P_X = \{a, ac, aca, acb, c, cc\} \quad \text{et} \quad S_X = \{a, ac, b, bb, c, ca, cac, cbb\}.$$

Le graphe d'incidence de X est alors le graphe



Proposition 5.1.4. *Soient S un ensemble acyclique, $X \subseteq S$ un code bifixe et G_X le graphe d'incidence de X.*

1. *Si $u_1, \dots, u_n \in P_X$ et $v_1, \dots, v_{n+1} \in S_X$ sont tels que $v_1, u_1, v_2, \dots, u_n, v_{n+1}$ est un chemin de G_X (1), alors le plus long préfixe commun à v_1 et v_{n+1} est préfixe propre de $v_1, \dots, v_n$ et v_{n+1}.*
2. *Si $u_1, \dots, u_{n+1} \in P_X$ et $v_1, \dots, v_n \in S_X$ sont tels que $u_1, v_1, u_2, \dots, v_n, u_{n+1}$ est un chemin de G_X (2), alors le plus long suffixe commun à u_1 et u_{n+1} est suffixe propre de $u_1, \dots, u_n$ et u_{n+1}.*

Démonstration. Montrons ces deux résultats en parallèle par récurrence sur $n \in \mathbb{N}_0$.

Si $n = 1$, alors, pour l'affirmation 1, il suffit de montrer que v_1 n'est pas préfixe de v_2 et inversement. Par définition de G_X, $u_1v_1, u_1v_2 \in X$ donc on peut conclure car X est un code préfixe et $v_1 \neq v_2$. On procède de même pour montrer le point 2 quand $n = 1$.

Passons maintenant à l'induction. Supposons les deux résultats vrais pour les chemins de longueur inférieure ou égale à $2n - 2$ et montrons le premier point, l'autre étant symétrique. Notons $U = \{u_1, \dots, u_n\}$, $V = \{v_1, \dots, v_n\}$ et $V' = \{v_2, \dots, v_{n+1}\}$. Pour tous $u, u' \in U$, il existe un chemin de longueur au plus $2n - 2$ les reliant (il suffit de prendre un tronçon du chemin principal) donc, par le point 2 de l'hypothèse de récurrence, le plus long suffixe commun à u et u' est un suffixe propre de ces deux mots. Par la Remarque 5.1.2, U est alors un code suffixe. Par un raisonnement similaire mais utilisant cette fois-ci le point 1 de la récurrence, V et V' sont des codes préfixes.

Notons p le plus long préfixe commun à v_1 et v_{n+1}. Si $v_1 = p$ (resp. $v_{n+1} = p$), alors $p \in V$ (resp. $p \in V'$). Comme S est factoriel et $X \subseteq S$, on a alors que $p, u_1, \dots, u_n, p$ est un cycle de $\mathcal{E}^{U,V}(\varepsilon)$ (resp. $\mathcal{E}^{U,V'}(\varepsilon)$), ce qui est absurde par le Théorème 4.1.8. On en déduit que p est bien un préfixe propre de v_1 et de v_{n+1}.

Montrons à présent que p est également un préfixe propre de $v_2, \dots, v_n$. Posons

$$W = (V \setminus pA^*) \cup \{p\}.$$

Il s'agit d'un code préfixe par la Remarque 2.1.4 car V en est un et que $p \in \mathrm{Pref}(v_1) \subseteq \mathrm{Pref}(V)$. Procédons par l'absurde et supposons que p ne soit pas préfixe propre d'un des v_k, autrement dit qu'il existe k tel que $v_k \in W$. Notons alors $i < j \leq n$ tels que $v_i, v_{j+1} \notin W$ et

(1). On suppose $u_i \neq u_{i+1}$ et $v_i \neq v_{i+1}$, de sorte que le chemin n'emprunte pas la même arête deux fois consécutives.

(2). Même remarque que dans la note (1).

$v_{i+1}, \dots, v_j \in W$ [(3)]. Comme v_i et v_{j+1} ont p pour préfixe propre [(4)], on a le cycle

$$p, u_i, v_{i+1}, \dots, v_j, u_j, p$$

dans $\mathcal{E}^{U,W}(\varepsilon)$, ce qui est absurde par le Théorème 4.1.8. □

Corollaire 5.1.5. *Si S est acyclique, alors pour tout code bifixe $X \subseteq S$, le graphe d'incidence de X est acyclique.*

Démonstration. La conclusion est immédiate car s'il existe un cycle $v_1, u_1, v_2, \dots, u_n, v_1$, la proposition précédente implique que v_1 est son propre préfixe propre, ce qui est absurde. □

Exemple 5.1.6. Bien que le graphe d'incidence de l'ensemble X de l'Exemple 5.1.3 soit acyclique, X n'est inclus dans aucun ensemble acyclique. En effet, X est un code bifixe et on a le chemin ca, c, b, a, cac dans G_X et ca (qui est le plus long préfixe commun à ca et cac) n'est préfixe propre ni de ca, ni de b, ce qui contredit le premier point de la Proposition 5.1.4.

5.2 Suite admissible et Théorème d'indépendance

Nous introduisons à présent la notion de suite admissible par rapport à des mots $y_1, \dots, y_n$. Il s'agit d'une façon de décrire les simplifications à faire pour trouver $\rho(y_1 \dots y_n)$. Ce concept est donc logiquement lié au caractère libre d'un ensemble et nous permet de montrer qu'un ensemble S est acyclique si et seulement si tous les codes bifixes inclus dans S sont libres. C'est ce que nous appelons le Théorème d'indépendance.

Définition 5.2.1. Soient X un ensemble et $y_1, \dots, y_n \in X \cup X^{-1}$. Une suite $(u_i, v_i, w_i)_{1 \leq i \leq n}$ où $u_i, v_i, w_i \in F_A$ est *admissible* par rapport à $y_1, \dots, y_n$ si

1. pour tout $i \leq n$, $y_i = u_i v_i w_i$,
2. $u_1 = \varepsilon = w_n$,
3. $v_1 \neq \varepsilon$ et $v_n \neq \varepsilon$,
4. pour tout $i \leq n-1$, $u_{i+1} = w_i^{-1}$,
5. pour tous $i < j \leq n$, si $v_i \neq \varepsilon$, $v_j \neq \varepsilon$ et si, pour tout $k \in \{i+1, \dots, j-1\}$, $v_k = \varepsilon$, alors $v_i v_j$ est réduit.

Remarque 5.2.2. Par construction, si la suite $(u_i, v_i, w_i)_{1 \leq i \leq n}$ est admissible par rapport à $y_1, \dots, y_n$, alors

$$y_1 \dots y_n \equiv v_1 \dots v_n.$$

De plus, $v_1 \dots v_n$ est réduit par la condition 5 donc $\rho(y_1 \dots y_n) = v_1 \dots v_n$. Par le point 3, on a en particulier $y_1 \dots y_n \not\equiv \varepsilon$.

(3). Ce qui est possible car $v_1, v_{n+1} \notin W$. Il suffit par exemple de prendre

$$i = \max\{i' < k \mid v_i \notin W\} \quad \text{et} \quad j = \min\{j' \geq k \mid v_{j+1} \notin W\}.$$

(4). En effet, $v_i \in V \setminus W = V \cap pA^+$. De même pour v_{j+1}, il faut simplement traiter à part le cas $j = n$ qui est immédiat.

Remarque 5.2.3. Au vu du point 4 de la Définition 5.2.1, si $X \subseteq A^*$ et si $w_i \neq \varepsilon$, alors soit $y_i \in X$ et $y_{i+1} \in X^{-1}$, soit l'inverse. On peut étendre cette réflexion si $w_i, \ldots, w_{i+k}$ sont tous non vides. Dans ce cas, si k est pair, alors y_i et y_{i+k+1} ne peuvent pas tous les deux provenir de X ou de X^{-1} et, si k est impair, ils doivent provenir du même ensemble.

Lemme 5.2.4. *Si $X \subseteq A^*$ est un code bifixe, alors, pour tous $x, y \in X$, $x \neq y$, $\rho(xy^{-1})$ n'est ni préfixe de x, ni suffixe de y^{-1}. De même, $\rho(y^{-1}x)$ n'est ni préfixe de y^{-1}, ni suffixe de x.*

Démonstration. On a

$$xy^{-1} \equiv \rho(xy^{-1}) \equiv \rho(xy^{-1})yy^{-1}$$

donc

$$x \equiv \rho(xy^{-1})y.$$

Si $\rho(xy^{-1})$ est un préfixe de x, x et $\rho(xy^{-1})y$ sont des mots sur A^*. Ils sont alors égaux. Cela signifie donc que y est un suffixe de x, ce qui est absurde. De même, on a

$$xy^{-1} \equiv \rho(xy^{-1}) \equiv xx^{-1}\rho(xy^{-1})$$

donc, par un raisonnement similaire, si $\rho(xy^{-1})$ est un suffixe de y, alors

$$y^{-1} = x^{-1}\rho(xy^{-1})$$

et

$$y = (\rho(xy^{-1}))^{-1}x$$

ce qui est absurde. Un raisonnement symétrique utilisant cette fois le caractère préfixe de X permet de conclure pour $\rho(y^{-1}x)$. □

Lemme 5.2.5. *Soit $X \subseteq A^*$ un code bifixe et soient $y_1, \ldots, y_n \in X \cup X^{-1}$ tels que $y_i \neq (y_{i+1})^{-1}$ pour tout $i < n$. Si $(u_i, v_i, w_i)_{1 \leq i \leq n}$ est un suite admissible par rapport à $y_1, \ldots, y_n$, alors, pour tout $i < n$,*

$$u_i v_i \neq \varepsilon \quad \text{et} \quad v_{i+1} w_{i+1} \neq \varepsilon.$$

En particulier, si $(u_i, v_i, w_i)_{1 \leq i \leq n}$ vérifie les points 1, 2 et 4 de la Définition 5.2.1, alors $v_1 \neq \varepsilon$ et $v_n \neq \varepsilon$.

Démonstration. Pour tout $i < n$, procédons par l'absurde. Si $u_i v_i = \varepsilon$, alors

$$\rho(y_i y_{i+1}) = \rho(w_i u_{i+1} v_{i+1} w_{i+1}) = v_{i+1} w_{i+1},$$

donc $\rho(y_i y_{i+1})$ est un suffixe de y_{i+1}. Or, X est un code bifixe donc $y_i \neq \varepsilon$ et on a alors $w_i \neq \varepsilon$. Par la Remarque 5.2.3, y_i et $(y_{i+1})^{-1}$ sont sur le même alphabet (soit A, soit A^{-1}). On peut donc leur appliquer le Lemme 5.2.4 avec X ou X^{-1}, ce qui prouve l'absurdité. De même, si $v_{i+1} w_{i+1} = \varepsilon$, alors

$$\rho(y_i y_{i+1}) = u_i v_i,$$

ce qui est également absurde.

Pour obtenir ce résultat, nous n'avons utilisé que les points 1 et 4 de la Définition 5.2.1. Si, de plus, $u_1 = \varepsilon = w_n$, on doit donc avoir $v_1 \neq \varepsilon$ et $v_n \neq \varepsilon$. □

Nous pouvons à présent prouver le résultat principal.

Théorème 5.2.6 (Théorème d'indépendance)**.** *Un ensemble biprolongeable S est acyclique si, et seulement si, tout code bifixe $X \subseteq S$ est un sous-ensemble libre de F_A.*

Démonstration. Commençons par montrer que si tout code bifixe $X \subseteq S$ est un sous-ensemble libre de F_A, alors S est acyclique. Par l'absurde, supposons qu'il existe $w \in S$ tel que $\mathcal{E}(w)$ contienne un cycle. Notons $a_1, b_1, a_2, \ldots, a_n, b_n, a_1$ ce cycle, $a_i \in L(w)$, $b_i \in R(w)$. Posons $X = AwA \cap S$. Comme tous les mots de X ont la même longueur, il s'agit d'un code bifixe. De plus, par définition,

$$a_1wb_1, a_2wb_1, a_2wb_2, \ldots, a_nwb_n, a_1wb_n \in X.$$

Or,

$$\begin{aligned} & a_1wb_1(a_2wb_1)^{-1}a_2wb_2\ldots(a_nwb_{n-1})^{-1}a_nwb_n(a_1wb_n)^{-1} \\ \equiv\ & a_1wb_1(b_1)^{-1}w^{-1}(a_2)^{-1}a_2wb_2\ldots(b_{n-1})^{-1}w^{-1}(a_n)^{-1}a_nwb_n(b_n)^{-1}w^{-1}(a_1)^{-1} \\ \equiv\ & a_1(a_2)^{-1}a_2\ldots(a_n)^{-1}a_n(a_1)^{-1} \\ \equiv\ & \varepsilon, \end{aligned}$$

ce qui est absurde car X est libre par hypothèse.

Supposons à présent avoir S acyclique et $X \subseteq S$ un code bifixe. Nous devons donc montrer que, pour tous $y_1, \ldots, y_n \in X \cup X^{-1}$ tels que $y_i \neq (y_{i+1})^{-1}$, on a $y_1 \ldots y_n \not\equiv \varepsilon$. Par la Remarque 5.2.2, il suffit de montrer qu'il existe une suite admissible par rapport à $y_1, \ldots, y_n$. Pour cela procédons par récurrence sur n. Si $n = 1$, alors on peut prendre la suite $(\varepsilon, y_1, \varepsilon)$.

Supposons le résultat vrai pour n et montrons-le pour $n+1$. Soient $y_1, \ldots, y_{n+1} \in X \cup X^{-1}$ tels que $y_i \neq (y_{i+1})^{-1}$ pour tout $i \leq n$. Par hypothèse de récurrence, il existe une suite admissible par rapport à $y_1, \ldots, y_n$. Notons $(u_i, v_i, w_i)_{1 \leq i \leq n}$ une telle suite choisie de sorte à maximiser $|v_n|$.

Soit u_{n+1} le préfixe de y_{n+1} de longueur maximale tel que $(u_{n+1})^{-1}$ soit un suffixe de v_n. Posons alors

$$w'_n = (u_{n+1})^{-1}, \quad v_n = v'_n w'_n \quad \text{et} \quad y_{n+1} = u_{n+1}v_{n+1}.$$

Il est immédiat que la suite

$$s = (u_1, v_1, w_1), \ldots, (u_{n-1}, v_{n-1}, w_{n-1}), (u_n, v'_n, w'_n), (u_{n+1}, v_{n+1}, \varepsilon)$$

vérifie les points 1, 2 et 4 de la Définition 5.2.1. Par le Lemme 5.2.5, le point 3 de la Définition 5.2.1 est donc également vérifié. Pour le point 5, il est uniquement nécessaire de le vérifier quand $j = n + 1$ car v'_n est un préfixe de v_n.

Si $v'_n \neq \varepsilon$, alors la suite s est admissible. En effet, $v'_n v_{n+1}$ est réduit car u_{n+1} a été choisi de longueur maximale.

Si $v'_n = \varepsilon$, notons $i < n$ maximal tel que $v_i \neq \varepsilon$. Par le Lemme 5.2.5, $w_{i+1}, \ldots, w_{n-1}, w'_n$ sont non vides. De plus, $u_{i+1} \neq \varepsilon$ donc on a également $w_i \neq \varepsilon$. Si $n - i$ est impair, par la Remarque 5.2.3, y_i et y_{n+1} sont soit tous les deux dans X, soit dans X^{-1}. En particulier, v_i et v_{n+1} sont sur le même alphabet donc $v_i v_{n+1}$ est réduit. Le point 5 de la Définition 5.2.1 étant alors vérifié, la suite s est admissible.

Si $n - i$ est pair, supposons $y_i \in X$ (l'autre cas se résout de façon symétrique). On a alors $y_{n+1} \in X^{-1}$. Dans le graphe d'incidence G_X de X, on a le chemin

$$u_i v_i \to w_i \to u_{i+2} \to w_{i+2} \to u_{i+4} \to \cdots \to u_n \to w'_n \to (v_{n+1})^{-1}.$$

En effet, $u_iv_iw_i = y_i$, $u_{i+2k}w_{i+2k} = y_{i+2k}$ $(k \neq 0)$ et $u_nw'_n = y_n$ sont tous des éléments de X. De plus, pour tout $0 \leq k < \frac{n-i}{2}$,

$$u_{i+2k+2}w_{i+2k} = \left((w_{i+2k})^{-1}(u_{i+2k+2})^{-1}\right)^{-1} = (u_{i+2k+1}w_{i+2k+1})^{-1} = (y_{i+2k+1})^{-1} \in X$$

donc (w_{i+2k}, u_{i+2k+2}) est une arête de G_X. Par un raisonnement similaire,

$$(v_{n+1})^{-1}w'_n = (u_{n+1}v_{n+1})^{-1} \in X.$$

Par la Proposition 5.1.4, si x est le plus long suffixe commun à u_iv_i et à $(v_{n+1})^{-1}$, il est suffixe propre de $u_iv_i, u_{i+2}, u_{i+4}, \ldots, u_n, (v_{n+1})^{-1}$.

Si v_i est un suffixe de x, cela signifie que v_i est lui-même un suffixe propre de ces mots et que v_i^{-1} est un préfixe propre de $w_{i+1}, w_{i+3}, \ldots, w_{n-1}, v_{n+1}$. La première partie implique que, pour $k > 0$, [(5)]

$$y_{i+2k} = u_{i+2k}w_{i+2k} = \rho(u_{i+2k}v_i^{-1})v_iw_{i+2k}.$$

et la seconde que, pour $k \geq 0$,

$$y_{i+2k+1} = u_{i+2k+1}w_{i+2k+1} = u_{i+2k+1}v_i^{-1}\rho(v_iw_{i+2k+1}).$$

Considérons la suite

$$\begin{gathered}(u_1, v_1, w_1),\ \ldots,\ (u_{i-1}, v_{i-1}, w_{i-1}),\\ (u_i, \varepsilon, v_iw_i),\ (u_{i+1}v_i^{-1}, \varepsilon, \rho(v_iw_{i+1})),\\ (\rho(u_{i+2}v_i^{-1}), \varepsilon, v_iw_{i+2}),\ \ldots,\ (\rho(u_nv_i^{-1}), v_iw'_n, \varepsilon).\end{gathered}$$

Il s'agit d'une suite admissible par rapport à $y_1, \ldots, y_n$. En effet, les points 1 à 4 de la Définition 5.2.1 sont facilement vérifiés. Pour le point 5, si $j < i$ est le plus grand indice tel que $v_j \neq \varepsilon$, alors par hypothèse sur la suite admissible de départ, v_jv_i est réduit. Comme v_i et w'_n sont sur le même alphabet, on en déduit que $v_jv_iw'_n$ est réduit. Or,

$$|v_iw'_n| = |v_iv_n| > |v_n|,$$

ce qui est impossible vu le choix de la suite $(u_k, v_k, w_k)_{1\leq k\leq n}$.

On peut donc en conclure que v_i a x pour suffixe propre. Rappelons également que x est suffixe propre de $(v_{n+1})^{-1}$. Notons $v_i = px$ et $(v_{n+1})^{-1} = qx$. Considérons la suite

$$\begin{gathered}(u_1, v_1, w_1),\ \ldots,\ (u_{i-1}, v_{i-1}, w_{i-1}),\\ (u_i, p, xw_i),\ (u_{i+1}x^{-1}, \varepsilon, \rho(xw_{i+1})),\\ (\rho(u_{i+2}x^{-1}), \varepsilon, xw_{i+2}),\ \ldots,\ (\rho(u_nx^{-1}), \varepsilon, xw'_n),\\ (u_{n+1}x^{-1}, q^{-1}, \varepsilon).\end{gathered}$$

Il s'agit d'une suite admissible par rapport à $y_1, \ldots, y_{n+1}$ car x a été choisi maximal donc pq^{-1} est réduit.

Nous avons donc trouvé une suite admissible par rapport à $y_1, \ldots, y_{n+1}$ dans tous les cas, ce qui permet de conclure l'induction et la démonstration. □

(5). Pour simplifier les notations, si $i + 2k = n$, nous supposons ici $w_{i+2k} = w'_n$.

5.3 Automates co-déterministes

Nous faisons à présent un retour sur les automates pour introduire les notions d'automate co-déterministe et d'automate minimal d'un ensemble. Nous nous intéressons aux propriétés de ce dernier lorsque l'ensemble en question est un code préfixe.

Définition 5.3.1. Un automate (déterministe) $\mathcal{A} = (Q, q_0, F, A, \delta)$ est *émondé* si, pour tout $q \in Q$, il existe un chemin de q_0 vers q et de q vers un état accepteur. Il est *simple* s'il est émondé et si $F = \{q_0\}$.

Définition 5.3.2. Un automate simple $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ est *co-déterministe* si, pour tous $a \in A$, $q \in Q$, il existe au plus un $p \in Q$ tel que $\delta(p, a) = q$. Dans ce cas, on peut définir l'automate *miroir* de $\mathcal{A}$ comme étant $\mathcal{A}^R = (Q, q_0, \{q_0\}, A, \delta')$ où

$$\delta'(q, a) = p \ \Leftrightarrow \ \delta(p, a) = q.$$

Proposition 5.3.3.

1. *Le langage accepté par un automate simple est engendré par un code préfixe.*
2. *Le langage accepté par un automate co-déterministe est engendré par un code bifixe.*

Démonstration. Soit $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ un automate simple. Notons X l'ensemble des mots non vides correspondant à un chemin de $\mathcal{A}$ commençant et se terminant en q_0 mais ne passant pas à un autre moment par q_0. On a directement que $L(\mathcal{A}) = X^*$.

1. Montrons que X est un code préfixe. Soient $x, y \in X$ tels que x soit un préfixe de y. Il existe donc $z \in A^*$ tel que $y = xz$. Le chemin correspondant à y passe alors par

   $$\delta(q_0, x) = q_0.$$

   Par définition de X, on doit avoir $z = \varepsilon$ et $x = y$ donc X est un code préfixe.

2. Montrons que si $\mathcal{A}$ est co-déterministe, X est également un code suffixe. Soient $x, y \in X$ tels que x soit un suffixe de y. Notons $y = zx$. Étant donné que $\mathcal{A}$ est co-déterministe, on a

   $$\delta(q_0, y) = q_0 = \delta(q_0, x) \Rightarrow \delta(q_0, zx_{[1,n-1]}) = \delta(q_0, x_{[1,n-1]}).$$

   En itérant, on trouve alors

   $$\delta(q_0, z) = q_0,$$

   ce qui implique que $z = \varepsilon$ et $x = y$.

□

Définition 5.3.4. L'automate *minimal* d'un ensemble non vide $X \subseteq A^*$ est l'automate $\mathcal{A}_X = (Q, q_0, F, A, \delta)$ où
- $Q = \{w^{-1}X \mid w \in A^*, w^{-1}X \neq \emptyset\}$,
- $q_0 = X = \varepsilon^{-1}X$,
- $F = \{w^{-1}X \mid w \in X\}$,
- $\delta = \{(w^{-1}X, a, (wa)^{-1}X) \mid a \in A \text{ et } w^{-1}X, (wa)^{-1}X \in Q\}$.

Un automate est *minimal* s'il est isomorphe à l'automate minimal du langage qu'il accepte.

Par construction, $\delta(q_0, w) = w^{-1}X$ si $X \cap wA^*$ est non vide et n'est pas défini sinon. On en déduit qu'un mot w est accepté par l'automate $\mathcal{A}_X$ si et seulement si $w \in X$.

Proposition 5.3.5. *Soit X un code préfixe.*

1. *Pour tout $x \in X^*$, $x^{-1}X^* = X^*$.*
2. *L'automate minimal de X^* est simple.*

Démonstration.

1. Il est évident que $X^* \subseteq x^{-1}X^*$ car, pour tout $y \in X^*$, on a également $xy \in X^*$. Montrons l'autre inclusion. Soient $x_1, \dots, x_n \in X$ tels que $x = x_1 \dots x_n$. Montrons que
$$x^{-1}X^* \subseteq (x_2 \dots x_n)^{-1}X^* \subseteq \cdots \subseteq x_n^{-1}X^* \subseteq X^*.$$
Toutes les inclusions se montrent de façon similaire donc ne montrons que la première. Soit $y \in x^{-1}X^*$. On a donc $xy \in X^*$. Notons $y_1, \dots, y_m \in X$ tels que $xy = y_1 \dots y_m$. Comme X est un code préfixe, on doit avoir $x_1 = y_1$ donc $x_2 \dots x_n y = y_2 \dots y_m \in X^*$ et $y \in (x_2 \dots x_n)^{-1}X^*$.
2. Notons $\mathcal{A}_{X^*} = (Q, q_0, F, A, \delta)$ l'automate minimal de X^*. Pour tout $p = w^{-1}X^* \in Q$, on a
$$\delta(q_0, w) = w^{-1}X^* = p.$$
De plus, $w^{-1}X^* \neq \emptyset$ par hypothèse donc, pour $v \in w^{-1}X^*$, on a $wv \in X^*$ donc
$$\delta(p, v) = (wv)^{-1}X^* = X^* = q_0.$$
En conséquence, l'automate $\mathcal{A}_{X^*}$ est émondé. Il est simple car
$$F = \{w^{-1}X^* \mid w \in X^*\} = \{X^*\}.$$

□

Les notions d'automate co-déterministe et d'automate minimal sont fortement liées, comme le montrent les résultats qui suivent.

Nous citons ici un résultat ([16, Proposition IV.3.10]) que nous ne redémontrons pas.

Proposition 5.3.6. *Soit $\mathcal{A} = (Q, q_0, F, A, \delta)$ un automate émondé. Il est minimal si, et seulement si, l'application*
$$\varphi : Q \to \mathcal{P}(A^*) \quad q \mapsto \{w \in A^* \mid \delta(q, w) \in F\}$$
est injective.

Proposition 5.3.7. *Tout automate co-déterministe est minimal.*

Démonstration. Soit $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ un automate co-déterministe. Montrons que l'application φ de la Proposition 5.3.6 est injective. Pour cela, montrons plus généralement que s'il existe $w \in A^*$ tel que $\delta(p, w) = \delta(q, w)$, alors $p = q$.

Procédons par récurrence sur $|w|$. Si $w = \varepsilon$, la conclusion est immédiate. Supposons le résultat vrai pour les mots de longueur $|w| - 1$ et montrons-le pour w. Notons $w = va$, $a \in A$. On a alors
$$\delta(\delta(p, v), a) = \delta(p, w) = \delta(q, w) = \delta(\delta(q, v), a).$$
Comme $\mathcal{A}$ est co-déterministe, on a $\delta(p, v) = \delta(q, v)$, ce qui permet de conclure par hypothèse de récurrence.

A présent, l'automate $\mathcal{A}$ étant émondé, $\varphi(q)$ est non vide pour tout $q \in Q$, ce qui signifie que si $\varphi(q) = \varphi(p)$, il existe un mot w tel que $\delta(p, w) = q_0 = \delta(q, w)$ donc $q = p$ et φ est injectif. □

Cependant, la réciproque n'est en général pas vraie. On a tout de même le résultat suivant.

Proposition 5.3.8. *Soit X un code préfixe. Les affirmations suivantes sont équivalentes :*

1. $X^* = \langle X \rangle \cap A^*$,
2. *l'automate minimal $\mathcal{A}_{X^*} = (Q, q_0, F, A, \delta)$ de X^* est co-déterministe*

Démonstration. Supposons $X^* = \langle X \rangle \cap A^*$. Par la Proposition 5.3.5, l'automate $\mathcal{A}_{X^*}$ est simple. Montrons qu'il est co-déterministe. Soient $a \in A$ et $q \in Q$. Supposons qu'il existe $p, p' \in Q$ tels que $\delta(p, a) = q = \delta(p', a)$. Par définition de Q, il existe $w, v \in A^*$ tels que $p = w^{-1}X^*$, $p' = v^{-1}X^*$. Montrons que $w^{-1}X^* = v^{-1}X^*$ donc que $p = p'$. L'automate $\mathcal{A}_{X^*}$ étant simple, il existe $u \in A^*$ tel que $\delta(q, u) = q_0$. On a alors

$$\delta(q_0, wau) = \delta(p, au) = \delta(q, u) = q_0$$

donc $wau \in X^*$. De même, $vau \in X^*$. Soit $x \in w^{-1}X^*$. On a $vx \in A^*$ donc

$$vx = \rho(vauu^{-1}a^{-1}w^{-1}wx) = \rho(vau(wau)^{-1}wx) \in \langle X \rangle.$$

Par hypothèse, on a alors $vx \in X^*$, ce qui signifie que $x \in v^{-1}X^*$ et $w^{-1}X^* \subseteq v^{-1}X^*$. On procède de façon symétrique pour montrer l'autre inclusion.

Supposons à présent $\mathcal{A}_{X^*}$ co-déterministe. On a immédiatement $X^* \subseteq \langle X \rangle \cap A^*$ donc montrons l'autre inclusion. Pour cela, notons $\mathcal{B}$ l'automate $(Q', q_0, \{q_0\}, A \cup A^{-1}, \delta')$ où

$$\delta' = \delta \cup \{(p, a^{-1}, q) \in Q \times A^{-1} \times Q \mid (q, a, p) \in \delta\}$$

qui est déterministe car $\mathcal{A}_{X^*}$ est co-déterministe. Pour tout $x = x_1 \dots x_n \in (X \cup X^{-1})^*$, on a $\delta'(q_0, x) = q_0$ car, si $x_i \in X$, le lire dans $\mathcal{B}$ est équivalent à le lire dans $\mathcal{A}_{X^*}$ donc correspond à un chemin de q_0 vers q_0 et si $x_i \in X^{-1}$, alors lire x_i dans $\mathcal{B}$ revient à prendre le chemin correspondant à x_i^{-1} dans $\mathcal{A}_{X^*}$ à l'envers.

Remarquons également que lire aa^{-1} à partir d'un état $p \in Q$ nous ramène à l'état p car si $q = \delta(p, a)$, il n'y a qu'à partir de p qu'on peut atteindre q en lisant a, et donc que vers p qu'on peut aller en lisant a^{-1} à partir de q.

En conclusion, tout mot $x \in \langle X \rangle$ est tel que

$$\delta'(q_0, x) = q_0$$

dans $\mathcal{B}$. Si, de plus, $x \in A^*$, alors le chemin emprunté est un chemin de $\mathcal{A}_{X^*}$ donc

$$x \in L(\mathcal{A}_{X^*}) = X^*.$$

□

Corollaire 5.3.9. *Soit $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ un automate co-déterministe. Si X est le code préfixe tel que $L(\mathcal{A}) = X^*$, alors*

$$X^* = \langle X \rangle \cap A^*.$$

Démonstration. Si $\mathcal{A}$ est co-déterministe, alors, par la Proposition 5.3.7, il est minimal. Par la Proposition 5.3.8, on en déduit que $X^* = \langle X \rangle \cap A^*$. □

Dans le cas d'un automate simple, on connaît exactement le sous-groupe qu'il décrit.

Proposition 5.3.10. *Soit $\mathcal{A} = (Q, q_0, \{q_0\}, A, \delta)$ un automate simple. Si X est le code préfixe tel que $L(\mathcal{A}) = X^*$, alors le sous-groupe décrit par $\mathcal{A}$ est $\langle X \rangle$.*

Démonstration. Notons H le sous-groupe décrit par $\mathcal{A}$. Comme $X \subseteq H$, on a $\langle X \rangle \subseteq H$. Montrons donc l'autre inclusion. Par définition, H est l'ensemble des réductions des mots acceptés par l'automate $\mathcal{B} = (Q, \{q_0\}, \{q_0\}, A, \Delta)$ où

$$\Delta = \delta \cup \{(p, a^{-1}, q) \in Q \times A^{-1} \times Q \mid (q, a, p) \in \delta\}.$$

Pour montrer que $H \subseteq \langle X \rangle$, procédons par récurrence sur le nombre de lettres de $w \in L(\mathcal{B})$ qui sont dans A^{-1}. Si $w \in A^*$, alors $w \in L(\mathcal{A}) = X^*$ donc $\rho(w) = w \in \langle X \rangle$.

Supposons à présent le résultat vrai pour les mots ayant strictement moins de lettres dans A^{-1} que $w \in (A \cup A^{-1})^n$. Notons i le premier indice tel que $w_i \in A^{-1}$ et notons $u = w_{[1,i-1]}$ et $v = w_{[i+1,n]}$. Sur le chemin de $\mathcal{B}$ qui permet d'accepter w, notons p l'état atteint en lisant u et q celui atteint en lisant uw_i. Par construction de $\mathcal{B}$, comme $u \in A^*$, on a donc

$$\delta(q_0, u) = p$$

et

$$\delta(q, w_i^{-1}) = p.$$

Étant donné que $\mathcal{A}$ est émondé, il existe un chemin de p vers q_0 étiqueté par $x \in A^*$ et un chemin de q_0 vers q étiqueté par $y \in A^*$. On a alors

$$ux, yw_i^{-1}x \in L(\mathcal{A}) = X^*.$$

De plus, yv est dans $L(\mathcal{B})$ et a strictement moins de lettres dans A^{-1} que w donc, par hypothèse de récurrence, $\rho(yv) \in \langle X \rangle$. On en déduit que

$$\rho(w) = \rho(uw_iv) = \rho(ux(yw_i^{-1}x)^{-1}yv) \in \langle X \rangle,$$

ce qui permet de conclure la récurrence. □

5.4 Automate littéral et automate quotient

Il est possible de définir d'autres automates liés à un ensemble X lorsque ce dernier est un code bifixe. Ces automates sont basés sur une relation définie sur les préfixes propres de X. Elle est intimement liée au graphe d'incidence de X défini dans la Section 5.1. Lorsque l'ensemble X est inclus dans un ensemble acyclique, on peut étudier les différentes propriétés de ces automates. Nous montrerons notamment que l'automate quotient décrit le sous-groupe engendré par X. Nous prouverons également le Théorème de saturation affirmant que, sous ces conditions,

$$X^* \cap S = \langle X \rangle \cap S.$$

Définition 5.4.1. Soient X un code bifixe et P l'ensemble de ses préfixes propres. Définissons la relation θ_X sur P comme étant la clôture transitive de la relation telle que $p \equiv q$ s'il existe $s \in A^+$ tel que $ps, qs \in X$.

Comme X est un code bifixe, ε n'est θ_X-équivalent à aucun autre mot car cela impliquerait qu'il existe $s, q \in A^+$ tels que $s, qs \in X$.

Sur $P \setminus \{\varepsilon\}$, la relation peut se traduire par $p \equiv q \mod \theta_X$ si, et seulement si, il existe un chemin de p vers q dans le graphe d'incidence de X. Les classes de P/θ_X correspondent donc à $\{\varepsilon\}$ et aux projections des composantes connexes du graphe G_X sur P.

Proposition 5.4.2. *Soient X un code bifixe et P l'ensemble de ses préfixes propres. Pour tous $p, q \in P$, si $p \equiv q \mod \theta_X$, alors* [(6)]

$$\langle X \rangle \cdot p = \langle X \rangle \cdot q.$$

Démonstration. Si p ou q est vide, le résultat est immédiat. Sinon, vu ce qui a été observé ci-dessus, il existe un chemin de p vers q dans le graphe d'incidence de X. Procédons par récurrence sur la longueur n de ce chemin ou, plus précisément, étant donné que la longueur est paire, nous effectuerons la récurrence sur k tel que $n = 2k$. Si $k = 0$, alors la conclusion est immédiate car $p = q$.

Supposons le résultat vrai pour les chemins de longueur $2(k-1)$ et montrons-le pour p et q reliés par un chemin de longueur $2k$. Notons $r \in P$ le premier sommet (autre que p) de P par lequel passe ce chemin. Par hypothèse de récurrence, $\langle X \rangle \cdot r = \langle X \rangle \cdot q$. De plus, par définition du graphe d'incidence, il existe $s \in A^+$ tel que $ps, rs \in X$. Dans ce cas,

$$p = \rho(ps(rs)^{-1}r) \in \langle X \rangle \cdot r$$

donc $\langle X \rangle \cdot p \subseteq \langle X \rangle \cdot r$. On procède de même pour montrer l'autre inclusion. □

Nous définissons à présent le premier des deux automates donnant leurs noms à cette section et montrons une première propriété le concernant.

Définition 5.4.3. Soit X un code préfixe. Notons P l'ensemble des préfixes propres de X. L'automate *littéral* de X^* est l'automate $\mathcal{A} = (P, \varepsilon, \{\varepsilon\}, A, \delta)$ où

$$\delta = \{(p, a, pa) \mid a \in A, pa \in P\} \cup \{(p, a, \varepsilon) \mid a \in A, pa \in X\}.$$

Cet automate est déterministe car X est un code préfixe donc $P \cap X = \emptyset$. De plus, il accepte le langage X^*. Notons également qu'il est simple car, pour tout $p \in P$, $\delta(\varepsilon, p) = p$ et par définition, il existe $x \in A^+$ tel que $px \in X$ donc $\delta(p, x) = \varepsilon$.

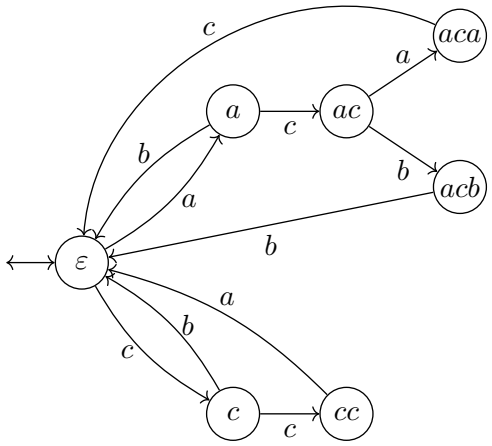
Exemple 5.4.4. Reprenons l'ensemble $X = \{ab, acac, acbb, cb, cca\}$ de l'Exemple 5.1.3. Il s'agit d'un code préfixe et $P = \{\varepsilon, a, ac, aca, acb, c, cc\}$. L'automate littéral de X^* est alors l'automate

(6). La notation $\langle X \rangle \cdot y$ représente

$$\{z \in F_A \mid \exists\, x \in X \text{ tq. } z = x \cdot y = \rho(xy)\},$$

de même que la notation Sw représente

$$\{u \in A^* \mid \exists\, v \in S \text{ tq. } u = vw\}.$$



Proposition 5.4.5. *Soient S un ensemble acyclique, $X \subseteq S$ un code bifixe et P l'ensemble de ses préfixes propres. Pour tous $p, q \in P$, $a \in A$, si $pa, qa \in P \cup X$, alors*

$$p \equiv q \mod \theta_X \Leftrightarrow \delta(p,a) \equiv \delta(q,a) \mod \theta_X$$

où δ est la relation de transition de l'automate littéral de X^*.

Démonstration. Supposons $p \equiv q \mod \theta_X$. Si $p = q$ [7], la conclusion est immédiate. Sinon, il existe un chemin $p, v_1, u_1, \ldots, u_{n-1}, v_n, q$ dans le graphe d'incidence G_X de X. On peut supposer que ce chemin ne passe pas deux fois par le même sommet. Comme $pa, qa \in P \cup X$, il existe également $x, y \in A^*$ tels que $pax, qay \in X$. On a donc le chemin

$$ax, p, v_1, \ldots, v_n, q, ay.$$

Si $n = 1$ et $ax = v_1 = ay$, alors $pax, qax \in X$ donc $pa \equiv qa \mod \theta_X$. Supposons ne pas être dans ce cas-là. Par la Proposition 5.1.4, cela signifie que a est un préfixe propre de ax, ay et de tous les v_i [8]. En particulier, x et y sont non vides donc $pa, qa \in P$ et on a $\delta(p,a) = pa$ et $\delta(q,a) = qa$. Notons $v_i = av_i'$. On a le chemin

$$pa, v_1', u_1 a, \ldots, u_{n-1} a, v_n', qa$$

dans G_X donc $pa \equiv qa \mod \theta_X$.

Réciproquement, supposons avoir $\delta(p,a) \equiv \delta(q,a)$. Si $pa \in X$, alors $\delta(p,a) = \varepsilon$ donc $qa \in X$ et inversement. Dans ce cas, on a directement $p \equiv q$. Si $pa, qa \in P$, alors il existe un chemin

$$pa, v_1, u_1, \ldots, u_{n-1}, v_n, qa$$

dans G_X. On suppose que ce chemin ne passe pas deux fois par le même sommet [9]. Par la Proposition 5.1.4, a est un suffixe propre de pa, qa et des u_i. Notons $u_i = u_i' a$. On a alors le chemin

$$p, av_1, u_1', \ldots, u_{n-1}', av_n, q$$

(7). Et donc également si p ou q est vide.

(8). En effet, soit $ax \neq v_1$ et $ay \neq v_n$ et on peut appliquer directement la Proposition 5.1.4, soit $ax = v_1$ ou $ay = v_n$ et, dans ce cas, on applique la proposition au chemin commençant en v_1 ou se terminant en v_n. Le chemin obtenu passera par au moins deux sommets différents parmi $ax, v_1, \ldots, v_n, ay$ car le cas dégénéré a été écarté.

(9). On suppose donc $pa \neq qa$, auquel cas la conclusion est immédiate.

dans G_X donc $p \equiv q \mod \theta_X$. □

L'automate quotient d'un code bifixe X est défini comme étant le quotient de l'automate littéral de X^* par la relation θ_X. Plus précisément, on a la définition suivante où nous notons $1_X := [\varepsilon]_{\theta_X}$ pour alléger les notations.

Définition 5.4.6. Soient X un code bifixe et P l'ensemble de ses préfixes propres. *L'automate quotient* de X est l'automate $\mathcal{B}_X = (P/\theta_X, 1_X, \{1_X\}, A, \delta')$ où $(P, a, Q) \in \delta'$ s'il existe $p \in P$, $q \in Q$ tels que $\delta(p, a) = q$ où δ est la relation de transition de l'automate littéral de X^*.

Remarquons qu'il s'agit bien d'un automate déterministe par la Proposition 5.4.5. En effet, si $p \equiv p' \mod \theta_X$ et si $\delta(p, a) = q$, alors $\delta(p', a)$ est soit indéfini, soit équivalent à q.

Remarque 5.4.7. À tout chemin de p vers q étiqueté par w dans l'automate littéral de X^* correspond un chemin de $[p]_{\theta_X}$ vers $[q]_{\theta_X}$ étiqueté par w dans $\mathcal{B}_X$. On peut le vérifier aisément par récurrence sur $|w|$, par exemple. En particulier, $\mathcal{B}_X$ est simple car l'automate littéral l'est.

Nous étudions ici quelques unes des propriétés de cet automate avant de démontrer le résultat principal de cette section, appelé Théorème de saturation.

Proposition 5.4.8. *Soient S un ensemble acyclique et $X \subseteq S$ un code bifixe.*

1. *L'automate quotient $\mathcal{B}_X$ est co-déterministe.*
2. *Si Z est le code bifixe engendrant $L(\mathcal{B}_X)$, alors $X \subseteq Z$.*
3. *Le sous-groupe décrit par $\mathcal{B}_X$ est $\langle X \rangle$.*

Démonstration. Notons $\mathcal{A} = (P, \varepsilon, \{\varepsilon\}, A, \delta)$ l'automate littéral de X^* et

$$\mathcal{B}_X = (Q, 1_X, \{1_X\}, A, \delta')$$

l'automate quotient de X.

1. Par la remarque précédente, $\mathcal{B}_X$ est simple. Il est alors immédiat que $\mathcal{B}_X$ est co-déterministe par la Proposition 5.4.5.
2. Étant donné que $\mathcal{B}_X$ est co-déterministe, un tel Z existe bien par la Proposition 5.3.3. Soit $x \in X$. Comme $x \in L(\mathcal{A})$, il correspond à un chemin $\mathcal{C}$ de ε vers ε dans $\mathcal{A}$. De plus, $\mathcal{C}$ ne passe pas par ε [(10)] car cela signifierait qu'un préfixe de x est dans X et contredirait le fait que X est un code bifixe. Le chemin correspondant à $\mathcal{C}$ dans $\mathcal{B}_X$ va de 1_X dans 1_X et est étiqueté par x donc $x \in L(\mathcal{B}_X) = Z^*$. Le mot ε n'étant équivalent à aucun autre pour la relation θ_X, ce chemin ne passe pas par 1_X [(11)] donc aucun préfixe de x n'est dans Z^*. On en conclut que $x \in Z$ car $x \neq \varepsilon$.
3. Notons H le sous-groupe décrit par $\mathcal{B}_X$. Par la Proposition 5.3.10, on a

$$H = \langle Z \rangle.$$

Comme $X \subseteq Z$, $\langle X \rangle \subseteq H$. Montrons l'autre inclusion. Pour cela, il suffit de montrer que, si $z \in Z$, $\langle X \rangle \cdot z = \langle X \rangle$. Plus généralement, montrons que si

$$\delta'([p]_{\theta_X}, w) = [q]_{\theta_X},$$

(10). On entend ici qu'il ne repasse pas par ε en dehors de son état de départ et de son état d'arrivée.
(11). Même remarque que dans la Note (10).

alors

$$\langle X\rangle \cdot pw = \langle X\rangle \cdot q,$$

ce qui permettra de conclure car $z \in L(\mathcal{B}_X)$ donc $\delta'([\varepsilon]_{\theta_X}, z) = [\varepsilon]_{\theta_X}$. Procédons par récurrence sur $|w|$. Si $w = \varepsilon$, alors cela signifie que $p \equiv q \mod \theta_X$ donc, par la Proposition 5.4.2, la conclusion est immédiate. Supposons à présent le résultat vrai pour les mots de longueur $|w| - 1$ et notons $w = w'a$. Notons également

$$r \in \delta'([p]_{\theta_X}, w')$$

tel que

$$\delta(r, a) \equiv q \mod \theta_X.$$

- Si $ra \in P$, alors $\delta(r, a) = ra$ donc par la Proposition 5.4.2,

$$\langle X\rangle \cdot q = \langle X\rangle \cdot ra.$$

- Si $ra \in X$, alors $\delta(r, a) = \varepsilon$ et, comme la classe d'équivalence de ε est réduite à lui-même, $q = \varepsilon$. On a alors

$$\langle X\rangle \cdot q = \langle X\rangle = \langle X\rangle \cdot ra.$$

Par hypothèse de récurrence,

$$\langle X\rangle \cdot pw' = \langle X\rangle \cdot r$$

donc on a

$$\langle X\rangle \cdot q = \langle X\rangle \cdot ra = \langle X\rangle \cdot pw'a = \langle X\rangle \cdot pw.$$

□

Théorème 5.4.9. *Soit S un ensemble acyclique. Pour tout code bifixe $X \subseteq S$,*

$$X^* \cap S = \langle X\rangle \cap S.$$

On dit alors que X^* est saturé *dans S.*

Démonstration. On a immédiatement $X^* \cap S \subseteq \langle X\rangle \cap S$ donc il suffit de montrer l'autre inclusion. Commençons par faire quelques observations. Notons Z le code bifixe engendrant $L(\mathcal{B}_X)$. L'automate $\mathcal{B}_X$ est co-déterministe donc minimal. Par la Proposition 5.3.8, on a alors

$$Z^* = \langle Z\rangle \cap A^*.$$

Or, par la Proposition 5.4.8, $X \subseteq Z$ donc

$$\langle X\rangle \cap S \subseteq \langle Z\rangle \cap S = \langle Z\rangle \cap A^* \cap S = Z^* \cap S.$$

Posons $Y = Z \cap S$. L'ensemble S étant factoriel,

$$Z^* \cap S = (Z \cap S)^* \cap S = Y^* \cap S.$$

Nous pouvons à présent montrer que $\langle X\rangle \cap S \subseteq X^*$. Soit $x \in \langle X\rangle \cap S$. Il existe donc $x_1, \ldots, x_n \in X \cup X^{-1}$ tels que

$$x = x_1 \cdot \ldots \cdot x_n.$$

Comme $\langle X\rangle \cap S \subseteq Y^* \cap S$, il existe également $y_1, \ldots, y_m \in Y$ tels que

$$x = y_1 \ldots y_m = y_1 \cdot \ldots \cdot y_m.$$

Or, $Y \subseteq S$ est un code bifixe donc, par le Théorème 5.2.6, Y est libre. De plus, X est inclus dans $Z \cap S = Y$. On a donc $n = m$ et $x_i = y_i$. En particulier, $x_i \in A^*$ donc $x_i \in X$. On obtient donc bien que $x \in X^*$. □

Chapitre 6

Propriété de base d'indice fini et stabilité des ensembles dendriques

Dans ce chapitre, nous établissons un lien entre le S-degré d'un code bifixe S-maximal et l'indice du sous-groupe qu'il engendre. Cela nous permettra d'obtenir une caractérisation des ensembles dendriques. Nous étudions ensuite la stabilité de la famille des ensembles dendriques pour une opération définie à partir des mots de retour. Grâce à la propriété d'indice fini, nous terminons aussi de prouver la stabilité par décodage bifixe maximal, commencée à la Section 4.2.

6.1 Indice d'un sous-groupe

Commençons par rappeler quelques définitions et résultats élémentaires d'algèbre concernant l'indice d'un sous-groupe d'un groupe G dont l'opération est notée $\cdot$.

Définition 6.1.1. Soit H un sous-groupe d'un groupe G. L'ensemble des *classes à gauche* selon H est l'ensemble

$$\{g \cdot H : g \in G\}.$$

Symétriquement, les *classes à droite* selon H sont les éléments de l'ensemble

$$\{H \cdot g : g \in G\}.$$

Proposition 6.1.2. *Soit H un sous-groupe d'un groupe G.*

1. *Les classes à gauche (resp. à droite) selon H forment une partition de G.*
2. *Il y a autant de classes à gauche selon H que de classes à droite selon H, i.e.*

$$|\{g \cdot H : g \in G\}| = |\{H \cdot g : g \in G\}|.$$

Définition 6.1.3. Soit H un sous-groupe d'un groupe G. *L'indice* de H est le cardinal de l'ensemble des classes à gauches selon H. On le note

$$[G : H] = |\{g \cdot H : g \in G\}|.$$

Proposition 6.1.4. *Soient H et H' deux sous-groupes d'un groupe G. Si $H \subseteq H'$ et si $[G : H] = [G : H']$, alors $H = H'$.*

Exemple 6.1.5. Remarquons que

$$\begin{aligned}\langle A^n\rangle &= \{\rho(w_1 \dots w_m) \mid w_i \in A^n \cup (A^{-1})^n, m \in \mathbb{N}\} \\ &= \{x \in F_A \mid |x|_A \equiv |x|_{A^{-1}} \mod n\}\end{aligned}$$

où $|x|_A$ représente le nombre de lettres de x dans l'alphabet A. De façon similaire, pour tout $y \in F_A$,

$$y \cdot \langle A^n\rangle = \{x \in F_A \mid |x|_A - |x|_{A^{-1}} \equiv |y|_A - |y|_{A^{-1}} \mod n\}.$$

On en déduit qu'il n'y a que n classes à gauches possibles et que

$$[F_A : \langle A^n\rangle] = n.$$

Mentionnons à présent deux résultats célèbres d'algèbre des groupes libres dûs à Jakob Nielsen et Otto Schreier. Rappelons que le rang d'un groupe libre est le cardinal de n'importe laquelle de ses bases.

Théorème 6.1.6 (Théorème de Nielsen-Schreier)**.** *Tout sous-groupe d'un groupe libre est un groupe libre.*

Théorème 6.1.7 (Formule de Schreier)**.** *Si G est un groupe libre de rang $n \in \mathbb{N}$ et si H est un sous-groupe de G d'indice $d \in \mathbb{N}$, alors H est un groupe libre de rang*

$$1 + d(n-1).$$

Nous utiliserons principalement cette formule dans le cas où $G = F_A$ et H admet X pour base. Elle devient alors

$$|X| = 1 + d(|A| - 1)$$

où d est l'indice de H.

Définition 6.1.8. Un ensemble récurrent S a la *propriété de base d'indice fini* si, pour tout code bifixe fini $X \subseteq S$, on a l'équivalence suivante : X est un code bifixe S-maximal de S-degré d si, et seulement si, X est la base d'un sous-groupe d'indice d du groupe libre F_A.

Remarque 6.1.9. En particulier, si S a la propriété de base d'indice fini, alors pour tout code bifixe X S-maximal fini, son S-degré est fini par la Proposition 2.5.8 et, par la formule de Schreier,

$$|X| = 1 + d_S(X)(|A| - 1).$$

On retrouve donc un résultat similaire au Théorème de cardinalité (Théorème 2.5.10).

6.2 Lien entre la propriété de base d'indice fini et les ensembles dendriques

Nous montrons ici qu'un ensemble récurrent est dendrique si, et seulement si, il a la propriété de base d'indice fini. Montrons tout d'abord que tout ensemble dendrique récurrent a la propriété d'indice fini. Pour cela, il faut montrer une équivalence. Le premier sens découle de la proposition suivante.

Proposition 6.2.1. *Soient S un ensemble dendrique récurrent et $X \subseteq S$ un code bifixe S-maximal fini. Si X est de S-degré d, alors $\langle X\rangle$ est d'indice d.*

Démonstration. L'ensemble X étant de S-degré d, il existe $w \in S$ tel que

$$\delta_X(w) = d.$$

Notons P l'ensemble des préfixes propres de X et Q l'ensemble des suffixes de w n'ayant pas de préfixe dans X ou, de façon équivalente car X est un code préfixe S-maximal,

$$Q = \mathrm{Suff}(w) \cap P.$$

Par la Proposition 2.5.3, $|Q| = d$. Montrons que

$$C = \{\langle X\rangle \cdot q \mid q \in Q\}$$

forme l'ensemble des classes à droite selon $\langle X\rangle$, ce qui permettra de conclure.

Montrons que les éléments de C sont distincts. Soient $p, q \in Q$ tels que $\langle X\rangle \cdot p = \langle X\rangle \cdot q$. Comme p et q sont tous deux des suffixes de w, supposons sans perte de généralité que $q \in \mathrm{Suff}(p)$ et notons $x \in S$ tel que $p = xq$. On a alors

$$\langle X\rangle \cdot xq = \langle X\rangle \cdot p = \langle X\rangle \cdot q$$

donc

$$\langle X\rangle \cdot x = \langle X\rangle,$$

ce qui signifie que $x \in \langle X\rangle$ et donc que $x \in \langle X\rangle \cap S$. Par le Théorème de saturation (Théorème 5.4.9), on a alors $x \in X^*$. Cependant, x est préfixe de $p \in Q \subseteq P$ donc x est préfixe propre d'un élément de X. L'ensemble X étant un code préfixe, la seule possibilité est d'avoir $x = \varepsilon$. Cela signifie donc que $p = q$ et on le résultat désiré.

Montrons à présent que les éléments de C recouvrent F_A tout entier. Pour cela, définissons

$$V = \{v \in F_A \mid Q \cdot v \subseteq \langle X\rangle \cdot Q\}$$

et montrons que $V = F_A$. Tout d'abord, montrons qu'il s'agit d'un sous-groupe de F_A.

- Comme $\varepsilon \in \langle X\rangle$, $Q \subseteq \langle X\rangle \cdot Q$ donc $\varepsilon \in V$,
- Soit $v \in V$. Montrons que $v^{-1} \in V$. Pour cela, montrons d'abord que l'application

$$\varphi : Q \to Q \qquad p \mapsto q \ \text{ tq. } \ p \cdot v \in \langle X\rangle \cdot q$$

est surjective. Elle est correctement définie car $v \in V$ donc un tel q existe pour tout $p \in Q$. De plus, les éléments de C sont disjoints donc ce q est unique. Remarquons que

$$\begin{aligned}
\varphi(p) = q &\Leftrightarrow p \cdot v \in \langle X\rangle \cdot q \\
&\Leftrightarrow \exists\, x \in \langle X\rangle \text{ tq. } p \cdot v = x \cdot q \\
&\Leftrightarrow \exists\, x \in \langle X\rangle \text{ tq. } x^{-1} \cdot p = q \cdot v^{-1} \\
&\Leftrightarrow q \cdot v^{-1} \in \langle X\rangle \cdot p
\end{aligned}$$

donc, pour tous $p, p' \in Q$, si $\varphi(p) = q = \varphi(p')$, alors

$$q \cdot v^{-1} \in \langle X\rangle \cdot p \cap \langle X\rangle \cdot p'$$

ce qui implique que $\langle X\rangle \cdot p = \langle X\rangle \cdot p'$ car les classes à droite selon $\langle X\rangle$ sont disjointes. Les éléments de C étant distincts, on a donc que $p = p'$. En conséquence, la fonction φ est injective. Elle va d'un ensemble fini dans lui-même donc elle est également surjective.

Il en découle que, pour tout $q \in Q$, il existe $p \in Q$ tel que $\varphi(p) = q$ donc tel que $q \cdot v^{-1} \in \langle X \rangle \cdot p$. On en conclut alors que

$$Q \cdot v^{-1} \subseteq \langle X \rangle \cdot Q$$

et que $v^{-1} \in V$.

- Soient $v, v' \in V$. On a

$$Q \cdot (v \cdot v') = (Q \cdot v) \cdot v' \subseteq \langle X \rangle \cdot Q \cdot v' \subseteq \langle X \rangle \cdot Q$$

donc $v \cdot v' \in V$.

Montrons maintenant que $\mathcal{R}_S(w) \subseteq V$ car alors

$$F_A = \langle \mathcal{R}_S(w) \rangle \subseteq \langle V \rangle = V \subseteq F_A$$

par le Théorème 3.5.1. Soient $u \in \mathcal{R}_S(w)$ et $q \in Q$. Comme $u, q \in A^*$, on souhaite montrer que $qu = q \cdot u \in \langle X \rangle \cdot Q$. Par définition de $\mathcal{R}_S(w)$, $wu \in S$ et il existe u' tel que $wu = u'w$. De plus, q est un suffixe de w par hypothèse donc

$$qu \in \mathrm{Suff}(wu) \subseteq S$$

car S est factoriel. Or, X est un code préfixe S-maximal donc qu est préfixe d'un mot de X^*. Notons

$$qu = x_1 \dots x_n p, \quad x_i \in X, p \in P.$$

Il nous suffit de montrer que $p \in \mathrm{Suff}(w)$ car, dans ce cas, $p \in Q$ et on pourra conclure étant donné que

$$qu \in X^* p \subseteq \langle X \rangle \cdot p.$$

Par construction,

$$p \in \mathrm{Suff}(qu) \subseteq \mathrm{Suff}(wu) = \mathrm{Suff}(u'w)$$

donc montrons que $|p| \leq |w|$. Procédons par l'absurde. Si $|p| > |w|$, alors w est un suffixe propre de $p \in P$ donc $w \in \mathrm{IFac}(X)$. Or, par définition de w, $\delta_X(w) = d = d_S(X)$, ce qui est absurde par la Proposition 2.5.8.

Nous avons donc montré que $F_A = V$. Autrement dit, pour tout $v \in F_A$,

$$Q \cdot v \subseteq \langle X \rangle \cdot Q.$$

Étant donné que $\varepsilon \in Q$, cela signifie que $F_A \subseteq \langle X \rangle \cdot Q$ et que C est bien l'ensemble des classes à droite selon $\langle X \rangle$. Vu le cardinal de C, on en conclut que l'indice de $\langle X \rangle$ est bien d. □

Avant de montrer la réciproque de ce résultat, revenons brièvement aux fonctions comptant le nombre de découpages par rapport à un ensemble. En effet, nous aurons besoin de savoir que tout code bifixe fini est inclus dans un code bifixe S-maximal S-fin et pour cela, nous avons besoin du lemme suivant.

Lemme 6.2.2. *Si $\delta : A^* \to \mathbb{N}$ est tel que*

1. $\delta(\varepsilon) = 1$,

2. *pour tous* $w \in A^*$ *et* $a \in A$,

$$\delta(aw) - \delta(w) \in \{0, 1\}$$

et

$$\delta(wa) - \delta(w) \in \{0, 1\},$$

3. *pour tous* $w \in A^*$ *et* $a, b \in A$,

$$\delta(w) + \delta(awb) \leq \delta(aw) + \delta(wb),$$

alors il existe un code bifixe X *tel que* $\delta = \delta_X$.

Réciproquement, si X *est un code bifixe,* δ_X *vérifie les conditions ci-dessus.*

Démonstration. Pour tout $w \in A^*$, définissons

$$\pi(w) = \begin{cases} \delta(\varepsilon) & \text{si } w = \varepsilon, \\ \delta(a) - 2\delta(\varepsilon) & \text{si } w = a \in A, \\ \delta(aw'b) - \delta(w'b) - \delta(aw') + \delta(w') & \text{si } w = aw'b,\ a, b \in A. \end{cases}$$

Remarquons que, si $w \neq \varepsilon$, $\pi(w) \in \{-1, 0\}$ car

- pour tout $a \in A$, par l'hypothèse 2,

$$\delta(a) - \delta(\varepsilon) \in \{0, 1\}$$

donc $\delta(a) - 2\delta(\varepsilon) \in \{-1, 0\}$,

- pour tous $a, b \in A$, $x \in A^*$, vu la deuxième hypothèse, $\pi(aw'b) \in \{-1, 0, 1\}$. De plus, par l'hypothèse 3, $\pi(aw'b) \leq 0$.

Posons

$$X = \{w \in A^* \mid \pi(w) = -1\}$$

et montrons qu'il s'agit d'un code bifixe.

Pour ce faire, notons

$$U = \{\varepsilon\} \cup \{aw \in A^+ \mid \delta(aw) - \delta(w) = 1\}.$$

On a alors

$$\begin{aligned} UA \setminus U &= \{wb \in A^+ \setminus U \mid w \in U\} \\ &= \{b \in A \mid \delta(b) - \delta(\varepsilon) \neq 1\} \\ &\quad \cup \{aw'b \in A^+ \mid \delta(aw'b) - \delta(w'b) \neq 1 \text{ et } \delta(aw') - \delta(w') = 1\} \\ &= \{a \in A \mid \delta(a) - \delta(\varepsilon) = 0\} \\ &\quad \cup \{aw'b \in A^+ \mid \delta(aw'b) - \delta(w'b) = 0 \text{ et } \delta(aw') - \delta(w') = 1\} \\ &= X. \end{aligned}$$

Montrons que $UA \setminus U$ est un code préfixe. Pour cela, commençons par montrer par récurrence sur la longueur de $w \in U$ que $\mathrm{Pref}(w) \subseteq U$. Si $|w| \leq 1$, la conclusion est immédiate. Si $w = aw'b$, alors

$$1 = \delta(aw'b) - \delta(w'b) \leq \delta(aw') - \delta(w') \leq 1$$

par hypothèse. Donc $aw' \in U$ et on peut conclure par hypothèse de récurrence sur aw' car

$$\mathrm{Pref}(w) = \{w\} \cup \mathrm{Pref}(aw').$$

À présent, soient $x, y \in UA \setminus U$ tels que $x \in \mathrm{Pref}(y)$. Notons $y = y'b$, $y' \in U$, $b \in A$. Si $x \neq y$, alors $x \in \mathrm{Pref}(y') \subseteq U$, ce qui est absurde. On en conclut que X est un code préfixe.

De façon symétrique, si

$$V = \{\varepsilon\} \cup \{wb \in A^* \mid \delta(wb) - \delta(w) = 1\},$$

alors

$$X = AV \setminus V$$

qui est un code suffixe. L'ensemble X est donc bien un code bifixe.

Montrons maintenant que $\delta(w) = \delta_X(w)$ pour tout $w \in A^*$. Si $w = \varepsilon$, le résultat est immédiat. Si $w = a \in A$, alors, par la Proposition 2.5.3,

$$\begin{aligned}\delta_X(a) &= \begin{cases} 2 & \text{si } a \notin X, \\ 1 & \text{sinon} \end{cases} \\ &= \begin{cases} 2 & \text{si } \pi(a) = 0, \\ 1 & \text{si } \pi(a) = -1 \end{cases} \\ &= \delta(a)\end{aligned}$$

car $\pi(a) = \delta(a) - 2$.

Supposons avoir l'égalité pour les mots de longueur au plus $|w|-1$, $|w| \geq 2$, et montrons-la pour w. Notons $w = aw'b$, $a, b \in A$. On a alors

$$\begin{aligned}\delta(w) &= \delta(aw') + \delta(w'b) - \delta(w') + \pi(w) \\ &= \delta_X(aw') + \delta_X(w'b) - \delta_X(w') - \mathbb{1}_X(w).\end{aligned}$$

Or, par la Proposition 2.5.3,

$$\delta_X(w'b) - \delta_X(w') = 1 - \mathbb{1}_{A^*X}(w'b)$$

et

$$\delta_X(aw') = \delta_X(w) - 1 + \mathbb{1}_{A^*X}(w)$$

donc

$$\begin{aligned}\delta(w) &= \delta_X(w) - 1 + \mathbb{1}_{A^*X}(w) + 1 - \mathbb{1}_{A^*X}(w'b) - \mathbb{1}_X(w) \\ &= \delta_X(w) + \mathbb{1}_{A^*X}(w) - \mathbb{1}_{A^*X}(w'b) - \mathbb{1}_X(w)\end{aligned}$$

Il suffit donc de montrer que

$$\mathbb{1}_{A^*X}(w) - \mathbb{1}_{A^*X}(w'b) - \mathbb{1}_X(w) = 0.$$

Si $w'b \in A^*X$, alors $w = aw'b \in A^+X$ et, comme X est un code bifixe, $w \notin X$. On a donc bien l'égalité dans ce cas. Si, par contre, $w'b \notin A^*X$, alors

$$w \in A^*X \;\Leftrightarrow\; w \in X$$

ce qui permet également d'obtenir l'égalité. En conclusion, X est un code bifixe tel que $\delta = \delta_X$.

Réciproquement, si X est un code bifixe, les conditions 1 et 2 sont trivialement vérifiées par δ_X par la Proposition 2.5.3. De plus,

$$\delta_X(aw) - \delta_X(w) = 1 - \mathbb{1}_{XA^*}(aw) \geq 1 - \mathbb{1}_{XA^*}(awb) = \delta_X(awb) - \delta_X(wb).$$

□

Proposition 6.2.3. *Soit S un ensemble récurrent. Tout code bifixe $X \subseteq S$ fini est inclus dans un code bifixe S-maximal S-fin.*

Démonstration. Tout d'abord, si X est S-maximal, la conclusion est immédiate car S est infini donc tout sous-ensemble fini est S-fin. Supposons donc que X n'est pas S-maximal. Posons

$$d = \max\{\delta_X(x) \mid x \in X\} + 1.$$

L'ensemble X étant fini, d l'est également. Définissons alors

$$\delta : A^* \to \mathbb{N} \quad w \mapsto \min\{d, \delta_X(w)\}$$

et montrons que δ vérifie les conditions du Lemme 6.2.2.

1. On a $\delta(\varepsilon) = \min\{d, 1\} = 1$.
2. Soient $w \in A^*$ et $a \in A$. Par construction,

   $$\delta(aw) - \delta(w) \in \{0, 1\}.$$

   En effet, si $\delta_X(w) \geq d$, alors $\delta_X(aw) \geq d$ donc $\delta(aw) - \delta(w) = 0$. Si, par contre, $\delta_X(w) < d$, alors $\delta_X(aw) \leq d$ par la seconde partie du Lemme 6.2.2 et

   $$\delta(aw) - \delta(w) = \delta_X(aw) - \delta_X(w) \in \{0, 1\}.$$

   Remarquons que, dans les deux cas,

   $$\delta(aw) - \delta(w) \leq \delta_X(aw) - \delta_X(w).$$

   On montre symétriquement que

   $$\delta(wa) - \delta(w) \in \{0, 1\}.$$

3. Soient $w \in A^*$ et $a, b \in A$. Si $\delta_X(w) \geq d$, alors

   $$\delta(w) + \delta(awb) = 2d = \delta(aw) + \delta(wb).$$

   Si $\delta_X(w) < d$, alors $\delta_X(aw) \leq d$ et on a

   $$\begin{aligned} \delta(awb) - \delta(wb) &\leq \delta_X(awb) - \delta_X(wb) \\ &\leq \delta_X(aw) - \delta_X(w) \\ &= \delta(aw) - \delta(w) \end{aligned}$$

   par la seconde partie du Lemme 6.2.2 donc

   $$\delta(w) + \delta(awb) \leq \delta(aw) + \delta(wb).$$

Par le Lemme 6.2.2, il existe un code bifixe Y tel que $\delta = \delta_Y$. Posons

$$Z = Y \cap S.$$

Montrons que Z est le code bifixe recherché, i.e. qu'il est S-maximal, S-fin et qu'il contient X. Comme Y est un code bifixe, Z également. De plus, S est factoriel donc, pour tout $w \in S$,

$$\delta_Z(w) = \delta_Y(w) = \delta(w).$$

On a alors

$$d_S(Z) = \max_{w \in S} \delta_Z(w) = \max_{w \in S} \delta(w) \leq d.$$

Par la Proposition 2.5.8, Z est S-maximal et S-fin car son S-degré est fini.

Par contraposition de ce même résultat, X n'est pas S-maximal donc son S-degré n'est pas fini. En particulier, il existe $w \in S$ tel que $\delta_X(w) \geq d$ donc tel que $\delta(w) = d$ et on a $d_S(Z) = d$. Pour tout $w \in X$, par définition de d,

$$\delta_X(w) = \delta(w) = \delta_Z(w)$$

donc

$$\delta_Z(w) < d = d_S(Z).$$

Par la Proposition 2.5.8, on en déduit que

$$X \subseteq \mathrm{IFac}(Z).$$

Pour montrer que $X \subseteq Z$ et pouvoir conclure, il suffit donc de montrer que, pour tout $w \in \mathrm{IFac}(Z)$, [(1)]

$$w \in X \;\Leftrightarrow\; w \in Z.$$

Procédons par récurrence sur $|w|$. Si $w = \varepsilon$, alors $w \notin X$ et $w \notin Z$ car ce sont des codes bifixes. Supposons la bi-implication vraie pour les mots de longueur strictement inférieure à $|w|$. Par la Proposition 2.5.8, on a

$$\delta_Z(w) < d$$

donc

$$\delta_Z(w) = \delta(w) = \delta_X(w).$$

On a alors, par la Proposition 2.5.3,

$$\begin{aligned} |w| + 1 - |\{(x,y,z) \mid y \in Z, xyz = w\}| &= \delta_Z(w) \\ &= \delta_X(w) \\ &= |w| + 1 - |\{(x,y,z) \mid y \in X, xyz = w\}|. \end{aligned}$$

Or, pour tout triplet (x,y,z) tel que $w = xyz$ et $xz \neq \varepsilon$, par hypothèse de récurrence,

$$y \in Z \;\Leftrightarrow\; y \in X$$

donc

$$\{(x,y,z) \mid y \in Z, xz \neq \varepsilon, xyz = w\} = \{(x,y,z) \mid y \in X, xz \neq \varepsilon, xyz = w\}.$$

On peut en conclure que $w \in Z$ si et seulement si $w \in X$. □

(1). En réalité, on montre même que $X = \mathrm{IFac}(Z) \cap Z$. On dit alors que X est le *noyau* de Z.

Nous pouvons à présent finir de montrer qu'un ensemble dendrique récurrent a la propriété de base d'indice fini.

Proposition 6.2.4. *Soient S un ensemble dendrique récurrent et $X \subseteq S$ un code bifixe fini. Si X est une base d'un sous-groupe de F_A d'indice fini d, alors X est S-maximal parmi les codes bifixes. De plus, X est de S-degré d.*

Démonstration. Par la proposition précédente, X est inclus dans un code bifixe Y S-maximal et S-fin. Par le Théorème de cardinalité (Théorème 2.5.10), on a alors

$$|Y| = 1 + d_S(Y)\,(|A| - 1)\,.$$

Or, par la formule de Schreier, on a également

$$|X| = 1 + d\,(|A| - 1)\,.$$

Montrons que $|X| = |Y|$. Si $|A| = 1$, la conclusion est évidente. Sinon, comme $X \subseteq Y$, on a

$$d \leq d_S(Y).$$

D'autre part, $\langle X\rangle \subseteq \langle Y\rangle$ donc l'indice de $\langle Y\rangle$ divise celui de $\langle X\rangle$. Or, par la Proposition 6.2.1, l'indice de $\langle Y\rangle$ est $d_S(Y)$ donc $d_S(Y)$ divise d et on a l'égalité. On a alors $|X| = |Y|$ donc $X = Y$ puisqu'on a une inclusion. On en déduit que X est S-maximal parmi les codes bifixes et que

$$d_S(X) = d_S(Y) = d.$$

□

Théorème 6.2.5. *Si S est dendrique et récurrent, alors il a la propriété de base d'indice fini.*

Démonstration. Soit $X \subseteq S$ un code bifixe fini. Montrons qu'il est S-maximal et de S-degré d si, et seulement si, c'est la base de $\langle X\rangle$ et $\langle X\rangle$ est d'indice d. Par le Théorème 5.2.6, X est une base de $\langle X\rangle$. De plus, s'il est S-maximal et de S-degré d, alors $\langle X\rangle$ est d'indice d par la Proposition 6.2.1. La réciproque n'est autre que la Proposition 6.2.4. □

Le résultat précédent admet une réciproque, que nous prouvons ici.

Théorème 6.2.6. *Si S est récurrent et a la propriété d'indice fini, alors il est dendrique.*

Démonstration. Commençons par rappeler que, par l'Exemple 2.5.6 pour tout $n \in \mathbb{N}_0$, $S \cap A^n$ est un code bifixe S-maximal de S-degré n. Comme l'ensemble S a la propriété de base d'indice fini, $S \cap A^n$ est la base d'un sous-groupe de F_A d'indice n. Par la formule de Schreier, on a alors

$$|S \cap A^n| = 1 + n(|A| - 1), \quad \forall n \in \mathbb{N}_0.$$

Soit $w \in S$. Montrons que le graphe $\mathcal{E}(w)$ est un arbre. Procédons par l'absurde et supposons avoir un cycle

$$(a_1, b_1, a_2, \ldots, b_k, a_1), \qquad a_i \in L(w), b_i \in R(w)\ \forall i \leq k$$

non dégénéré, i.e. où $a_i \neq a_{i+1}$, $a_k \neq a_1$ et $b_i \neq b_{i+1}$. Posons $X = AwA \cap S$. On a alors

$$a_1wb_1, a_2wb_1, a_2wb_2, \ldots, a_kwb_k, a_1wb_k \in X.$$

De plus,

$$a_1wb_1 \cdot (a_2wb_1)^{-1} \cdot a_2wb_2 \cdot \ldots \cdot a_kwb_k \cdot (a_1wb_k)^{-1} = \varepsilon.$$

Autrement dit, X n'est pas libre. Or, $X \subset A^{|w|+2} \cap S$ qui est libre car c'est la base d'un sous-groupe d'indice $|w|+2$. C'est donc absurde et $\mathcal{E}(w)$ est acyclique.

Montrons à présent que S est neutre, ce qui permettra de conclure par la Proposition 1.2.7. Soit $k \in \mathbb{N}$. On a

$$\begin{aligned}\sum_{w \in A^k \cap S} e(w) &= p_S(k+2) \\ &= |S \cap A^{k+2}| \\ &= 1 + (k+2)(|A|-1) \\ &= 2(1 + (k+1)(|A|-1)) - (1 + k(|A|-1)) \\ &= 2|S \cap A^{k+1}| - |S \cap A^k| \\ &= \sum_{w \in A^k \cap S} (l(w) + r(w) - 1).\end{aligned}$$

Or, pour tout $w \in A^k \cap S$, $\mathcal{E}(w)$ est acyclique donc on a

$$e(w) \leq l(w) + r(w) - 1.$$

Au vu de l'égalité ci-dessus, on doit alors avoir

$$e(w) = l(w) + r(w) - 1$$

pour tout $w \in A^k \cap S$ et pour tout $k \in \mathbb{N}$. En conclusion, l'ensemble S est neutre donc dendrique par la Proposition 1.2.7. □

6.3 Ensembles dérivés

Dans cette section, nous introduisons la notion d'ensemble dérivé par rapport à un morphisme codant pour les mots de premier retour d'un mot fixé. Comme nous le verrons dans le chapitre suivant, cette notion est utile pour trouver une représentation S-adique d'un mot dendrique.

Nous nous intéressons ici aux propriétés stables pour le passage à l'ensemble dérivé qui seront utiles pour la section suivante.

Définition 6.3.1. Soient S un ensemble récurrent, $w \in S \setminus \{\varepsilon\}$ et f un morphisme codant pour $\mathcal{R}_S(w)$. *L'ensemble dérivé* de S par rapport à f est

$$D_f(S) = f^{-1}(w^{-1}S)$$

Tout comme dans la Section 4.2, si $\mathcal{R}_S(w)$ n'est pas fini, le morphisme f est défini sur un alphabet infini et l'ensemble $D_f(S)$ contient alors des mots construits sur ce même alphabet. Cependant, dans ce travail nous considérons les ensembles dérivés dans le cas d'un ensemble uniformément récurrent S, ce qui nous permet de rester dans le cadre des alphabets finis. Seuls les deux premiers points de la Proposition 6.3.2 ne suivent pas cette règle.

Rappelons que $\Gamma_S(w)$ est l'ensemble des mots de retour de w dans S.

Proposition 6.3.2. *Soient S est un ensemble récurrent et $w \in S \setminus \{\varepsilon\}$. Si $f : B^* \to A^*$ est un morphisme codant pour $\mathcal{R}_S(w)$, alors*

1. *$D_f(S) = \{\varepsilon\} \cup f^{-1}(\Gamma_S(w))$,*
2. *$D_f(S)$ est récurrent,*
3. *si S est dendrique, $D_f(S)$ est dendrique.*

Démonstration.

1. Montrons que
$$\{\varepsilon\} \cup f^{-1}(\Gamma_S(w)) \subseteq D_f(S).$$
On a immédiatement $\varepsilon \in D_f(S)$ car
$$f(\varepsilon) = \varepsilon \in w^{-1}S$$
étant donné que $w \in S$. À présent, si $x \in f^{-1}(\Gamma_S(w))$, alors $wf(x) \in S$ donc $x \in f^{-1}(w^{-1}S)$.
Pour l'autre inclusion, si $f(x) \in w^{-1}S$, alors $wf(x) \in S$. Soit $x = \varepsilon$, auquel cas la conclusion est immédiate, soit $f(x) \in (\mathcal{R}_S(w))^+$ et dans ce cas $w \in \mathrm{Suff}(wf(x))$ donc $f(x) \in \Gamma_S(w)$ car $wf(x) \in S$.
2. Soient $x, y \in D_f(S)$. Par définition, $wf(x), wf(y) \in S$. Or S est récurrent donc il existe $u \in S$ tel que $wf(x)uwf(y) \in S$. Par le point 1., $w \in \mathrm{Suff}(wf(x))$ donc uw est un mot de retour pour w. En particulier, $uw \in (\mathcal{R}_S(w))^+$ donc il existe z tel que $f(z) = uw$ par la Proposition 4.2.2. On a alors $xzy \in D_f(S)$ car
$$wf(xzy) = wf(x)uwf(y) \in S.$$
L'ensemble $D_f(S)$ est donc récurrent.
3. Supposons S dendrique et considérons $\mathcal{E}_{D_f(S)}(x)$ pour $x \in D_f(S)$. Rappelons que
$$\mathcal{R}'_S(w) = \{u \in A^+ \mid uw \in w\mathcal{R}_S(w)\} = \{u \in A^+ \mid uw \in S, w \in \mathrm{Pref}(uw) \setminus \mathrm{IFac}(uw)\}$$
et que, similairement à ce qui a été fait pour $\mathcal{R}_S(w)$ dans la Remarque 2.3.3, on peut montrer que $\mathcal{R}'_S(w)$ est un code suffixe Sw^{-1}-maximal.
Si $f' : B^* \to A^*$ est le morphisme tel que, pour tout $a \in B$,
$$f'(a)w = wf(a).$$
alors, pour tous $a, b \in B$,
$$\begin{aligned}(a,b) \in \mathcal{E}_{D_f(S)}(x) &\Leftrightarrow axb \in D_f(S)\\ &\Leftrightarrow f(axb) \in \Gamma_S(w)\\ &\Leftrightarrow wf(a)f(x)f(b) \in S\\ &\Leftrightarrow f'(a)wf(x)f(b) \in S\\ &\Leftrightarrow (f'(a), f(b)) \in \mathcal{E}_S^{\mathcal{R}'_S(w),\mathcal{R}_S(w)}(wf(x)).\end{aligned}$$
Or, $wf(x) \in S$ donc, par le Théorème 4.1.7, $\mathcal{E}_S^{\mathcal{R}'_S(w),\mathcal{R}_S(w)}(wf(x))$ est un arbre et $\mathcal{E}_{D_f(S)}(x)$ aussi.

□

6.4 Fin de la stabilité par décodage bifixe

Nous pouvons à présent finir de prouver la stabilité de la famille des ensembles dendriques récurrents pour l'opération de décodage bifixe commencée à la Section 4.2. Pour cela, nous prouvons d'abord quelques propositions concernant un morphisme surjectif φ de F_A dans un groupe fini G. Plus précisément, nous prouvons trois résultats qui montrent que la surjectivité est conservée sur des ensembles de plus en plus petits.

Proposition 6.4.1. *Soit G un groupe fini. Si $\varphi : F_A \to G$ est un morphisme surjectif, alors*

$$\varphi(A^*) = G.$$

Démonstration. Il suffit de montrer que

$$\varphi(A^{-1}) \subseteq \varphi(A^*)$$

car dans ce cas, pour tout $x \in F_A$, on a $\varphi(x) \in \varphi(A^*)$.

Soit $a \in A$. Comme G est fini et que $\{\varphi(a^k) \mid k \in \mathbb{N}\} \subseteq G$, il existe $m, n \in \mathbb{N}$, $m < n$ tels que

$$\varphi(a^m) = \varphi(a^n).$$

Dans ce cas,

$$\begin{aligned}\varphi(a^{-1}) &= (\varphi(a^m))^{-1} \cdot \varphi(a^n) \cdot \varphi(a^{-1}) \\ &= \varphi(a^{-m+n-1}) \\ &\in \varphi(A^*)\end{aligned}$$

car $n - m - 1 \geq 0$. □

Le second résultat nécessite les deux lemmes suivants. Il utilise également la propriété de base d'indice fini.

Lemme 6.4.2. *Soient X un sous-monoïde de A^* et $X' = X \setminus \{\varepsilon\}$. Si X est unitaire à droite, i.e si*

$$x, xy \in X \Rightarrow y \in X,$$

alors $Y = X' \setminus X'X'$ est un code préfixe tel que $X = Y^*$.

Démonstration. On a

$$Y^* = (X')^* = X^* = X$$

car X est un sous-monoïde. Montrons que Y est un code préfixe. Soient $x, z \in Y$ tels que $x \in \mathrm{Pref}(z)$. Notons $z = xy$. On a alors $x, xy \in X' \subseteq X$ donc $y \in X$. Si $y \neq \varepsilon$, alors $y \in X'$ donc $z = xy \in X'X'$, ce qui est absurde. On a donc $y = \varepsilon$ et $x = z$. □

Par symétrie, si X est unitaire à gauche, i.e. si

$$x, yx \in X \Rightarrow y \in X,$$

alors Y est un code suffixe.

Lemme 6.4.3. *Soient G un groupe fini et $\varphi : F_A \to G$ un morphisme surjectif. Le sous-monoïde $\varphi^{-1}(1_G) \cap A^*$ est engendré par un code bifixe de A^*-degré $|G|$.*

Démonstration. Notons $Z = \varphi^{-1}(1_G) \cap A^*$. Il s'agit d'un sous-monoïde car $\varphi(\varepsilon) = 1_G$ et si $x, y \in Z$, alors

$$\varphi(xy) = \varphi(x)\varphi(y) = 1_G.$$

De plus, Z est unitaire à droite car si $x, xy \in Z$, alors

$$\varphi(y) = \varphi(x^{-1} \cdot xy) = \varphi(x)^{-1}\varphi(xy) = 1_G$$

donc $y \in Z$. On montre symétriquement que Z est unitaire à gauche. Dans ce cas, si $Z' = Z \backslash \{\varepsilon\}$, alors $Y := Z' \setminus Z'Z'$ est un code bifixe tel que $Z = Y^*$. Pour conclure, on doit donc montrer que $d_{A^*}(Y) = |G|$.

Soit $w \in A^*$. Montrons que

$$\delta_Y(w) \leq |G|.$$

Par la Proposition 2.5.3, il suffit de montrer que tous les suffixes de w qui n'ont pas de préfixe dans Y ont des images différentes par φ. Soit x, y de tels suffixes de w. Sans perte de généralité, supposons que $x \in \mathrm{Suff}(y)$ donc qu'il existe z tel que $y = zx$. On a alors

$$\begin{aligned} \varphi(x) = \varphi(y) &\Leftrightarrow \varphi(z) = 1_G \\ &\Leftrightarrow z \in Z \\ &\Leftrightarrow z = \varepsilon \end{aligned}$$

car $\mathrm{Pref}(z) \subseteq \mathrm{Pref}(y)$ et que y ne peut pas avoir de préfixe dans Y. On a donc bien $\delta_Y(w) \leq |G|$ pour tout $w \in A^*$.

Supposons à présent que w ne soit pas un facteur interne de Y [(2)] et montrons que $\delta_Y(w) \geq |G|$, ce qui suffira pour conclure. Soit $g \in G$. Notons $g' = g\varphi(w)^{-1}$. On a alors

$$g'\varphi(w)g^{-1} = 1_G.$$

Par la Proposition 6.4.1, φ est surjectif sur A^* donc il existe $y, z \in A^*$ de longueurs minimales tels que $\varphi(y) = g'$ et $\varphi(z) = g^{-1}$. On a alors

$$ywz \in Z = Y^*.$$

Montrons qu'on peut trouver un découpage de w qui sera différent pour chaque $g \in G$.

- Si $g = \varphi(w)$, l'ensemble Y étant un code bifixe, il existe d'uniques x, p tels que p n'ait pas de préfixe dans Y, $x \in Y^* = Z$ et $w = xp$. Un découpage de w est alors donné par (ε, x, p). De plus, par définition, on a $\varphi(x) = 1_G$ donc $\varphi(p) = \varphi(w) = g$.
- Si $g = 1_G$, on procède de même pour obtenir le découpage (s, x, ε) où $\varphi(s) = \varphi(w)$ et $\varphi(\varepsilon) = g$.
- Sinon, comme Y est un code bifixe, il existe une unique factorisation de ywz en mots de Y. Posons s le plus court préfixe de w tel que $ys \in Y^*$ et p le plus court suffixe de w tel que $pz \in Y^*$. De tels préfixes et suffixes existent car w n'est facteur interne d'aucun mot de Y. De plus, s et p sont non vides car $y, z \notin Y^*$ (sinon, nous serions dans un des deux cas ci-dessus). Si x est tel que $w = sxp$, on a alors que (s, x, p) est un découpage de w. En effet, vu le choix de s, il n'a aucun suffixe dans Y et de même pour p. On a de plus,

$$\begin{aligned} \varphi(p) &= \varphi(pz)(\varphi(z))^{-1} \\ &= g. \end{aligned}$$

(2). Un tel w existe toujours. En effet, par la Proposition 2.5.8, Y est un code bifixe A^*-fin.

On a donc bien un découpage de w pour chaque $g \in G$ et tous les découpages seront différents, autrement dit,

$$d_{A^*}(Y) = \delta_Y(w) = |G|.$$

□

Proposition 6.4.4. *Soient S un ensemble dendrique récurrent et G un groupe fini. Si le morphisme $\varphi : F_A \to G$ est surjectif, alors*

$$\varphi(S) = G.$$

Démonstration. Reprenons les notations du lemme précédent et posons

$$X = Y \cap S.$$

Remarquons que

$$\begin{aligned} d_S(X) &= d_S(Y \cap S) \\ &= \max_{w \in S} |\{u \in \mathrm{Pref}(w) \mid \mathrm{Suff}(u) \cap Y \cap S = \emptyset\}| \\ &= \max_{w \in S} |\{u \in \mathrm{Pref}(w) \mid \mathrm{Suff}(u) \cap Y = \emptyset\}| \\ &= \max_{w \in S} \delta_Y(w) \\ &\leq \max_{w \in A^*} \delta_Y(w) \\ &= |G|. \end{aligned}$$

Le S-degré de X est donc fini et, par la Proposition 2.5.8, X est un code bifixe S-maximal fini. Etant donné que S a la propriété de base d'indice fini, X est la base d'un sous-groupe d'indice $d_S(X)$. Or,

$$\langle X \rangle \subseteq \langle Y \rangle$$

donc l'indice de $\langle Y \rangle$ qui vaut $|G|$ divise celui de $\langle X \rangle$ qui vaut $d_S(X) \leq |G|$. On a donc

$$d_S(X) = |G|.$$

Soit $x \in X \setminus \mathrm{IFac}(X)$ [3]. Notons T l'ensemble des suffixes de x n'ayant pas de préfixe dans X. Par les Propositions 2.5.3 et 2.5.8, on a

$$|T| = \delta_X(x) = d_S(X) = |G|.$$

Montrons que $\varphi_{|T}$ est injectif pour en déduire que $\varphi(T) = G$. Soient $u, v \in T$ tels que $\varphi(u) = \varphi(v)$. On peut supposer sans perte de généralité que $u \in \mathrm{Suff}(v)$ donc qu'il existe w tel que $v = wu$. Dans ce cas,

$$\varphi(w) = \varphi(v)\varphi(u)^{-1} = 1_G$$

donc $w \in Z = Y^*$. De plus, $w \in \mathrm{Fac}(x) \subseteq S$ car S est factoriel et $x \in S$. On a donc $w \in X^*$. Par définition de T, la seule possibilité est d'avoir $w = \varepsilon$ donc $u = v$. En conclusion, on a

$$G = \varphi(T) \subseteq \varphi(S) \subseteq G$$

donc on a bien l'égalité. □

(3). On peut, par exemple, prendre $x \in X$ de longueur maximale.

Le troisième résultat, quant à lui, utilise la stabilité du caractère dendrique pour le passage à l'ensemble dérivé.

Proposition 6.4.5. *Soient S un ensemble dendrique récurrent, G un groupe fini et $w \in S\backslash\{\varepsilon\}$. Si $\varphi : F_A \to G$ est un morphisme surjectif, alors*

$$\varphi(\Gamma_S(w) \cup \{\varepsilon\}) = G.$$

Démonstration. Soient B un alphabet et $\alpha : B^* \to A^*$ un morphisme codant pour $\mathcal{R}_S(w)$. Étendons α pour en faire un morphisme de groupes de F_B dans F_A. Posons

$$\beta : F_B \to G \quad x \mapsto \varphi(\alpha(x)).$$

On a alors

$$\beta(F_B) = \varphi(\alpha(\langle B\rangle)) = \varphi(\langle\alpha(B)\rangle) = \varphi(\langle\mathcal{R}_S(w)\rangle) = \varphi(F_A) = G$$

donc β est un morphisme surjectif de F_B dans G. Par la Proposition 6.3.2, l'ensemble

$$D_\alpha(S) = \{\varepsilon\} \cup \alpha^{-1}(\Gamma_S(w))$$

est dendrique et récurrent. Par la proposition précédente, on en déduit que

$$G = \beta(\{\varepsilon\} \cup \alpha^{-1}(\Gamma_S(w))) = \varphi(\{\alpha(\varepsilon)\} \cup \Gamma_S(w)) = \varphi(\{\varepsilon\} \cup \Gamma_S(w)).$$

□

Nous passons maintenant au résultat principal de cette section.

Théorème 6.4.6. *Soient S un ensemble dendrique récurrent, X un code bifixe fini S-maximal et $f : B^* \to A^*$ un morphisme codant pour X. L'ensemble $f^{-1}(S)$ est récurrent.*

Démonstration. Nous avons déjà montré que l'ensemble $f^{-1}(S)$ était factoriel. Montrons à présent que, pour tous $u, v \in f^{-1}(S)$, il existe $w \in f^{-1}(S)$ tel que $f(uwv) \in S$.

Pour cela, commençons par introduire plusieurs notations. Posons

$$Q = \{\langle X\rangle \cdot x \mid x \in F_A\}$$

et, pour tout $u \in F_A$,

$$\varphi_u : Q \to Q \quad \langle X\rangle \cdot x \mapsto \langle X\rangle \cdot x \cdot u.$$

Comme X est un code bifixe S-maximal fini, son S-degré est fini par la Proposition 2.5.8. L'ensemble S a la propriété d'indice fini donc ça signifie que $\langle X\rangle$ est alors d'indice fini, autrement dit, que Q est fini. Or, la fonction φ_u est injective donc il s'agit d'une simple permutation de Q. Ces permutations sont en nombre fini donc l'ensemble

$$G = \{\varphi_u \mid u \in F_A\}$$

est fini. On peut le munir d'une structure de groupe via l'opération

$$\diamond : G \times G \to G \quad (\varphi_u, \varphi_v) \mapsto \varphi_u \diamond \varphi_v = \varphi_{u\cdot v}.$$

Dans ce cas, la fonction

$$\varphi : F_A \to G \quad u \mapsto \varphi_u$$

est un morphisme surjectif.

Maintenant que nous savons cela, passons à la preuve proprement dite. Soient $u, v \in f^{-1}(S)$. Comme S est récurrent, il existe $w \in S$ tel que

$$x := f(u)wf(v) \in S \setminus \{\varepsilon\}.$$

Par la Proposition 6.4.5,

$$\varphi(\Gamma_S(x) \cup \{\varepsilon\}) = G.$$

Il existe donc $y \in \Gamma_S(x) \cup \{\varepsilon\}$ tel que $\varphi(y) = (\varphi_x)^{-1}$. Dans ce cas, $\varphi(xy) = 1_G$ et, en particulier,

$$\langle X \rangle \cdot xy = \varphi(xy)(\langle X \rangle) = \langle X \rangle$$

donc $xy \in \langle X \rangle$. De plus, par définition de y, $xy \in S$.

Or, par le Théorème de saturation (Théorème 5.4.9), $\langle X \rangle \cap S = X^* \cap S$ donc $xy \in X^*$. On a également

$$f(u), f(v) \in f(B^*) = X^*.$$

Rappelons de plus que $y \in \Gamma_S(x)$ donc $f(v) \in \mathrm{Suff}(x) \subseteq \mathrm{Suff}(xy)$ et il existe $z \in S$ tel que

$$f(u)wf(v)y = xy = f(u)zf(v).$$

L'ensemble X étant un code bifixe, X^* est unitaire à droite et à gauche donc on a les implications suivantes :

$$f(u), f(u)zf(v) \in X^* \Rightarrow zf(v) \in X^*,$$

$$f(v), zf(v) \in X^* \Rightarrow z \in X^*.$$

Il existe alors $t \in B^*$ tel que $f(t) = z$. On en déduit que

$$f(utv) = xy \in S$$

donc que $utv \in f^{-1}(S)$, ce qui permet de conclure que $f^{-1}(S)$ est récurrent. □

En combinant ce résultat avec le Théorème 4.2.5 obtenu précédemment, on obtient le théorème suivant :

Théorème 6.4.7. *La famille des ensembles dendriques récurrents est stable par décodage bifixe maximal.*

Chapitre 7

Représentations $\mathcal{S}$-adiques de mots dendriques

Dans ce chapitre, nous abordons deux dernières propriétés des ensembles dendriques liées à leurs représentations $\mathcal{S}$-adiques et aux bases tame. Remarquons que, dans ce chapitre, nous n'utiliserons plus la notation S pour désigner l'ensemble (dendrique ou non) de mots que nous considérons pour éviter la confusion avec l'ensemble $\mathcal{S}$ de morphismes qui intervient dans la définition de $\mathcal{S}$-adique. Nous introduisons d'abord les représentations $\mathcal{S}$-adiques ainsi qu'une propriété d'existence d'une telle représentation pour un ensemble uniformément récurrent T. Nous nous intéressons ensuite aux morphismes et aux bases tame. Enfin, nous montrons que tout ensemble dendrique récurrent a une représentation $\mathcal{S}$-adique particulière.

7.1 Définition des représentations $\mathcal{S}$-adiques

Bien que déjà étudiées avant (dans [2] notamment), les représentations $\mathcal{S}$-adiques a été introduites avec leur terminologie actuelle en 1996 par Sébastien Ferenczi dans [12]. Elles généralisent les itérations infinies d'endomorphismes en autorisant d'utiliser des morphismes différents lors de chaque itération. Les morphismes sont cependant choisis dans un ensemble fixé.

Une représentation $\mathcal{S}$-adique d'un ensemble T est définie comme suit.

Définition 7.1.1. Soient $\mathcal{S}$ un ensemble de morphismes et $T \subseteq A^*$ un ensemble factoriel. Une *représentation $\mathcal{S}$-adique* de T est la donnée d'une suite $\boldsymbol{s} = (\sigma_n)_{n\in\mathbb{N}}$ de morphismes de $\mathcal{S}$ tels que

$$\sigma_n : A_{n+1}^* \to A_n^*$$

où $A_0 = A$ et pour lesquels

$$T = \bigcup_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \dots \sigma_n(A_{n+1})).$$

Si une telle suite existe, on dit que T est *$\mathcal{S}$-adique*.

Les représentations $\mathcal{S}$-adiques sont cependant généralement étudiées pour des mots infinis et avec la définition suivante.

Définition 7.1.2. Soient $\mathcal{S}$ un ensemble de morphismes et $x \in A^{\mathbb{N}}$ un mot infini. Une *représentation $\mathcal{S}$-adique* de x est la donnée d'une suite $\boldsymbol{s} = (\sigma_n)_{n\in\mathbb{N}}$ de morphismes de $\mathcal{S}$ tels que

$$\sigma_n : A_{n+1}^* \to A_n^*$$

où $A_0 = A$ et d'une suite $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$ de lettres telles que $a_n \in A_n$ pour lesquelles [(1)]

$$x = \lim_{n\to\infty} \sigma_0 \dots \sigma_n(a_{n+1}).$$

Nous allons montrer que ces deux définitions sont évidemment liées. Pour ça, nous commençons par définir deux propriétés des suites de morphismes.

Définition 7.1.3. Une suite de morphismes $(\sigma_n)_{n\in\mathbb{N}}$ telle que $\sigma_n : A_{n+1}^* \to A_n^*$ est *croissante partout* si

$$\lim_{n\to\infty} \min_{a\in A_{n+1}} |\sigma_0 \dots \sigma_n(a)| = +\infty.$$

Remarquons que, si la limite

$$\lim_{n\to\infty} \max_{a\in A_{n+1}} |\sigma_0 \dots \sigma_n(a)|$$

est finie, la suite $\boldsymbol{s} = (\sigma_n)_{n\in\mathbb{N}}$ ne peut pas être une représentation $\mathcal{S}$-adique d'un ensemble infini ni faire partie d'une représentation $\mathcal{S}$-adique d'un mot infini.

Définition 7.1.4. Une suite de morphisme $(\sigma_n)_{n\in\mathbb{N}}$ telle que $\sigma_n : A_{n+1}^* \to A_n^*$ est *primitive* si, pour tout $n \in \mathbb{N}$, il existe $m > n$ tel que, pour chaque $a \in A_{m+1}$, toutes les lettres de A_n soient présentes dans

$$\sigma_n \dots \sigma_m(a).$$

Une représentation est *primitive* si sa suite de morphismes est primitive.

Remarque 7.1.5. Toute suite de morphismes $(\sigma_n)_{n\in\mathbb{N}}$ primitive et pour laquelle il existe une infinité d'indices $i_1 < i_2 < \dots$ tels que $|A_{i_k}| > 1$ est croissante partout. En effet, comme la suite est primitive, les morphismes sont non-effaçants, i.e.

$$\forall\, n \in \mathbb{N}, \forall\, a \in A_{n+1} \quad \sigma_n(a) \neq \varepsilon.$$

En particulier, pour tout $n \in \mathbb{N}$ et tout $w \in A_{n+1}^*$,

$$|\sigma_n(w)| \geq |w|.$$

La suite étant primitive, il existe $j_1 > i_1$ tel que toutes les lettres de A_{i_1} soient dans $\sigma_{i_1} \dots \sigma_{j_1}(a)$ pour tout $a \in A_{j_1+1}$. En particulier,

$$\min_{a\in A_{j_1+1}} |\sigma_0 \dots \sigma_{j_1}(a)| \geq \min_{a\in A_{j_1+1}} |\sigma_{i_1} \dots \sigma_{j_1}(a)| \geq |A_{i_1}|.$$

Quitte à considérer une sous-suite, on peut supposer $i_2 > j_1$. Par le même raisonnement, il existe alors $j_2 > i_2$ tel que

$$\min_{a\in A_{j_2+1}} |\sigma_0 \dots \sigma_{j_2}(a)| \geq |A_i| \min_{a\in A_{j_2+1}} |\sigma_{j_1+1} \dots \sigma_{j_2}(a)| \geq |A_{i_1}| \cdot |A_{i_2}|.$$

(1). Pour rappel, une suite $(u_n)_{n\in\mathbb{N}}$ de mots finis converge vers un mot infini u si, pour tout $k \in \mathbb{N}$, il existe $N \in \mathbb{N}$ tel que, pour tout $n \geq N$, u_n ait le même préfixe de longueur k que u.

On montre ainsi qu'il existe une suite $(j_k)_{k\in\mathbb{N}}$ strictement croissante telle que

$$\min_{a\in A_{j_k+1}} |\sigma_0 \dots \sigma_{j_k}(a)|$$

converge vers $+\infty$, ce qui suffit pour conclure.

Nous prouvons maintenant que, dans le cas des représentations $\mathcal{S}$-adiques primitives, il est équivalent de considérer les mots ou les ensembles.

Proposition 7.1.6. *Si $(\boldsymbol{s}, \boldsymbol{a})$ est une représentation $\mathcal{S}$-adique primitive de $x \in A^{\mathbb{N}}$ alors $\boldsymbol{s}$ est une représentation $\mathcal{S}$-adique de* $\mathrm{Fac}(x)$.

Démonstration. Tout facteur de x est facteur de $\sigma_0 \dots \sigma_n(a_{n+1})$ pour n assez grand donc

$$\mathrm{Fac}(x) \subseteq \bigcup_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \dots \sigma_n(A_{n+1})).$$

Pour l'autre inclusion, prenons $u \in \bigcup_{n\in\mathbb{N}} \mathrm{Fac}(\sigma_0 \dots \sigma_n(A_{n+1}))$. Notons $n \in \mathbb{N}$ et $a \in A_{n+1}$ tels que $u \in \mathrm{Fac}(\sigma_0 \dots \sigma_n(a))$. Comme la représentation est primitive, il existe $m > n+1$ tel que a apparaisse dans $\sigma_{n+1} \dots \sigma_m(b)$ pour tout $b \in A_{m+1}$. En particulier,

$$u \in \mathrm{Fac}(\sigma_0 \dots \sigma_m(b)) \quad \forall\, b \in A_{m+1}.$$

Notons $L = \max_{b\in A_{m+1}} |\sigma_0 \dots \sigma_m(b)|$. Par construction, u apparaît comme facteur du préfixe de longueur L de $\sigma_0 \dots \sigma_{n'}(a_{n'+1})$, et ce pour tout $n' \geq m$. Pour n' assez grand, ce préfixe est également préfixe de x donc $u \in \mathrm{Fac}(x)$. □

La réciproque fait appel à un résultat de théorie des graphes dû au mathématicien hongrois Dénes König.

Lemme 7.1.7 (Lemme de König)**.** *Soit $\mathcal{T}$ un arbre infini dont chaque sommet a un degré fini. Pour tout sommet v, il existe un chemin simple infini à partir de ce sommet.*

Démonstration. Comme le graphe est connexe, tout sommet est relié à v par un (unique) chemin. Or le graphe est infini et v n'a qu'un nombre fini de voisins donc il existe un voisin v_1 de v par lequel démarre un infinité de ces chemins. Considérons le sous-arbre de $T \setminus \{v\}$ contenant v_1. On peut réitérer le raisonnement sur v_1 pour obtenir v_2, et ainsi de suite. La suite de sommets obtenue fourni un chemin infini de $\mathcal{T}$. □

Lemme 7.1.8. *Soit $\boldsymbol{s}$ une suite de morphismes non-effaçants. Il existe une suite $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}_0}$ telle que, pour tout $n \in \mathbb{N}$, $\sigma_n(a_{n+1})$ commence par a_n.*

Démonstration. Construisons par récurrence l'arbre enraciné $\mathcal{T}$ de la façon suivante :

1. la racine est étiquetée par ε,
2. les sommets du niveau $n+1$ sont les lettres de A_n,
3. tous les sommets du premier niveau sont reliés à la racine par une arête,
4. pour tout $n \geq 1$, un sommet a du niveau $n+1$ est relié au sommet b du niveau n si $\sigma_{n-1}(a)$ commence par b.

Le graphe obtenu est effectivement un arbre car tout sommet autre que la racine est relié à exactement un sommet de niveau inférieur (et à aucun sommet du même niveau). Cet arbre a une infinité de sommets mais chaque sommet est de degré fini. Par le lemme de König, il existe un chemin infini commençant à la racine. Notons $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$ la suite des sommets de ce chemin (sans la racine). Pour tout $n \in \mathbb{N}$, on a alors $a_n \in A_n$ et $\sigma_n(a_{n+1})$ commence par a_n. □

Proposition 7.1.9. *Si $\boldsymbol{s}$ est une représentation $\mathcal{S}$-adique primitive d'un ensemble infini $T \subseteq A^*$, alors il existe une suite $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$ et un mot $x \in A^{\mathbb{N}}$ tels que* $\mathrm{Fac}(x) = T$ *et $(\boldsymbol{s}, \boldsymbol{a})$ soit une représentation $\mathcal{S}$-adique de x.*

Démonstration. Par le lemme précédent, il existe une suite $\boldsymbol{a}$ telle que $\sigma_n(a_{n+1})$ commence a_n pour tout $n \in \mathbb{N}$. Dans ce cas, pour tout $n \in \mathbb{N}$, $\sigma_0 \dots \sigma_{n-1}(a_n)$ est un préfixe de $\sigma_0 \dots \sigma_n(a_{n+1})$. La suite

$$(\sigma_0 \dots \sigma_n(a_{n+1}))_{n\in\mathbb{N}}$$

converge donc. Comme la suite $\boldsymbol{s}$ est croissante partout, la limite est un mot infini $x \in A^{\mathbb{N}}$. Le couple $(\boldsymbol{s}, \boldsymbol{a})$ est une représentation $\mathcal{S}$-adique de x par construction. De plus, par la Proposition 7.1.6, $\boldsymbol{s}$ est une représentation $\mathcal{S}$-adique de $\mathrm{Fac}(x)$ donc $\mathrm{Fac}(x) = T$. □

7.2 Automorphismes et bases tame

Dans la suite, nous ne considérons pas tous les morphismes mais nous restreignons plutôt aux automorphismes les plus simples qui restent néanmoins intéressants. Ces automorphismes vont également permettre de définir des bases particulières appelées « tame ». Nous montrons notamment que toute base incluse dans un ensemble dendrique récurrent est tame.

Définition 7.2.1. Un automorphisme σ de F_A est *positif* si, pour toute lettre $a \in A$, $\sigma(a)$ est un mot non vide sur A.

L'ensemble des automorphismes positifs forme un monoïde pour la composition. Parmi tous les automorphismes positifs de F_A, on peut en distinguer certains à partir desquels sont construits les automorphismes tame.

Définition 7.2.2. Pour toutes lettres $a, b \in A$ distinctes, on définit les endomorphismes $\alpha_{a,b}$ et $\tilde{\alpha}_{a,b}$ sur F_A de sorte que

$$\alpha_{a,b}(a) = ab, \qquad \tilde{\alpha}_{a,b}(a) = ba$$

et pour tout $c \in A \setminus \{a\}$,

$$\alpha_{a,b}(c) = c = \tilde{\alpha}_{a,b}(c).$$

On se convainc rapidement que $\alpha_{a,b}(A) = (A \setminus \{a\}) \cup \{ab\}$ est une base de F_A donc que $\alpha_{a,b}$ est un automorphisme positif, de même pour $\tilde{\alpha}_{a,b}$.

Définition 7.2.3. Pour toute permutation ν de A, on définit aussi l'automorphisme α_ν de F_A tel que

$$\alpha_\nu(a) = \nu(a).$$

Ces automorphismes ainsi que les $\alpha_{a,b}$, $\tilde{\alpha}_{a,b}$ définis précédemment forment l'ensemble des automorphismes positifs *élémentaires* de A noté $\mathcal{S}_e$.

Définition 7.2.4. Un automorphisme positif de F_A est *tame* s'il est dans le sous-monoïde engendré par les automorphismes positifs élémentaires de A.

On a une terminologie similaire pour les bases du groupe libre F_A.

Définition 7.2.5. Une base X de F_A est *positive* si $X \subseteq A^*$. Elle est *tame* s'il existe un automorphisme tame α tel que $X = \alpha(A)$.

Remarquons que toute base tame est positive car, si $Y \subseteq A^*$, alors $\alpha(Y) \subseteq A^*$ pour tout automorphisme élémentaire α.

Exemple 7.2.6. Si $A = \{a, b, c\}$, l'ensemble $X = \{ac, acb, cacb\}$ engendre F_A donc il s'agit d'une base. C'est même une base tame car

$$X = \alpha_{a,c}(a, ab, cab) = \alpha_{a,c}\tilde{\alpha}_{b,a}(a, b, cb) = \alpha_{a,c}\tilde{\alpha}_{b,a}\alpha_{c,b}(A).$$

L'ensemble $Y = \{abb, abc, ac\}$ est également une base de F_A car

$$c = (abb)^{-1}abc(ac)^{-1}abc$$

donc $c \in \langle Y \rangle$ et il en découle rapidement que $a, b \in \langle Y \rangle$. Cependant, il ne peut pas être obtenu comme image d'un ensemble par un des morphismes $\alpha_{\beta,\gamma}$ ou $\tilde{\alpha}_{\beta,\gamma}$, $\beta, \gamma \in A$ car aucune des lettres a, b ou c n'est tout le temps précédée ou suivie de la même lettre. Nous allons effectivement montrer qu'il ne s'agit pas d'une base tame de F_A.

Proposition 7.2.7. *Un ensemble $X \subseteq A^+$ est une base tame de F_A si, et seulement si, $X = A$ ou il existe une base tame Y de F_A et des mots $u, v \in Y$ tels que*

$$X = (Y \setminus \{u\}) \cup \{uv\} \quad \textit{ou} \quad X = (Y \setminus \{v\}) \cup \{uv\}.$$

Démonstration. Si X est une base tame de F_A, alors il existe un automorphisme tame α tel que $X = \alpha(A)$. Par définition, il existe alors $\alpha_1, \ldots, \alpha_n$ des automorphismes élémentaires tels que

$$X = \alpha_1 \ldots \alpha_n(A).$$

Montrons le résultat par récurrence sur $n \in \mathbb{N}$. Si $n = 0$, alors $X = A$ et on peut conclure. Supposons à présent le résultat vrai pour $Y = \beta(A)$ où $\beta = \alpha_1 \ldots \alpha_{n-1}$ est un automorphisme tame. Si α_n était un automorphisme obtenu à partir d'une permutation, alors $X = Y$ donc on peut conclure par hypothèse de récurrence. Si $\alpha_n = \alpha_{a,b}$, alors on a

$$\begin{aligned} X &= \alpha(A) \\ &= \beta\alpha_{a,b}(A \setminus \{a\}) \cup \{\beta(\alpha_{a,b}(a))\} \\ &= \beta(A \setminus \{a\}) \cup \{\beta(ab)\} \\ &= (\beta(A) \setminus \{\beta(a)\}) \cup \{\beta(a)\beta(b)\} \\ &= (Y \setminus \{u\}) \cup \{uv\} \end{aligned}$$

où $u = \beta(a)$ et $v = \beta(b)$ sont deux éléments de Y. Si $\alpha_n = \tilde{\alpha}_{a,b}$, on trouve par le même raisonnement que

$$X = (Y \setminus \{\beta(a)\}) \cup \{\beta(b)\beta(a)\}.$$

On peut donc conclure la récurrence.

Pour la réciproque, si $X = A$, alors il est évident que X est une base tame de F_A. Si $X = (Y \setminus \{u\}) \cup \{uv\}$ où Y est une base tame de F_A et où $u, v \in Y$, alors il existe un automorphisme tame β tel que $Y = \beta(A)$. Notons

$$a = \beta^{-1}(u) \quad \text{et} \quad b = \beta^{-1}(v).$$

On a alors

$$\begin{aligned} X &= (Y \setminus \{u\}) \cup \{uv\} \\ &= (\beta(A) \setminus \{\beta(a)\}) \cup \{\beta(a)\beta(b)\} \\ &= \beta\alpha_{a,b}(A) \end{aligned}$$

donc X est une base tame. On conclut pour le cas où $X = (Y \setminus \{v\}) \cup \{uv\}$ par un raisonnement symétrique. □

Corollaire 7.2.8. *Si X est une base tame de F_A et un code bifixe, alors $X = A$.*

Démonstration. En effet, si $X = (Y \setminus \{u\}) \cup \{uv\}$, alors X n'est pas un code suffixe car $v \in Y \setminus \{u\} \subseteq X$. Dans l'autre cas, X n'est pas un code préfixe car $u, uv \in X$. □

Exemple 7.2.9. Pour revenir à l'exemple précédent, on a

$$X = \{ac, acb, cacb\} = (\{ac, acb, c\} \setminus \{c\}) \cup \{cacb\},$$

$$\{ac, acb, c\} = (\{ac, b, c\} \setminus \{b\}) \cup \{acb\}$$

et

$$\{ac, b, c\} = (A \setminus \{a\}) \cup \{ac\}.$$

Ceci confirme le fait que X est une base tame de F_A. De plus, grâce à la preuve de la Proposition 7.2.7, on retrouve les mêmes automorphismes élémentaires permettant d'obtenir X à partir de A.

Par contre, l'ensemble Y étant un code bifixe, il ne peut pas être une base tame.

Le théorème suivant montre que, dans le cadre d'un ensemble dendrique récurrent, toutes les bases sont tame.

Théorème 7.2.10. *Toute base de F_A incluse dans un ensemble dendrique récurrent est tame.*

Démonstration. Soient S un ensemble dendrique récurrent. Montrons que, pour toute base $X \subseteq S$ de F_A, X est tame. Comme toute base de F_A a le même cardinal que A et est donc en particulier finie, on peut définir

$$l(X) = \sum_{x \in X} |x|.$$

Procédons par récurrence sur $l(X)$. Soit $X \subseteq S$ une base de F_A pour laquelle $l(X)$ est minimal. Etant donné que $\varepsilon \notin X$,

$$l(X) \geq |A|$$

donc le cas de base est le cas où $X = A$ qui est trivialement tame.

Supposons à présent le résultat vrai pour toute base $Y \subseteq S$ de F_A telle que $l(Y) < l(X)$.

Si X est un code bifixe, alors, comme il s'agit d'une base de F_A qui est un sous-groupe d'indice 1 et que S a la propriété d'indice fini, X est un code bifixe S-maximal de S-degré 1. Par l'Exemple 2.5.7, on en conclut que $X = A$ qui est tame.

Si, maintenant, X n'est pas un code suffixe, alors il existe $x, y \in A^+$ tels que $y, xy \in X$. Dans ce cas, posons

$$Y = (X \setminus \{xy\}) \cup \{x\}.$$

On a donc $X = (Y \setminus \{x\}) \cup \{xy\}$ où $x, y \in Y$. Remarquons que

$$Y \subseteq X \cup \{x\} \subseteq S,$$

que $\langle Y \rangle = \langle X \rangle = F_A$ et que $|Y| = |X| = |A|$ donc Y est une base de F_A incluse dans S. De plus,

$$l(Y) = l(X) - |xy| + |x| = l(X) - |y| < l(X)$$

donc, par hypothèse de récurrence, Y est une base tame. Par la Proposition 7.2.7, X aussi.

Si X n'est pas un code préfixe, on procède de façon symétrique. □

7.3 Une représentation $\mathcal{S}$-adique particulière

Dans cette section, nous présentons une représentation $\mathcal{S}$-adique particulière pour les mots uniformément récurrents. Nous montrons que, dans le cas d'un mot dendrique, une représentation composée uniquement d'automorphismes élémentaires en découle.

Tout d'abord, rappelons la notation suivante :

$$\mathcal{R}'_T(w) = \{u \in A^+ \mid uw \in w\mathcal{R}_T(w)\}$$

pour tout $w \in T$ où T est un ensemble de mots. Comme nous travaillons ici avec des mots infinis, nous utiliserons plutôt

$$\mathcal{R}'_x(w) = \mathcal{R}'_{\mathrm{Fac}(x)}(w)$$

pour tout $w \in \mathrm{Fac}(x)$ où x est un mot infini.

Comme pour l'ensemble des mots de premier retour à droite $\mathcal{R}_T(w)$, on peut montrer que, si x est uniformément récurrent, $\mathcal{R}'_x(w)$ est fini pour tout $w \in \mathrm{Fac}(x)$ et, si x est un mot dendrique récurrent, $\mathcal{R}'_x(w)$ est une base de F_A.

Remarquons que, si x_0 est la première lettre d'un mot infini x et si elle apparaît une infinité de fois dans x, alors

$$x \in (\mathcal{R}'_x(x_0))^{\mathbb{N}}.$$

Cette décomposition est unique car x_0 ne peut apparaître que comme préfixe des mots de $\mathcal{R}'_x(x_0)$.

Si x est uniformément récurrent, notons $k = |\mathcal{R}'_x(x_0)|$ et

$$\mathcal{R}'_x(x_0) = \{w_1, \ldots, w_k\}$$

où les w_i sont numérotés selon leur ordre d'apparition comme mots de premier retour à gauche dans x. Dans ce cas, on a les définitions suivantes.

Définition 7.3.1. Si $x \in A^{\mathbb{N}}$ est uniformément récurrent, le *morphisme de retour* de x est le morphisme

$$\lambda_x : \{1, \dots, k\}^* \to (\mathcal{R}'_x(x_0))^*$$

tel que

$$\lambda_x(i) = w_i.$$

L'image dérivée de x est alors le mot $D(x) \in \{1, \dots, k\}^{\mathbb{N}}$ tel que

$$x = \lambda_x(D(x)).$$

En particulier, par construction, $D(x)$ commence par 1. Remarquons de plus que λ_x est un morphisme codant pour $\mathcal{R}'_x(x_0)$ et que

$$\mathrm{Fac}(D(x)) = D'_{\lambda_x}(\mathrm{Fac}(x))$$

si on définit D'_f en inversant gauche et droite dans la définition d'ensemble dérivé, i.e. si f est un morphisme codant pour $\mathcal{R}'_T(w)$,

$$D'_f(T) = f^{-1}(Tw^{-1}).$$

Par un raisonnement similaire à ce qui a été fait dans la démonstration de la Proposition 6.3.2, on peut alors montrer que, si x est dendrique et récurrent, $D(x)$ l'est aussi.

Exemple 7.3.2. Rappelons que le mot de Tribonacci (Exemple 1.4.7) est donné par

$$x = abacabaabacababacabaabacabac\cdots$$

Les mots de retour à gauche pour a sont, dans leur ordre d'apparition, ab, ac et a. Le morphisme λ_x est alors défini par

$$\lambda_x(1) = ab, \quad \lambda_x(2) = ac \quad \text{et} \quad \lambda_x(3) = a.$$

On a également

$$D(x) = 121312112131212\cdots$$

Nous allons utiliser ces définitions pour construire une représentation $\mathcal{S}_e$-adique de n'importe quel mot dendrique récurrent.

Définition 7.3.3. Soit $x \in A^{\mathbb{N}}$ un mot dendrique récurrent. On définit la suite de morphismes $\boldsymbol{\lambda} = (\lambda_n)_{n\in\mathbb{N}}$ et la suite de mots infinis $(u_n)_{n\in\mathbb{N}}$ par

$$u_0 = x, \quad u_{n+1} = D(u_n) \quad \text{et} \quad \lambda_n = \lambda_{u_n}.$$

Par le Théorème de retour, on constate par récurrence sur $n \in \mathbb{N}$ que l'alphabet sur lequel est défini λ_n ou, de façon équivalente, l'alphabet de u_{n+1} est exactement $\{1, \dots, |A|\}$.

Proposition 7.3.4. *Soit $x \in A^{\mathbb{N}}$ un mot dendrique récurrent*[(2)]*. Avec les notations précédentes, on a*

$$\lim_{n\to\infty} |\lambda_0 \dots \lambda_n(1)| = +\infty,$$

$$x = \lim_{n\to\infty} \lambda_0 \dots \lambda_n(1)$$

et, si on suppose $A = \{1, \dots, |A|\}$, alors, pour tout $n \in \mathbb{N}$, λ_n peut s'étendre en un automorphisme tame de F_A.

(2). Nous écartons ici le cas dégénéré où $|A| = 1$.

Démonstration.

1. Procédons par l'absurde et supposons que les longueurs soient bornées. Par définition de λ_n, pour tout mot w et pour tout $n \in \mathbb{N}$, on a

$$|\lambda_n(w)| \geq |w|.$$

Cela signifie donc qu'il existe alors $N \in \mathbb{N}$ tel que, pour tout $n \geq N$,

$$|\lambda_0 \dots \lambda_n(1)| = |\lambda_0 \dots \lambda_N(1)|.$$

Or, pour tout $n \geq N$, par construction,

$$1 \in \mathrm{Pref}(\lambda_{N+1} \dots \lambda_n(1))$$

donc on doit avoir

$$1 = \lambda_{N+1} \dots \lambda_n(1)$$

ou encore

$$\lambda_m(1) = 1$$

pour tout $m > N$.
Pour tout $k \in \mathbb{N}$ et pour tout $n \in \mathbb{N}_0$, si $\lambda_n(1) = 1$ et $1^k \in \mathrm{Pref}(u_{n+1})$, alors $1^{k+1} \in \mathrm{Pref}(u_n)$. En effet, par définition, $u_n = \lambda_n(u_{n+1})$ et l'image de toute lettre par λ_n commence par 1 donc

$$1^{k+1} = \lambda_n(1^k)1 \in \mathrm{Pref}(u_n).$$

Par récurrence sur $k \in \mathbb{N}$, on montre alors que

$$1^{k-i+1} \in \mathrm{Pref}(u_{N+i+1}), \quad \forall\, i \leq k.$$

En effet, c'est immédiat pour $k = 0$ et pour $k > 0$, on procède par récurrence sur $k - i$.
En particulier, $1^k \in \mathrm{Pref}(u_{N+1})$ pour tout $k \in \mathbb{N}$ donc $u_{N+1} = 1^\omega$. C'est absurde car l'alphabet minimal de u_{N+1} doit être $\{1, \dots, |A|\}$.

2. Par construction, pour tout $n \in \mathbb{N}$,

$$u_n = \lambda_n(u_{n+1})$$

donc

$$x = u_0 = \lambda_0(u_1) = \dots = \lambda_0 \dots \lambda_n(u_{n+1}).$$

De plus, 1 est un préfixe de u_{n+1} donc $\lambda_0 \dots \lambda_n(1)$ est un préfixe de x et ce, pour tout $n \in \mathbb{N}$. On peut donc conclure que

$$x = \lim_{n \to \infty} \lambda_0 \dots \lambda_n(1)$$

grâce au point précédent.

3. Comme expliqué précédemment, l'alphabet sur lequel est défini λ_n est $\{1, \dots, |A|\} = A$ donc on peut étendre λ_n en un morphisme pour avoir

$$\lambda_n : F_A \to \langle \mathcal{R}'_{u_n}(1) \rangle \subseteq F_A$$

pour tout $n \in \mathbb{N}$. Le mot u_n étant dendrique récurrent, $\mathcal{R}'_{u_n}(1)$ est une base de F_A qui, de plus, est incluse dans $\mathrm{Fac}(u_n)$ qui est un ensemble dendrique. Par le Théorème 7.2.10, $\mathcal{R}'_{u_n}(1)$ est donc une base tame de F_A. Quitte à composer l'automorphisme tame permettant d'obtenir $\mathcal{R}'_{u_n}(1)$ à partir de A avec une permutation, on peut en déduire que λ_n est un automorphisme tame pour tout $n \in \mathbb{N}$.

□

Définition 7.3.5. La *représentation* Λ*-adique* d'un mot x dendrique récurrent est le couple $(\boldsymbol{\lambda}, \mathbf{1})$ où $\mathbf{1} = (1)_{n \in \mathbb{N}_0}$.

Exemple 7.3.6. Le lecteur observateur aura remarqué que, dans l'Exemple 7.3.2, si on assimile les deux alphabets, i.e. $1 \equiv a$, $2 \equiv b$ et $3 \equiv c$, alors λ_x est exactement le morphisme σ utilisé pour construire le mot de Tribonacci et $D(x) = x$. La représentation Λ-adique du mot de Tribonacci n'est alors rien d'autre que l'itération infinie de σ, i.e. $\lambda_n = \sigma$ pour tout $n \in \mathbb{N}$.

Bibliographie

[1] B. Adamczewski et Y. Bugeaud, *On the complexity of algebraic numbers I. Expansions in integer bases*, Ann. Math. (2) **165** (2007), n° 2, 547–565.

[2] P. Arnoux et G. Rauzy, *Représentation géométrique de suites de complexité* $2n+1$, Bull. Soc. Math. Fr. **119** (1991), n° 2, 199–215.

[3] J. Berstel *et al.*, *Bifix codes and Sturmian words*, J. Algebra **369** (2012), 146–202.

[4] J. Berstel, D. Perrin, et C. Reutenauer, *Codes and Automata*, Encyclopedia of Mathematics and Its Applications, n° 129, Cambridge University Press, Cambridge, 2010.

[5] V. Berthé *et al.*, *Acyclic, connected and tree sets*, Monatsh. Math. **176** (2015), n° 4, 521–550.

[6] ———, *Bifix codes and interval exchanges*, J. Pure Appl. Algebra **219** (2015), n° 7, 2781–2798.

[7] ———, *Maximal bifix decoding*, Discrete Math. **338** (2015), n° 5, 725–742.

[8] ———, *The finite index basis property*, J. Pure Appl. Algebra **219** (2015), n° 7, 2521–2537.

[9] ———, *Rigidity and substitutive dendric words*, Int. J. Found. Comput. Sci. **29** (2018), n° 5, 705–720.

[10] F. Dolce et D. Perrin, *Neutral and tree sets of arbitrary characteristic*, Theor. Comput. Sci. **658** (2017), 159–174.

[11] ———, *Eventually dendric shifts*, Computer science – Theory and applications, CSR 2019 (R. van Bevern et G. Kucherov, éds), Springer, 2019, p. 106–118.

[12] S. Ferenczi, *Rank and symbolic complexity*, Ergodic Theory Dyn. Syst. **16** (1996), n° 4, 663–682.

[13] J. Justin et L. Vuillon, *Return words in Sturmian and episturmian words*, Theor. Inform. Appl. **34** (2000), n° 5, 343–356.

[14] M. Keane, *Interval exchange transformations*, Math. Z. **141** (1975), 25–31.

[15] R. Lyndon et P. Schupp, *Combinatorial Group Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, n° 89, Springer-Verlag, Berlin, 1977.

[16] M. Rigo, *Théorie des automates et langages formels*, Année académique 2009-2010, disponible via l'URL <http://www.discmath.ulg.ac.be/cours/main_autom.pdf>.